A Folk Theorem for Stochastic Games

Dutta, Prajit

Working Paper No. 293
August 1991

A Folk Theorem for Stochastic Games

Prajit K. Dutta

Rochester Center for Economic Research
Working Paper No. 293

August 1991

A Folk Theorem for Stochastic Games*

Prajit K. Dutta*
Columbia University

and

The Rochester Center for Economic Research
Working Paper No. 293

November1990
Revised July 1991

Abstract: In many dynamic economic applications, the appropriate game theoretic structure is that of a stochastic game. We present a folk theorem for such games. The result subsumes a variety 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.

* 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. A preliminary version of this paper was presented at the International Game Theory Conference, July 1989, at Ohio State University. I thank the seminar audience at that conference as well as that of the theory seminar at Columbia University for helpful comments. I have especially benefitted from the comments of Drew Fudenberg and Abraham Neyman. All errors of course remain my responsibility. This paper was completed while I was visiting the University of Rochester, and I thank them for the logistical and other support.

A Folk Theorem for Stochastic Games*

Prajit K. Dutta*
Columbia University

and

The Rochester Center for Economic Research
Working Paper No. 293

November1990
Revised July 1991

Abstract: In many dynamic economic applications, the appropriate game theoretic structure is that of a stochastic game. We present a folk theorem for such games. The result subsumes a variety 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.

* 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. A preliminary version of this paper was presented at the International Game Theory Conference, July 1989, at Ohio State University. I thank the seminar audience at that conference as well as that of the theory seminar at Columbia University for helpful comments. I have especially benefitted from the comments of Drew Fudenberg and Abraham Neyman. All errors of course remain my responsibility. This paper was completed while I was visiting the University of Rochester, and I thank them for the logistical and other support.

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 (1976), Rubinstein (1979), Abreu (1988) and Fudenberg and Maskin (1986) 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 premissed 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 human and natural resources are the relevant productive assets[1], financial models and 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 typically 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, players' 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 focussed on the issue of existence of (perfect) equilibria in Markovian strategies (see, for example, Parthasarathy (1973), Himmelberg et al (1976), Nowak (1985) and Parthasarathy and Sinha (1990)).[3] This paper provides instead a characterization of equilibrium payoffs when players are very patient, dropping the assumption of Markovian behavior. The latter restriction appears to be arbitrary; indeed in the strictly repeated context, it is seldom suggested that Markovian behavior is strategically salient (see however Maskin and Tirole (1988)). Recently the folk theorem question in non-repeated settings has been also investigated by Friedman (1987) and Lockwood (1990). Their results are discussed in Section 7.

[1]Strategic formulations include Benhabib and Radner (1988), Stokey (1990), Bernheim and Ray (1986) and Sundaram (1989).
[2]For instance, see Maskin and Tirole (1988) and Dutta and Madhavan (1991).
[3]Mertens and Parthasarathy (1988) have shown the existence of perfect equilibria in a more general class of strategies.

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 has, at least asymptotically, an easy resolution. 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 one. The folk theorems for repeated games with perfect monitoring may be extended to this setting. These include the theorems of Aumann and Shapley (1976), Rubinstein (1979) and Fudenberg and Maskin (1986). We provide an analog of the last result, which is suitable for the applications we discuss. The argument also yields a modest generalization of the Fudenberg and Maskin discounted folk theorem for strictly repeated games.

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 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 $< S, A_i, r_i, q, \delta;\ i=1,..n >$ where i is the player index, S is the set of states and A_i is the i-th player's set of actions. The sets S, A_i, i=1,..n are finite. Assume, 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 i-th player's one-period reward is $r_i{:}S\times A \to \mathbb{R}$. It associates with every vector of players' 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 vector a, then the game moves to state s' next period with probability $q(s' \mid s,a)$. Further, $\delta \leq 1$ is the

---

[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 Partharasarathy (1973).

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.

A behavior strategy for player i is denoted Π_i. It is a sequence $\Pi_{i0}, \Pi_{i1},..\Pi_{it}...$ where Π_{it} selects a distribution, at 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 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. we analyze a game with unobservable (private) mixed strategies. However players can coordinate on a public randomization device, i.e. players are allowed to randomize publicly.

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 trivially 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).

The folowing 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 in the discussion we will talk of a "punishment regime" for player i during which regime player j's action will be denoted a_j^i. Finally, $\| \cdot \|$ will denote any one of the equivalent norms in $\mathbb{R}^n$.

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 $\Pi = \Pi_1....\Pi_n$ and initial state $s_0 = s$. The discounted average (expected) returns to player i if the initial state is s, the players employ strategy Π and the discount factor is $\delta < 1$ is:

$$W_i(s ; \Pi,\delta) = (1-\delta) \sum_0^\infty \delta^t r_i(t ; \Pi,s) \tag{1}$$

[5] In the literature such strategies have sometimes been called stationary (for example, see Bewley and Kohlberg (1976), Himmelberg et. al. (1976), Partharasarathy (1973)) whereas more recent usage has called them Markov (for example, Maskin and Tirole (1988)). We adopt the latter convention.

The long-run average expected returns for the same setting is:

$$W_i\,(s\,;\,\Pi) = \liminf_{T\to\infty} \frac{1}{T}\sum_{0}^{T-1} r_i\,(t\,;\,\Pi,s) \qquad (2)$$

For a given initial state s, a strategy choice is a Nash equilibrium if no player profits from unilateral deviation, i.e. $W_i\,(s\,;\,\Pi_i, \Pi_{-i}, \delta) \geq W_i\,(s\,;\,\Pi_i', \Pi_{-i}, \delta)$ for all Π_i' and all i (and similarly for the long-run average). A (subgame) perfect equilibrium is a strategy choice such that after every history, the strategy continuations constitute a Nash equilibrium.

The min-max level[6] of player i, for initial state s and discount factor δ (respectively, long-run average) will be denoted $\underline{v}_i^i(s,\delta)$ (respectively $\underline{v}_i^i(s)$) and is given by

$$\underline{v}_i^i(s,\delta) = \inf_{\Pi_{-i}} \sup_{\Pi_i} W_i\,(s;\, \Pi_i,\Pi_{-i}\,,\delta) \qquad (3)$$

$\underline{v}_i^i(s)$ is defined similarly. 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 to analyze is the set of average (discounted or long-run) expected returns in the infinite horizon game. Unlike a repeated game there is no unique set of (average) payoffs which can be achieved at every discount factor and from every initial state. Nor can all feasible payoffs be achieved by convexifying over strategies which involve the infinite repetition of a constant action (with associated constant immediate rewards). In this section we investigate two issues: the relation between the feasible payoff sets for different discount factors and the existence of a set of simple strategies which would realize all feasible payoffs.

3.1 Pure, Markov Strategies Suffice

Let F(s,δ) (respectively F(s)) denote the set of feasible discounted (respectively long-run) average expected returns, i.e.

