JOURNAL OF ECONOMIC THEORY 66, 1–32 (1995)

A Folk Theorem for Stochastic Games*

PRAJIT K. DUTTA

Department of Economics, University of Wisconsin, Madison, Wisconsin 53706; and Department of Economics, Columbia University, New York, New York 10027

Received August 26, 1991; revised November 29, 1993

In many dynamic economic applications, the appropriate game theoretic structure is that of a stochastic game. A folk theorem for such games is presented. The result subsumes a number of results obtained earlier and applies to a wide range of games studied in the economics literature. The result further establishes an underlying unity between stochastic and purely repeated games from the point of view of asymptotic analysis, even though stochastic games offer a much richer set of deviation possibilities. *Journal of Economic Literature* Classification Number: C73, D90.

1. INTRODUCTION

In recent years dynamic strategic interaction has been extensively studied, particularly within the context of repeated games. See, for instance, Aumann and Shapley [3], Rubinstein [26], Abreu [1], and Fudenberg and Maskin [13] for analyses of the basic repeated game model with complete information and perfect monitoring. A drawback of the repeated game paradigm is that it is premised upon a *completely unchanging environment*. In many applications, such an assumption is not even approximately correct. For instance, in economic models with stock variables, current and future action possibilities and payoffs are directly a function of the available stocks. Cases in point are growth models, in which capital or

* This research was initiated with Dilip Abreu and without his continual questions and substantive suggestions it may well not have been completed. Needless to say, I am extremely grateful to him for all his help. Parts of this paper were presented at seminars at Queen's, UCLA, UC Riverside, CalTech., Cornell, Columbia, NYU, Penn State, University of Texas, Austin, and University of Wisconsin–Madison. A preliminary version of this paper was presented at the International Game Theory Conference, July 1989, at Ohio State University. I thank the various seminar audiences for helpful comments. I have especially benefited from the comments of Drew Fudenberg, Abraham Neyman, an anonymous associate editor, and an anonymous referee. All errors, of course, remain my responsibility. This paper was completed while I was visiting the University of Rochester, and I thank them for their support.

0022-0531/95 $6.00

human and natural resources are the relevant productive assets,[1] financial models with price competition, in which accumulated wealth or historical prices are determinants of current and future action possibilities and payoffs.[2] Intertemporal links may also be present through other payoff relevant factors as demand and cost conditions or level of innovations, representing "shocks" to the system (which may persist across periods). The appropriate model in these cases is a *stochastic game* in which a state variable represents the environment of the game and its evolution is determined by the initial conditions, player's actions, and the transition law. The abstract model of a stochastic game is, of course, very general. In particular, the transition rule from the current state to the subsequent state(s) may be either probabilistic or deterministic. The purely deterministic case is sometimes referred to as a dynamic game and a special case of it is the repeated game.

Earlier work on stochastic games has focused on the issue of existence of (perfect) equilibria in Markovian strategies (see, for example, [16, 22–24].[3] This paper provides instead a characterization of equilibrium payoffs when players are very patient, dropping the assumption of Markovian behaviour. The latter restriction appears to be arbitrary; indeed in the strictly repeated context, it is seldom suggested that Markovian behaviour is strategically salient (see, however, Maskin and Tirole [19]). Recently, the folk theorem question in non-repeated settings has been also investigated by Friedman [12] and Lockwood [18]. Their results are discussed in Section 8. The approach to folk theorem analysis that is taken in this paper has some implications for the purely repeated game case as well; these implications are developed in detail in Abreu, *et al.* [2] and briefly discussed in Section 8 of this paper.

A major difficulty in analysing stochastic games is that deviations not only alter current payoffs but also change the distribution over future states. A central observation of this paper is that for a variety of cases this difficulty does have a resolution, at least asymptotically. Indeed, it is shown that for games ranging from completely communicating stochastic games to deterministic capital accumulation games, both immediate gain and state manipulation incentives may be deterred as the discount factor goes to 1. The folk theorems for repeated games with perfect monitoring may be extended to this setting. These include the theorems of Aumann and Shapley [13], Rubinstein [26], and Fudenberg and Maskin [13].

[1] Strategic formulations include Benhabib and Radner [4], Bernheim and Ray [5], Dutta and Sundaram [11], and Sundaram [27].

[2] For instance, see Maskin and Tirole [19], Holden and Subrahmanyam [17] and Dutta and Madhavan [10].

[3] Mertens and Partharasarathy [21] have shown the existence of perfect equilibria in a more general class of strategies.

I provide an analog of the last result, which is suitable for the applications discussed.

Section 2 describes the model. Preliminary results on feasible payoffs and min–max levels are contained in Sections 3 and 4. Section 5 presents and discusses the assumptions. Section 6 contains the main theorem and related results while Section 7 presents two examples which illustrate the necessity of the hypotheses. The concluding section, Section 8, contains a discussion of some applications and the relationship of the theorem to other available results.

2. The Model

This paper considers infinite horizon stochastic games with perfect monitoring. These games are defined by a quintuple $\langle S, A_i, r_i, q, \delta; i = 1, ..., n \rangle$, where i is the player index, S is the set of *states*, and A_i is the ith player's set of *actions*. The sets S, A_i, $i = 1, ..., n$ are assumed to be finite. Assume also, without loss of generality and only to save on notation, that each player has available to him the same set of actions in every state.[4] Denote $A = \prod_{i=1}^{n} A_i$. The ith player's *one-period reward* is $r_i: S \times A \to \mathbb{R}$. It associates with every vector of player's actions a and the current state s an immediate reward $r_i(s, a)$. q is the *law of motion* of the system—it associates with each (s, a) in period t a distribution over the $(t+1)$ period's state, $q(\cdot \mid s, a)$. If the game is in state s and the players choose the action a, then the game moves to state s' next period with probability $q(s' \mid s, a)$. Further, $\delta \leqslant 1$ is the common *discount factor* which the players employ in evaluating payoff streams. Finally, all past states, the current state, and all players' past actions are assumed to be observable.

The following notation will be used: s will refer to a generic state and s_t will be the state in period t while (the generic) player i's action in that period will be denoted a_{it}. a_t will describe the action vector $(a_{1t}, ..., a_{nt})$. In all statements pertaining to i, j will index "another" player while $-i$ will refer to the group of players other than i. At various points I will talk of a "punishment regime" for player i during which regime player j's action will be denoted a_j^i. Finally, $\|.\|$ will denote any one of the equivalent norms in $\mathbb{R}^n$.

A behaviour *strategy* for player i is denoted Π_i. It is sequence of maps, $\Pi_{i0}, \Pi_{i1}, ..., \Pi_{it}, ...$, where Π_{it} selects a distribution, in period t, over the set of actions A_i as a function of the previous history $h_t = (s_0, a_0, ..., s_{t-1}, a_{t-1}, s_t)$. If the distribution depends only on the current

[4] If this requirement is violated, one can define "dummy" action variables and add these to the available set of actions appropriately in order to arrive at a problem in which this condition is met. For a more detailed discussion of this issue, see [23].

state and further if this choice is independent of t, then the strategy is said to be *Markov*. If the distributions are degenerate we have a *pure Markov* strategy.[5] Player i's randomization device is assumed to be unobservable to other players, i.e., I analyse a game with *unobservable (private) mixed strategies*. However, players can coordinate on a *public randomisation* device.

Note that although the games are called stochastic, there is no requirement that the transition probabilities $q(\cdot \mid s, a)$ be non-degenerate. The class of games in which the transitions are deterministic are sometimes called dynamic games. In particular, complete information repeated games are trivial examples of stochastic games (under the restrictions that $q(s \mid s, a) = 1$, for all s, a and r_i is independent of s).

A strategy for each player, and the initial state, determines a distribution over finite period histories and by extension a distribution over infinite histories. Let $r_i(t; \Pi, s)$ denote the expected returns of player i at period t under the strategy vector $\Pi = \Pi_1, ..., \Pi_n$ and initial state $s_0 = s$. The discounted average (expected) return vector, if the initial state is s and the players employ strategy Π and the discount factor is $\delta < 1$, is

$$\left\{(W_1, W_2, ..., W_n)\colon W_i(s; \Pi, \delta) = (1-\delta) \sum_0^{\infty} \delta^t r_i(t; \Pi, s)\right\}. \tag{1}$$

A long-run average expected return vector, again for initial state s and strategy vector Π, is

$$\left\{(W_1, W_2, ..., W_n)\colon \exists (T_k)_{k=0}^{\infty} \text{ s.t. } W_i(s; \Pi) = \lim_{T_k \to \infty} \frac{1}{T_k} \sum_0^{T-1} r_i(t; \Pi, s)\right\}.^6 \tag{2}$$

For a given initial state s and discount factor $\delta < 1$, a strategy choice Π is a *Nash equilibrium* if no player profits from unilateral deviation, i.e., $W_i(s; \Pi, \delta) \geqslant W_i(s; \Pi_i', \Pi_{\hat{i}}, \delta)$ for all Π_i' and all i. A (subgame) *perfect equilibrium* is a strategy choice such that after every history, the strategy continuations constitute a Nash equilibrium.

[5] In the literature such strategies have sometimes been called stationary (for example, see [6, 16, 23], whereas more recent usage has called them Markov (for example, [19]). I adopt the latter convention.

[6] Note that in the above definition of a feasible long-run average payoff vector there is a requirement that there exists *some* (common) subsequence of finite horizons, $T_0, T_1, T_2, ...$, along which the per period payoff of each player has a limit; hence, every such subsequential limit is admissible as a long-run average payoff. When $n = 1$, this is a more permissive definition than the standard one which takes the *smallest* such limit as *the* long-run average payoff. On the other hand, for $n > 1$ the definition is more restrictive than an alternative one which would allow the horizon subsequences to differ across players. Of course, all of this is an issue only because the finite horizon averages need not have a limit along the sequence $T = 0, 1, 2, ...$.

The *min–max level*[7] of player i, for *initial state* s *and discount factor* δ, will be denoted $m_i(s, \delta)$ and is defined by

$$m_i(s, \delta) = \inf_{\Pi_{-i}} \sup_{\Pi_i} W_i(s; \Pi_i, \Pi_{-i}, \delta). \tag{3}$$

For the long-run average criterion the min–max level, $m_i(s)$, is defined as

$$m_i(s, \delta) = \inf_{\Pi_{-i}} \sup_{\Pi_i} \inf_{(T_k)} W_i(s; \Pi_i, \Pi_{-i}, \delta).^{8} \tag{4}$$

Note that, in general, min–max levels will vary with the initial state and the discount factor. Max–min levels can be defined analogously.

