Reprinted from JOURNAL OF ECONOMIC THEORY

Vol. 54, No. 1, June 1991
Printed in Belgium

An Approximate Folk Theorem with Imperfect Private Information*

DREW FUDENBERG

Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

AND

DAVID K. LEVINE†

Department of Economics, University of California, Los Angeles, California 90024

Received April 24, 1989; revised September 7, 1990

This paper studies repeated games in which players are imperfectly informed about the uncertain consequences of their opponents stage game actions. We show that if the game is informationally connected, the set of sequential equilibrium payoffs includes the enforceable mutually punishable set, if the intertemporal criterion is (i) the lim inf time average, or (ii) the limit of ε-equilibria with discount factor δ, and $\varepsilon \to 0$ as $\delta \to 1$. We also explore the link between these two criterion functions. *Journal of Economic Literature* Classification Numbers: 021, 022, 026.

1. INTRODUCTION

We give a partial folk theorem for the approximate equilibria of a class of repeated games with imperfect information satisfying an informational linkage condition. Every payoff vector which exceeds a mutual threat point and is generated by one-shot mixed strategies from which no player can profitably deviate without having some effect on other players, is approximately an approximate sequential equilibrium payoff if the discount factor is close enough to one.

The class of repeated games we consider has complete but imperfect information. Each period, player i chooses an action a_i, then observes his

* We are grateful to Andreu Mas-Colell and the referee. NSF Grants 89-07999, 87-08616 and 86-09697 and a grant from the UCLA Academic Senate provided financial support.
† Part of this research was conducted while at the University of Minnesota.

0022-0531/91 $3.00

own payoff and also a signal z_i of the play of his opponents. This model includes as a special case games of imperfect public information as defined by Fudenberg, Levine, and Maskin [5]. In these games, there is a publicly observed variable $\tilde{y}$, and each player i observes $z_i = (\tilde{y}, a_i)$. The public information case includes the literature on repeated oligopoly (Green and Porter [10], Abreu, Pearce, and Stacchetti [1]), repeated principal-agent games (Radner [21, 22], Rubinstein and Yaari [25]), repeated partnerships (Radner [23], Radner, Myerson, and Maskin [24]), as well as the classic case of observable actions (Aumann and Shapky [29], Rubinstein [30], Friedman [3], Fudenberg and Maskin [6]). It also includes the examples studied in the literature of repeated games with "semi-standard information," as discussed in Sorin [26].

While examples of repeated games that have been formally studied have public information, players have private information in some situations of economic interest. Indeed, the central feature in Stigler's [28] model of secret price-cutting is that each firm's sales depend on the prices of its rivals, but firms only observe their own demand. The goal of this paper is to show that a partial folk theorem extends to these cases, provided we relax the notion of equilibrium slightly.

The key element in any folk theorem is that deviators must be punished. This requires, with three or more players, that non-deviators coordinate their punishments, and may even require a player to cooperate in his own punishment. With public information, the necessary coordination can be accomplished by conditioning play on the commonly observed outcome. When players receive different private signals they may disagree on the need for punishment. This raises the possibility that some players may believe that punishment is required while others do not realize this. The most straightforward way to present such confusion from dominating play is to periodically "restart" the strategies at commonly known times, ending confusion, and permitting players to recoordinate their play. This, however, makes it difficult to punish deviators near the point at which play restarts and forces us toward approximate equilibrium.

The use of approximate equilibrium leads to an important simplification, because it allows the use of review strategies of the type introduced by Radner [21]. Under these strategies, each player calculates a statistic indicating whether a deviation has occurred. If the player's statistic crosses a threshold level at a commonly know time, he communicates this fact during a communication stage of the type introduced by Lehrer [12]. Our informational linkage condition ensures that all players correctly interpret this communication, so that the communications stage allows coordination of punishments. The importance of approximate equilibrium is that Radner-type review strategies do not punish small deviations, that is, those that do not cause players statistics to cross the threshold level. Consequently, there will typically be small gains to deviating. A second consequence of approximate equilibrium is that sequentiality loses its force, and the review strategies are perfect whenever they are Nash. This is because punishments end when the strategies are restarted so the cost to punishers is small. Thus the punishments are "credible" in the sense that carrying them out is an approximate equilibrium.

Our results are closely connected to those of Lehrer [12–17] who considers time-average Nash equilibrium of various classes of games with private information and imperfect monitoring but with non-stochastic outcomes. In addition to the deterministic nature of outcomes, Lehrer's work has a different emphasis than ours, focusing on completely characterizing equilibrium payoffs under alternative specifications of what happens when a time average fails to exist.

A secondary goal of our paper is to clarify the connection between Lehrer's work and other work on repeated time-average games, and work on discounted games such as ours. In particular we exposit the connection between approximate discounted equilibrium and exact time-average equilibrium, which is to some extent already known to those who have studied repeated games. We emphasize the fact that the time-average equilibrium payoff set includes limits of approximate as well as exact discounted equilibria. Although we show that the converse is false in general, the type of construction used by Lehrer (and us) yields both an approximate discounted and a time-average theorem. For economists, who are typically hostile to the notion of time-average equilibrium, we hope to make clear that results on repeated games with time-average payoffs are relevant, provided the logic of approximate equilibrium is accepted.