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

$$F(s,\delta) = \{w \in \mathbb{R}^n: \exists\, \Pi \text{ s.t. } w_i = W_i(s; \Pi,\delta),\ i=1,..n\} \qquad (4)$$

(and similarly F(s)). (Note that Π may be a correlated strategy). 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 $\Sigma_i \lambda_i w_i$ where $w \in F(s,\delta)$ and $\lambda_i \in R$. From Blackwell (1965) it then follows that, in the discounted stochastic game, the extreme points of $F(s,\delta)$ are generated by pure Markov strategies. Since public randomization is admissible, 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 i) $F(s,\delta) = \text{co}\ \phi(s,\delta)$, $\forall\ s \in S,\ \delta < 1$

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

Proof: In the appendix. •

The lemma simplifies the analysis in the sequel considerably.[7] The restriction, without loss of generality, to pure (publicly randomized) strategies will make the detection of deviation from such strategies immediate. Further, this result allows 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

For any two closed sets B and C in $\mathbb{R}^n$, define the Hausdorff distance as

$$d(B,C) = \max\ (\sup_{x \in B} \rho(x,C),\ \sup_{y \in C} \rho(y,B))$$

where $\rho(x,C) = \inf \gamma(x,z),\ z \in C$ and γ is any metric on $\mathbb{R}^n$.

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), we have

[7] It will be used to show that time randomized pure Markov strategies in fact are sufficient in an even stronger sense (see Lemma 8).

Lemma 2 $F(s,\delta) \to F(s)$, as $\delta \to 1$, for every $s \in S$; the convergence is to be understood to be in the Hausdorff metric.

Proof: In the appendix. •

4. Individual Rationality

Unlike repeated games, 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? A natural benchmark is the limit of the (state-dependent) discounted average min-max payoffs, as the discount factor goes to one. It follows from results of Bewley and Kohlberg (1976) and Mertens and Neyman (1981) that this limit 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 (1976, Theorem 3.1) show that $\lim \underline{v}_i^i(s,\delta)$ (as $\delta \uparrow 1$) exists for all i and s in S. For this same class of games Mertens and Neyman (1981) then showed that this limit is in fact the long-run average min-max. If we think of player i and the group of players -i as constituting a "two person" game, the Mertens-Neyman theorem yields[8]:

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 $\infty \geq T \geq N$ and every strategy Π_i,

$$W_i(s; \Pi_i, \Pi_{-i}^*, T) \leq \lim_{\delta \uparrow 1} \underline{v}_i^i(s,\delta) + \eta$$

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).

$\underline{v}_i^i(s,\delta)$ is both the min-max and max-min level of player i, by a result of Parthasarathy (1973). From Proposition 3 it is clear that $\lim \underline{v}_i^i(s,\delta)$ is the long-run average

[8] 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 the "ability to act as one". In particular they may not have access to (n-1) player randomization. However, the Mertens-Neyman result is valid in this context as well.

min-max and max-min level for player i (henceforth $\underline{v}_i^i(s)$). This will be the relevant security level of player i in the analysis that follows.

4.2 Individually Rational Payoffs

In this subsection we define two alternative notions of individual rationality. Fix a discount factor $\delta \leq 1$ and consider any initial state s. We 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 $w(s,\delta) \geq \underline{v}_i^i(s,\delta)$ for all i (respectively $w(s) \geq \underline{v}_i^i(s)$ for all i). Let $F^*(s,\delta)$ (respectively $F^*(s)$) denote the discounted (respectively long-run) average strictly individually rational (ex-ante sense) payoff sets, i.e.

$$F^*(s,\delta) = \{ w \in F(s,\delta): w_i > \underline{v}_i^i(s,\delta),\ i=1,..n\}$$

(and similarly $F^*(s)$). It should be remembered however that feasible payoffs are not in general generated by (convexification over) constant action strategies. Hence, we will say that a payoff vector $w(s,\delta)$ is individually rational in the ex-post sense if it is generated by a strategy Π such that all of its continuation payoffs are individually rational in the ex-ante sense after all histories. We 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 For all $\varepsilon > 0$, there is $\underline{\delta} < 1$, s.t. for $\delta \geq \underline{\delta}$,

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

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 individually rational we also have for the set of equilibrium payoffs, $V(s,\delta)$, the following corollary:

Corollary 5 For all $\varepsilon > 0$, there is $\underline{\delta} < 1$ and ε-neighborhoods of $F^*(s)$, say $B(\varepsilon,F^*(s))$, s.t. for $\delta \geq \underline{\delta}$,

$$B(\varepsilon, F^*(s)) \supset V(s,\delta) \quad \forall s.$$

Corollary 5 implies that the folk theorem proved in the sequel provides a complete characterization.

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: Firstly, asymptotic state independence:

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

(A2) The long-run average min-max $\underline{v}_i^i(s)$ is independent of s, for all i, say $\underline{v}_i^i(s) = \underline{v}_i^i$

Secondly, we will make one of the following assumptions. Denote by F_{ij} the projection of F on the i-j axes.

(PA) Payoff Asymmetry There is $\bar{v}^i \in F, i=1,..n$ s.t. $\bar{v}_i^i < \bar{v}_i^j$, $\forall i, j, i \neq j$

(PF) Pairwise Full-dimensionality For all i,j, $i \neq j$, $\dim(F_{ij}) = 2$.

(FD) Full Dimensionality $\dim(F) = n$.

The three conditions are obviously related intimately. Full dimensionality (FD) clearly implies the other two. Pairwise full dimensionality (PF) can, in turn, be shown to imply payoff asymmetry (PA).[9] The main theorem below will be proved under (FD) but for some interesting special cases, the weaker condition (PF) or even the weakest, (PA), will be seen to suffice.

The assumptions above 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. We briefly discuss these assumptions and their implications now and return in Section 7 to a fuller discussion of primitive models in which the conditions are satisfied.

---

[9]The condition (PF) has been used in independent work by Smith (1990). His focus was on generalizations of folk theorems in the purely repeated game. He has also proved that (PF) implies (PA). See Section 7 for a contextual discussion.

5.1 Asymptotic State Independence

Future feasible and equilibrium payoffs in a stochastic game depend on the current state. Hence, for a folk theorem to hold, there must be some similarity in the possibilities from different states.[10] The issue then is how restrictive must these conditions be? (A1) is a mild requirement as Section 7 will make clear. (A2) is stronger and we defer to Section 7 a discussion of primitive conditions on the game which guarantee this. If one or the other of these assumptions is not satisfied, our 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 randomization over pure Markov strategies realizes 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. However, in the presence of (A1), a one-shot randomization can be replicated by a scheme of time-averaging or cycling which moves repeatedly between different pure Markov strategies in a manner consistent with the one-shot convexification and in such a way that the continuation payoffs are approximately the same after all histories.

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 says:

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.