3. Feasible Payoffs in the Game

Since there is no stage game, the relevant set of feasible payoffs is the set of (discounted or long-run) average expected returns in the infinite horizon game. Unlike a repeated game, however, not all feasible payoffs can be achieved by (convexifying over) strategies which repeatedly play a constant action (and hence yield a constant payoff every period). Indeed, the set of strategies in a stochastic game is very large; for example, there are strategies which condition on history in arbitrarily complex ways. Furthermore, there is no unique set of (average) payoffs which can be achieved at every discount factor and from every initial state. In this section I consequently investigate two issues: (a) the existence of a set of simple strategies whose payoffs "span" the set of feasible payoffs and (b) the relation between the sets of feasible payoffs at different discount factors.

3.1. *Pure, Markov Strategies Suffice*

Let $F(s, \delta)$ denote the set of publicly randomised discounted average returns that are feasible in the stochastic game, i.e., let $F(s, \delta)$ be the convex hull of the set of discounted average payoffs that can be realised by behaviour strategy tuples:

$$F(s, \delta) = \Big\{ w \in \mathbb{R}^n \colon \exists \beta^j \geqslant 0, \sum_j \beta^j = 1,$$
$$\text{and } \Pi^j \text{ s.t. } w_i = \sum_j \beta^j W_i(s; \Pi^j, \delta),\ i = 1, \ldots, n \Big\}$$

[7] Since there is no stage game, min–max levels are naturally defined according to the returns over the entire game.

[8] Recall, from (2) above, that a long-run average payoff is any subsequential limit of per period payoffs. The min–max payoff then is defined with respect to the smallest such limit.

Similarly we can define the set of long-run average returns that are feasible; denote this set $F(s)$.[9] Let $\phi(s, \delta)$ (respectively $\phi(s)$) denote the discounted (respectively long-run) average expected returns when *only pure Markov strategies* are used. The extreme points of the (convex) set of feasible payoffs $F(s, \delta)$ are clearly the solutions to Max $\sum_i \lambda_i W_i$, where W is the payoff (vector) to some behavior strategy and $\lambda_i \in R$, $\lambda \neq 0$. From [7] it then follows that the solution to such an optimization problem is realized by a pure Markov strategy, i.e., all of the extreme points of $F(s, \delta)$ are generated by pure Markov strategies. It readily follows from the above observation that public randomization over pure Markov strategies recovers *all* feasible payoffs in the discounted game. For the undiscounted stochastic game, a limiting argument yields the same spanning result.

LEMMA 1. *All feasible payoffs, in the discounted as well as the undiscounted game, can be realised by one-shot public randomization over pure Markov strategies*:

$$\text{(i)} \quad F(s, \delta) = \text{co}\, \phi(s, \delta), \qquad \forall s \in S, \delta < 1$$

$$\text{(ii)} \quad F(s) = \text{co}\, \phi(s)$$

Proof. In the appendix. •

The lemma simplifies the analysis in the rest of the paper considerably. The restriction, without loss of generality, to pure (publicly randomized) strategies will make the detection of deviation from such strategies immediate. Further, this result makes possible a simple resolution of the related question (which is important for asymptotic analysis): is the set of feasible payoffs continuous in the discount factor at $\delta = 1$?

3.2. *Continuity of Feasible Payoffs*

Since the set of feasible payoffs is spanned by public randomization over pure Markov strategies and since the payoff to such strategies can be shown to be continuous at $\delta = 1$ (under the finiteness of state and action spaces assumed here), it follows that:

[9] From (4) it may seem that the only time players are allowed to publicly randomise their actions is at date 0. However, it can be shown by standard arguments that the set of feasible payoffs when players publicly randomise their actions after any number of arbitrary histories is also $F(s, \delta)$.

LEMMA 2. *The set of feasible payoffs is continuous at* $\delta = 1$, *i.e.*, $F(s, \delta) \to F(s)$, *as* $\delta \to 1$, *for every* $s \in S$; *the convergence is to be understood to be in the Hausdorff metric.*[10]

Proof. In the Appendix. •

4. INDIVIDUAL RATIONALITY

As discussed above, the min–max level in a stochastic game varies with the discount factor and the initial state. What then is the relevant security level which should be the benchmark for folk theorem analysis? Since the folk theorem is an asymptotic result, a natural benchmark would be the limit of the (state-dependent) discounted average min–max payoffs, as the discount factor goes to 1. It follows from results of Bewley and Kohlberg [6] and Mertens and Neyman [20] that this limits exists and furthermore equals the long-run average min–max.

4.1. *Continuity of Min–Max*

For *two-person zero-sum* games, Bewley and Kohlberg [6, Theorem 3.1] show that $\lim m_i(s, \delta)$(as $\delta \uparrow 1$) exists for all i and s in S. For this same class of games, Mertens and Neyman [20] then showed that this limit is, in fact, the long-run average min–max level. If we think of player i and the group of players $-i$ as constituting a "two-person" game, the Mertens-Neyman theorem yields[11]:

PROPOSITION 3. *For all* $\eta > 0$, *there is a strategy of players other than i, say* Π^*_{-i}, *and* $N > 0$, *s.t. for all* $\infty \geqslant T \geqslant N$ *and every strategy* Π_i,

$$W_i(s; \Pi_i, \Pi^*_{-i}, T) \leqslant \lim_{\delta \uparrow 1} m_i(s, \delta) + \eta,$$

where $W_i(s; \Pi, T)$ *is the T-period time average of expected returns from strategy* Π *and initial state* s ($T = \infty$ *refers to the* limsup *of such finite period averages*).

[10] For any two closed sets B and C in R^n, the Hausdorff distance between the sets is defined as $d(B, C) = \max(\sup_x \rho(x, C), \sup_y \rho(y, B))$, where $\rho(x, C) = \inf \|x - z\|$ $x \in B$, $z \in C$, likewise for $\rho(y, B)$.

[11] The game with i and the group $-i$ as "two players" is different from a standard two-person game in that the players $-i$ may not have "ability to act as one". In particular they may not have access to $(n-1)$ player randomisation. However, the Mertens–Neyman result is valid in this context as well.

$m_i(s, \delta)$ is both the min–max *and* max–min level of player i, by a result of Partharasarathy [23]. From Proposition 3, it is clear that the limit of $m_i(s, \delta)$(as $\delta \uparrow 1$) denoted $m_i(s)$, is the long-run average min–max *and* max–min level for player i. This will be the relevant security level of player i in the analysis that follows.

4.2. *Individually Rational Payoffs*

An appropriate definition of an individually rational payoff is a somewhat delicate matter for stochastic games. This is because the continuation payoffs to any strategy are, typically, different for different histories and, in particular, are different from the lifetime payoff computed at period zero. Hence, the comparison with a min–max level can be made at several different time points and after different histories. In this subsection I present two alternative definitions for a payoff to be deemed individually rational: the first definition is with respect to the payoff computed at period zero alone whereas the second definition incorporates the continuation payoffs after all histories.

Fix a discount factor $\delta \leqslant 1$ and consider any initial state s. I will say that a discounted average payoff $w(s, \delta)$ (respectively, a long-run average payoff $w(s)$) is individually rational in the *ex ante* sense if $w_i(s, \delta \geqslant m_i(s, \delta)$ for all i (respectively $w_i(s) \geqslant m_i(s)$ for all i), where s is the period zero state. Let $F^*(s, \delta)$ (respectively $F^*(s)$) denote the discounted (respectively long-run) average strictly individually rational (ex ante sense) payoffs sets, i.e.,

$$F^*(s, \delta) = \{w \in F(s, \delta) \colon w_i > m_i(s, \delta), i = 1, ..., n\}$$

(and similarly $F^*(s)$). On the other hand a payoff vector w is said to be individually rational in the *ex post* sense if it is generated by a strategy Π such that, after all histories, continuation payoffs are individually rational in the ex ante sense, i.e., the inequalities above hold for all states s_t and periods t. I return to the connection between these concepts in Section 5.

From the continuity of the min–max levels (Proposition 3) and the convergence of feasible payoff sets (Lemma 2) it clearly follows that the set of strictly individually rational payoffs (in the ex ante sense) converge.

LEMMA 4. *The set of ex ante individually rational payoffs is continuous at $\delta = 1$; for all $\varepsilon > 0$, there is $\underline{\delta} < 1$, s.t. for $\delta \geqslant \underline{\delta}$,*

$$d(F^*(s, \delta), F^*(s)) < \varepsilon, \forall s \in S,$$

where d is the Hausdorff distance.

An implication of such continuity is, of course, that if a payoff vector is strictly individually rational in the long-run average sense, then it can be arbitrarily closely approximated by strictly individually rational discounted average payoffs. Since equilibrium payoffs are ex ante individually rational we also have the following corollary, which applies to the set of *equilibrium payoffs*, $V(s, \delta)$:

COROLLARY 5. *For all* $\varepsilon > 0$, *there is* $\underline{\delta} < 1$ *s.t. for* $\delta \geqslant \underline{\delta}$,

$$B_{\varepsilon}(F^*(s)) \supset V(s, \delta) \; \forall s.$$

where $B_{\varepsilon}(F^*(s))$ *is the* ε*-neighbourhood of* $F^*(s)$.

The folk theorem, to be proved shortly, will assert that all strictly individually rational long-run average payoffs can be approximated by an equilibrium discounted average payoff, for sufficiently large discount factors. Evidently, the theorem and Corollary 5 together provide a complete characterisation of the equilibrium payoff set when the discount factor is close to 1.

5. ASSUMPTIONS AND IMPLICATIONS

In the next two sections the following (folk theorem) question is investigated: under what conditions on the stochastic game will any strictly individually rational payoff (in the ex-post sense) arise as a subgame perfect equilibrium payoff for sufficiently high discount factors? Two types of assumptions will be made: First, *asymptotic state independence*:

(A1) The set of feasible long-run average payoffs $F(s)$ is independent of s, says $F(s) = F$.

(A2) The long-run average min–max $m_i(s)$ is independent of s, for all i, say $m_i(s) = m_i$.

Second, I will make *one* of the following two assumptions on the set of feasible long-run average payoffs F.

Payoff Asymmetry (PA). There are n payoff vectors $\bar{v}^i \in F$, $i = 1, ..., n$, such that player i's payoff is the least under $\bar{v}^i$, i.e., $\bar{v}^i_i < \bar{v}^j_i$, $\forall i, j$, $i \neq j$.

Full Dimensionality (FD). The dimension of the set of feasible payoffs is the same as the number of players, i.e., dim $(F) = n$.

The two conditions are obviously related; full dimensionality (FD) clearly implies pairwise asymmetry (PA) but the converse is not true. The main theorem below will be proved under (FD) but for an interesting special case which has been much discussed in the literature, the weaker condition (PA) will be seen to suffice.

