

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 777

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

ON FINITELY REPEATED GAMES AND PSEUDO-NASH EQUILIBRIA

Chien Chou and John Geanakoplos

October 1985

ON FINITELY REPEATED GAMES AND PSEUDO-NASH EQUILIBRIA

by

Chien Chou and John Geanakoplos

1. INTRODUCTION

It is by now a well-known paradox that if a game is repeated infinitely often (to construct a "supergame") then it is possible to achieve cooperation between the players as a (perfect) Nash equilibrium, whereas for many games, like the prisoner's dilemma, a finite repetition of the game, no matter how long, will not provide for any more cooperation than the one-shot game itself. So far two modifications of the equilibrium notion for finitely repeated games have been introduced into the literature in order to restore "continuity at infinity," i.e., to explain the anecdotal and experimental evidence suggesting that in long finitely repeated games players do seem to behave as if the games were to be infinitely repeated. In this paper we shall provide a third pseudo-Nash equilibrium concept.

It is clear that strategic rivalry in a long term relationship might modify behavior in each of the one-shot struggles. A player who considers the subsequent reactions of his opponents may be led by the fear of retaliation to exhibit cooperative behavior which otherwise would not be to his advantage to show. In an infinitely repeated game (without discounting) the potential for punishment is always infinite, while in a finitely repeated game it declines as time reaches its end. In the last period, there is no future to consider, and so the players must play a Nash equilibrium. In the prisoner's dilemma, the players work backwards; knowing that the worst will happen to them anyway in the final period, they have no more incentive

Geanakoplos' work supported by NSF Grant 82-10034 and SES-8411149. Geanakoplos would like to acknowledge a helpful conversation with Paul Milgrom.

to act cooperatively in the second to last period than in the one-shot game. Repeating the argument shows that the players will rationally act myopically even in the first period of an N-repeated prisoner's dilemma.

Radner¹ (1980) broke the yoke of backward induction by defining an ϵ -equilibrium in which a player is satisfied with his strategy unless he can gain at least ϵ per period more by deviating. If T is large, then deviations which begin near the end cannot gain $T\epsilon$, while for deviations which begin earlier there is more time left for retaliation. More recently Kreps-Wilson (1982), Milgrom-Roberts (1982), and Kreps-Milgrom-Roberts-Wilson (1982) proposed an " ϵ -crazy" equilibrium in which with probability ϵ a player may behave in some arbitrarily specified "crazy" manner. They have shown (see also Fudenberg-Maskin (1983)) that given any average payoff per period (x_1, x_2) which can be sustained in an infinitely repeated two player game, there is a specification of "craziness" for the two players such that (x_1, x_2) can nearly be obtained as the average payoff of a T times repeated game, for T sufficiently large.

Both the ϵ -equilibrium and the ϵ -crazy equilibrium allow for the suspension of the Nash optimality test along the game tree, especially near the end. In principle both definitions allow for a large amount of total suboptimality if not average suboptimality. We might ask, how much fury doth hell need to induce the players to cooperate?

In this paper we propose a pseudo-Nash equilibrium for N-person games in which very simply we allow play in the last period to be arbitrary, but otherwise it must conform to the (perfect) Nash optimality criterion. This is analogous both to Radner's definition in the sense that for the first

¹See also Fudenberg-Levine (1983).

T-1 periods both players must play an ϵ -equilibrium, but since only one period is arbitrary, $\epsilon = K/(T-1) \rightarrow 0$, where K is some constant corresponding to the value of cooperation in the one-shot game at stage T , and to the K-M-R-W "crazy-equilibrium" in the sense that in period T we allow players to be crazy.*

Our definition also permits us to draw an analogy between repeated games and overlapping generations (OG) economies, and between money and cooperation. Rational, self-interested agents hold money because they expect future money to be valuable; similarly, they cooperate because they expect future cooperation to be valuable. It is a well-known phenomenon that in an infinite-horizon OG economy rational agents may hold money and the equilibrium set may often be a continuum, while in finite horizon economies it is typically finite, and no agent holds money. Just as infinite horizon OG economies can be understood as limits of finite horizon OG economies where the markets do not need to clear at the last time period, so we are suggesting that infinitely repeated games can be understood as limits of finitely-repeated games in which the Nash test need not be applied in the last period.

To see the force of our modification of Nash equilibrium, consider the prisoner's dilemma game:

$$G_1 = \begin{pmatrix} (1/5, 1/5) & (1, 0) \\ (0, 1) & (4/5, 4/5) \end{pmatrix}.$$

*To put it another way, we suppose that an $N+1^{\text{st}}$ "principal" announces a punishment scheme that he will deliver in period T as a function of how the players behaved until $T-1$. If the punishment is limited in size to what the N "agents" could inflict upon themselves in a one-shot game, we ask whether the principal, by varying his limited punishment scheme, can induce the agents to achieve every cooperative outcome (in payoff space) as a perfect Nash equilibrium, as $T \rightarrow \infty$ (and therefore as his threat becomes arbitrarily small relative to the total payoff).

The outcome $(4/5, 4/5)$ cannot be sustained as a Nash equilibrium, but it can as a pseudo-Nash equilibrium. If player one deviates at period $T-1$, he gains $1 - 4/5 = 1/5$. In period T , however, he can be punished (say, by $4/5 - 1/5 = 3/5$) so that on the whole he loses. On the other hand in the game

$$G_2 = \begin{pmatrix} (1, 0) & (0, 1) \\ (0, .8) & (.9, 0) \end{pmatrix}$$

the average payoff $(1/2, 1/2)$ is sustainable in the infinitely repeated game (a famous folk-theorem guarantees that any individually rational payoff in the convex set of the payoff pairs of G is sustainable in G^∞) yet it would seem hard to sustain as a pseudo-Nash equilibrium. The only way to make it feasible is to alternate $(1,0)$ and $(0,1)$ and if in period T player two stands the most to gain, what will prevent player one from cheating when the payoff is $(0,1)$?

In this paper we show that for a generic N -player smooth game G , such as Cournot's model of quantity competition, any payoff pair sustainable in G^∞ is the limit of T period (perfect) pseudo-Nash equilibrium payoffs. The same property holds for any two player $m \times n$ game G that has a non-degenerate mixed strategy Nash equilibrium, such as G_2 , above, and for any 2 player 2×2 matrix game that has any Pareto inferior Nash equilibrium, such as G_1 above.

The difference between smooth games and matrix games is that in the latter, when mixed strategies are used, cheating is difficult to detect. Thus we remark in Section 6 that for matrix games the force of the entire last period threat is required for the proof, while for smooth games period T can be discounted arbitrarily (just short of totally) without

affecting the limiting results. To put it differently, for generic smooth games Radner's proposition about ϵ -average equilibria can be strengthened to ϵ -total equilibria--no matter how small $\epsilon > 0$ is, if no player will deviate unless he can gain at least ϵ in total, then any individually rational payoff vector can be achieved arbitrarily closely as $T \rightarrow \infty$.