Given the continuity of the min-max (Proposition 3) and feasible payoffs (Lemma 2) from the construction of the proof of Lemma 6 it also follows that

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 $\delta \geq \underline{\delta}$ 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 $\delta \geq \underline{\delta}$ and all s.

[10]If no conditions are placed it is easy to construct counter-examples to the full folk theorem. See Lockwood (1990) for some illuminating examples.

5.2 Payoff Asymmetry, Pairwise Full Dimensionality and Full Dimensionality

These conditions are necessitated of course by a counter-example in Fudenberg and Maskin (1986) which shows that if there is perfect congruence of interests among the players, punishments to deter deviations from individually rational paths may not be credible. 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. The condition may be interesting not so much because it is weaker than full dimensionality[11] but because it can be shown that within the class of strategies analyzed in this paper and Fudenberg and Maskin (1986), it is additionally almost a necessary condition for the folk theorem.[12] If mixed strategies are unobservable, we will need to strengthen (PA) to the (pairwise) full dimensionality assumptions (PF) or (FD). As will become clearer in the sequel, these are possibly not the weakest assumptions that will work although we have not been able to find an alternative one nor show that payoff asymmetry by itself suffices.

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 $\delta \geq \underline{\delta}$, there is a perfect equilibrium whose payoff $v(s,\delta)$ satisfies $\|v(s, \delta) - w\| < \varepsilon$ for all s.**

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):

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 $w_i > \underline{v}_i^i(s)$ $\forall$ i, s.**

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

[11] 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.

[12] The necessary condition allows weak inequalities in (PA), with an additional restriction in case of an equality.

Corollary 9.2 **Suppose that (FD) is satisfied by F. Then, the conclusions of Theorem 9 hold for any average payoff $w \in F$, such that $w_i > \underline{v}_i^i(s)$ $\forall$ i,s.**

The full dimensionality condition can be weakened in some interesting special cases. We report the analog of Theorem 9 in each case and note at this point that the analogs of Corollaries 9.1 and 9.2 also hold.

Observable Mixed Strategies

Many authors have analyzed the simpler model in which mixed strategies are observable (see Section 7 for a discussion). In that context we show:

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

Dynamic and Repeated Games

Suppose that mixed strategies are unobservable but we confine attention to games with deterministic transitions. We then show[13]:

Proposition 9.4 **Suppose that (A1), (A2) and (PF) hold in a dynamic game. Then, the conclusions of Theorem 9 are valid for any $w \in F^*$.**

Since a repeated game is a simple example of a dynamic game, an immediate corollary is[14]

Corollary 9.5 **In a repeated game, (PF) 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.**

The principal reason that folk theorem analysis is more difficult for stochastic games is that deviation yields one-shot gains and in addition allows a player to manipulate the distribution of the state next period. This incentive can be asymptotically deterred globally, if (A1)-(A2) hold (Theorem 9, Propositions 9.3-9.4), or locally if one or the other assumption does not hold (Corollaries 9.1 and 9.2). The assumptions (A1), (A2)

[13]Actually (PF) is sufficient for a class of stochastic games larger than just the dynamic ones (see the remark following the proof of Proposition 9.4).
[14]Since there is no distinction between ex-ante and ex-post payoffs in a repeated game, the payoff w can be realized exactly.

and (FD) (or their local versions) will allow a logic of proof that is similar to the purely repeated case except for two additional sets of arguments necessitated by the state manipulation possibility.[15] The arguments will refer, respectively, to observable and unobservable deviations and will be discussed in the course of an informal presentation of the proof (steps 2 and 4-5 below).

6.1 An Informal Discussion of the Proofs

Step 1: There exist (n+1) pure cyclic strategies $\tilde{\Pi}$, Π^i, i=1,..n (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), such that for sufficiently high discount factors the associated payoffs [w(s,δ), $V^i(s,\delta)$, i=1,..n] are i) asymmetric uniformly across states ($V_i^i(s,\delta) < V_i^j(s',\delta)$ for all s,s',i,j), ii) strictly individually rational and iii) $\tilde{\Pi}$-dominated ($\underline{v}_i^i(s,\delta) < V_i^i(s',\delta) < w(s'',\delta)$, $\forall$ s, s', s"). Further, $\|w(s,\delta) - w\| < \varepsilon, \forall$ s.

Remark Π^i is going to be 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)). Step 1 addresses these simple state-manipulation issues.

Since the game is non-repeated, ex-post continuation payoffs are history-dependent and will typically differ from the ex-ante payoffs from an initial state. Consider the play of i's worst strategy Π^i. Suppose history h_t is such that player i's continuation payoffs $v_i^i(h_t,\delta)$ are less than $v_i^i(s,\delta)$. Deviation, which involves a finite min-max period followed by $E[v_i^i(s,\delta)]$ may then be profitable.[16] Step 2 deals with this ex-post incentives problem.

Step 2: Let g^i be the pure Markov strategy that maximizes player i's long-run average payoffs. Also, denote $T^i = \Sigma_p T_p^i$. The strategy Π^i is modified as follows: at the beginning of each T^i cycle, play proceeds to (g_p^i, p=1,..P) with probability $\mu^i(s)$ and to g^i with the

[15]Of course, even under (A1)-(A2) it will still be the case that the discounted feasible payoff set F(s,δ) and the discounted min-max $\underline{v}_j^i(s,\delta)$, will be state dependent. From the proof it will be clear that in order to deter deviations 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 the appropriate rates.

[16]In the repeated game even in the absence of public randomization 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 (almost surely) greater at t+1 than at t.

remaining probability. The probabilities are conditioned on the state at the beginning of the cycle and are chosen in a way such that player i's payoffs over the cycle are independent of this initial state. We retain notation and call this strategy Π^i as well.

Step 3: Consider the following strategy: play $\tilde{\Pi}$ till such time as player i deviates.[17] Then switch to "i's punishment regime": players -i play the strategy that min-maxes i, in the long-run average sense, for an appropriate number (T_m) of periods followed by a move to Π^i. Player i plays a best response over the min-max period and his component of Π^i thereafter. Upon observable deviation by any player j, in the course of i's punishment regime, start j's punishment regime. Call this strategy Π^*.

Steps 1-3 suffice to prove the folk theorem if mixed strategies are observable (i.e. Proposition 9.4 will have been proved at this point). However, if mixed strategies are unobservable, players -i may deviate unobserved while min-maxing i. To prevent this, we construct differential continuation payoffs and probabilistic punishments for each player $j \neq i$ (after the min-max phase) in such a way that he is indifferent, in expected terms, between all of his actions <u>during</u> the min-max phase.

Step 4: There exist pure cyclic strategies Π^{ij} $\forall$ $i \neq j$ (with associated payoffs $U^{ij}(s,\delta)$) which have the following properties for sufficiently high discount factors: $U^{ij}_i(s,\delta) = V^i_i(s',\delta)$, at the beginning of each T^i cycle (indifference for i across states and strategies Π^{ij}, Π^i), $\| U^{ij}_j(s,\delta) - V^i_j(s',\delta) \| > 0$ (differential payoffs for j but not necessarily for $k \neq j$), and $U^{ij}_j(s,\delta) > V^j_j(s',\delta)$ (asymmetry); in each case $\forall$ s,s'.