The assumptions above, and especially (A1)–(A2), are not expressed in terms of primitives. A statement based on primitives would be unwieldy because the variety of conditions under which the assumptions are satisfied could not be succinctly encompassed in a single theorem. I briefly discuss these assumptions and their implications now and return in Section 8 to a fuller discussion of primitive models in which they are satisfied.

5.1. *Asymptotic State Independence*

Future feasible (and hence, equilibrium payoffs) in a stochastic game depend on the current state. Therefore, for a folk theorem to hold, there must be some similarity in the possibilities from different states.[12] The issue then is how restrictive must these conditions be? (A1) is a mild requirement as Section 8 will make clear. (A2) is stronger and I defer to Section 8 a discussion of primitive conditions on the game which guarantee this assumption. If one or the other of these assumptions is not satisfied, the method of proof will illustrate the appropriate subset of the feasible payoff space on which state manipulation incentives can be deterred (see Corollaries 9.1 and 9.2 below).

From Lemma 1 we know that an initial one-shot public randomisation over pure Markov strategies realises all feasible long-run average payoffs. Such a scheme does not guarantee that the expected long-run average after *all* histories is, approximately, the same (and this will be seen to be important for a player's ex post incentives). However, in the presence of (A1), any one-shot randomisation can be replicated by the following "time-averaged" strategy: its components are pure Markov strategies and it cycles repeatedly between these pure strategies with the cycle lengths consistent with the one-shot convexification. Furthermore, the cycle lengths are chosen in such a way that the continuation payoffs are approximately the same after all histories. Therefore, we get:

LEMMA 6. *Under* (A1), *for any* $w \in F$ *and* $\varepsilon > 0$, *there is a pure strategy whose long-run average payoff is within* ε *of* w, *after all histories.*

Proof. In the Appendix.

Given Lemma 6, the asymptotic state independence min–max assumption (A2) then implies:

LEMMA 7. *Under* (A1)–(A2), *a long-run average payoff* $w \in F$ *is strictly individually rational in the ex post sense if and only if it is strictly ex ante individually rational.*

[12] In Section 7, I present two examples which demonstrates the necessity of these asymptotic state invariance assumptions.

Given the continuity, at $\delta = 1$, of the min–max levels (Proposition 3) and the feasible payoff set (Lemma 2), the construction of the proof of Lemma 6 yields:

LEMMA 8. *Under* (A1) *and* (A2), *for any* $w \in F^*$ *and* $\varepsilon > 0$, *there is a pure strategy and* $\underline{\delta} < 1$, *s.t. for all* $\delta \geqslant \underline{\delta}$ *and all initial states* s, *its discounted average payoff is within* ε *of* w *after all histories. Consequently such a payoff is strictly individually rational in the ex-post sense, for all* $\delta \geqslant \underline{\delta}$ *and all* s.

5.2. *Payoff Asymmetry and Full Dimensionality*

Fudenberg and Maskin [13] have shown in a repeated game that if there is *perfect congruence of interests* among the players, punishments to deter deviations from individually rational paths may not be credible. Hence, some asymmetry in payoff possibilities is required for a discounted folk theorem to hold for repeated games, and, by implications, for stochastic games. The appropriate expression of this asymmetry are the two conditions PA and FD. Payoff asymmetry is an easy condition to check; it is guaranteed by the existence, in all states, of an action tuple $a^i(s)$ which is strictly worse for player i than any other action tuple. This condition may be interesting not so much because it is weaker than full dimensionality[13] but because it can be shown that within the class of strategies analysed in this paper and Fudenberg and Maskin [13] it is additionally almost a necessary condition for the folk theorem.[14] If mixed strategies are unobservable, I will need to strengthen (PA) to (FD).

6. RESULTS

THEOREM 9. *Under* $(A1)$, $(A2)$, *and* (FD), *any* $w \in F^*$ *can be arbitrarily approximated as an equilibrium payoff, for sufficiently high discounting; for all* $\varepsilon > 0$, *there is* $\underline{\delta} < 1$ *s.t. for any* $\delta \geqslant \underline{\delta}$, *there is a perfect equilibrium strategy whose payoff, after all histories that are reached with positive probability, is within* ε *of* w.

If either asymptotic state independence condition, (A1) or (A2), does not hold, the following results still hold (and are immediate corollaries of the proof of the theorem):

[13] It might be worth noting that for $n > 3$ it can be shown that it is even weaker than $n - 1$ dimensionality of the feasible payoff set. Examples can be constructed for higher dimensions where the payoff set is simply a two-dimensional plane.

[14] The necessary condition allows weak inequalities in (PA), with an additional restriction in case of an equality. See also Abreu *et al.* [2] which shows that in repeated games payoff asymmetry is in fact an almost necessary condition for the folk theorem. That paper also contains a further discussion of the relationship between (PA) and (FD).

COROLLARY 9.1. *Suppose that* (A1) *and* (FD) *hold. Then the conclusions of Theorem 9 hold for any long-run average payoff that is strictly individually rational from all states, i.e., for any* $w \in F$ *such that* $w_i > m_i(s)\ \forall i, s$.

In the absence of (A1), define $F = \bigcap_s F(s)$.

COROLLARY 9.2. *Suppose that* (FD) *is satisfied by* $F = \bigcap_s F(s)$. *Then the conclusions of Theorem 9 hold for any long-run average payoff that is feasible and strictly individually rational from all states, i.e., for any* $w \in F$ *such that* $w_i > m_i(s)\ \forall i, s$.

The full dimensionality condition can be weakened in a interesting special case that has been widely studied in the literature (see Section 8 for a discussion). I report the analog of Theorem 9 in this case and note that the analogs of Corollaries 9.1 and 9.2 also hold (but are not reported). This is the simpler model in which *mixed strategies are observable*. In that context I show:

PROPOSITION 9.3. *Suppose that* (A1), (A2) *and* (PA) *hold and suppose further that mixed strategies are observable. Then the conclusions of Theorem 9 are valid for any* $w \in F^*$.

Remark. Abreu, *et al.* [2] have shown that Proposition 9.3 can be generalised to allow for unobservable mixed strategies provided we restrict attention to repeated games. Since their general approach is very similar to one critical step in the proof of the main theorem here, I report, without proof, their result:[15]

PROPOSITION 9.4. *In a repeated game,* (PA) *implies that for any* $w \in F^*$, *there is* $\underline{\delta} < 1$, *s.t. for all* $\delta \geq \underline{\delta}$ *there is a perfect equilibrium whose discounted average payoff is* w.

6.1 *An Overview and An Informal Discussion of the Proof*

The presence of a state variable makes folk theorem analysis difficult because a deviation by any player now has two consequences: (i) it yields, as in a repeated game, one-shot gains but, and more importantly, (ii) it changes the distribution of the state in the next period and therefore, in all future periods). The second effect has several implications; first, the effect need not vanish even when players are very patient if the strategic possibilities differ markedly across states—since it can then linger for an

[15] Since there is no distinction between ex ante and ex post payoffs in a repeated game, the payoff w can be realised exactly.

infinite future. Thus one *necessary* condition for a folk theorem to hold in stochastic games is that strategic possibilities should be, at least asymptotically, state-invariant. The results demonstrate that this additional incentive can be deterred—globally, if (A1)–(A2) hold (Theorem 9 and Proposition 9.3), or locally, if one or the other assumption does not hold (Corollaries 9.1 and 9.2)—as players become patient.[16]

A second implication is that a player may deviate only to "improve" the state distribution (and this will be true whenever continuation payoffs differ across states). Imagine that, in the play of a particular strategy vector, the continuation payoffs are higher for player i from state s_t than from state s_t'. Then player i has every incentive, in period $t-1$, to choose a current action which make s_t more likely (even if it involves some immediate loss).[17] Note, furthermore, that this incentive is stronger the *more* patient are the players (unlike the incentive to deviate for immediate gain alone). Hence, in the method of proof, it will be important that strategies be constructed which have very similar continuation payoffs after all histories.

It turns out that the folk theorem proof is much simpler when deviations are observable (or equivalently, all mixed strategies are observable). Thus I separate the proof into two parts in order to address separately observable and unobservable deviations. The proof when deviations are observable is presented in what follows whereas the proof for the unobservable deviation case is presented in the Appendix. Furthermore, to clarify the somewhat technical arguments, in each case I first present an informal discussion of the proof and then the details. The assumptions (A1), (A2), and (FD) (or their local versions) will allow a logic of proof that is similar to that for the purely repeated case as in Fudenberg and Maskin [13] except for two additional sets of arguments necessitated by the state manipulation possibility. (See Steps 1 and 2 below and Steps 3 and 4 in the Appendix).

An Informal Discussion of the Proof When Mixed Strategies Are Observable. The proof is as follows: I construct a strategy whose components are $(n+1)$ pure strategies $\tilde{\Pi}$, Π^i, $i=1, ..., n$. Play starts with $\tilde{\Pi}$ and proceeds according to this strategy till player i deviates. Then play proceeds to T_m periods of min–maxing i, in the long-run average sense, followed by Π^i.

[16] Of course, even under (A1)–(A2) it is still the case that the discounted feasible payoff set $F(s, \delta)$ and the discounted min–max level $m_i(s, \delta)$ *will* be state dependent. From the proof it will be clear that in order to deter deviations in order to effect state manipulation, one needs not just the fact that these sets and security levels converge but additional arguments that they can be made to converge at appropriate rates.

[17] Of course, this incentive for deviating from a proposed strategy can be deterred if the deviation can be detected immediately (i.e., a pure strategy is being played) *and* punished severely. In other words this cause for deviation will be problematical whenever (a) mixed strategies are being played or (b) the worst equilibrium strategy is being played (and hence no more severe a punishment can be imposed than that which is currently in play).

During the min–max phase, player i plays a (dynamic) best response. Every unilateral deviation by a player j immediately initiates T_m periods of min–maxing j followed by Π^j. In order to deter deviations for the purpose of improving the state distribution, the strategies $\tilde{\Pi}, \Pi^i, i = 1, ..., n$ are constructed in the following manner:

Step 1. The strategies are *cyclic* (e.g., Π^i involves playing pure Markov strategies $g_1^i, g_2^i, ..., g_p^i$, for $T_1^i, T_2^i, ..., T_p^i$ periods respectively and then restarting the same sequence again and again). Denote the length of each cycle T^i, i.e., $T^i = \sum_p T_p^i$. Evidently, the lifetime payoffs and deviation possibilities for any player in the play of Π^i are the same, for a fixed state, at period t $(< T^i)$ as at period $t + NT^i$ (where N is any positive integer). Furthermore, the cycle lengths are so chosen that, for sufficiently high discount factors, the associated payoffs, say $[w(s_t, \delta)\ V^i(s_t, \delta),\ i = 1, ..., n,\ t < T^i]$, are, uniformly across states and periods, (i) *asymmetric*, (ii) *strictly individually rational* and (iii) *$\tilde{\Pi}$-dominated*, i.e., $V_i^i(s_t, \delta) < V_i^j(s'_\tau, \delta)$ and $m_i(s_t, \delta) < V_i^i(s'_\tau, \delta) < w(s''_\tau, \delta)$, in each case for all s_t, s'_τ, s_τ, and for all i, j. Further, $\|w(s, \delta) - w\| < \varepsilon$, $\forall s$.