If we consider the set of equilibrium payoffs as potential contracts in a mechanism design problem, there is another interpretation of approximate equilibrium. The set of equilibrium payoffs in our theorem in some cases strictly exceeds that for exact equilibria. This means that exact equilibrium may be substantially too pessimistic. The efficiency frontier for contracts may be substantially improved if people can be persuaded to forego the pursuit of very small private benefits. For example, even if ethical standards have only a small impact on behavior, they may significantly improve contracting possibilities.

2. The Model

In the *stage game*, each player i, $i=1$ to N, simultaneously chooses an action a_i from a finite set A_i with m_i elements. Each player observes an outcome $z_i \in Z_i$, a finite set with M_i elements. We let $z=(z_1, ..., z_n)$ and $Z = \times_{i=1}^{n} Z_i$. Each action profile $a \in A = \times_{i=1}^{n} A_i$ induces a probability distribution $\pi_z(a)$ over outcomes z. Each player i's realized payoff $r_i(z_i)$ depends on his own observed outcome only; the opponents' actions matter only in their influence on the distribution over outcomes.

Player i's expected payoff to an action profile a is

$$g_i(a) = \sum_{z \in Z} \pi_z(a)\, r_i(z_i).$$

We also define $\bar{d} = \max_{i, a_i, a_i', a_{-i}} g_i(a_i, a_{-i}) - g_i(a_i', a_{-i})$ to be the maximum one-period gain any player could obtain by playing one of his actions a_i instead of another a_i'.

We will also wish to consider various types of correlated profiles. A *correlated action profile* α is simply a probability distribution over A. We write α_{-i} for the induced probability distribution over $A_{-i} \equiv \times_{j \neq i} A_j$. A *belief* by player i is such a probability distribution, and the vector (a_i, α_{-i}) represents the correlated action in which player i plays a_i and the other players correlate according to α_{-i}. We also write α_i for the marginal induced over A_i by α. If the play of players is independent, we refer to α as a *mixed action profile*, in which case α is characterized by and may be identified with $(\alpha_1, ..., \alpha_n)$, the vector of marginals.

For any correlated action profile α, we can calculate the induced distribution over outcomes

$$\pi_z(\alpha) = \sum_{a \in A} \pi_z(a)\, \alpha(a).$$

We can also calculate the marginal distribution over player i's outcomes

$$\pi_i(z_i, \alpha) = \sum_{\{\tilde{z} \in Z \mid \tilde{z}_i = z_i\}} \pi_{\tilde{z}}(\alpha).$$

Finally, we may calculate the expected payoff to player i

$$g_i(\alpha) = \sum_{z_i \in Z_i} \pi_i(z_i, \alpha)\, r_i(z_i) = \sum_{z \in Z} \pi_z(\alpha)\, r_i(z_i).$$

In the repeated game, in each period $t = 1, 2, ...,$ the stage game is played, and the corresponding outcome is then revealed. The *history for player i* at time t is $h_i(t) = (a_i(1), z_i(1), ..., a_i(t), z_i(t))$. We also let $h_i(0)$ denote the null history existing before play begins. A *strategy for player i* is a sequence of maps $\sigma_i(t)$ mapping his private history $h_i(t-1)$ to probability distrtibutions over A_i.

A *system of beliefs* b_i for player i specifies for each time t and private history $h_i(t)$, a probability distribution $b_i(h_i(t))$ over private histories of other players of length t. A profile of beliefs b for all players is *consistent* with the strategy profile σ, if for every finite time horizon the truncation of b is consistent with the truncation of σ in the sense of Kreps and Wilson [11]. This requires that there exists for each truncation T a sequence of truncated strategy profiles $\sigma_n^T \to \sigma^T$, putting strictly positive probability on every action, such that the unique sequence of truncated belief profiles derived from Bayes law, $b_n^T \to b^T$.

Given a system of beliefs b_i and a history $h_i(t)$ a distribution is induced over the history of all players play. Given this distribution and the strategy profile σ, a corresponding distribution is induced over play at all times τ. In turn this gives an expected payoff in period τ conditional $h_i(t)$ to player i, which we denote by $G_i(\tau, h_i(t), b_i, \sigma)$. The corresponding normalized present value to player i at discount factor $0 \leqslant \delta < 1$ is

$$W_i(h_i(t), b_i, \sigma, \delta) = (1-\delta) \sum_{\tau=1}^{\infty} \delta^{\tau-1} G(\tau, h_i(t), b_i, \sigma).$$

While if $\delta = 1$

$$W_i^T(h_i(t), b_i, \sigma) = (1/T) \sum_{\tau=1}^{T} G_i(\tau, h_i(t), b_i, \sigma),$$

and

$$W_i(h_i(t), b_i, \sigma, 1) = \liminf_{T \to \infty} W_i^T(h_i(t), b_i, \sigma).$$

Notice that in the case of the initial null history $h_i(0)$, the present value of the time-average expected payoff is the same for all beliefs consistent with σ. The emphasize this, we write $W_i(\sigma, \delta)$ and $W_i^T(\sigma)$ in place of $W_i(h_i(0), b_i, \sigma, \delta)$ and $W_i^T(h_i(0), b_i, \sigma)$.