Consider now the following modification of the strategy Π^* (retain notation). After min-maxing i for T_m periods, play proceeds to Π^{ij} with probability $P^{ij}(h_T)$ and to Π^i with probability $1 - \Sigma_j P^{ij}(h_T)$. In order to selectively affect player j's incentives, P^{ij} is conditioned <u>directly</u> only on his (observed) actions and the states during the min-maxing phase. A problem remains however that player $k \neq j$ can manipulate P^{ij} and his own continuation payoffs by the influence his actions have on the distribution of the state after the min-maxing phase. To deter this we need

Step 5: The strategies Π^i and Π^{ij} can be chosen in such a way that in the component cycles (of length T^i and T^{ij}) the payoffs of each player is independent of the initial state of

[17] Since the strategy $\tilde{\Pi}$ invoves only pure actions such a deviation is observable.

the cycle. Then, probabilities P^{ij} exist under which every player j, at each node of the min-maxing phase, has the same expected reward from all actions.[18]

Remark For repeated games, Fudenberg and Maskin (1986) are able to construct continuation payoffs such that players doing the punishing are indifferent between their actions since every sample path during the min-maxing phase has the same lifetime reward. Even with a full dimensionality assumption such constructions do not seem possible for stochastic games.

Steps 1-5 are then used to show that appropriate values can be chosen for T_m and T^i_p such that the strategy Π^* is in fact a subgame perfect equilibrium.

6.2 The Details of the Proof

Let $\underline{w}^i$ (respectively M_i) denote the worst (respectively best) long-run average payoff to player i in the game, i.e. $\underline{w}^i_i \equiv \min \{v_i: (v_{-i}, v_i) \in F\}$ (respectively $M^i_i \equiv \max \{v_i: (v_{-i}, v_i) \in F\}$). 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 $\underline{v}^i_i \equiv 0$. Pick convexification weights $\beta_1 > 0$, $\beta_2 > 0$, and define

$$V^i = \beta_1 \underline{w}^i + \beta_2 \bar{v}^i + (1-\beta_1-\beta_2)\, w \qquad (5)$$

Clearly one can pick the convexification weights to prove

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

a) strict individual rationality	$V^i >> 0,$	
b) asymmetry	$V^i_i < V^j_i$	$i \neq j$
c) target payoff domination	$V^i_i < w_i$	

From Lemma 1 and the proof of Lemma 6 it 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_1, g^i_2,....g^i_p, played succesively for T^i_1, T^i_2,....T^i_p periods and then repeated infinitely many times. Of course, the ratio $T^i_p/\Sigma_p T^i_p$ reflects the convexification weights induced by (5) and the bigger is T^i_p, the closer the approximation. From the continuity of payoffs to pure cyclic strategies at $\delta = 1$, it follows from Lemma 10 that Step 1 holds.

[18]For the dynamic game, a variation of Step 5 will be employed.

For any Markov strategy g, let W(s;T) denote the T-period discounted average for initial state s, i.e. $W_i(s;T) = [(1-\delta)/(1-\delta^T)] \sum_0^{T-1} \delta^t r_i(t; g,s)$. Let T' be a cycle length such that for all pure Markov strategies g, $\|W(s;T) - W(s;g)\| < \varepsilon$ for all $T \geq T'$, $\delta \geq \delta_1$ and s, where W(s;g) is of course the long-run average from initial state s, under the strategy g. Denote similarly the payoffs over the T^i cycle as $V^i(s;T^i)$.

<u>Lemma 11</u> There are probabilities $\mu^i(s)$, i=1,..n and $s \in S$, such that for all $\delta \geq \delta_1$,s,s'

$$\mu^i(s)V_i^i(s;T^i) + [1-\mu^i(s)]M_i^i(s;T^i) = \mu^i(s')V_i^i(s';T^i) + [1-\mu^i(s')]M_i^i(s';T^i) \quad (6)$$

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

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

Proof: Pick any $\varepsilon > 0$ with the property that $V_i^i + \varepsilon < V_i^j - \varepsilon$. Take $T_p^i > T'$. Hence, we have $\|V^i(s;T^i) - V^i\| < \varepsilon$, for all s, i, $\delta \geq \delta_1$ or equivalently, $\max_s V_i^i(s;T^i) < \min_s M_i^i(s;T^i)$. So we can find probabilities $\mu^i(s)$ as defined in (6), with in fact the added property that $\max_s V_i^i(s;T^i) = \mu^i(s)V_i^i(s;T^i) + [1-\mu^i(s)]M_i^i(s;T^i)$, for all s. As ε goes to zero, $\mu^i(s)$ clearly goes to one. For sufficiently small ε, (7) holds. The lemma folows.•

For future reference, let $v^i(s;\delta)$ denote the infinite horizon discounted average payoffs to the strategy Π^i (with public randomization 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. $v_i^i(s;T^i) = v_i^i(s';T^i) = v_i^i(\delta)$. Define the strategy Π^* as detailed in Step 3 in Section 6.1. It is clear that we have completed Steps 1-3 in the proof.

We now show that if mixed strategies are observable, then Proposition 9.3 follows by the constructions above. Let the best (respectively the worst) one-shot payoff of player i be denoted b_i (respectively $\underline{m}_i$). Pick $\eta < V_i^i - \varepsilon$. From Proposition 3 it follows that there is $\delta_2 < 1$ and T' s.t. upon min-maxing for at least T' periods, i's T'-period discounted average payoffs can be held below η. Let $T_m \geq T'$ satisfy for $\delta \geq \delta_3 \geq \max(\delta_1, \delta_2)$,

$$(1-\delta^{T^i})\underline{m}_i + \delta^{T^i} v_i^i(\delta) > (1-\delta)b_i + (1-\delta^{T_m})\eta + \delta^{T_m} v_i^i(\delta) \quad (8)$$

(8) can clearly be satisfied by choosing T_m to be large relative to T^i. (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. 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 given Step 1iii).

Unobservable Mixed Strategies:

From (PF) and Lemma 6, it follows that there are pure strategies Π^{ij}, i,j =1,..n, $i \neq j$ (which are in fact time randomizations 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 $\qquad U^{ij}_i > \underline{v}^i_i$

b) asymmetry $\qquad V^k_k < U^{ij}_k \qquad$ (9)

c) differential incentives for j $\qquad U^{ij}_j > V^i_j \qquad$ (10)

d) indifference for i $\qquad V^i_i = U^{ij}_i \qquad$ (11)

The following lemma will be used repeatedly in what follows. Let $B_\theta(W)$ denote the θ-neighborhood of $W \in R^n$.