The paper is organized as follows. Section 2 gives definitions, and Sections 3 and 4 introduce useful devices called the continuation principle and reusable reward systems. The difficulty in implementing multiperiod (perfect) cooperation schemes is that each player might deviate many times, and each time he must be threatened with a new punishment. This tends to put a heavy burden on the force of the final retribution. Our method of coping with this is to devise schemes where (1) no player will deviate consecutively and (2) when player i deviates he brings upon himself a punishment to be delivered at the end, but at the same time a reward to the previous player $j \neq i$ that deviated! Thus the total net punishment (or reward) that must be delivered at the end can be kept small.

In Section 5 we use this technique to prove that $\lim_{T \rightarrow \infty} (\text{perfect}) \text{pNE}(G^T) = \text{NE}(G^\infty)$ for smooth games G . In Section 6 we give the corresponding proof for matrix games with nondegenerate mixed strategy Nash equilibria, and in Section 7 we state our theorem for $2 \times 2 \times 2$ games.

A related result by Benoit and Krishna shows that when a game G has multiple Nash equilibria, then the perfect NE of G^T converge to the NE of G^∞ . Of course many games of great interest, like the Cournot game, often have a unique Nash equilibrium. Nevertheless, in view of the importance of the Benoit-Krishna result, we have shown in Section 8 how our continuation principle can be used to give a short demonstration of their theorem.

2. General Definitions

2.1. Pseudo-Nash Equilibria

Definition 1. An N-person game G is defined by $G = [\Sigma_i, \Pi_i, i = 1, \dots, N]$, where the Σ_i are the strategy spaces and the Π_i are the payoff functions for players $i = 1, \dots, N$. We assume that each Σ_i is compact and convex.

Let $\Sigma = \prod_{i=1}^N \Sigma_i$. We also assume that $\Pi_i : \Sigma \rightarrow \mathbb{R}$ is continuous, and concave in the i^{th} coordinate.

We have thus restricted our attention to "one-shot" games which have Nash equilibria. The standard matrix games, where each player has a finite number of pure strategies, can be regarded as a special case if we include all the "randomized" strategies for each player. We shall discuss these games in Section 6. A canonical example of a game G satisfying our definition is the Cournot game, where each player must choose a quantity $q_i \in [0, Q_i] = \Sigma_i$, and the payoffs are given by the profits (equal to revenue minus costs) that are obtained in some market.¹ Typically in a Cournot game the Nash payoffs to the sellers are less than the monopoly profits that could be shared if they agreed to cooperate and produce less. (A still worse situation for player i occurs when all the other players play $q_j = Q_j$, the "minmax strategies" against player i .)

Definition 2. The repeated game G^T is defined to be the game that repeats G T times, and whose payoff is the sum of the individual period payoffs. A strategy $\sigma_i^T \in \Sigma_i^T$ for player i in G^T can be represented by

¹Of course the inverse demand function $P = P(q_1 + q_2 + \dots + q_N)$ must be well-behaved and similarly for the cost functions $c_i(q_i)$ so that the profit function for firm i , $\Pi_i(q_1, \dots, q_N) = q_i P(q_1, \dots, q_N) - c_i(q_i)$ will be concave in q_i , for each $i = 1, \dots, N$.

$$\sigma_i^T = (\sigma_i^{(1)}, \sigma_i^{(2)}, \dots, \sigma_i^{(T)})$$

where $\sigma_i^{(1)} \in \Sigma_i$, and $\sigma_i^{(t)}$ is a function from $\prod_{T=1}^{t-1} \Sigma$ to Σ_i ,

$t = 2, \dots, T$. Given the N-tuple $\sigma^T = (\sigma_1^T, \dots, \sigma_N^T)$, let

$\bar{\sigma}_i^T = (\bar{\sigma}_i^{(1)}, \bar{\sigma}_i^{(2)}, \dots, \bar{\sigma}_i^{(T)})$ be the realized sequence of actions for player i . The payoff function for G^T is

$$\Pi_i^T = \Pi_i^T(\sigma^T) = \sum_{t=1}^T \Pi_i^T(t) = \sum_{t=1}^T \Pi_i(\bar{\sigma}_1^{(t)}, \dots, \bar{\sigma}_N^{(t)}) .$$

Notice that if we regard Σ_i as the randomized mixtures over a finite set of pure strategies, then the definition we have just given for G^T allows the t -period choice of player i to depend on the randomizing device used by player $j \neq i$ in period $t-1$. We change our definition when we consider matrix games in Section 6.

Definition 3. An N-tuple $\sigma^* = (\sigma_1^*, \dots, \sigma_N^*)$ of strategies for the repeated game G^T is a Nash Equilibrium (NE) if and only if

$$\Pi_i^T(\sigma_1^*, \dots, \sigma_i^*, \dots, \sigma_N^*) \geq \Pi_i^T(\sigma_1^*, \dots, \sigma_i, \dots, \sigma_N^*)$$

for all $\sigma_i \in \Sigma_i^T$, and all $i = 1, \dots, N$.

We are now ready for our main definition:

Definition 4. The N-tuple σ^* of strategies for the repeated game G^T is a pseudo-Nash equilibrium (PNE) if and only if

$$\Pi_i^T(\sigma_1^*, \dots, \sigma_i^*, \dots, \sigma_N^*) \geq \Pi_i^T(\sigma_1^*, \dots, \sigma_i, \dots, \sigma_N^*)$$

for all $\sigma_i \in \Sigma_i^T$ such that

$$\sigma_i^{(T)} = \sigma_i^{*(T)}, \quad i = 1, \dots, N.$$

Notice that any NE is a PNE.

It is also possible to define a perfect Nash equilibrium (perfect NE) and a perfect PNE of the repeated game G^T . Let $\bar{\sigma}^\tau = (\bar{\sigma}^{(1)}, \dots, \bar{\sigma}^{(\tau)})$ be any sequence of joint moves in $\underbrace{\Sigma \times \dots \times \Sigma}_\tau$, where $\tau < T$. Then the N-tuple of strategies $^*\sigma^T = (^*\sigma_1^T, \dots, ^*\sigma_N^T)$ for the game G^T , together with the given τ -period history $\bar{\sigma}^\tau$, defines an N-tuple of strategies $(\bar{\sigma}^\tau, ^*\sigma^T)$ for the game $G^{T-\tau}$ in the obvious way. If this is an NE (PNE) for $G^{T-\tau}$, no matter what is τ or $\bar{\sigma}^\tau$, then we say that σ^* is a perfect NE (perfect PNE) of G^T .