Remark. Π^i is part of player i's "punishment regime." For a punishment strategy to be credible, an obvious necessary condition is that punishing a deviant must not take the game into a state which is unfavorable for the players doing the punishing (hence (*i*)). Furthermore, a deviant player must be unable to take the game into states from which his worst individually rational payoff is better than continuation payoffs to non-deviation (hence (ii)).

The incentive for player i to deviate solely to improve the state is particularly strong when i's worst strategy Π^i in play (since play will eventually return to Π^i itself but possibly at a more favourable state). Suppose, for example, that the continuation payoff at some period t, within the cycle $0, ..., T^i$, is such that $V_i^i(s_t, \delta)$ is less than $E[V_i^i(s_0, \delta)]$. Deviation, which involves a finite min–max period followed by $E[V_i^i(s_0, \delta)]$ may then be profitable, especially since the finite min–max period is increasingly irrelevant as δ goes to 1.[18] Step 2 deals with this ex post incentives problem.

Step 2. Let g^i be the pure Markov strategy that maximises player i's long-run average payoffs. The strategy Π^i can be modified to make the payoffs over a cycle independent of the initial state: at the beginning of each T^i cycle, play proceeds to $(g_p^i, p = 1, ..., P)$ with probability $\mu^i(s_0)$ and

[18] In a repeated game this problem is avoided since there exist (publicly randomised) constant actions in each period which realise any feasible payoff. In fact, even if public randomisation is not allowed, one can construct a strategy such that continuation payoffs are always monotonically increasing in time. In the presence of a state variable it is not possible to ensure that continuation payoffs are constant or (almost surely) greater at $t + 1$ than at t.

to g^i with the remaining probability. The probabilities are chosen in a way such that i) player i's payoffs over the cycle are independent of s_0 and ii) are asymmetric, i.e., are strictly less than i's payoffs under the strategy which represents a similar modification of Π^j. I retain notation and call these strategies Π^i, $i = 1, ..., n$, as well.

Now consider exactly the strategy defined by $\tilde{\Pi}$, Π^i, $i = 1, ..., n$ above. Call this strategy Π^*. I show, first, that it is possible to construct such a strategy (given (A1)–(A2) and (PA) and, second, that it is a subgame perfect equilibrium (for sufficiently) high discount factors); hence, Proposition 9.3 holds.

6.2. *The Details of the Proof When Mixed Strategies Are Observable*

Proof of Step 1. Let l^i (respectively, b^i) denote the long-run average payoff vector in which player i's payoff is the *least* (respectively, the *best*), i.e. $l^i_i \equiv \min\{v_i: (v_{-i}, v_i) \in F\}$ (respectively, $b^i_i \equiv \max\{v_i: (v_{-i}, v_i) \in F\}$). Recall that w is the given strictly individually rational long-run average payoff and the asymmetric payoffs (whose existence is asserted by (PA)) are denoted $\bar{v}^i$, $i = 1, ..., n$. Further, let $m_i \equiv 0$. Pick convexification weights $\beta_1 > 0$, $\beta_2 > 0$, and define[19]

$$V^i \equiv \beta_1 l^i + \beta_2 \bar{v}^i + (1 - \beta_1 - \beta_2) w. \tag{5}$$

Clearly one can pick the convexification weights appropriately to prove:

LEMMA 10. *There are feasible long-run average payoffs* V^i, $i = 1, ..., n$, *satisfying* $\forall i, j$

(a) *strict individual rationality* $V^i \gg 0$,

(b) *asymmetry* $V^i_i < V^j_i$ $i \neq j$

(c) *target payoff domination* $V^i_i < w_i$.

From Lemma 1 it follows that each of the payoffs l^i, $\bar{v}^i$, and w is itself a convex combination of payoffs to pure Markov strategies. Hence (and this is made precise in the proof of Lemma 6), it then follows that that there is a pure cyclic strategy Π^i which approximates V^i; let it be defined by pure Markov strategies $g^i_i, g^i_2, ..., g^i_p$, played successively for $T^i_1, T^i_2, ..., T^i_p$ periods and then repeated infinitely many times. The ratio $T^i_p / \sum_p T^i_p$ reflects the convexification weights induced by (5), and the bigger is T^i_p, the closer is the approximation. Now we can appeal to the

[19] Since $l_i \leqslant m_i < w_i$, the two payoff vectors, l^i and w are distinct. Also note that the convexification weights do not depend on the player index i.

continuity of payoffs to pure cyclic strategies at $\delta = 1$ (this is made precise in Lemma 2).

Step 1 has been proved. •

Proof of Step 2. For any Markov strategy g, let $W(s; T)$ denote the T-period discounted average from initial state s; $W_i(s; T) = [(1-\delta)/(1-\delta^T)] \sum_0^{T-1} \delta^t r_i(t; g, s)$.[20] (When $T = \infty$, the notation will be $W_i(s)$). Let $T(\varepsilon)$ be a period length and $\delta(\varepsilon)$ be a discount factor such that for all pure Markov strategies g and all initial states s, $\| W(s; T) - W(s) \| < \varepsilon$ whenever $T \geqslant T(\varepsilon)$, $\delta \geqslant \delta(\varepsilon)$. Denote similarly the payoffs to Π^i over the T^i cycle as $V^i(s; T^i)$.

LEMMA 11. *There are probabilities $\mu^i(s)$, $i = 1, ..., n$ and $s \in S$, such that for all $\delta \geqslant \delta(\varepsilon)$, s, s',*

$$\mu^i(s)\, V_i^i(s; T^i) + [1 - \mu^i(s)]\, b_i^i(s; T^i) = \mu^i(s')\, V_i^i(s'; T^i) + [1 - \mu^i(s')]\, b_i^i(s'; T^i). \tag{6}$$

Further, writing $v_i^j(s; T^j) \equiv \mu^j(s)\, V_i^j(s; T^j) + [1 - \mu^j(s)]\, b_i^j(s; T^j)$,

$$v_i^i(s; T^i) < v_i^j(s'; T^j), \quad i \neq j,\ s, s' \in S. \tag{7}$$

Proof. Pick any $\varepsilon > 0$ with the property that $V_i^i + \varepsilon < V_i^j - \varepsilon$. Take $T_p^i > T(\varepsilon)$, for all i and p. Hence, we have $\| V^i(s; T^i) - V^i \| < \varepsilon$, for all s, i, $\delta \geqslant \delta(\varepsilon)$ or equivalently, $\max_s V_i^i(s; T^i) < \min_s b_i^i(s; T^i)$. Thus we can find probabilities $\mu^i(s)$ as defined in (6), with, in fact, the added property that $\max_s V_i^i(s; T^i) = \mu^i(s)\, V_i^i(s; T^i) + [1 - \mu^i(s)]\, b_i^i(s; T^i)$, for all s. As ε goes to 0, $\mu^i(s)$ clearly goes to 1. For sufficiently small ε, (7) holds. The lemma follows. •

For future reference, let $v^i(s; \delta)$ denote the infinite horizon discounted average payoffs to the strategy Π^i (with public randomisation according to μ^i at the end of every T^i periods), if the state at the beginning of Π^i is s. In particular, player i's payoffs within each T^i cycle are independent of the state at the beginning of that cycle, i.e., $v_i^i(s; T^i) = v_i^i(s'; T^i) \equiv v_i^i(\delta)$.

Step 2 has been proved. •

Proof of proposition 9.3. Let the best (respectively, the least) one-shot payoff of player i be denoted b_i (respectively l_i).[21] Pick $0 < \eta < V_i^i - \varepsilon$. From

[20] Note that I suppress the arguments (g, δ) in the notation for the average payoffs, i.e., instead of writing $W(s; g, \delta, T)$ (as would be suggested by Eq. (1) for example). I am simply writing $W(s; T)$.

[21] I use this notation since it is suggestive of the notation employed for the best and least long-run average payoffs, b_i^i and l_i^i.

Proposition 3 it follows that there is $\delta_1 < 1$ and T' such that upon min–maxing for at least T' periods, player i's T'-period discounted average payoffs can be held below η. Let $T_m \geqslant T'$ satisfy for $\delta \geqslant \delta_2 \geqslant \max(\delta_1, \delta(\varepsilon))$,

$$(1-\delta_i^T) l_i + \delta_i^T v_i^i(\delta) > (1-\delta) b_i + \delta[(1-\delta^{T_m})\eta + \delta^{T_m} v_i^i(\delta)]. \qquad (8)$$

Equation (8) can clearly be satisfied by choosing T_m to be large relative to T^i. Let us first check the i's punishment regime, T_m periods of min–maxing followed by Π^i, is a subgame perfect equilibrium. Note that the left-hand side of eq. (8) is a lower bound on the lifetime payoffs of player i from the play of Π^i (since l_i is a lower bound on the flow payoffs during any cycle). Likewise, the right-hand side of (8) is an upper bound on the lifetime payoffs of player i from deviating against Π^i since by Proposition 3 his average payoffs during a min–max phase is at most η. Hence, (8) implies that player i has no profitable deviation once the play of Π^i is initiated. By definition, he has no incentives to deviate during the min–max phase. From (7) it follows that, for sufficiently high δ, players $j \neq i$ have no profitable deviation after any history, either during the min–max phase or during the play of Π^i; this is because, as δ goes to 1, player j's payoffs from not deviating are at least (approximately) $\min_s v_j^i(s; T^i)$ which is strictly greater than his lifetime payoffs from deviating since this is at most (approximately) $\max_s v_j^j(s; T_j)$. Hence, i's punishment regime is a perfect equilibrium in the subgame after any deviation of player i.

Deviation from $\tilde{\Pi}$ is unprofitable for any player, for sufficiently high discount factors, given Step 1 iii) (again since the lifetime payoffs from not deviating are (arbitrarily close to) $\min_s w_i(s, \delta)$, whereas the payoffs to deviation are bounded above by $\max_s v_i^i(s; T^i)$. Proposition 9.3 is proved. •

The only part of the above strategy where players take privately observed random actions is in the phase where players $j \neq i$ min–max player i. Since these players may not be indifferent between the pure actions in the support of their mixed strategies, the difficulty now, in moving to the unobservable mixed strategies case, is to induce these players to play the pure actions with the appropriate probabilities. As noted by Fudenberg and Maskin [13], the only way to do so is to make them indifferent across their pure actions.