For $\delta < 1$, a strategy profile σ is an *ε-Nash equilibrium* if for each player i and strategy σ_i'

$$W_i((\sigma_i', \sigma_{-i}), \delta) \leqslant W_i(\sigma, \delta) + \varepsilon. \tag{2.1}$$

There is also a corresponding notion of a *truncated ε-Nash equilibrium*, where we require

$$W_i^T(\sigma_i', \sigma_{-i}) \leqslant W_i^T(\sigma) + \varepsilon.$$

An *ε-sequential equilibrium* is defined in a similar way, except that we require there exist beliefs b consistent with σ such that

$$W_i(h_i(t), b_i, (\sigma_i', \sigma_{-i}), \delta) \leqslant W_i(h_i(t), b_i, \sigma, \delta) + \delta^{t-1}\varepsilon \tag{2.2}$$

for all players i, and all histories $h_i(t)$ (not merely the null initial history). Notice that W_i is measured as a present value at time 1, rather than time t. Consequently, an ε-sequential equilibrium is defined so that the gain to deviating at time t is of order ε, in time-t units. Notice that this definition is stronger than the usual version, found in Fudenberg and Levine [9], for example: ordinarily the gain to deviating at time t is measured in time-0 units, not time-t units.

For $\delta = 1$, we will impose a regularity condition on the equilibrium. A *uniform Nash equilibrium* is a profile σ such that

$$W_i^T(\sigma) \text{ converges,} \tag{2.3}$$

and

$$\text{for all } \rho > 0, \text{ and } \tau \text{ there exists a } T \geqslant \tau \text{ such that}$$
$$W_i^T(\sigma_i', \sigma_{-i}) \leqslant W_i^T(\sigma) + \rho. \tag{2.4}$$

A *uniform sequential equilibrium* is a profile σ such that

$$W_i^T(h_i(t), b_i, \sigma) \text{ converges for all histories } h_i(t) \text{ and consistent beliefs } b, \tag{2.5}$$

and

$$\text{there exist consistent beliefs } b \text{ such that for all } \rho > 0 \text{ and } \tau \text{ there exists a } T \geqslant \tau \text{ such that for all histories } h_i(t) \text{ and strategies } \sigma_i,$$
$$W_i^T(h_i(t), b_i, (\sigma_i', \sigma_{-i})) \leqslant W_i^T(h_i(t), b_i, \sigma) + \rho. \tag{2.6}$$

When σ satisfies condition (2.3) we call it *regular*; if it satisfies (2.5) we say it is *sequentially regular*.

Our definitions of time-average equilibrium require that the appropriate regularity condition be satisfied. In other words, we have required that the time-average payoff exist if no player deviates, and that there be a uniform bound on the gain to deviating in any sufficiently long subgame. Both conditions are responses to our unease in comparing lim infs or lim sups of payoffs when the limit does not exist. We impose the conditions because we interpret time averaging as the idealized version of a game with long finite horizon or very little discounting. Below we show that non-uniform equilibria cannot always be interpreted in this way.

This notion of equilibrium strengthens that of *lower equilibrium*, which assigns the lim inf to time averages that do not converge by adding a uniformity criterion. If the equilibrium time averages converge, as we require, a stronger notion is that of *upper equilibrium*, which assigns the lim sup to deviations with time averages that do not converge. With our uniformity criterion, this would involve replacing "for all τ there exists a $T \geqslant \tau$ such that" in (2.4) and (2.6) with "there exists τ for all $T \geqslant \tau$ such that." We would have preferred to prove the stronger criterion was satisfied, but have not been able to do so.

3. Discounting and Time Averaging

Our goal is to characterize the set of equilibrium payoffs when players are very patient. In the next section we characterize truncated ε-Nash equilibrium payoffs with small ε as containing a certain set V^*. Since this notion of equilibrium is not the most economically meaningful, we first show that this is in fact a strong characterization of payoffs.