Definition 5. Let $(\text{perfect}) \text{PNE}(G^T) = \{x = (x_1, \dots, x_N) \mid \exists \text{ a } (\text{perfect}) \text{PNE } ^*\sigma^T \text{ of } G^T \text{ such that } \frac{1}{T}\Pi^T(G^T) = x\}$.

Example 1. Prisoners' Dilemma Game

The payoff matrix of a prisoners' dilemma game can be represented by

	Left	Right
Top	(2/3, 2/3)	(0, 1)
Bottom	(1, 0)	(1/3, 1/3)

The pure strategy pair $(\tilde{\sigma}_1^T, \tilde{\sigma}_2^T)$ defined below is a perfect PNE but not a NE of G repeated T times:

$$\tilde{\sigma}_1^T(t) = \begin{cases} \text{Top} & \text{if } \bar{\sigma}_2^{(\tau)} \neq \text{Right for all } \tau < t, \\ \text{Bottom} & \text{otherwise} \end{cases}$$

and

$$\tilde{\sigma}_2^T(t) = \begin{cases} \text{Left} & \text{if } \bar{\sigma}_1^{(\tau)} \neq \text{Bottom for all } \tau < t, \\ \text{Right} & \text{otherwise.} \end{cases}$$

Hence, $(2/3, 2/3) \in \text{perfect PNE}(G^T)$, whereas $\text{NE}(G^T) = \{(1/3, 1/3)\}$ for all T .

2.2. The Folk Theorem

Let $\Sigma_{-j} = \prod_{\substack{i=1 \\ i \neq j}}^N \Sigma_i$, for $j = 1, \dots, N$. The minmax value v_i for player i in the (one-shot) game G is $\text{Min}_{\sigma_{-i} \in \Sigma_{-i}} \text{Max}_{\sigma_i \in \Sigma_i} \Pi_i(\sigma_i, \sigma_{-i})$
 $= \text{Max}_{\sigma_i \in \Sigma_i} \text{Min}_{\sigma_{-i} \in \Sigma_{-i}} \Pi_i(\sigma_i, \sigma_{-i}) = \Pi_i(\mu_i)$. Let $C(G)$ be the convex hull of all the payoff N -vectors $(\Pi_1(\sigma), \dots, \Pi_N(\sigma))$ as σ varies over Σ .

Definition 6. We let $\text{NE}(G^\infty) = C(G) \cap \{(x_1, \dots, x_N) \mid x_i \geq v_i, i = 1, \dots, N\}$.

The folk theorem asserts that the Nash equilibrium average payoffs (and also the perfect NE payoffs) to the infinite repeated game G^∞ are given by the set $\text{NE}(G^\infty)$.

Definition 7. $x = (x_1, \dots, x_N)$ is called a (perfect) PNE (average) payoff, and we let $x \in \lim_{T \rightarrow \infty} \text{PNE}(G^T)$, if and only if there exists a sequence x^T where x^T is a (perfect) PNE average payoff for G^T , and $\lim_{T \rightarrow \infty} x^T = x$.

3. The Continuation Principle

3.1. Enforceability and Consecutive Deviations

Let (r_1, \dots, r_N) be a vector of nonnegative numbers which we call rewards. Clearly we shall use rewards to enforce cooperation. However, if we wish to enforce "perfect" cooperation, then we must be prepared to punish deviations off the equilibrium path, and to punish deviations from deviations, etc. With this in mind, we give:

Definition 8. The strategies $*\sigma^T$ for the game G^T are said to be enforceable by the reward structure (r_1, \dots, r_N) if $\Pi_i^T(\sigma_i^T, *\sigma_{-i}^T) - \Pi_i(*\sigma^T) \leq r_i$ for all $\sigma_i^T \in \Sigma_i^T$, and all $i = 1, \dots, N$. Similarly, if for any τ period history $\bar{\sigma}^\tau$, with $\tau \leq T$, if $(\bar{\sigma}^\tau, *\sigma^T)$ is enforced in $G^{T-\tau}$ by the reward structure (r_1, \dots, r_N) , then we say that (r_1, \dots, r_N) perfectly enforces $*\sigma^T$.

Note that $*\sigma^T$ is a (perfect) NE of G^T if and only if it is (perfectly) enforceable by the reward structure $(0, 0, \dots, 0)$.

As we have just said, when dealing with perfect equilibria it may be necessary to be able to "punish" many players, and some players more than once. It would then appear that the prizes (r_1, \dots, r_N) would have to be given out over and over, and hence the total reward might need to be arbitrarily large. We shall show, however, that once consecutive profitable offenses are eliminated, the rewards need only be given out once.

Let $*\sigma^T$ be an N-vector of strategies for G^T . Given a history $\bar{\sigma}^\tau$, up until $\tau \leq T$, we shall say that player i deviated from $*\sigma_i^T$ at time τ if $*\sigma_i^{(\tau)}(\bar{\sigma}^{\tau-1}) \neq \bar{\sigma}_i^{(\tau)}$ and $\sigma_j^{(\tau)}(\bar{\sigma}^{\tau-1}) = \bar{\sigma}_j^{(\tau)}$ for $j < i$. Furthermore, we shall say that player i was the last deviator from $*\sigma^T$ up until time τ if there is a time $t \leq \tau$ at which he was the deviator,

and if there is no time t' , $t < t' \leq \tau$ at which some other player $j \neq i$ was the deviator.

Definition 9. The strategies $^*\sigma^T$ for the game G^T are said to prevent consecutive deviations if for any $\tau \leq T$ period history $\bar{\sigma}^\tau$, if i was the last deviator from $^*\sigma^T$ up until τ , then

$$\Pi_i^{T-\tau}(\bar{\sigma}^\tau, (\sigma_i^T, ^*\sigma_{-i}^T)) \leq \Pi_i^{T-\tau}(\bar{\sigma}^\tau, ^*\sigma^T) \text{ for all } \sigma_i^T \in \Sigma_i^T.$$

"Trigger strategies" are the most famous example of such strategies. Let $(^*s_1, \dots, ^*s_N) = ^*s$ be a one-shot Nash equilibrium for G , and let $\Pi_i(s_1, \dots, s_N) > \Pi_i(^*s)$. The trigger strategy $^*\sigma_i$ for player i in G^T is $^*\sigma_i^{(\tau)}(\bar{\sigma}^{\tau-1}) = s_i$ if $\bar{\sigma}(t) = s$ for all $t \leq \tau-1$, and $^*\sigma_i^{(\tau)}(\bar{\sigma}^{\tau-1}) = ^*s_i$ otherwise, for $\tau = 1, \dots, T$. Trigger strategies have the special property that once one player has deviated, then no player can subsequently deviate advantageously.