This is done by modifying the strategy Π^* in the following way: I construct strategies $\Pi^{ij}, j \neq i$, which are "close" to Π^i. The strategy Π^{ij} can be interpreted as player j's "punishment" *within* the punishment regime of player i; he gets a payoff from it that is smaller than his payoff from Π^i. After min–maxing player i, play proceeds to Π^i "most of the time," but with a small probability it goes to Π^{ij}, for every $j \neq i$. Furthermore, the probabilities are so chosen that every player $j \neq i$ is indifferent, in expected terms, between each pure action in the support of his mixed min–maxing strategy.[22] Note, finally, that deviations outside the support of the mixed strategies that min–max player i can be deterred by the same considerations as those used above for the observable deviations case. All of these arguments are made precise, and Theorem 9 is proved, in the Appendix.

7. Examples

In this section I present two examples which illustrate the necessity of the asymptotic state-invariance conditions, (A1) and (A2), in order for a stochastic game folk theorem to hold. The first counter-example violates (A2).

Example 1. The set of long-run average feasible payoffs are invariant across states, i.e., (A1) holds, but the long-run average min–max levels are state-dependent, i.e., (A2) does not hold. Consequently, there are ex post individually rational payoffs which are not equilibrium payoffs.

Details. Suppose there are two states, s and σ, two players, 1 and 2, and each player has two actions (in both states); player 1's actions (respectively, player 2's actions) are denoted (a_1, a_2) (respectively, (b_1, b_2)). The law of motion is such that (a_1, b_2) takes the game from state s today to σ tomorrow and likewise from state σ today to s tomorrow. Every other action vector leaves the state unchanged. The one-shot payoffs are defined as follows:

State s	b_1	b_2
a_1	$(0, 1)$	$(0, -1)$*
a_2	$(2, 2)$	$(-1, 0)$

State ϕ	b_1	b_2
a_1	$(2, 2)$	$(0, 1)$*
a_2	$(-1, 0)$	$(0, -1)$

(Note that * signifies the fact that when this action vector is played, the game moves with probability 1 to the other state).

Since there is an action vector, (a_1, b_2), by use of which we can go between the states, it is immediate that $F(s) = F(\sigma)$. However (the column) player 2's long-run average min–max level is state-dependent; $m_2(s) = 1 \neq 0 = m_2(\sigma)$.

Consider now any pair of long-run average payoffs, $[W_1(s), W_2(s)]$, $[W_1(\sigma), W_2(\sigma)]$, feasible from states s and σ respectively, that give player

[22] For repeated games, Fudenberg and Maskin [13] induce the required indifference by constructing continuation payoffs such that every sample path during the min–max phase has the same lifetime reward, i.e., a player is indifferent with probability one (rather than in expectation alone). Continuation payoffs satisfying such a (strong) condition appear to be impossible to construct for stochastic games.

2 a payoff from state s, $W_2(s) \in (1, 2)$ and from state σ a payoff, $W_2(\sigma) \in (0, 1)$. (This is possible to do; e.g., the strategy which publicly randomizes between (a_1, b_1) and (a_2, b_1) in both states, with appropriate probabilities, yields such payoff vectors). Evidently these payoffs are ex post strictly individually rational.[23]

However, these payoffs cannot be equilibrium payoffs if the game starts from σ. To see this note that in order to give player 2 a strictly positive payoff, player 1 must (occassionally) play a_1. If he does so player 2 can guarantee himself a long-run average payoff of 1 by playing (the deviant action) b_2 in state σ and with probability 1 (eventually) taking the game into state s. Evidently the same argument works for high discount factors. Hence, not all individually rational payoffs are subgame perfect equilibria even at discount factors close to 1.

Remark . Note that Corollary 9.1 holds in this example. Hence, any (long-run average) feasible payoffs in which player 2's payoffs are strictly greater than 1 are indeed equilibrium payoffs for patient players.[24]

I now turn to a counter-example in which (A1) is violated (although not (A2)).

EXAMPLE 2. The long-run average min–max levels are invariant across states, i.e., (A2) holds, but the set of long-run average feasible payoffs is state-dependent, i.e., (A1) does not hold. Consequently, there are ex post individually rational payoffs which are not equilibrium payoffs.

Details. As in Example 1 there are two states, s and σ, two players, 1 and 2, and each player has two actions (in both states); again, player 1's actions (respectively, player 2's actions) are denoted a_1, a_2) (respectively, (b_1, b_2)). The law of motion is such that (a_2, b_1) takes the game from state σ today to s tomorrow. State s is an absorbing state; no action vector takes the game out of state s once it is reached. The one-shot payoffs are defined as follows:

State s	b_1	b_2	State ϕ	b_1	b_2
a_1	$(0, 2)$	$(3, 3)$	a_1	$(3, 0)$	$(0, -1)$
a_2	$(-1, 0)$	$(-1, 0)$	a_2	$(1, 0)$*	$(0, 1)$

(Note that, as before, * signifies the fact that when this action vector is played, the game moves with probability 1 to state s).

[23] In the payoffs $W(s)$ and $W(\sigma)$ suppose that player 1's payoffs are strictly positive, i.e., $W_1(s) > 0$, $W_1(\sigma) > 0$. We can find such feasible payoffs; e.g., the strategy described above does, in fact, yield player 1 strictly positive payoffs. Note that these payoffs are strictly individually rational for player 1 as well since his state-independent min–max level is easily verified to be 0.

[24] Again player 1's payoffs need to be strictly positive, of course.

It is straightforward to check that $m_i(s) = m_i(\sigma) = 0$, for players $i = 1, 2$; hence, (A2) is satisfied. Evidently, $F(s) = \text{co } \phi(s)$, where $\phi(s) = \{(0, 2), (3, 3), (-1, 0)\}$. But, since the game can transit from state σ to s, $F(\sigma) = \text{co } (F(s) \cup G)$ where $G = \text{co}\{(3, 0), (0, -1), (0, 1)\}$. In particular, $F(s) \neq F(\sigma)$.

Now consider any feasible pair of strictly positive long-run average payoffs, $[W_1(s), W_2(s)]$, $[W_1(\sigma), W_2(\sigma)]$, where $[W_1(\sigma), W_2(\sigma)]$ is additionally a vector which belongs to G but not to $F(s)$ and $W_2(\sigma) < 3/4$; in other words, the payoff is generated by some public randomisation over (a_1, b_1), (a_1, b_2) and (a_2, b_2) (which ensures that play never proceeds to state s). Evidently these payoffs are ex post strictly individually rational.

However, any such payoff vector necessarily involves the play of a_2 with positive probability. But then player 2 can (deviate and) play b_1, thereby taking the game to state s. From state s, his payoffs cannot be any less than 3/4, in equilibrium; else, player 1 makes less than his min–max level 0. Thus player 2 will find it profitable to deviate. Evidently the same argument works for high discount factors. Hence, not all individually rational payoffs are subgame perfect equilibria even at discount factors close to 1.

Remark. Note that Corollary 9.2 holds in this example. Hence, any (long-run average) payoff that is feasible from all states (and strictly individually rational), i.e., any strictly positive payoff vector in $F(s)$, is an equilibrium payoff for patient players.

8. Applications and Discussion

The principal structural restriction that was imposed was the finiteness of state and action sets. I believe that this restriction can be dispensed with, at the expense of a more technical analysis. Finiteness was critically used in establishing continuity of payoff sets and min–max levels at $\delta = 1$. Dutta [8] (and Mertens and Neyman [20]) give conditions under which such continuity of feasible payoffs sets (and min–max levels) would hold under general specifications of state and action spaces. Finiteness was also used in the asymmetry and state independence arguments of Lemmas 10 and A.2; the modifications here would be in the nature of uniformity conditions. In discussing whether the other hypotheses of our game are satisfied by various economic models, I will momentarily ignore the fact that the state–action spaces there are typically non-finite.

8.1. *Asymptotic State Independence of Payoffs*

There are two general conditions, special cases of which are satisfied by many economic models, which imply that feasible long-run average payoff sets are state independent. It is useful to remember, incidentally, that the long-run average criterion ignores all finite period returns and thus condition (A1) is equivalent to a requirement that payoff possibilities from any two states are eventually the same.

By analogy with the theory of Markov chains let us define:

DEFINITION. A stochastic game is said to be *communicating* if for each pair of states (s, s'), there is some strategy Π and an integer N such that the probability of going from s to s' in N steps, $q_{\Pi}^{N}(s, s') > 0$.

LEMMA 12. *In a communicating stochastic game, the set of feasible long-run average payoffs is independent of the initial state.*

Proof. In the Appendix. •

Cyclic or fully communicating games [15] i.e., games in which $q(s' \mid s, a) > 0$ for all s, s', a, are immediate examples of communicating games. So are dynamic games in which any one state can eventually (deterministically) transit to any other state, through some appropriate strategy. In economic structures like growth or oligopoly capital accumulation models, investment models in macroeconomics, or financial models, *pure accumulation* strategies (which involve zero consumption) typically allow the appropriate state to increase, and eventually to any desired level. Conversely, *free disposal* ensures that the state can also decrease. Communication is a consequence in such models. In models with sticky prices or other historical variables, typically the full communication condition is met.[25] Models in which there are exhaustible resources are examples of non-communicating systems.

A second general class of models in which asymptotic state independence holds are strictly stochastic games, i.e., those with "noisy" transition laws. The noise ensures that eventually the effect of the initial state disappears. There are many ways in which to formalise this idea. I report here a class of structures called scrambling models which have been recently studied by Lockwood [18].

DEFINITION. A stochastic game is called *scrambling* if the transition probabilities defined by any pure Markov strategy g have the following property: for all pairs of states s, s' there is a state s'' such that $q_g(s, s'') > 0$ and $q_g(s', s'') > 0$.

LEMMA 13. *Scrambling games satisfy* (A1).

Proof. See [18]. •

[25] The references in Footnotes 1 and 2 are covered by these remarks.

8.2. *Min–Max State Invariance*

Long-run average min–max levels will be independent of the initial state from which the game starts if the system communicates *independently* of the actions of any one player. Gillette [15], e.g., shows that in a cyclic game min–max values in the long-run average sense are state independent. Clearly, a weaker requirement is that there be some strategy choice of $(n-1)$ players, which generates a communicating system, regardless of the ith player's strategy. Such $(n-1)$ *state controllability* is exhibited by many *common state* resource games, in which players extract simultaneously from some common property resource and there is an upper bound on feasible extraction levels. $(n-1)$ players can make the resource grow or shrink by appropriately altering their own extraction rates. A somewhat different reason for long-run average min–max values to be state invariant is $(n-1)$ *eventual payoff controllability*, that similar returns be enforceable, eventually, from a number of alternative states and that *one* theses states be reachable by $(n-1)$ players. As an example, consider *separate-state* games, where the state $s = s_1, ..., s_n$, is n-dimensional and each player controls his own dimension. Although the ith player controls his own state, his worst payoff may be realised by the $(n-1)$ players (eventually) achieving some s_{-i} and playing some catastrophic action (for i) thereafter. Capital accumulation games offer an example, where above critical capital levels $(n-1)$ players can continuously drive the ith player's profits to zero by overproduction.