Lemma 12 Suppose that $W^i \in R^n$, i=0,1,..n and further that dim $co(W^0,...W^n) = n$. 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 satisfying $W^i(q) \in B_\theta(W^i)$, $\forall$ i, q, it follows that

$$B_\varepsilon(W^i) \bigcap co[W^0(1),..W^n(1)] \ldots \bigcap co[W^0(Q),..W^n(Q)] \neq \emptyset \qquad (12)$$

Proof: A contradiction to the claim implies the existence of $\varepsilon > 0$ and sequences $[W^0(q;p),..W^n(q;p)]$, q=1,..Q, $p \geq 0$ (with $\lim_{p\to\infty} W^i(q;p) = W^i$, for all i, q), such that $B_\varepsilon(W^i) \bigcap co[W^0(1;p),..W^n(1;p)] \bigcap ....co[W^0(Q;p),..W^n(Q;p)] = \emptyset$, for all p. This is impossible given the full dimensionality of $co(W^0,...W^n)$.•

Consider then the vectors, M^i, $V^1,...V^n$ and suppose without loss of generality that dim $co(M^i, V^1,...V^n) = n$. Let ε be defined again by the requirement that $V^i_i + \varepsilon < V^j_i - \varepsilon$. By the arguments preceding Lemma 11, there is $\delta_1 < 1$ and cycle length T' such that $\|V^i(s;T) - V^i\| < \theta$, $\|M^i(s;T) - M^i\| < \theta$, whenever $T \geq T'$ and $\delta \geq \delta_1$. Fix an index i. Then, for every s, $[M^i(s;T), V^1(s;T),....V^n(s;T)]$ is a (n+1) set of vectors each drawn from the θ-neighborhood of M^i, $V^1,...V^n$ respectively. There is one such set for each s. By

Lemma 12 it then follows that there are probabilities $\rho_j^i(s)$, $s \in S$, $j=0,..n$, such that for all s,s' and k=1,..n

$$\rho_0^i(s)M_k^i(s;T) + \Sigma_{j\neq 0,i}\, \rho_j^i(s)V_k^j(s;T) + [1-\Sigma_j\rho_j^i(s)]\, V_k^i(s;T) =$$

$$\rho_0^i(s')M_k^i(s';T) + \Sigma_{j\neq 0,i}\, \rho_j^i(s')V_k^j(s';T) + [1-\Sigma_j\rho_j^i(s')]\, V_k^i(s';T) \qquad (13)$$

In other words, there is a public randomization at the beginning of every T^i cycle over the strategies yielding M^i, $V^1,..V^n$ (as long-run average payoffs), such that <u>each</u> player's T^i period discounted average payoffs are independent of the initial state of the cycle. Let these constant payoffs be denoted $\hat{V}_j^i(\delta)$, i,j =1,..n.

Clearly identical arguments can be applied to the long-run average vectors M^i, $V^1,..U^{ij},..V^n$. Let the implied (state-independent) payoffs be denoted $\hat{U}_k^{ij}(\delta)$, i,j,k=1,..n, $i \neq j$. Finally, let $m^i(\delta)$ be state-independent payoffs with the property that $m_i^i(\delta) > \max[\hat{V}_i^i(\delta), \hat{U}_i^{ij}(\delta)]$ (such a payoff can be constructed by starting with an asymmetric payoff V'^i such that $V'^i_i > V_i^i$). Exactly as in (6) let μ^i and μ^{ij} be convexifications such that

$$\mu^i\, \hat{V}_i^i(\delta) + [1-\mu^i]\, m_i^i(\delta) = \mu^{ij}\, \hat{U}_i^{ij}(\delta) + [1-\mu^{ij}]\, m_i^i(\delta) \qquad (14)$$

Denote the strategy that succesively plays $g_1^i,....g_P^i$ over T^i periods, $\hat{\Pi}^i$ (respectively that which generates U^{ij}, $\hat{\Pi}^{ij}$). Collecting all of the above arguments we have

<u>Lemma 13</u> Define the strategy Π^i as public randomization between $\hat{\Pi}^1,..\hat{\Pi}^i,..\hat{\Pi}^n$ and g^i every T^i periods, using the probabilities defined by (13)-(14) (respectively Π^{ij} as public randomization between $\hat{\Pi}^1,..\hat{\Pi}^{ij},..\hat{\Pi}^n$ and g^i) and denote the associated payoffs $v^i(\delta)$ and $u^{ij}(\delta)$. Then, there is a cycle length T^i and $\delta_1 < 1$ s.t. for $\delta \geq \delta_1$, i,j,k, $i \neq k$, $i \neq j$

a) asymmetry $\qquad v_k^k(\delta) < u_k^{ij}(\delta) \qquad (15)$

b) differential incentives for j $\qquad v_j^i(\delta) < u_j^{ij}(\delta) \qquad (16)$

c) indifference for i $\qquad v_i^i(\delta) = u_i^{ij}(\delta) \qquad (17)$

Some additional notation is required before we discuss the probabilistic construction of Steps 4 and 5. Let $\theta^i(a_{-j};h_t)$ (respectively $\theta^i(a;h_t)$) denote the probability with which, after history h_t, -j play the vector a_{-j} (respectively all players play a) while (long-run average) min-maxing player i. Denote the one-period conditional expected reward of player j in playing a_j, while players -j min-max i and i plays a best-response, by $\bar{r}_j(a_j;h_t)$; i.e. $\bar{r}_j(a_j;h_t) = \Sigma_{a_{-j}} r_j(s_t,a_j,a_{-j})\ \theta^i(a_{-j};h_t)$ (respectively $\bar{r}_j(h_t)$ is the one-period reward if all players correctly min-max i). Let $\bar{q}(h_{t+\tau}|\ a_j)$ denote the distribution over histories $h_{t+\tau}$, for $\tau>0$, if the action at t by player j is a_j and all other players (and j himself after period t) use the correct min-maxing probabilities; e.g. $\bar{q}(h_{t+1}\ |\ a_j) = \Sigma_{a_{-j}} q(s_{t+1}\ |s_t,a_j,a_{-j})\ \theta^i(a_{-j};h_t)$. We shall use $R_j(a_j;h_t)$ to denote the expected returns of player j from period t till the end of the min-maxing phase, under the supposition that he plays a_j in period t and according to min-maxing probabilities thereafter; i.e. $R_j(a_j;h_t) = \bar{r}_j\ (a_j;h_t) + \delta\Sigma_{h_{t+1}} \bar{q}(h_{t+1}\ |\ a_j)\ \bar{r}_j(h_{t+1}) + ...\delta^{T-t}\Sigma_{h_T} \bar{q}(h_T\ |\ a_j)\ \bar{r}_j(h_T)$. To conserve notation, normalize the period of deviation to zero.