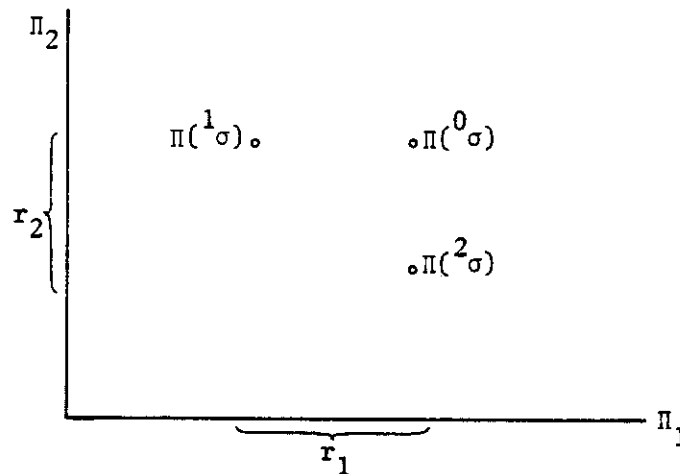
3.2. Reusable Reward Systems and the Continuation Property

Definition 10. A reusable, (perfect), T -period reward system R^T with reward structure (r_1, \dots, r_T) is a set of $N+1$ (perfect) PNE's $(^0\sigma, ^i\sigma, i = 1, \dots, N)$ for the game G^T which satisfies

$$\Pi_i^T(^j\sigma) - \Pi_i^T(^i\sigma) \geq r_i \text{ for all } i = 1, \dots, N, j = 0, 1, \dots, N, i \neq j. *$$

The following diagram displays the payoffs to two players of 3 pNE's forming a reusable reward system: Note that each $^i\sigma = (^i\sigma_1, \dots, ^i\sigma_N)$ is itself an N -vector of strategies, for $i = 0, \dots, N$.

*Our reusable reward system is in some sense the finite-time horizon analogue of Abreu's [1982] simple penal code. A similar device for infinitely repeated games was also used by Fudenberg-Maskin [1983].



Notice that as long as the intended perfect PNE is $\sigma \in \{^0\sigma, ^1\sigma, ^2\sigma\}$ but $\sigma \neq ^i\sigma$, then player i can be threatened with a loss of r_i if play switches to $^i\sigma$. Since $^i\sigma$ is itself a (perfect) PNE, this threat is credible.

Suppose that the strategies $\sigma^{T'}$ for the game $G^{T'}$ do not allow consecutive deviations, and suppose furthermore that they can be perfectly enforced by the reward structure (r_1, \dots, r_N) . Let $R^{T''}$ be a reusable, perfect, T'' -period reward system with reward structure (r_1, \dots, r_N) . Define the strategies $^*\sigma^{T'+T''}$ in the game $G^{T'+T''}$ by $^*\sigma_i(\tau) = \sigma_i(\tau)$ for all players i and $\tau \leq T'$. Furthermore, from $T'+1$ to $T'+T''$ the players follow $^i\sigma$ if i was the last player to diverge from $\sigma^{T'}$ up until time T' . Otherwise they follow $^0\sigma$.

Definition: The continuation property asserts that if $^*\sigma^{T'}$ prevents consecutive deviations in the game $G^{T'}$ and is (perfectly) enforceable by the reward structure (r_1, \dots, r_N) , and if $\{^0\sigma, ^1\sigma, \dots, ^N\sigma\}$ is a (perfect) reusable reward system for the game $G^{T''}$ with reward structure (r_1, \dots, r_N) , then their combination as above in the game $G^T = G^{T'+T''}$ is a (perfect) pNE. Moreover, if the $^j\sigma$ are all (perfect) NE's, then the combination is

a (perfect) NE in the game G^T . We denote the combined strategy for the game $G^{T'+T''}$ by $(*_{\sigma}^{T'}, R^{T''})$.

The continuation property hardly needs proof. However, it suggests the strategy we shall use for the proof of our main theorem. Take an arbitrary N-player game G , and payoffs $x = (x_1, \dots, x_N) = \Pi(\sigma) = (\Pi_1(\sigma), \dots, \Pi_N(\sigma))$ strictly greater than the maxmin payoffs (v_1, \dots, v_N) , where σ is some N-tuple of one shot moves by all the players. Lemma 1 shows that there is some reward $\bar{r} > 0$ such that for all T' there is an N-tuple of strategies $*_{\sigma}^{T'}$ in $G^{T'}$ that prevents consecutive deviations, that is perfectly enforceable with reward structure $(\bar{r}, \dots, \bar{r})$, and that yields payoffs (x_1, \dots, x_N) in almost every period if there are no deviations.

In view of Lemma 1, the proof of our main theorem for a specific game G is reduced to showing that for given (r_1, \dots, r_N) , there is some T'' and a reusable reward system $\{^0_{\sigma}, ^1_{\sigma}, \dots, ^N_{\sigma}\}$ for $G^{T''}$ with rewards at least (r_1, \dots, r_N) . Note that since \bar{r} is fixed independent of T' , so is T'' , and hence as $T' \rightarrow \infty$, the average payoffs of $(*_{\sigma}^{T'}, R^{T''})$ converge to x .

Observe also that for a PNE, the last period can be arbitrary. Thus even when $T' = 1$, there is a reusable reward system with rewards on the order of $(1/N, \dots, 1/N)$. It turns out that for smooth games one can easily build this into an arbitrarily harsh reward system. The same is true when there are multiple one-shot Nash equilibria, even without benefit of the last period. For matrix games, the proof is a bit more difficult.

4. The Basic Lemma

Consider the one-shot game G with maxmin payoffs $v = (v_1, \dots, v_N)$, $v_i = \Pi_i(\mu_i^i, \mu_{-i}^i) \geq \Pi_i(\sigma, \mu_{-i}^i)$. $\forall \sigma \in \Sigma_i$, and a Nash equilibrium $*s$ with payoffs $\Pi_i(*s)$. Without loss in generality we take $\text{Min}_{\sigma \in \Sigma} \Pi_i(\sigma) = 0$ and $\text{Max}_{\sigma \in \Sigma} \Pi_i(\sigma) = 1$, for all $i = 1, \dots, N$.

Lemma 1. Let \vec{s} be an N -tuple of one-shot strategies for G with payoffs $\Pi_i(\vec{s}) = x_i$ satisfying $x_i > v_i$ for all $i = 1, \dots, N$. Let $W = \{i \in N \mid \Pi_i(*s) > v_i\}$, let $\bar{x}_i = \text{Min}(x_i, \Pi_i(*s))$, and let $K = \text{Max}_{i \in W} \left[\frac{1}{\bar{x}_i - v_i} \right] + 1$. Then for any period T it is possible to devise strategies σ^T that (1) do not permit consecutive deviations and (2) are enforced by the reward structure (K, \dots, K) and (3) yield a history $\vec{\sigma}^T$ with realization \vec{s} in at least $T-K$ periods.