Payoff asymmetry of long-run average payoffs (or even full dimensionality) are satisfied in many of the economic models mentioned above. A simple sufficient condition is that there is some steady state of the system in which players have asymmetric (or full dimensional) one-shot rewards.

8.3. *Other Results*

The two papers closest to this one are Friedman [12] and Lockwood [18].[26] Friedman studies a class of non-repeated games in which there are no explicit state variables and period t returns depend on current and immediately preceding action, in his notation $P_i(a_{t-1}, a_t)$. This set-up is formally a dynamic game as can be seen by writing $s_t \equiv a_{t-1}$ and $r_i(s_t, a_t) \equiv P_i(a_{t-1}, a_t)$. Define $V(a) \equiv \{v\colon \exists a' \text{ s.t. } v_i = P_i(a, a')\}$ and $V = \bigcap_a V(a)$. It is immediate that $\bigcap_s F(s) \supseteq V$. Friedman then defines a notion of (state-independent) min–max, call it v^m, which has the property

[26] I am also aware of a result of Neyman, but so far have been unable to get a copy of his paper.

that $m_i(s) \leqslant v_i^m$.[27] With a full-dimensionality assumption on V, he then proves the asymptotic equilibrium sustainability of all $v \in V$ such that $v \gg v^m$. This result follows then from Corollary 9.2 (indeed with payoff asymmetry, since mixed strategies are inadmissible in the Friedman analysis).

Lockwood [18] analyses a stochastic game in which the transition matrix has the scrambling property defined above. Consequently (A1) and (A2) follow (see Lemmas 2.1 and 2.2 in his paper).[28] He imposes full-dimensionality on the (state-independent) long-run average payoff set[29] and establishes a folk theorem. All mixed strategies are observable in his framework. Thus his result is implied by Theorem 9 (and indeed can be strengthened to admit unobservable mixed strategies). Alternatively, maintaining observability of mixed strategies, his result is true under payoff asymmetry (Proposition 9.3)[30]

Recently Abreu *et al.* [12] by using a logic of proof similar to that employed in this paper, have proved the folk theorem for discounted repeated games under the payoff asymmetry condition (PA). Furthermore, they show that in repeated games, (PA) is also (almost) a necessary condition for the theorem.

There is also an extensive literature in non-repeated models, especially for specific applications, which investigates the sustainability of first-best or collusive outcomes alone (for example, Benhabib and Radner [4]). Dutta [9] shows that on this question, the predictions of repeated and non-repeated games may be dramatically different (in contrast to the above folk theorem conclusions).

APPENDIX

Proof of Lemma 1. The number of pure Markov strategies is finite and hence co $\phi(s, \delta)$ is a closed convex set. Suppose $w \in F(s, \delta)$ and

[27] The inequality is driven by the facts that (a) Friedman restricts himself to pure strategies and (b) that the state-independent min–max level is defined by taking the supremum over the state-dependent levels. Note also that the model considers action sets that are convex, compact subsets of R^n and thus our results do not immediately apply. The comments that follow should be interpreted as applying to either the infinite version of our model or the finite version of Friedman's.

[28] The scrambling assumption has the strong implication that finite period state distributions converge to an initial-state independent invariant distribution at a geometric rate that is uniform over all strategies.

[29] Actually Lockwood assumes the stronger condition that the payoff set formed by cycling over pure Markov strategies is full-dimensional.

[30] For repeated games, Fudenberg and Maskin [14] have a demonstrated the dispensability of public randomization in folk theorem analysis. The critical issue in deriving a similar conclusion for stochastic games is: can any feasible correlated long-run average payoff be exactly generated by high discount factors? Without full dimensionality, the answer is no. It remains an open question whether, given (FD), public randomization is inessential.

$w \notin \text{co}\, \phi(s, \delta)$. Then, by the strong separating hyperplane theorem [25, Corollary 11.4.2], w and co $\phi(s, \delta)$ lie in opposite open half-spaces of some hyperplane. But Blackwell [7, Theorem 7b] shows that for any extreme point of $F(s, \delta)$ there is a pure Markov strategy that generates it. We clearly have a contradiction.

Let us now show that the extreme points of $F(s)$ are also generated by pure Markov strategies. In other word, I prove:

LEMMA A.1. *For all* λ_i, $i = 1, ..., n$, *and initial state* s, *there is a pure Markov strategy* g^* *s.t.*

$$\sum_i \lambda_i W_i(s; g^*) \geqslant \sum_i \lambda_i W_i(s; \Pi). \tag{A.1}$$

for any feasible strategy Π.

Proof. Consider any pure Markov strategy, say g, and let $(r_i(t; g, s))_{t \geqslant 0}$ be the sequence of t-period expected returns for initial state s. Let p_t be the associated probability distribution, i.e., $r_i(t; g, s) = \sum_s r_i(s_t, g(s_t)s)\, p_t(s)$. For notational simplicity let me suppress all indices except time; consequently, the expected flow payoff in period t will be written as r_t.

I first show that $1/T \sum_{t=0}^{T-1} r_t$ has a limit as $T \to \infty$. Since the dynamic system formed by the strategy g is a finite Markov chain, we can partition the state space into a subset of transient states, say B, and a finite number of closed sets, $C_1, C_2, ..., C_p$. If $s \in C_p$ for some closed set, then a standard argument establishes the existence of a limit to $1/T \sum_{t=0}^{T-1} r_t$. On the other hand, if $s \in B$, then $W_i(s; g) = \sum_{s' \in B} p_t(s')\, W_i(s'; g) + \sum_{s' \in C} p_t(s')\, W_i(s'; g)$. Since $p_t(s') \to 0$ for all transient states, as $t \uparrow \infty$, it then follows that a limit exists for $1/T \sum_{t=0}^{T-1} r_t$ even when the initial state is transient.

It then follows by Abel's theorem[31] that the sequence of discounted average returns generated by the Markov strategy g also has a limit and indeed that the two limits are equal:

$$\lim_{T \uparrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} r_t = \lim_{\delta \uparrow 1} (1 - \delta) \sum_{t=0}^{\infty} \delta^t r_t. \tag{A.2}$$

Now pick an arbitrary strategy Π. Recall that its period t expected returns are denoted $r_i(t; \Pi)$. Let T^k be a sequence such that $\lim_{T^k \uparrow \infty} 1/T^k \sum_{t=0}^{T^k-1} r_i(t; \Pi) = W_i(s; \Pi)$, for all i. It is well known that we can find a particular sequence $(\delta_m)_{m \geqslant 0}$ and $\delta_m \uparrow 1$ with the property that $\lim_{T^k \uparrow \infty} 1/T^k \sum_{t=0}^{T^k-1} r_i(t; \Pi) = \lim_{\delta m \uparrow 1} (1 - \delta_m) \sum_{t=0}^{\infty} \delta^t r_i(t; \Pi)$, for

[31] Abel's theorem: for any sequence $(b_t)_{t \geqslant 0}$, $\lim_{\delta \to 1}(1 - \delta) \sum_{t=0}^{\infty} \delta^t b_t = \lim_{T \to \infty} 1/T \sum_{t=0}^{T-1} b_t$, if either limit exists.

all $i = 1, \ldots, n$. Since the number of pure Markov strategies is finite, for any $\delta_m \uparrow 1$, there is some pure Markov strategy g which maximises $\sum_i \lambda_i W_i(s; \Pi', \delta)$, over all feasible strategies Π', along a subsequence of δ_m. It then follows that

$$\begin{aligned}
\sum_i \lambda_i W_i(s; \Pi) &= \sum_i \lambda_i \left[\lim_{T^k \uparrow \infty} \frac{1}{T^k} \sum_{t=0}^{Tk-1} r_i(t; \Pi) \right] \\
&= \sum_i \lambda_i \left[\lim_{m \uparrow 1} (1 - \delta_m) \sum_{t=0}^{\infty} \delta_m^t r_i(t; \Pi) \right] \\
&= \lim_{m \uparrow 1} (1 - \delta_m) \sum_{t=0}^{\infty} \delta_m^t \sum_i \lambda_i r_i(t; \Pi) \\
&\leqslant \lim_{m \uparrow 1} (1 - \delta_m) \sum_{t=0}^{\infty} \delta_m^t \sum_i \lambda_i r_i(t; g) && \text{(A.3)} \\
&= \sum_i \lambda_i \left[\lim_{m \uparrow 1} (1 - \delta_m) \sum_{t=0}^{\infty} \delta_m^t r_i(t; g) \right] && \text{(A.4)} \\
&= \sum_i \lambda_i \left[\lim_{T \uparrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} r_i(t; g) \right] = \sum_i \lambda_i W_i(s; g). && \text{(A.5)}
\end{aligned}$$

Equation (A.3) follows from the optimality of g, in the discounted problems, while (A.4) and (A.5) follow from the arguments in the preceding paragraphs. Lemma A.1 is therefore proved. The remaining arguments left in order to establish Lemma 1 (ii) are identical to those used in proving Lemma 1 (i). ∎

Proof of Lemma 2. It is necessary and sufficient to show that

(a) $\forall w$ for which there is a sequence $\delta_n \to 1$, and $w_n \in F(s, \delta_n)$ with $w_n \to w$, $w \in F(s)$

(b) $\forall w$ in $F(s)$, as $\delta \to 1$, there is $w_\delta \to w$, $w_\delta \in F(s, \delta)$.

Invoking Lemma 1, all of these statements can be made for co $\phi(s, \delta)$ and co $\phi(s)$. Then, both (a) and (b) follow from the continuity of the returns to pure Markov strategies, at $\delta = 1$, and that has been proved within the proof of Lemma A.1 ∎

Proof of Lemma 6. Let w be a payoff in F. By Lemma 1, it follows that $w = \sum_{j=1}^{k} \lambda_j w_j$, where w_j is the long-run average return to some pure Markov strategy g_j.[32] Consider the following strategy tuple: the strategy g_1

[32] Strictly speaking for two different initial states s and s', w_i may be generated by different pure Markov strategies $g_i(s)$ and $g_i(s')$. The argument that follows can be modified in the obvious way to account for this.

is used for T_1 periods, followed by g_2 for T_2 periods and so on. After $T=\sum_{j=1}^{k} T_j$ periods, the cycle is repeated. T_j are chosen such that (a) T_j/T is arbitrarily close to λ_j and (b) $W_i(g_j; s, T_j) > w_i - \varepsilon$, for all s and i. Clearly, this strategy suffices to prove the lemma. ∎

Proof of Theorem 9. I have already proved the result under the simplification that mixed strategies are observable (Proposition 9.3). Hence, all that remains to be done is to extend that proof to accommodate unobservable mixed strategies.