The probabilities $P^{ij}\ (h_{T+1})$ will have two important properties: i) <u>aggregation</u>- there will be component probabilities $p^{ij}(a_{jt};h_t)$ such that $P^{ij}(h_{T+1}) = \Sigma_{t\leq T+1}\ p^{ij}(a_{jt};h_t)$ and ii) <u>targetting</u>- p^{ij} will depend only on the action of player j at period t (although it will depend on the actions of other players at previous dates). The existence of probabilities satisfying requisite properties will be established by way of a backward induction argument. Let $\hat{p}^{ij}(a_{jt};h_t)$ be probabilities; i.e. $\hat{p}^{ij}(a_{jt};h_t) \geq 0$, for all a_{jt},h_t and $\Sigma_{j\neq i}\Sigma_t\ \hat{p}^{ij}(a_j;h_{jt}) \leq 1$, for all h_{T+1}. We shall construct associated "weights" $p^{ij}(a_j;h_t)$ which satisfy: after every $h_{t'}$, each player j is indifferent between all of his actions if the "probability" with which play proceeds to Π^{ij} at period T is given by $\Sigma_{t<t'}\ \hat{p}^{ij}\ (a_{jt};h_t) + \Sigma_{t\geq t'}\ p^{ij}\ (a_{jt};h_t) \equiv \hat{P}^{ij}\ (a_{jt};h_t)$. $p^{ij}(a_{jt};h_t)$ are referred to as "probability weights" (with quotation marks) since $\Sigma_{j\neq i}\Sigma_t$ $p^{ij}(a_{jt};h_t)$ may not be less than one, for arbitrary δ. But as $\delta \uparrow 1$, they will be and hence the construction will indeed have yielded probabilities. Let $\Gamma(a_j;h_t)$ denote the lifetime expected returns for player j from using action a_j at period t; i.e.

$$\Gamma(a_j;h_t) = (1-\delta)\ R_j(a_j;h_t) + \delta^{T-t}\ \Sigma_{k\neq i}\ \left[\Sigma_{h_{T+1}}\ \hat{P}^{ik}\ (a_j;h_t)\bar{q}(h_{T+1}|a_j;h_t)\right]\ u_j^{ik}(\delta)$$

$$+\delta^{T-t}\left\{\Sigma_{h_{T+1}}\ \left[1 - \Sigma_{k\neq i}\ \hat{P}^{ik}(a_j;h_t)\right]\bar{q}(h_{T+1}|a_j;h_t)\right\}v_j^i(\delta) \qquad (18)$$

Suppose we solve recursively the system of equations given by $\Gamma(a_j;h_t) = \Gamma(\bar{a}_j;h_t)$, for all $a_j,\bar{a}_j$. For t=T, this involves finding $p^{ij}\ (a_{jT};h_T)$ such that

$$(1-\delta)\ [R_j(a_j;h_T) - R(\bar{a}_j;h_T)] + \delta\ [p^{ij}\ (a_j;h_T) - p^{ij}\ (\bar{a}_j;h_T)][u_j^{ij}(\delta) - v_j^i(\delta)] = 0 \quad (19)$$

Since the first term in (19) goes to zero as $\delta\uparrow 1$, while the second term is strictly bigger than zero (by (16)), it follows that for all $\delta\geq\delta_1$, (19) defines probabilities.

Moreover, these probabilities go to zero as $\delta\uparrow 1$. Suppose then that we have solved for $p^{ij}(a_{j\tau};h_\tau)$, for all $\tau > t$; these are probabilities for all $\delta \geq \delta_{t+1}$, and as $\delta\uparrow 1$ they tend to zero. For any history h_{T+1}, let $\phi^{ik}(h_{T+1})$ denote the incremental probability of play proceeding, at period T+1, to Π^{ij} if the observable history thus far is h_t; i.e. $\phi^{ik}(h_{T+1}) = \Sigma_{\tau>t}$ $p^{ik}(a_{k\tau};h_\tau)$, if h_t is the sub-history of h_{T+1} and zero otherwise. Then,

$$\Gamma(a_j;h_t) - \Gamma(\bar{a}_j;h_t) = (1-\delta)[R_j(a_j;h_t) - R(\bar{a}_j;h_t)] +$$

$$\delta^{T-t}\,\Sigma_{k\neq i}\{[\Sigma_{h_{T+1}}\,\phi^{ik}(h_{T+1})\bar{q}(h_{T+1}|a_j;h_t)] - [\Sigma_{h_{T+1}}\,\phi^{ik}(h_{T+1})\bar{q}(h_{T+1}|\bar{a}_j;h_t)]\}\,[u^{ik}_j(\delta) - v^i_j(\delta)]$$

$$+\ \delta^{T-t}\,[p^{ij}(a_j;h_t) - p^{ij}(\bar{a}_j;h_t)][u^{ij}_j(\delta) - v^i_j(\delta)] \qquad (20)$$

Since $\phi^{ik}(h_{T+1})$ goes to zero, as $\delta\uparrow 1$, (20) defines probabilities for $p^{ij}(a_j;h_t)$ and $p^{ij}(\bar{a}_j;h_t)$ provided $\delta \geq \delta_t$ say. Further, these probabilities themselves go to zero as $\delta\uparrow 1$. For any finite T then, there is $\delta_T < 1$ and associated probabilities for every $\delta \geq \delta_T$ such that 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. In particular, min-maxing i is a best response for j during this phase.

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

<u>Dynamic Games</u>

If the transitions are deterministic a folk theorem obtains under the weaker hypothesis of pairwise full dimensionality (PF) rather than full dimensionality (FD). The latter assumption was used twice in the proof above: firstly, to assert the existence of U^{ij} satisfying (9)-(11) and secondly in the derivation of state-independent payoffs for all players, u^{ij} and v^i. The first argument involves only the payoffs of players i and j and hence can be equivalently derived from pairwise full dimensionality. Replacing the second argument is more problematical. The continuation payoffs u^{ij} and v^i cannot be significantly state dependent or else every player j has an incentive (which does not disappear asymptotically) to deviate during the min-max phase in order to manipulate the state distribution (and the other players' punishment probabilities). However, it is not necessary to eliminate state-dependence altogether. The cycle lengths T^i and T^{ij} (to generate v^i and u^{ij}), can be made sufficiently long in order to make payoffs insufficiently sensitive to the initial state of each cycle. This generally creates a different problem: player i may wish to deviate after some histories within this long cycle if his continuation payoffs $v_i^i(h_t,\delta)$ are less than $v_i^i(s,\delta)$. He suffers min-maxing for a short period and then receives $E[v_i^i(s,\delta)]$ thereafter.[19] For dynamic games it can be shown that i's incentives are not deletriously affected by making T^i and T^{ij} long. In particular one can ignore the state-independence construction of Lemmas 12 and 13; after min-maxing i, play proceeds directly to $\hat{\Pi}^{ij}$ or $\hat{\Pi}^i$. The appropriate probablities are constructed exactly as in (19)-(20). The details of the proof are in the appendix.