Proof. We shall give a sketch--the proof is obvious except for notation.

Let $T > K$ (otherwise there is nothing to prove). The intended path means each player i plays \vec{s}_i at each date $\tau \leq T-K$, and plays $*s_i$ for $T-K < \tau \leq T$.

Let W be the set of players with $v_i < \Pi_i(*s)$, and let B be the rest, all with $v_i = \Pi_i(*s)$.

If a player $i \in W$ deviates from the intended path at time $\tau \leq T-K$, then from $\tau+1$ to $\tau+K$ all other players should play μ^i . After $\tau+K$, all players should return to the intended path. If during period $\tau+1$ to $\tau+K$ a player $j \in W$, $j \neq i$ deviates from the intended punishment of i (playing anything different from that specified by μ_j^i) then play returns immediately to the intended path. If i himself deviates from his own punishment phase, then the punishment continues as before. It is easy

to see that no player $i \in W$ can advantageously deviate from the intended path, or from his own punishment. Once i deviates from the punishment of $j \in W$, play immediately returns to the intended path. Hence $i \in W$ cannot deviate consecutively, advantageously, provided that $i \in W$ has no opportunity to deviate following a deviation of $k \in B$.

Since the most that $i \in W$ can gain by deviating from the punishment phase of $j \in W$ is K , it follows that the behavior of $i \in W$ is perfectly enforced by the reward K .

If a player $i \in B$ deviates from the intended path, or from the punishment phase of a player $j \in W$, then all players should play s^* until the end. Deviations from this path by any player are ignored. Clearly once player $i \in B$ has deviated, no player can again advantageously deviate. Note that the behavior of $i \in B$ is perfectly enforced by the reward $1 < K$.

Q.E.D.

Lemma 1 can easily be extended to average payoffs

$(x_1, \dots, x_N) \gg (v_1, \dots, v_N)$ that are rational combinations of payoffs from strategy N -tuples \vec{s} and \vec{s}^* for the one-shot game G . Let

$x = \frac{n_1}{C}\Pi(\vec{s}) + \frac{n_2}{C}\Pi(\vec{s}^*)$. The intended path must now be a sequence of C -cycles, in any one of which \vec{s} occurs n_1 times and \vec{s}^* n_2 times. The punishment phase shall last $K = \left\lceil \max_{i \in W} \left[\frac{1}{x_i - v_i} \right] + 1 \right\rceil$ cycles, i.e. be KC periods long, and r_i must be at least KC .

As T gets large, the K (or KC) periods during which x is not realized becomes negligible, as does the payoff from the reusable reward system phase at the end. Thus to conclude the proof of the general proposition for a class of games G requires only that one can demonstrate that for each such game it is possible to construct a reusable reward system with arbitrarily large reward structure.

5. Smooth Games

Let Σ_i be an ℓ_i dimensional rectangle, $i = 1, \dots, N$, and let S_i be its interior. Let θ be an $\ell_0 \geq \ell = \ell_1 + \dots + \ell_N$ dimensional rectangle of parameters. Let the games G_θ be defined by $(\Sigma_i, i = 1, \dots, N)$ and the functions $\Pi_i : \Sigma_1 \times \dots \times \Sigma_N \times \theta \rightarrow \mathbb{R}$, where each Π_i is three times differentiable, and strictly concave in Σ_i , for any $(\sigma_{-i}, \theta) \in \Sigma_{-i} \times \theta$. The classic example of such a game is the Cournot game, where Π_i is the profit function of firm i , depending on the outputs $\sigma_j \in \Sigma_j = [0, \bar{Q}_j]$ of all the players $j \in N$, and the constant marginal cost θ_i of player i ($\theta = (\theta_1, \dots, \theta_N)$).

We shall make the following assumptions:

- A1) For any $\theta \in \theta$ all the Nash equilibria of G_θ involve strategies lying in the interior, i.e. in $S_1 \times \dots \times S_N$.
- A2) For any choice of $\bar{\theta} \in \theta$ and $\bar{\sigma}_1 \in S_1, \dots, \bar{\sigma}_N \in S_N$, and for any player $i \in N$, and coordinate $k \leq \ell_i$ it is possible to choose a direction Δ in θ space such that (1) $\partial^2 \Pi_j / \partial \Delta \partial \sigma_r^i \Big|_{\bar{\theta}, \bar{\sigma}} = 0$, $j \neq i$, $r = 1, \dots, \ell_j$, (2) $\partial^2 \Pi_i / \partial \Delta \partial \sigma_r^i \Big|_{\bar{\theta}, \bar{\sigma}} = 0$, $r \neq k$, and (3) $\partial^2 \Pi_i / \partial \Delta \partial \sigma_k^i \Big|_{\bar{\theta}, \bar{\sigma}} \neq 0$.

Assumption A2 means that we have a rich set of games. For example, in the Cournot model where $\ell_i = 1$ for all $i \in N$ and $\theta = (\theta_1, \dots, \theta_N)$ is the vector of cost parameters $c_i(\theta, \sigma) = \theta_i \sigma_i$, then $\partial^2 \Pi_i / \partial \theta_i \partial \sigma_i = -1$, while $\partial^2 \Pi_j / \partial \theta_i \partial \sigma_j = 0$ for all $j \neq i$.

Let \mathcal{D} be a subset of θ . We say \mathcal{D} is generic if it is open and dense in θ , and if its complement in θ has measure 0.

Theorem 1: There is a generic set $\mathcal{D} \subset \theta$, such that for any game G_θ with $\theta \in \mathcal{D}$, $\lim_{T \rightarrow \infty} \text{PNE}(G_\theta^T) = \text{NE}(G_\theta^\infty)$.

Proof. From Lemma 1 and the continuation property it suffices to show the existence of $N+1$ PNE's as follows.

It is easy to show that for a generic set \mathcal{D} , any $\bar{\theta} \in \mathcal{D}$ and any Nash equilibrium $\bar{\sigma}$ of $G_{\bar{\theta}}$ satisfies the following property:

Write the matrix of derivatives of profit Π_i in every strategy σ_k^j , $i = 1, \dots, N$, $j = 1, \dots, N$, $k = 1, \dots, \ell_j$,

$$A = \begin{bmatrix} \frac{\partial \Pi_1}{\partial \sigma_1^1} & \dots & \frac{\partial \Pi_1}{\partial \sigma_{\ell_N}^N} \\ \vdots & & \\ \frac{\partial \Pi_N}{\partial \sigma_1^1} & \dots & \frac{\partial \Pi_N}{\partial \sigma_{\ell_N}^N} \end{bmatrix}$$

evaluated at $(\bar{\sigma}, \bar{\theta})$. Then A has full row rank. This is proved in Dubey and Rogawski [4]; it is derived directly from the transversality theorem.