From the arguments in Section 6 leading up to the proof of Proposition 9.3 it is clear that mixed strategies are only employed when players $j \neq i$ min–max player i; the issue then is how to deter these players from deviating (unobserved) within the support of their mixed min–maxing strategies. To repeat the discussion at the end of Section 6, I ensure that players $j \neq i$ are indifferent between all available pure actions by modifying the strategy Π^* in the following way: I construct strategies $\Pi^{ij}, j \neq i$, which are "close" to Π^i; each player $j \neq i$ prefers Π^i to Π^{ij}. After min–maxing player i, play proceeds to Π^i "most of the time", but with a small probability it goes to Π^{ij}, and this is true for every $j \neq i$. Furthermore, the probabilities are so chosen that every player $j \neq i$ is indifferent, in expected terms, between each pure action in the support of his mixed min–maxing strategy. The argument will be carried out in two steps:

Step 3. There exist pure cyclic strategies $\Pi^{ij}\,\forall j \neq i$ (with associated payoffs $U^{ij}(s, \delta)$) which have the following properties for sufficiently high discount factors and for any two initial states of a cycle, s and s': (i) player i is indifferent between the strategies Π^{ij} and Π^i, i.e., $U_i^{ij}(s, \delta) = V_i^i(s', \delta)$, (ii) player $j \neq i$ prefers that strategy Π^i be initiated rather than strategy Π^{ij}, i.e., $U_j^{ij}(s, \delta) < V_j^i(s', \delta)$, (iii) all players $k \neq i$ prefer Π^{ij} to their punishment strategy Π^k, i.e., $U_k^{ij}(s, \delta) > V_k^k(s', \delta)$, and (iv) all payoffs are strictly individually rational.

Condition (ii) will allow the required probabilistic construction; since player j prefers Π^i to Π^{ij} he will be deterred from taking pure actions that he prefers (in terms of one-shot payoffs) by the fact that doing so will increase the likelihood of Π^{ij} being initiated. Moreover, in order to selectively affect player j's incentives, the probability that play proceeds to Π^{ij}, say P^{ij}, will be conditioned *directly* only on his (observed) actions (and the states visited during the min–maxing phase). Despite this targeting, there is a further source of trouble; a player $k \neq j$ can manipulate P^{ij}, and therefore his own continuation payoffs, by influencing (in an unobserved manner) the distribution of the state at the end of the min–maxing phase. To deter this we need to ensure that:

Step 4. The strategies Π^i and Π^{ij} can be chosen in such a way that in the component cycles of Π^i and Π^{ij} (of length T^i and T^{ij}) the payoffs of each player is independent of the initial state of the cycle (and continues to satisfy (i)–(iv) of Step 3).

Steps 3 and 4 are then used to complete the proof of Theorem 9 by showing that there exist probabilities P^{ij} under which every player $j \neq i$, at each node of the min–maxing phase, has the same *expected* reward from all actions.

Proof of Step 3. From (FD) and Lemma 6, it follows that there are pure strategies Π^{ij}, $i, j = 1, ..., n$, $i \neq j$ (which are in fact cycles over finite sets of pure Markov strategies) such that their associated long-run average payoffs U^{ij} satisfy: for all i, j, k, $i \neq k$, $i \neq j$,

(a)	strict individual rationality	$U_i^{ij} > 0$	
(b)	asymmetry	$V_k^k < U_k^{ij}$	(A.6)
(c)	differential incentives for j	$U_j^{ij} < V_j^i$	(A.7)
(d)	indifference for i	$V_i^i = U_i^{ij}$	(A.8)

For appropriately high discount factors, Step 3 (iii) (asymmetry) follows straightforwardly from (A.6), whereas Step 3 (ii) (player j has different incentives under Π^i and Π^{ij}) follows from (A.7). Step 3 (i) (player i is indifferent across the strategies) follows from (A.8).[33] Finally, Step 3 (iv) is immediate from (a) (and Proposition 3). Thus Step 3 has been proved. ▮

Proof of Step 4. The proof of Step 4 is somewhat involved and will be accomplished by way of proving two lemmas (Lemmas A.2 and A.3). Let $B_\theta(W)$ denote the θ-neighbourhood of a payoff $W \in R^n$.

LEMMA A.2. *Fix an integer $Q \geqslant 1$ and a set of $(n+1)$ vectors, $W^i \in R^n$, $i = 0, 1, ..., n$ and suppose that* $\dim \operatorname{co}(W^0, ..., W^n) = n$. *Then, for all $\varepsilon > 0$, there is $\theta > 0$ such that for any finite collection of $(n+1)$ vectors, $[W^0(q), W^1(q), ..., W^n(q)]$, $q = 1, ..., Q$ which satisfy the condition that $W^i(q) \in B_\theta(W^i)$, $\forall i, q$, it follows that*

$$B_\varepsilon(W^0) \cap \operatorname{co}[W^0(1), ..., W^n(1)] \cap \operatorname{co}[W^0(Q), ..., W^n(Q)] \neq \varnothing. \quad \text{(A.9)}$$

[33] Strictly speaking it follows by an argument identical to that used in proving Step 2 in the text (and that argument is also implicit in (A.8)).

Proof. If the lemma were untrue, then there would exist an $\varepsilon > 0$ and sequences $[W^0(q; p), ..., W^n(q; p)]$, $q = 1, ..., Q, p \geqslant 0$ (with $\lim_{p \to \infty} W^i(q; p) = W^i$, for all i, q), such that $B_\varepsilon(W^i) \cap \text{co}[W^0(1; p), ..., W^n(1; p)] \cap \cdots \text{co}[W^0(Q; p), ..., W^n(Q; p)] = \varnothing$, for all p. This is impossible given the full dimensionality of $\text{co}(W^0, ..., W^n)$. •

In what follows, Lemma A.2 will be used as follows: the primitive set of vectors will be $W^0, ..., W^n$, where W^i is the long-run average payoff to a cyclic strategy Π^i. The collection of $(n+1)$ vectors will be $W^0(s; T), ..., W^n(s; T)$, where each vector, say $W^i(s; T)$, is a T-period discounted average payoff to Π^i, when the initial state of the cycle is s. As this initial state s varies, we get different T-period average payoff vectors; the number of such state-dependent vectors is exactly the same as the number of states and that number has been denoted Q in the lemma.

Consider first the $(n+1)$ long-run average payoff vectors, $V^i, b^i, V^j, j \neq i$ (recall that b^i is the best long-run average payoff for i and the asymmetric payoffs $V^j, j = 1, ..., n$ are defined in Eq. (5)). Suppose, after invoking (FD), that $\dim \text{co}(V^i, b^i, V^j, j \neq i) = n$. Let ε be defined by the requirement that $V^i_i + \varepsilon < V^j_i - \varepsilon$. By arguments preceding Lemma 11, there is $\delta_1 < 1$ and cycle length T' such that the T'-period discounted averages satisfy: $\| V^i(s; T) - V^i \| < \theta$, $\| b^i(s; T) - b^i \| < \theta$ and $\| V^j(s; T) - V^j \| < \theta$ whenever $T \geqslant T'$ and $\delta \geqslant \delta_1, j \neq i$ and for all initial states s. Evidently, the collection $[V^i(s; T), b^i(s; T), V^j(s; T), j \neq i, s \in S]$ together with the long-run average payoffs $[V^i, b^i, V^j, j \neq i]$ satisfy the hypothesis of Lemma A.2. Hence, it follows that there are probabilities $p^i_j(s), s \in S, j = 1, ..., n$, such that for al s, s' and $k = 1, ..., n$

$$\begin{aligned} p^i_i(s)\, b^i_k(s; T) + \sum_{j \neq i} p^i_j(s)\, V^j_k(s; T) + \left[1 - \sum_j p)^i_j(s) \right] V^i_k(s; T) \\ = p^i_0(s')\, b^i_k(s'; T) + \sum_{j \neq i} p^i_j(s')\, V^j_k(s'; T) + \left[1 - \sum_j p^i_j(s') \right] V^i_k(s'; T). \end{aligned} \tag{A.10}$$

To repeat, there is a public randomisation at the beginning of every T^i cycle, over the strategies yielding $b^i, V^1, ..., V^n$ (as long-run average payoffs), such that *each* player's T^i period discounted average payoffs are independent of the initial state of the cycle. Let this constant payoff vector be denoted $\hat{V}^i(\delta)$. Furthermore, $\| \hat{V}^i(\delta) - V^i \| < \varepsilon$.

Identical arguments apply to the $(n+1)$ long-run average vectors $[U^{ij}, b^i, V^j, j \neq i]$. Let the implied (state-independent) payoff vector, in this case, be denoted $\hat{U}^{ij}(\delta), j \neq i$. The vectors $\hat{V}^i(\delta)$ and $\hat{U}^{ij}(\delta)$ are (almost) the vectors whose existence is asserted by Step 4. They satisfy all the requirements, asserted in Steps 3 (i)–(iv) except for the fact that player i's payoffs may not be the same under $\hat{V}^i(\delta)$ and $\hat{U}^{ij}(\delta)$. To fix that let $B^i(\delta)$ be any state-independent payoff with the property that $B^i_i(\delta) > \max[\hat{V}^i_j(\delta), \hat{U}^{ij}(\delta)]$ (such a payoff can be constructed by arguments similar to those above). Let μ^i and μ^{ij} be convexifications such that a randomisation with B^i leaves player i appropriately indifferent, i.e.,

$$\mu^i \hat{V}^i_i(\delta) + [1-\mu^i]\, B^i_i(\delta) = \mu^{ij} \hat{U}^{ij}_i(\delta) + [1-\mu^{ij}]\, B^i_i(\delta). \tag{A.11}$$

Collecting all of the above arguments we can define the strategy Π^i (retaining notation) as a publicly randomised cyclic strategy in which at the beginning of a cycle the players randomize between the strategies which yield V^i, b^i, V^j, and B^i, using the probabilities defined by (A.10) and (A.11). Likewise we can define a cyclic strategy Π^{ij} as one that randomises appropriately between the strategies that yield U^{ij}, b^i, V^j, and B^i. Hence, I have proved:

LEMMA A.3. *There are publicly randomised cyclic strategies $\Pi^i, \Pi^{ij}, j \neq i$ with associated payoffs, at the beginning of each cycle, $v^i(\delta)$ and $u^{ij}(\delta)$ and there is a cycle length T^i and a discount factor $\delta_1 < 1$ s.t. for $\delta \geqslant \delta_1$, $i, j, k, k \neq i, j \neq i$,*

(a)	*asymmetry*	$v^k_k(\delta) < u^{ij}_k(\delta)$	(A.12)
(b)	*differential incentives for j*	$v^i_j(\delta) > u^{ij}_j(\delta)$	(A.13)
(c)	*indifference for i*	$v^i_j(\delta) = u^{ij}_i(\delta)$	(A.14)

Thus Step 4 has been proved. •

The Construction of Probabilistic Punishments

For simplicity let us normalise the period of player i's deviation to 0. Min–maxing will take place for T_m periods (and for notational simplicity I drop the subscript on T_m). The probability with which play proceeds to Π^{ij}, if history h_T has been observed, will be denoted $P^{ij}(h_T)$ (and for notational simplicity, I will write that as $P^j(h_T)$). In fact, I will construct the probability as a sum of component probabilities $p^j(a_{jt})$; $P^j(h_T) = \sum_{t \leqslant T} p^j(a_{jt})$, $t \leqslant T$. Furthermore, as the notation suggest, p^j will depend only on the action of player j at period t.