Specifically we show

Theorem 3.1. *Suppose there exists a sequence of times T^n, non-negative numbers $\varepsilon^n \to 0$, and strategy profiles σ^n such that σ^n is a T^n-truncated time-average ε^n-Nash equilibrium and $W_i^{T_n}(\sigma^n) \to v_i$. Then*

(A) *There exists a sequence of discount factors $\delta^n \to 1$, non-negative numbers $\varepsilon^n \to 0$, and strategy profiles σ^n such that σ^n is an ε^n-sequential equilibrium for δ^n and $W_i(\sigma^n, \delta^n) \to v_i$.*

(B) *There exists a uniform sequential equilibrium strategy profile σ such that $W_i(\sigma, 1) = v_i$.*

Remark 1. It is well known that the hypotheses of the theorem imply that if $\sigma^n = \sigma$, σ is a time-average Nash equilibrium (see, for example, Sorin [26]). The connection between the hypothesis of the theorem and (A) has been studied in context of dynamic programming by Lehrer and Sorin [19]. They show that the uniform convergence of the discounted maximized values is equivalent to the uniform convergence of the finite time-average maximized values.

Remark 2. The hypothesis is weak in the sense that only approximate Nash equilibria are required. However, as long as we consider approximate equilibrium, or infinite time-average equilibrium, (A) and (B) show that sequentiality has little force. Roughly, the reason is that after a deviation, no further deviations are anticipated, so the cost of punishment must be only paid once. This has negligible cost if players are very patient. It is easily seen in the following example, which is a special case of the main theorem in the next section. Consider a game in which players perfectly observe each other's past play, and suppose there is a pure action profile a for which $g(a)$ strictly exceeds the minmax for all players. Clearly there is a discount factor $1 > \underline{\delta} \geqslant 0$ and number of periods K, such that if $\delta \geqslant \underline{\delta}$ the loss to each player of being minmaxed by his opponents for K periods exceeds any possible one-shot gain of deviating from a. Consider then a strategy profile consisting of a review phase in which all players play a, and n different punishment phases in the ith of which, player i is minmaxed for K periods. Initially, the game is in the review phase, and the review phase continues as long as no player deviates, or more than one player deviates simultaneously. Whenever a single player deviates, a punishment phase against that player immediately begins. The punishment phase continues K periods, regardless of whether any further deviation occurs, then the game restarts with a new review phase. Since no player can profitably deviate during a review phase for $\delta \geqslant \underline{\delta}$, these strategies form a Nash equilibrium with payoffs $g(a)$. It is equally clear that they are not generally sequential (subgame perfect), because players may profitably deviate during punishment phases.

In proving the folk theorem, Fudenberg and Maskin [6] construct strategies that induce the players to carry out punishments by providing "rewards" for doing so. But such rewards are not provided by our review strategies. Indeed, the Fudenberg and Maskin construction is only possibly if the game satisfies a "full-dimensionality" condition that we have not imposed. Thus the conclusion of (A) cannot in general be strengthened to $\varepsilon^n = 0$.

The review strategies do, however, form an $\varepsilon(\delta)$-sequential equilibrium with $\varepsilon(\delta) \to 0$ as $\delta \to 1$. (Note that with observable actions, sequential equilibrium and subgame-perfect equilibrium are equivalent.) To show this, we need only calculate the greatest gain to deviating during a punishment phase. Recall that $\bar{d}$ is the greatest one-shot gain to any player of deviating from any profile. Consequently, the gain is at most $(1-\delta) K\bar{d}$, which clearly goes to zero as $\delta \to 1$. The point is that the punishment lasts only K periods, and players never expect to engage in punishment again. Consequently, with extreme patience, the cost of the punishment is negligible. In particular, with time averaging this is an exact equilibrium.

Proof of (A). Consider the strategy $\hat{\sigma}^n$ of playing σ^n from 1 to T^n, then starting over and playing σ^n from $T^n + 1$ to $2T^n$, and so forth. We call period 1 to T^n "round 1," $T^n + 1$ to $2T^n$ "round 2," and so on. Fix $\delta < 1$ and any beliefs consistent with $\hat{\sigma}^n$. Let us calculate an upper bound on any player's gain to deviating in time-t normalized present value. Fix a time t and a history $(h_1(t), ..., h_n(t))$, and choose k so that $kT^n < t \leqslant (k+1)\, T^n$, so that t is in round k. Since the maximum of per period gain to deviating is

$\bar{d}$, we know that regardless of the history, no player by deviating in round k, can gain more than $(1-\delta)\,T^n\bar{d}$. Further, play from time $(k+1)\,T^n+1$ on is independent of what happened during the previous rounds. Since round $k+1$ is a truncated time-average ε^n-Nash equilibrium, at most $(1-\delta)\,T^n\varepsilon^n$ can be gained by deviating in this round. During round $k+2$ at most $\delta(1-\delta)\,T^n\varepsilon^n$ can be gained, and so forth. Consequently, no player can gain more than $(1-\delta)\,T^n(\bar{d}+\varepsilon^n/(1-\delta^{T_n}))$. As $\delta \to 1$, this quantity approaches ε^n by l'Hôpital's rule. Clearly, then, we may choose δ^n close enough to 1 that $\hat{\sigma}^n$ is a $2\varepsilon^n$-sequential equilibrium. In addition, because of the repeated structure of $\hat{\sigma}^n$, it is clear that $W_i(\hat{\sigma}^n, \delta^n) \to W_i^{T_n}(\sigma^n)$ as $\delta^n \to 1$ so $W_i(\hat{\sigma}^n, \delta^n) \to v_i$. ∎