The fact that A has full row rank means that the Nash equilibrium $\bar{\sigma}$ is not Pareto optimal. There are small finite changes $d\sigma_1, \dots, d\sigma_N$ which increase the payoffs to all the players. In the Cournot case it is no surprise to learn that all the firms can improve their profits if they collude.

Let $\bar{\Pi}$ be the N -vector of payoffs to all the players at the Nash equilibrium $\bar{\sigma}$ in the game $G_{\bar{\theta}}$. Since A has full row rank, it follows that for small ϵ , it is possible to find small finite changes $d\sigma_1^0, \dots, d\sigma_N^0$ such that $\Pi_{\bar{\theta}}(\bar{\sigma} + d\sigma^0) = \bar{\Pi} + \epsilon e$, where $e = (1, 1, \dots, 1)$. Similarly for any j , we can find small $(d\sigma_1^j, \dots, d\sigma_N^j) = d\sigma^j$ such that $\Pi_{\bar{\theta}}(\bar{\sigma} + d\sigma^j) = \bar{\Pi} + \epsilon e - \epsilon e_j$. Lastly, we can also find μ such that $\Pi_{\bar{\theta}}(\mu) \ll \bar{\Pi}$.

Observe now that for ϵ small, any of the $N+1$ average payoffs $\bar{\Pi} + \epsilon e - \epsilon e_j$ ($j = 0, 1, \dots, N$; $e_0 = 0$) can be supported as an average pseudo-perfect Nash equilibrium for any $T > 1$. Consider for example $\bar{\Pi} + \epsilon e$. Let T be given. Let the intended path for all τ be for each player i to play $\bar{\sigma}_i + d\sigma_i^0$. If any player i deviates from the intended path at time $\tau < T$, then from $t = \tau+1$ to $t = T-1$, all players play $\bar{\sigma}$. At time T they all play μ . For ϵ small, the facts that $\bar{\sigma}$ is a Nash equilibrium and Σ_i is compact and Π_i is continuous implies that

$$\max_{\sigma_i \in \Sigma_i} \Pi_i(\sigma_i, \bar{\sigma}_{-i} + d\sigma_{-i}^0) - \bar{\Pi}_i < \bar{\Pi}_i - \Pi_i(\mu).$$

The theorem now follows from Lemma 1 and the continuity property. Q.E.D.

Corollary (of Proof): Let G be a generic smooth game. Let $\epsilon > 0$ and $x = (x_1, \dots, x_N) \gg (v_1, \dots, v_N)$ be given, with $x \in C(G)$. Then for every T there is an N -tuple of strategies $*\sigma^T$ such that

- (1) the total gains to deviating are less than ϵ , i.e. $*\sigma^T$ is perfectly enforced by the reward structure $(\epsilon, \epsilon, \dots, \epsilon)$ and
- (2) $\lim_{T \rightarrow \infty} \Pi_i(*\sigma^T) = x_i$.

The above corollary shows that for generic smooth games, Radner's ϵ -average-equilibrium theorem can be strengthened to an ϵ -total-equilibrium theorem.

6. Two-Person Matrix Games with Mixed Strategy Nash Equilibria

Consider now a two person $m \times n$ matrix game $G = A, B$, like the one given in the introduction:

$$G = \begin{bmatrix} (1, 0) & (0, 1) \\ (0, .8) & (.9, 0) \end{bmatrix}$$

It would seem difficult to support the alternation of payoffs $(0, 1)$ and $(1, 0)$ in a perfect PNE, because in period T it is possible to reward one player or the other, but not both. In fact, it can be shown that if the last period is discounted by any factor $\delta > 0$, then there is always an open set of matrix games for which $\lim_{T \rightarrow \infty} \text{PNE}(G^T) \neq \text{NE}(G^\infty)$.

Observe now that G has a (unique) mixed strategy Nash equilibrium $(p^* = (.8/1.8, 1/1.8), q^* = (.9/1.9, 1/1.9))$. Furthermore, this Nash equilibrium is nondegenerate, or not perfectly safe: by choosing q near q^* player II can change player I's payoffs, and vice versa. Nondegeneracy of course is a generic property: for almost any matrix game G , if G has any mixed strategy Nash equilibria, then all of them are nondegenerate.

Thus the "mixed extension" of the typical matrix game with mixed strategy Nash equilibria has properties very similar to the smooth games of the last section. There is an essential difference, however, in that in matrix games only the realized outcome is observable. In the proof for smooth games, all cheating was immediately punished. When a player's "intended" move in a given period is a mixture, he can always pretend that his personally favorable realization was a matter of luck. It is this difficulty that we must deal with, and it is for this reason that the full last period is needed. For smooth games, the last period can be discounted by any finite δ without affecting the result.

Theorem 2. Let $G = [A, B]$ be any $m \times n$ matrix game with a nondegenerate mixed strategy Nash equilibrium. Let G also be such that either there is a unique (i, j) with $A_{ij} = \text{Max}_{k, \ell} A_{k\ell}$ or else a unique (i', j') with $B_{i', j'} = \text{Max}_{k, \ell} B_{k\ell}$. Then $\lim_{T \rightarrow \infty} \text{perfect PNE}(G^T) = \text{NE}(G^\infty)$.

We defer the proof to the appendix.

7. The $2 \times 2 \times 2$ Case

Theorem 3. Let $G = [A, B]$ be a two player, 2×2 matrix game. Suppose there are at least two Nash equilibria of G with distinct payoffs, or one Nash equilibrium that can be strictly Pareto dominated by something in the convex hull of the entries of G . Then $\lim_{T \rightarrow \infty} \text{perfect PNE}(G^T) = \text{NE}(G^\infty)$.

This theorem covers games like the Prisoner's dilemma, which has a unique, pure strategy Nash equilibrium. We do not bother to give the proof, which consists of an exhaustive examination of cases. The theorem is false in general for $2 \times 3 \times 3$ games.

8. Perfect Nash Equilibria

The continuation principle and the reusable reward system we introduced earlier apply not only to our pseudo Nash equilibria, but also to Radner's ϵ -equilibria, the ϵ -crazy equilibrium notion, perfect Nash equilibria, and Nash equilibria. In this section we use our constructions to give a very simple, alternative proof of the important theorem by Benoit and Krishna.