Some additional notation is, unfortunately, required. Let $\bar{r}(a_{jt}; h_t)$ denote the conditional expectation of player j's tth period return (where $t \leqslant T$) if all players other than j play the correct mixed actions, in min–maxing player i, while player j plays the pure action a_{jt}. Let $\bar{q}(h_{t+\tau} | a_{jt}$ denote the distribution over histories $h_{t+\tau}$, for $\tau > 0$, if the action at t by player j is a_{jt})

and all other players (and j himself after period t) use the correct min-maxing probabilities. Finally, I will use $R(a_{jt}; h_t)$ to denote the expected returns of player j from period t till the end of the min-maxing phase T, if he plays a_{jt} in period t and according to the correct min-maxing probabilities thereafter; i.e., $R(a_j; h_t) = \bar{r}(a_{jt}; h_t) + \delta \sum h_{t+1}\, \bar{q}(h_{t+1} \mid a_{jt})\, \bar{r}(h_{t+1}) + \cdots \delta^{T-t} \sum h_t \bar{q}(h_T \mid a_{jt})\, \bar{r}(h_T)$ (where $\bar{r}(h_{t+\tau})$ is player j's conditional expected reward in period $t+\tau$ of the min-maxing phase).

In particular, the gain to player j over the periods $t+1, ..., T$ from using the pure action a_j rather than $\bar{a}_j$ is simply $(1-\delta)[R(a_j; h_t) - R(\bar{a}_j; h_t)]$. The probabilities of going to j's less-preferred strategy Π^{ij}, $p^i(a_j)$ and $p^i(\bar{a}_j)$, will be chosen such that player j is indifferent between the two actions: if $(1-\delta)[R(\bar{a}_j; h_t)] > 0$, then we shall ensure that $p^j(a_j) > p^j(\bar{a}_j)$. The existence of such probabilities will be established by way of a backward induction argument.

In fact, suppose for a moment that the probabilities $p^j(a_{j\tau}$, for $\tau > t$ have been determined (and we would like to determine $p^j(a_{jt})$. Let $\phi^j(h_T) = \sum_{\tau > t} p^j(a_{j\tau}$, the probability of play eventually proceeding to Π^{ij} if the sample path during the min-max phase is h_T. Hence, the conditional probability that player j will in fact be penalised at the end of the min-max phase, if he plays action a_j in period t, is $[\sum_{h_T} \phi^j(h_T)\, \bar{q}(h_T \mid a_j; h_t)]$. Thus the lifetime difference in payoffs from using actions a_j and $\bar{a}_j$ is 0 only if

$$
\begin{aligned}
&(1-\delta)[R(a_j; h_t) - R(\bar{a}_j; h_t] \\
&\quad + \delta^{T-t} \sum_{k \neq i} [u_j^{ik}(\delta) - v_j^i(\delta)] \left\{ \sum_{h_T} [\bar{q}(h_T \mid a_j; h_t) - \bar{q}(h_T \mid \bar{a}_j; h_t]\, \phi^k(h_T) \right\} \\
&\quad + \delta^{T-t}[p^j(a_j; h_t) - p^j(\bar{a}_j; h_t)][u_j^{ij}(\delta) - v_j^i(\delta)] = 0. \qquad \text{(A.15)}
\end{aligned}
$$

The equations (A.15), one for every pair of actions a_j and $\bar{a}_j$ and every time-period t, can be solved by backward induction. When $t = T$, the second term in (A.15) drops out to yield the equation $(1-\delta)[R(a_j; h_T) - R(\bar{a}_j; h_T)] + [p^j(a_j; h_T) - p^j(\bar{a}_j; h_T)][u_j^{ij}(\delta) - v_j^i(\delta)] = 0$. There is evidently a solution to this equation for every pair of actions a_j and $\bar{a}_j$ and moreover, as $\delta \uparrow 1$, the solutions $p^i(a_j; h_T)$ and $p^j(\bar{a}_j; h_T)$ go to 0 and hence define probabilities (in that their sum is less than 1)[34]

We can then proceed to period $T-1$ and thus on back to period 0. For sufficiently high δ we do, in fact, create probabilities which also satisfy (A.15). Since player j is indifferent between all of his actions, provided other players continue to min-max i and play proceeds after the min-maxing phase to Π^{ik} or Π^i with these probabilities, player j has a best response which is to min-max player i with the correct probabilities.

[34] The residual probability is attached to the option "go to the strategy Π^{i}"; this probability goes to 1, as δ goes to 1.

The arguments that remain to show that the grand strategy Π^* is a subgame effect equilibrium are identical to the observable mixed strategy case. The proof of Theorem 9 is complete. ∎

Remark. The idea of constructing probabilities punishments to deter unobserved deviations is also used in Abreu, *et al.* [2] in their study of folk theorems for repeated games. Since there is no state variable in those games, Steps 3 and 4 of the proof above are not required. Hence, the full dimensionality condition, which is only used in these two steps, is dispensable and the corresponding result can be proved with only (PA).

Proof of Lemma 12. It is not difficult to see that a consequence of the definition is the (ostensibly) stronger condition: there is a (possibly mixed) Markov strategy $\tilde{\Pi}$ s.t. for all (s, s') there is N s.t. $q^N(s, s') > 0$, i.e., that we have a stationary Markov chain. Since the number of states is finite, by standard results they are all persistent. Let $v(s')$ be a feasible long-run average payoff from initial state s'. By Lemma 2, it is realised by ex ante randomization over Markov strategies. Starting from $s \neq s'$, a strategy, which follows $\tilde{\Pi}$ until the first time s' is reached and then follows the Markov strategies that generate $v(s')$, clearly generates the same long-run average payoff. ∎

REFERENCES

1. D. Abreu, On the theory of infinitely repeated games with discounting, *Econometrica* **56** (1988), 383–396.
2. D. Abreu, P. K. Dutta, and L. Smith, The folk theorem for repeated games: A NEU condition, *Econometrica* **62** (1994), 939–948.
3. R. Aumann and L. Shapley, Long term competition: A game theoretic analysis, mimeo, Hebrew University, 1976.
4. J. Benhabib and R. Radner, Joint exploitation of a productive asset, *Econ. Theory* **2** (1991), 165–191.
5. B. D. Bernheim and D. Ray, Economic growth with intergenerational altruism, *Rev. Econ. Stud.* **54** (1987), 227–242.
6. T. Bewley and E. Kohlberg, The asymptotic theory of stochastic games, *Math. Operations Res.* **1** (1976), 197–208.
7. D. Blackwell, Discounted dynamic programming, *Ann. Math. Statist.* **36** (1965), 226–235.
8. P. K. Dutta, What do discounted optima converge to? A theory of discount rate asymptotics in economic models, *J. Econ. Theory* **55** (1991), 64–94.
9. P. K. Dutta, Collusion, discounting and dynamic games, University of Rochester Working Paper 272; forthcoming, *Journal of Economic Theory*.
10. P. K. Dutta and A. Madhavan, Information aggregation and strategic trading in speculative markets, University of Rochester Working Paper 339, 1992.
11. P. K. Dutta and R. Sundaram, The tragedy of the commons?, *Econ. Theory* **3** (1993), 413–432.
12. J. Friedman, A modified folk theorem for time-dependent supergames, Center for Interdisciplinary Studies, University of Bielfield, Working Paper 2, 1987.

13. D. Fudenberg and E. Maskin, The folk theorem in repeated games with discounting of incomplete information, *Econometrica* **54** (1986), 533–554.
14. D. Fudenberg and E. Maskin, On the dispensability of public randomisation in discounted repeated games, *J. Econ. Theory* **53** (1991), 428–431.
15. D. Gillette, Stochastic games with zero-stop probabilities, *in* "Contributions to the Theory Games," 3 (Annals of Mathematical Studies 39), pp. 179–187, Princeton Univ. Press, 1957.
16. C. J. Himmelberg, T. Partharasarathy, T. E. S. Raghavan, and F. S. Van Vleck, Existence of p-equilibrium and optimal stationary strategies in stochastic games, *Trans. Amer. Math. Soc.* **60** (1976) 245–261.
17. C. Holden and A. Subrahmanyam, Long-lived private information and competition, *J. Finance* **26** (1992).
18. B. Lockwood, The folk theorem in stochastic games with and without discounting, mimeo, University of London, 1990.
19. E. Maskin and J. Tirole, A theory of dynamic oligopoly II: Price competition, kinked demand curves and edgeworth cycles, *Econometrica* **56** (1988), 571–601.
20. J.-F. Mertens and A. Neyman, Stochastic games, *Int. J. Game Theory* **10** (1981), 53–66.
21. J.-F. Mertens and T. Partharasarathy, Stochastic games, CORE Working Paper 8750, Louvain-la-Neuve, 1987.
22. A. S. Nowak, Existence of equilibrium stationary strategies in discounted non-cooperative stochastic games with uncountable state space, *J. Opt. Theory Appl.* **45** (1985), 561–602.
23. T. Partharasarathy, Discounted, positive and non-cooperative stochastic games, *Int. J. Game Theory* **2** (1973), 25–37.
24. T. Partharasarathy and S. Sinha, Existence of stationary equilibrium strategies in non-zero sum discounted stochastic games with uncountable state space and state independent transitions, *Int. J. Game Theory* **19** (1990).
25. R. T. Rockefeller, "Convex Analysis," Princeton Univ. Press, Princeton, NJ, 1970.
26. A. Rubinstein, Equilibrium in supergames with the overtaking criterion, *J. Econ. Theory* **31** (1979), 227–250.
27. R. Sundaram, Perfect equilibrium in a class of symmetric dynamic games, *J. Econ. Theory* **47** (1989), 153–177.