Proof of (B). Consider the strategy σ of playing σ^1 from 1 to T^1, σ^2 from T^1+1 to T^1+T^2, and so forth. Fix any beliefs consistent with σ. Let a history $(h_1(t), ..., h_n(t))$ be given and let t be in round k. By deviating in the current round, player i can gain at most $\bar{d}$ per period as play in the future is independent of play in this round. The per-period gain at $k+1$ is at most ε^{k+1}, in round $k+2$, ε^{k+2}. Since $\varepsilon^k \to 0$, this implies that the time-average gain is zero. Since $W_i^{T_k}(\sigma^k) \to v_i$, σ is a uniform sequential equilibrium, and this yields the desired conclusion. ∎

Each of the two conclusions of Theorem 3.1 has strengths and weaknesses. The limit of discounting is more appealing than time averaging, which suggests the interpretation (A), but exact equilibrium is more appealing than approximate equilibrium, which suggests the interpretation (B). In this context, it is worth emphasizing reiterating that the conclusion of (A) cannot be strengthened to $\varepsilon^n = 0$. One interpretation of this fact is that the set of time-average equilibria includes not only the limits of exact discounted equilibria, but the limits of approximate discounted equilibria as well.

If we weaken definition (2.6) to allow the length of the gain to deviating, T, to depend on the history and deviation, we have a *time-average sequential equilibrium*, rather than a uniform one. These equilibria include all limits of discounted sequential equilibria as shown by

PROPOSITION 3.2. *If σ is sequentially regular, and is an $\varepsilon(\delta)$-sequential equilibrium for discount factor δ with $\varepsilon(\delta) \to 0$, then σ is a time-average sequential equilibrium.*

Proof. This effectively follows from the fact that the limit points, as $\delta \to 1$, of the normalized present value of a sequence of payoffs are contained in the set of limit points of the finite horizon time averages. This implies that if σ is sequentially regular, then the discounted value of payoffs along the continuation equilibrium approaches the value with time averaging. Deviations when evaluated with the discounting criterion yield values whose limits are no less than the lim inf of the time average. ∎

The results of Lehrer and Monderer [18] imply that if the convergence of the discounted payoffs is uniform in the history, the type of equilibrium in the conclusion may be strengthened to the type of uniform upper equilibrium described above.

It is worth noting that the converse of this proposition is not true. An example of a time-average equilibrium that is not even an approximate equilibrium with discounting can be found in the following "guru" game: Player 1 is the guru. He chooses one of three actions 0, 1, or 2, each period. Player 2 is a disciple. He also chooses one of two actions. Each period the guru receives zero, while the disciple receives an amount equal to the action chosen by the guru. In other words, the guru is completely indifferent, while the disciple cares only about what the guru does. Notice that sequentiality reduces to subgame perfection in this example, and that subgame perfection has no force. Any strategy by the guru is optimal in any subgame, while any Nash equilibrium need only be adjusted for the disciple so that he chooses an optimum in subgames not actually reached with positive probability in equilibrium. (There is a trivial complication in that the guru has strategies such that no optimum for the disciple exists.)

Consider the following equilibrium with time averaging. As long as the disciple never played 2 in the past, the guru plays 1. Let t denote the first period in which the disciple plays 2. Then the guru plays 2 in period $(t+1)$ to $2t$, plays 0, in periods $2t+1, ..., 3t+1$, and in period $3t+2$ the guru reverts back to 1 forever. Regardless of the history the disciple always plays 1. With time averaging the disciple is completely indifferent between all his deviations, since they all yield a time average of 1. Similarly the guru is clearly indifferent between all his actions. Consequently this is a time-average equilibrium. (In fact, it is also an equilibrium when the players' time preference is represented by the overtaking criterion.) On the other hand, with discount factor $\delta \geqslant 1/2$, we can always find a time t such that δ^t is between 1/2 and 1/4, and by deviating at such a time player 2 always gains at least 1/32. Consequently this configuration is only an ε-equilibrium for $\varepsilon \geqslant 1/32$, and in particular, ε does not converge to zero as δ goes to 1.

Note that the counterexample is to the strategy profile σ not being an approximate equilibrium with discounting. From the folk theorem, we know that there are other strategies that are exact equilibria and yield approximately the same payoff. This is not true in games with a Markov structure and absorbing states, as shown by the example of Sorin [27]. In fact, his example has equilibrium payoffs with time averaging that are not even payoffs of approximate equilibria with discounting.

We would argue that time averaging is of interest only insofar as its captures some sort of limit of discounting. Consequently, we argue that this equilibrium of the guru game does not make good economic sense.