Theorem 4 (Benoit-Krishna). Let G satisfy the following properties:

- (1) There are multiple Nash equilibrium strategy tuples, $l_\sigma^1, \dots, l_\sigma^L$ whose payoff N -tuples $*\Pi^1, \dots, *\Pi^L$; $*\Pi^l = \Pi(l_\sigma^l)$ satisfy

$$*\Pi_i \equiv \frac{1}{L} \sum_{l=1}^L *\Pi_i^l > \text{Min}_{l=1, \dots, L} *\Pi_i^l = *\Pi_i^{l(i)}, \quad i = 1, \dots, N, \text{ and}$$

- (2) There are $N+1$ strategy N -tuples $l_\sigma^1, \dots, l_\sigma^{N+1}$ whose payoffs $\Pi(l_\sigma^1), \dots, \Pi(l_\sigma^{N+1})$ have convex hull in \mathbb{R}^N with nonempty interior.

Then $\lim_{T \rightarrow \infty} \text{(perfect) NE}(G^T) = \text{NE}(G^\infty)$.

Proof. The idea of the proof is very simple. If there are $N+1$ one-shot Nash equilibria $\{\sigma^0, \sigma^1, \dots, \sigma^{N+1}\}$ among the L Nash equilibria which satisfy $\Pi_i(\sigma^j) - \Pi_i(\sigma^i) > 0$ for all $i = 1, \dots, N$, $j = 0, \dots, N$, $j \neq i$, then a T'' -period reusable reward system can easily be constructed by repeating each one-shot NE T'' times. In this way the reward structure can be made arbitrarily large. Slightly more generally, if the span of the payoffs of the one-shot NE's is full (i.e. N -) dimensional, then one can alternate them in the right proportions to construct $N+1$ different T'' period perfect NE's which also form a reusable reward system. Again the reward structure can be made arbitrarily large.

Furthermore, even if the one-shot NE payoffs do not span a full dimensional set, if their payoffs are distinct, then it is always possible to use them to make a large threat. (The intended sequence of one-shot NE's alternates the L NE's in cyclical fashion. Any deviation by player i is punished by switching to his worst NE. If there are enough such L -cycles, the threat is arbitrarily large). But then any one-shot move can be enforced in period 1 if it is followed by a sequence of one-shot Nash equilibria as above. But this is simply to say (in view of assumption 2) that for some K , the perfect NE's of G^K are full dimensional. But we have already said that in this case there is nothing more to show beyond Lemma 1.

BIBLIOGRAPHY

- [1] Abreu, D., "Infinitely repeated games with discounting: A general theory," Harvard mimeo, 1984, Princeton thesis, 1982.
- [2] Aumann, R., "Lecture notes on game theory," (1977).
- [3] Benoit, J., and V. J. Krishna, "Finitely repeated games," Econometrica, 53 (1985), 905-922.
- [4] Chou, C., Two Essays on Game Theory, Ph.D Dissertation, Department of Economics, Yale University, 1984.
- [5] Dubey, P. and A. Rogawski, "Inefficiency of Nash equilibria," CFDP No. 622, March 1982.
- [6] Fudenberg, D. and D. Levine, "Subgame-perfect equilibria of finite- and infinite-horizon games," JET, 31 (1983), 251-268.
- [7] _____ and E. Maskin, "The folk theorem in two-person repeated games with discounting and incomplete information," mimeo, July 1983.
- [8] Geanakoplos, J. and D. Brown, "Understanding overlapping generations economies as lack of market clearing at infinity," mimeo, 1983.
- [9] Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson, "Rational cooperation in the finitely repeated prisoners' dilemma," JET, 27 (1982), 245-252.
- [10] Kreps, D. M. and R. Wilson, "Sequential equilibria," Econometrica, 50 (1982), 863-894.
- [11] Luce, R. P. and H. Raiffa, Games and Decisions, Wiley, New York, 1957.
- [12] Milgrom, P. and J. Roberts, "Limit pricing and entry under incomplete information: An equilibrium analysis," Econometrica, 50 (1982), 443-459.
- [13] Radner, R., "Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives," JET, 22 (1980), 136-154.
- [14] Rubinstein, A., "Equilibrium in supergames with the overtaking criterion," JET, 21 (1979), 1-9.
- [15] _____, "Perfect equilibrium in a bargaining model," Econometrica, 50 (1982), 97-110.
- [16] _____, "Strong perfect equilibrium in supergames," International Journal of Game Theory, 9 (1979), 1-12.
- [17] Selten, R., "Re-examination of the perfectness concept for equilibrium points in extensive games," International Journal of Game Theory, 1 (1975), 25-55.

APPENDIX

Let us give a proof of Theorem 2. Let $G = [A, B]$ be a matrix game with a mixed strategy Nash equilibrium p^* , q^* . Let A and B be normalized so $\max_{ij} A_{ij} = 1 = \max_{kl} B_{kl}$, and $\min_{ij} A_{ij} = 0 = \min_{kl} B_{kl}$. Let $w_1^* = p^* A q^*$, and let $w_2^* = p^* B q^*$. We shall prove in Part I that for any $\delta > 0$ it is possible to construct for some T , a T -period perfect PNE μ^T with residue $\Pi_i^T(\mu^*) - (T-1)w_i^* > 1-\delta$, for $i = 1, 2$. In Part II we use this perfect PNE to construct the usual reusable reward system.

Proof of Part I:

Let us assume for now that $p^* \gg 0$, and $q^* \gg 0$.

Since (p^*, q^*) is non-degenerate, let us suppose that by perturbing player two's strategy we can improve player one, hence, there exists a $q = (q_1, q_2, \dots, q_n)$ such that $\sum_{j=1}^n q_j = 0$ (hence, $p^* B q = 0$) and $p^* A q > 0$. For q sufficiently small, $\tilde{q} = q^* + q$ is still a completely mixed strategy. Later we shall make q even smaller by multiplying by $1/k$. Consider the one-shot strategy pair $(p^*, \tilde{q}) = (p^*, q^* + q)$. The one-shot payoffs are

$$\Pi_1 = p^* A (q^* + q) > p^* A q^* = w_1^*$$

and

$$\Pi_2 = p^* B (q^* + q) = p^* B q^* + p^* B q = w_2^* + 0 = w_2^*.$$

Since player one still plays the NE strategy, p^* , player two is indifferent between any two strategies in the one-shot game. But p^* is not an optimal strategy for player one when player two plays \tilde{q} .

Consider the repeated game G^T . Suppose player one plays p^* and

player two plays \tilde{q} all the time except the last move. If we use the payoff of the last move T to compensate player one such that he is indifferent between all strategies, and at the same time guarantee player two an expected payoff arbitrarily close to 1 on the last move, then we can derive a PNE with a sufficiently large residue for both players. Against p^* player two has no incentive to cheat, and player one is steadily gaining his residue. One problem is that after paying two's residue in period N we have only a very limited resource, that is, $1 - \text{two's residue}$, to compensate player one for not cheating in the earlier periods. Another problem is that we can observe only the realization of one's randomization at each move, not how he randomizes.