7. Applications and Discussion

The principal structural restriction that was imposed was the finiteness of state and action sets. We 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 (1990) (and Mertens and Neyman (1981)) give conditions under which such continuity of feasible payoff 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 12; 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, we will momentarily ignore the fact that the state-action spaces there are typically non-finite.

7.1 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 so 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:

[19] Making the min-max phase T_m longer requires making the cycles T^i yet longer which needs making T_m longer still....etc. The problem stems of course from the fact that we have not made any assumptions about the rate at which discounted state-dependent payoffs converge to long-run average state independent payoffs.

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

Lemma 14 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 (Gillette (1957)), i.e. games in which $q(s'|s,a) > 0$ 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.[20] 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 formalize this idea. We report here a class of structures called scrambling models which have been recently studied by Lockwood (1990).

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 15 Scrambling games satisfy (A1).
Proof: See Lockwood (1990).•

[20]The references in footnotes 1 and 2 are covered by these remarks.

7.2 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 (1957) 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 i-th players' 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 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 of these states be reachable by (n-1) players. As an example, consider separate-state games, where the state $s{=}s_1,\ldots s_n$, is n-dimensional and each player controls his own dimension. Although the i-th player controls his own state, his worst payoffs may be realized 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 i-th 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.

7.3 Other Results

The two papers closest to ours are Friedman (1987) and Lockwood (1990).[21] 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 setup 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)

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

min-max, call it v^m, which has the property that $\underline{v}_i^i(s) \leq v_i^m$.[22] With a full-dimensionality assumption on V, he then proves the asymptotic equilibrium sustainability of all $v \in V$ such that $v >> v^m$. This result follows then from Corollary 9.2 (indeed with payoff asymmetry, since mixed strategies are inadmissible in the Friedman analysis).

Lockwood (1990) analyzes 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).[23] He imposes full-dimensionality on the (state-independent) long-run average payoff set[24] and establishes a folk theorem. All mixed strategies are observable in his framework. So 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).

Finally, Corollary 9.5 represents a modest generalization of the Fudenberg and Maskin folk theorem for purely repeated games in that full dimensionality is replaced with pairwise full dimensionality. This condition has also been used by Smith (1990). Unlike him, we allow unobservable mixed strategies and our contribution here is the construction of probabilistic punishments which deter min-maxing players from unobserved deviations.[25]

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 (1988)). Dutta (1991) 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).

---

[22]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 so 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.

[23]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.

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

[25]For repeated games, Fudenberg and Maskin (1991) have 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.

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 $w \notin$ co $\phi(s,\delta)$. Then, by the strong separating hyperplane theorem (Rockefeller (1970) Corollary 11·4·2), w and co $\phi(s,\delta)$ lie in opposite open half-spaces of some hyperplane. But Blackwell (1965, 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 words, we prove

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

$$\sum_i \lambda_i W_i(s;g^*) \geq \sum_i \lambda_i W_i(s;\Pi) \qquad \text{(A.1)}$$

for any feasible strategy Π.

Pf. Consider any pure Markov strategy, say g, and let $(r_t)_{t\geq 0}$ be the sequence of t-period expected returns for initial state s. Let p_t be the associated probability distribution, i.e. $r_t = \sum_s r_g(s)p_t(s)$. We first show that $\frac{1}{T}\sum_{t=0}^{T-1} r_t$ 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 $\frac{1}{T}\sum_{t=0}^{T-1} r_t$. 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 it then follows that a limit exists for $\frac{1}{T}\sum_{t=0}^{T-1} r_t$ even when the initial state is transient. It then follows by Abel's theorem[26] that

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

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

Now pick an arbitrary strategy Π. Recall that its period t expected returns are denoted $r_i(t;\Pi)$. Let T^i be a sequence such that $\lim_{T^i\uparrow\infty}\frac{1}{T^i}\sum_{t=0}^{T^i-1} r_i(t;\Pi) = W_i(s;\Pi)$. It is well known that we can find a particular sequence $(\delta_m)_{m\geq 0}$ and $\delta_m\uparrow 1$ with the property that $\lim_{T^i\uparrow\infty}\frac{1}{T^i}\sum_{t=0}^{T^i-1} r_i(t;\Pi) = \lim_{\delta_m\uparrow 1}(1-\delta_m)\sum_{t=0}^{\infty}\delta^t r_i(t;\Pi)$, for all i=1,..n. Since the number of pure Markov strategies is finite, for any $\delta_m\uparrow 1$, there is some pure Markov strategy g which maximizes $\sum_i \lambda_i W_i(s;\Pi',\delta)$, over all feasible strategies Π', along a subsequence of δ_m. It then follows that

$$\sum_i \lambda_i W_i(s;\Pi) = \sum_i \lambda_i \left[\lim_{T^i\uparrow\infty}\frac{1}{T^i}\sum_{t=0}^{T^i-1} r_i(t;\Pi)\right]$$

$$= \sum_i \lambda_i \left[\lim_{\delta_m\uparrow 1}(1-\delta_m)\sum_{t=0}^{\infty}\delta^t r_i(t;\Pi)\right]$$

$$= \lim_{\delta_m\uparrow 1}(1-\delta_m)\sum_{t=0}^{\infty}\delta^t \sum_i \lambda_i r_i(t;\Pi)$$

$$\leq \lim_{\delta_m\uparrow 1}(1-\delta_m)\sum_{t=0}^{\infty}\delta^t \sum_i \lambda_i r_i(t;g) \tag{A.3}$$

$$= \sum_i \lambda_i \left[\lim_{\delta_m\uparrow 1}(1-\delta_m)\sum_{t=0}^{\infty}\delta^t r_i(t;g)\right] \tag{A.4}$$

$$= \sum_i \lambda_i \left[\lim_{T\uparrow\infty}\frac{1}{T}\sum_{t=0}^{T-1} r_i(t;g)\right] = \sum_i \lambda_i W_i(s;g) \tag{A.5}$$

(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 1ii) are identical to those used in proving Lemma1i).•

<u>Proof of Lemma 2</u>: It is necessary and sufficient to show

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), 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 $\delta=1$.•

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.[27] Consider the following strategy tuple: the strategy g_1 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) $\frac{T_j}{T}$ 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.•

Proof of Proposition 9.4 Consider $\hat{\Pi}^i$; i.e. infinite repetitions of the cycle g_1^i, $g_2^i,\ldots.g_P^i$, for $T_1^i, T_2^i,\ldots.T_P^i$ periods respectively. Consider the following modification of the strategy to ensure that player i's payoffs are independent of the initial state, s, of the cycle. At the end of the cycle, play proceeds to i's most favorable strategy g^i for T_{P+1}^i periods and T_{P+1}^i is conditioned on s in such a way that state independence holds. As before, this constant payoff is denoted $v_i^i(\delta)$. Recall that $v_i^i(s;g_p^i,T_p^i)$ is i's T_p^i period discounted average payoffs in the play of g_p^i if the state at the beginning of that sub-cycle is s. Let $v_i^i(s;p)$ denote player i's infinite horizon payoffs evaluated from the beginning of the g_p^i sub-cycle. Suppose, without loss of generality, that $v_i^i(s';g_{p+1}^i,T_{p+1}^i) \geq v_i^i(s;g_p^i,T_p^i)$, for all p,s,s'. In particular, $v_i^i(s;p) \geq v_i^i(\delta)$, for all p.