4. A Folk Theorem

We now restrict attention to a limited, but important class of games, which we call *informationally connected games.* All two-player games are informationally connected. If $N>2$, we say that player i is *directly connected* to player $j \neq i$ despite player $k \neq i, j$ if there exists a mixed action profile α and a mixed action $\hat{\alpha}_i$ for player i such that $\pi_i(\cdot, \hat{\alpha}_i, \alpha'_k, \alpha_{-i-k}) \neq \pi_j(\cdot, \alpha)$ regardless of the play α'_k of player k. In other words, at α, player i has a deviation that can be potentially detected by player j regardless of how player k plays. We say that player i is *connected* to player j if for every player $k \neq i, j$, there exists a sequence of players $i_1, ..., i_n$ with $i_1 = i$, $i_n = j$, and $i_p \neq k$ for any p, and such that player i_p is directly connected to player i_{p+1} despite player k. In other words, a message can always be passed from i to j, regardless of which single other player might try to interfere. A game is *informationally connected* if every player is connected to every other player. For an example of such a game, suppose that each player determines an output level of zero or one. "Average" price is a strictly decreasing function of total output, and each player observes an "own price" that is an independent random function of average price, with the property that the distribution corresponding to a strictly higher average price strictly stochastically dominates that corresponding to the lower price. In this case, all players can choose zero output levels, and if any player produces a unit of output, this changes the distribution of prices for all other players, with a deviation by a second player merely enhancing the signal by lowering the distribution of prices still further.

A *mutual threat point* v is a payoff vector for which there exists a *mutual punishment action* α such that $g_i(\alpha'_i, \alpha_{-i}) \leqslant v_i$ for all players i and mixed actions α'_i. With three or more players the vector v^* of minmax values need not be a mutual threat point, as there may not be a single action profile that simultaneously holds all of the players to their minmax values. When such a profile exists, the game is said to satisfy the "mutual minmax property" (Fudenberg and Maskin [6]). One example where this property obtains is a represented quantity-setting oligopoly with capacity constraints where the profile "all players produce to capacity" serves to minmax all of the players.

A payoff vector that weakly Pareto-dominates a mutual threat point is called *mutually punishable*; the closure of the convex hull of such payoffs is the *mutually punishable set* $\underline{V}$.

We say that a payoff vector v is (independently) *enforceable* if there exists a mixed action profile α with $g(\alpha)=v$, and such that if for some player i and mixed action α_i', $g_i(\alpha_i', \alpha_{-i})>v_i$, then for some other player $j\neq i$, $\pi_j(\cdot, \alpha_i', \alpha_{-i})\neq \pi_j(\cdot, \alpha)$. In other words, any improving deviation for player i can potentially be detected by some other player j. The *enforceable set* $\tilde{V}$ is the closure of the convex hull of the enforceable payoffs. Notice that in addition to the static Nash equilibrium, which is clearly enforceable, every extremal Pareto efficient payoff is. This is because any extremal payoff is generated by a pure action profile, and any efficient pure action profile is enforceable: if an action profile is not enforceable, one player can strictly improve himself without anyone else knowing or caring. Note also that if the unconditional play in any Nash equilibrium at time t is a mixed (rather than correlated) action profile, it is clear that the profile must be enforceable (except in the infinite time average case, for a negligible fraction of periods). However, it is possible that unenforceable payoffs can be achieved through correlation.

Finally, we define the *enforceable mutually punishable set* $V^*=\underline{V}\cap \tilde{V}$, which is closed, convex, and contains at least the convex hull of static Nash equilibrium payoffs. We can now prove:

THEOREM 4.1 (Folk Theorem). *In an informationally connected game, if* $v\in V^*$, *there exists a sequence of times* T^n, *of non-negative numbers* $\varepsilon^n\to 0$, *and strategy profiles* σ^n *such that* σ^n *is a* T^n*-truncated time-average* ε^n*-Nash equilibrium and* $W_i^{T_n}(\sigma_n)\to v_i$.

Remark. This is a partial folk theorem in that, as remarked earlier, the mutually enforceable payoffs may be a strictly smaller set than the individually rational socially feasible ones. Moreover, in games with imperfect observation of the opponents' actions and three or more players, even the individually rational payoffs may not be a lower bound on the set of payoffs that can arise as equilibria (see Fudenberg, Levine, and Maskin [5] for an example). Moreover, Sorin [27] and Lehrer [12, 14] show that the set of equilibria can include payoffs not in V^*.

The proof constructs strategies that have three stages. In review stages, players play to obtain the target payoff v. At the end of review stages players "test" to see if a statistic they compute exceeds a threshold. Then follows a "communication stage" where players use their actions to "communicate" whether or not the review was passed; the assumption of information connectedness is used to ensure that such communication is possible. Finally, if players learn that the test was failed, they revert to a "punishment state" for the remainder of the game.

The idea of using strategies with reviews and punishments was introduced by Radner [21, 23] who did not need the communication phase because he studied models with publicly observed outcomes. Lehrer [12] introduces the idea of a communications phase to coordinate play. Our proof is in some ways simpler than Radner's as we establish only the existence of ε-equilibria of the finite-horizon games and then appeal to Proposition 3.1. Lehrer's [12] proof is more complex than ours because he obtains a larger set of payoffs.