Let

$$c_j = (\text{expected one-shot payoff for player 1 from strategy } j \\ \text{against } \tilde{q}) - w_1^*$$

$$(c_1, c_2, \dots, c_m) = Aq,$$

and let $Y^j(t)$ be the total number of times player one plays strategy j in rounds $1, 2, \dots, t$. Define $X(0) = 0$ and

$$X(t) = \sum_{j=1}^m Y^j(t) c_j / k, \quad t = 1, 2, \dots, T-1.$$

$X(t) + tw_1^*$ is player one's expected payoff through period t , given his choices for $\tau \leq t$, and given that player two randomizes according to $\tilde{q}_k = q^* + q/k$ at each move. Let $s > 0$ be given. For large enough k , $|X(t+1) - X(t)|$ will always be less than s . Let $\bar{a} = \text{Max}\{a_{ij} | b_{ij} = 1\}$. It follows from continuity that for any c , with $\bar{a} < c < 1$, there are one shot strategies $p(c)$ and $q(c)$ with $p(c)Aq(c) = c$; moreover, for c near \bar{a} , $p(c)Bq(c)$ may be taken near 1.

Define a PNE, (μ_1^*, μ_2^*) as follows:

- (1) At the t^{th} move ($1 \leq t < T$),
- i) if $-s < X(\tau) < 1 - \bar{a} - 3s$ for all $\tau < t$, then

$$(p^*, q_k)$$
 is played, where $q_k = q^* + q/k$;
 - ii) if $-s \geq X(\tau)$ or $X(\tau) \geq 1 - \bar{a} - 3s$ for some $\tau < t$, (p^*, q^*) is played.

Define

\bar{t} = the first time $X(t)$ crosses either one of the boundaries,
 $-s$ or $1 - \bar{a} - 3s$,

$T-1$ if $X(t)$ never cross the boundaries.

- (2) At T^{th} move, suppose $X(\bar{t}) = 1 - 2s - c$. Then $(p(c), q(c))$ with $p(c)Aq(c) = c$ as defined above is played. Observe that for large enough k , c satisfies $\bar{a} < c < 1$.

The above rules specify strategies for the two players.

Since player one always plays p^* except the last move, and since player two cannot affect the last move, player two is indifferent between all strategies. Hence, to test that (μ_1^*, μ_2^*) is a pseudo-Nash equilibrium, we need only check that player one cannot gain by deviating during $t = 1, 2, \dots, T-1$. But it is obvious that no matter what strategy player one adopts, if player two plays according to μ_2^* and player one does not vary at time T , then his expected payoff is:

$$(T-1)r_1^* + 1 - 2s.$$

Thus, (μ_1^*, μ_2^*) is certainly a PNE. Hence, for k sufficiently large so that $|X(t+1) - X(t)| < s$,

$$r_1(\mu_1^*, \mu_2^*) = 1 - 2s.$$

Notice that when the players stick to the strategies μ_1^* and μ_2^* , the path of $X(t)$ is a drifting process with mean drift p^*Aq/k and variance proportional to $1/k^2$. Hence, for k sufficiently large, the probability of crossing the top boundary if player one always player p^* approaches 1. At the top boundary, the period T payoff goes almost entirely to player two. Hence, given any $h > 0$, for k sufficiently large and s sufficiently small, player two's expected payoffs is

$$\Pi_2^N(\mu_1^*, \mu_2^*) \geq (T-1)w_2^* + (1-h)(1-s)$$

and the residue is

$$r_2(\mu_1^*, \mu_2^*) \geq (1-h)(1-s) .$$

Finally we note that dropping the hypothesis that $p^* \gg 0$ and $q^* \gg 0$ changes almost nothing in the proof. Player two is still indifferent, against μ_1^* , to using any strategy μ_2 which only involves playing one shot strategies j with $q_j^* > 0$. Doing anything else can only make player two worse off. As for player one, we still define $(c_1, c_2, \dots, c_m) = Aq$ and $X(t)$ as before. If p^* is not strictly positive, then player one's expected payoff through period t is bounded above by $X(t) + tv_1$; equality necessarily holds only when player one uses one shot strategies i with $p_i^* > 0$.
Q.E.D.

Proof of Part II

We can now show how the perfect PNE μ^* can be used to construct the reusable reward system posited in the continuity principle.

Since (p^*, q^*) is a nondegenerate mixed strategy, it is strictly Pareto dominated, and there are strategies σ^A and σ^B and integers $k \leq K$

such that $x = \frac{k}{K}\Pi(\sigma^A) + \frac{K-k}{K}\Pi(\sigma^B) \gg (w_1^*, w_2^*)$. Let $\Pi(\sigma^A) \equiv (a, b) \equiv ab$, and $\Pi(\sigma^B) \equiv (a', b') \equiv a'b'$. We shall support the average payoff x as a perfect PNE for a game with arbitrarily large T . Once one has two perfect PNE's, whose total difference in payoffs can be made arbitrarily large, then the proof is finished as in our proof of Benoit-Krishna.

Of course the idea is to alternate the payoffs ab and $a'b'$ over each cycle of K periods so that ab occurs k times and $a'b'$ $K-k$ times. At the end comes the PNE μ^* . In case there is any deviation, both players get (w_1^*, w_2^*) till the very end. Thus by cheating a player loses $1-\delta$ from the final phase.

Lemma. Let (a, b) and (a', b') be given. Let $x = \frac{k}{K}(a, b) + \frac{K-k}{K}(a', b')$. Then there exists a function $f : \{1, \dots, K\} \rightarrow \{(a, b), (a', b')\}$ such that $\#f^{-1}(a, b) = k$, and such that for any $1 \leq t \leq K$,

$$(1) \quad \sum_{\tau=t}^K f_1(\tau) \geq (K-t+1)x_1 \quad \text{and}$$

$$(2) \quad \sum_{\tau=t}^K f_2(\tau) \geq (K-t)x_2 .$$

Suppose player 1 deviates at some time t . He gets at most $1 + (K-t)w_1^*$ until the end of the cycle, instead of $(K-t+1)x_1$. His net gain is $(1 - x_1) - (K-t)(x_1 - w_1)$. Similarly player 2's maximum gain is $1 - (K-t)(x_2 - w_2)$. Since $x_i > w_i$, these numbers must be less than $1-\delta$ (for sufficiently small δ), except when $t = K$.

Thus the only time a player can gain as much as 1 is on the last period, if the payoff is supposed to be $a0$ (or $0b'$). We can always replace $a0$ and $0b'$ with $\bar{1}\bar{b}$ and $\bar{a}1$. If there is a unique ij with $A_{ij} = 1$, or a unique $k\ell$ with $B_{k\ell} = 1$, then either $\bar{1}\bar{b}$ or $\bar{a}1$ has the property that neither player can change his strategy alone and gain 1. Q.E.D.