Since the number of states are finite there is a period, say γ, after which all pure Markov strategies begin to cycle. Fix a min-max period T_m satisfying for all $\delta \geq \delta_1$,

$$v_i^i(\delta) > \frac{1}{T_m+1}[\, b_i(1+\gamma)] + \eta \qquad \text{(A.6)}$$

---

[27]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.

Now consider starting with the sub-cycle g_p^i (and renormalize the initial period to zero) followed by g_{p+1}^i, g_{p+2}^i,...etc. We will show that player i has no profitable deviation in this subgame and this is true regardless of the length T_p^i. If $t \leq \gamma$, then

$$v_i^i(s_t) = \frac{1}{\delta^t}[v_i^i(s;p) - (1-\delta) \sum_{\tau=1}^{t} \delta^\tau r_j(\tau)]$$

$$\geq \frac{1}{\delta^t}[v_i^i(\delta) - (1-\delta^\gamma)\, b_i] \qquad \text{(A.7)}$$

We need to show that $v_i^i(s_t) \geq (1-\delta)b_i + (1-\delta^{Tm})\eta + \delta^{Tm+1}v_i^i(\delta)$, but that is seen to follow from (A.6) and (A.7). If $t > \gamma$, g_p^i is already in a cycle of states and actions. Let the length of this cycle be λ and player i's constant (per period) rewards during this cycle c_i. Writing T for the remaining periods of strategy g_p^i, i.e. $T = T_p^i - t$, we have

$$v_i^i(s_t) = (1-\delta^T)c_i + \delta^T v_i^i(s;p)$$

$$\geq (1-\delta^T)c_i + \delta^T v_i^i(\delta) \qquad \text{(A.8)}$$

From (A.8) it is clear that if there was no profitable deviation at the beginning of the cycle, i.e. at $t = \gamma$, then there is no such deviation later. Making T_p^i arbitrarily long makes the payoffs of <u>all</u> players, $v_j^i(s;p)$, sufficiently insensitive to the initial state of the sub-cycle. Hence, the probabilistic construction of (18)-(20) is possible for the strategies $\hat{\Pi}^{ij}$ and $\hat{\Pi}^i$. Proposition 9.4 follows.•

<u>Remark</u>: It can be checked that a similar method of proof works for fully stochastic games or more generally for games whose transitions satisfy the following: for every pure Markov strategy g and state s, the smallest closed set containing s has a single closed subset.

<u>Proof of Lemma 14</u> 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 realized by ex-ante randomization over Markov strategies. Starting from $s \neq s'$, a strategy that 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

Abreu, D., 1988, On the Theory of Infinitely Repeated Games with Discounting, Econometrica 56, 383 - 396.

Aumann, R. and L.Shapley, 1976, Long Term Competition: A Game Theoretic Analysis, mimeo, Hebrew University.

Benhabib, J. and R. Radner, 1988, Joint Exploitation of a Productive Asset, mimeo, AT&T Bell Laboratories, Murray Hill, N.J.

Bernheim, B. D. and D. Ray, 1987, Economic Growth with Intergenerational Altruism, Review of Economic Studies, 54, 227-242.

Bewley, T. and E.Kohlberg, 1976, The Asymptotic Theory of Stochastic Games, Mathematics of Operations Research, 1, 197 - 208.

Blackwell, D., 1965, Discounted Dynamic Programming, Annals of Mathematical Statistics, 36, 226 - 235.

Dutta, P.K., 1990, What do Discounted Optima Converge to? A Theory of Discount Rate Asymptotics in Economic Models, University of Rochester Working Paper # 264, forthcoming Journal of Economic Theory.

Dutta, P.K., 1991, Collusion, Discounting and Dynamic Games, University of Rochester Working Paper # 272.

Dutta, P.K. and A. Madhavan, 1991, Dynamic Insider Trading, University of Rochester Working Paper # 270.

Friedman, J., 1987, A Modified Folk Theorem for Time-Dependent Supergames, Center for Interdisciplinary Studies, University of Bielfeld, Working Paper #2.

Fudenberg, D. and E.Maskin, 1986, The Folk Theorem in Repeated Games with Discounting or Incomplete Information, Econometrica, 54, 533 - 554.

Fudenberg, D. and E.Maskin, 1991, On the Dispensability of Public Randomization in Discounted Repeated Games, Journal of Economic Theory, 53, 428-431.

Gillette,D., 1957, Stochastic Games with Zero-Stop Probabilities, Contributions to the Theory of Games, # 3, (Annals of Mathematical Studies 39) Princeton, 179 - 187.

Himmelberg, C.J., T. Partharasarathy, T.E.S.Raghavan and F.S.VanVleck, 1976, Existence of p-Equilibrium and Optimal Stationary Strategies in Stochastic Games, Transactions of the American Mathematical Society, 60, 245 - 261.

Lockwood, B., 1990, The Folk Theorem in Stochastic Games with and without Discounting, mimeo, University of London.

Maskin, E. and J. Tirole, 1988, A Theory of Dynamic Oligopoly II: Price Competition, Kinked Demand Curves and Edgeworth Cycles, Econometrica, 56, 571-601.

Mertens, J-F and A.Neyman, 1981, Stochastic Games, International Journal of Game Theory, 10, 53 - 66.

Mertens, J-F and T. Partharasarathy, Stochastic Games, 1987, CORE Working Paper 8750, Louvain-la-Neuve.

Nowak, A.S., 1985, Existence of Equilibrium Stationary Strategies in Discounted Non-Cooperative Stochastic Games with Uncountable State Space, Journal of Optimization Theory and Applications, 45, 591 - 602.

Partharasarathy, T., 1973, Discounted, Positive and Non-Cooperative Stochastic Games, International Journal of Game Theory, 2.

Partharasarathy, T. and S.Sinha, 1990, Existence of Stationary Equilibrium Strategies in Non-Zero Sum Discounted Stochastic Games with Uncountable State Space and State Independent Transitions, International Journal of Game Theory , 19.

Rockafellar, R.T., 1970, Convex Analysis, Princeton University Press, Princeton, NJ.

Rubinstein, A., 1979, Equilibrium in Supergames with the Overtaking Criterion, Journal of Economic Theory, 31, 227 - 250.

Smith, L., 1990, Folk Theorems: Two Dimensionality is (Almost) Enough, mimeo, University of Chicago.

Stokey, N., 1990, Credible Public Policy, mimeo, University of Chicago.

Sundaram, R., 1989, Perfect Equilibrium in a Class of Symmetric Dynamic Games, Journal of Economic Theory, 47, 153-177.