The key to proving Theorem 4.1 is a lemma proven in the Appendix:

LEMMA 4.2. *If $\bar{\alpha}^h$ is enforceable, then for $j \neq i$ there exists m_j-vectors of weights λ_j such that for all α_i, $\sum_{j \neq i} \lambda_j \pi_j(\cdot, \alpha_i, \bar{\alpha}^h_{-i}) \geq g_i(\alpha_i, \bar{\alpha}^h_{-i}) + \eta$ and $\sum_{j \neq i} \lambda_j \pi_j(\cdot, \bar{\alpha}^h) = g_i(\bar{\alpha}^h) + \eta$.*

This shows how we can reduce the problem of detecting profitable deviations by player i to a one-dimensional linear test. We need only deter α_i's that lead to the information vectors $\bar{\pi}_j(\cdot, \alpha_i, \bar{\alpha}^n_{-1})$ that lie above the half-space whose existence is asserted in the Lemma. This makes the problem very similar to one with publicly observed signals.

Proof of Theorem 4.1. Fix $v \in V^*$, and suppose $\varepsilon > 0$ is given. We show how to find T and σ such that σ is a T-truncated time-average 100ε-Nash equilibrium, and $|W_i^T(\sigma) - v_i| \leq 100\varepsilon$. This clearly suffices. The proof proceeds in several steps. First we construct a class of strategy profiles that depends on a vector of constants L determined by the game and v, and constants l that are free parameters. The constants L will index the relative lengths of various phases of play, and the l's will determine both the absolute length of the phases and the number of times the phases are repeated. The length of the game T is implicitly determined by these constants. We then show how to choose the constants l to make a 100ε-equilibrium yielding payoffs within 100ε of v.

Step 1 (Payoffs). Since $v \in V^*$, $v \in \bar{V}$, and $v \in \underline{V}$. This means we can find finite L'-dimensional vectors of non-negative coefficients $\bar{\mu}^h$, $\underline{\mu}^h$ summing to 1, and of payoffs $\bar{v}^h$, $\underline{v}^h$, with $\sum_{h=1}^{L'} \bar{\mu}^h \bar{v}^h = v$, $\sum_{h=1}^{L'} \underline{\mu}^h \underline{v}^h = v$ such that $\bar{v}^h$ is enforceable and $\underline{v}^h$ is mutually punishable. Corresponding to these are mixed action profiles $\bar{\alpha}^h$ and $\underline{\alpha}^h$, where $\bar{\alpha}^h$ yields payoff $\bar{v}^h$ and $\underline{\alpha}^h$ is a mutual punishment that enforces $\underline{v}^h$. Moreover, it is clear that we can find L^R and non-negative L'-dimensional vectors of integers $\bar{L}^h$, $\underline{L}^h$ that sum to L^R and such that $|\sum_{h=1}^{L'} (\bar{L}^h/L^R)\, \bar{v}^h - v| \leq \varepsilon$ and $|\sum_{h=1}^{L'} (\underline{L}^h/L^R)\, \underline{v}^h - v| \leq \varepsilon$.

Step 2 (Temporal Structure). We first describe the temporal structure of the game, which consists of l' repetitions of a review stage followed by a communications stage. Set $L^C = N(N-2)(N-1)!$. A review stage lasts $l^R L'$ periods, and a communications stage lasts $l^C L^C$ periods. Each stage is further subdivided into phases: a review stage into L^R phases indexed by $h = 1, ..., L'$, and a communications stage into L^C phases, indexed as described below. The temporal structure is outlined in Fig. 1; the game lasts $T = l'(l^R L^R + l^C L^C)$ periods.

The length of a phase of a review stage depends on whether it is regarded as a reward or punishment stage; if it is a reward stage, the hth phase lasts $l^R \bar{L}^h$ periods; if it is a punishment stage it last $l^R \underline{L}^h$ periods. Each phase of a communications stage lasts l^C periods.

Each communications phase is assigned an index (i, j, k) corresponding to a triple of players. (Each index generally occurs more than once.) The first $(N-2)(N-1)!$ phases have indices $(i, j, 1)$; the next $(N-2)(N-1)!$ have indices $(i, j, 2)$, and so forth. Fixing k, the third index, the (i, j) indices are determined by taking the set of all players but k, and calculating every permutation of the set, giving rise to $(N-1)!$ blocks of $(N-2)$ periods. Fix a permutation, $i_1, i_2, ..., i_{N-1}$. Then the first index in the block is (i_1, i_2, k), the second (i_2, i_3, k), and so forth up to (i_{n-2}, i_{n-1}, k). The point of this rather complex structure is that because the game is informationally connected, each player has an opportunity to send a signal to all other players, without being blocked by any other single player.

Using the fact that the game is informationally connected, we associate distinct triples (i, j, k) with certain mixed strategies. If i is connected to j despite k, we say that (i, j, k) is an active link, and let α^{ijk} be the action allowing communication and $\hat{\alpha}_i^{ijk}$ the deviation for player i that enables him

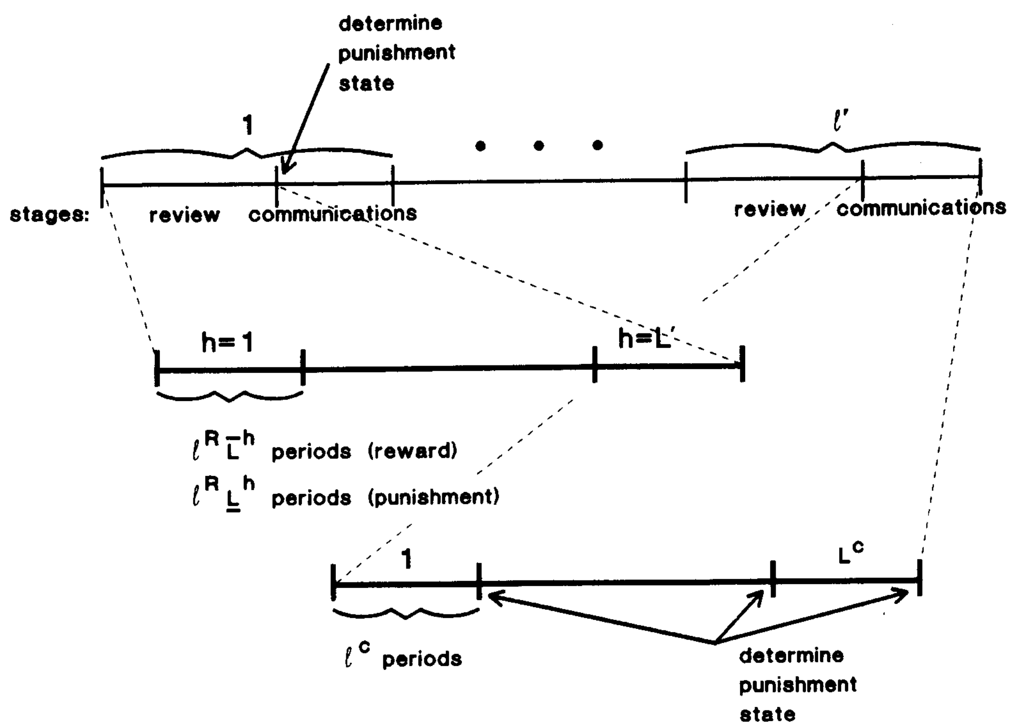

FIG. 1. Temporal structure.

to communicate with j. If i is not connected to j despite k we say that (i, j, k) is an inactive link, and arbitrarily define α^{ijk} to be $\bar{\alpha}^1$, and $\hat{\alpha}_i^{ijk}$ to be $\bar{\alpha}_i^1$.

Step 3 (Strategies). We now describe player i's strategy. He may be in one of two states, a reward state or a punishment state. He begins the game in the reward state. The punishment state is absorbing, so once reached, player i remains there forever. We must describe how a player plays in each state, and when he moves from the reward to punishment state.

In the reward state and review stage, player i regards the stage as divided into L' phases, with phase h lasting $l^R\bar{L}^h$ periods. During the hth phase he plays $\bar{\alpha}_i^h$. In the reward state and communications stage, in the phase indexed by (i', j, k) he plays $\alpha_i^{i'jk}$.

In the punishment state and review stage, player i regards the stage as subdivided into L' phases, with phase h lasting $l^R\underline{L}^h$ periods. During the hth phase he plays $\underline{\alpha}_i^h$. In the punishment state and communications stage, in the phase indexed by (i', j, k) he plays $\alpha_i^{i'jk}$ if $i \neq i'$, and he plays $\hat{\alpha}_i^{ijk}$ if $i = i'$.

Player i can change states only at the end of a review stage, or at the end of a communication phase (j, i, k) that is an active link. The transition is determined by a parameter ε'.

At the end of each reward phase of the review stage player i calculates $\bar{\pi}_i^h(z_i)$ to be the fraction of the $l^R\bar{L}^h$ periods in which z_i occurred. If at the end of the review stage $|\bar{\pi}_i^h(\cdot) - \pi_i(\cdot, \bar{\alpha}^h)| > \varepsilon'$ for any h, player i switches to the punishment state.

At the end of the pth communications phase corresponding to a triple (j, i, k) that is an active link, player i calculates $\bar{\pi}_i^p(z_i)$ to be the fraction of the l^C periods in which z_i occurred. If $|\bar{\pi}_i^p(\cdot) - \pi_i(\cdot, \alpha^{jik})| > \varepsilon'$, player i switches to the punishment state.

In all other cases player i's state does not change; in particular the punishment state is absorbing.

We must show how to choose l', l^R, l^C, and ε' to give expected payoffs within 10ε of v, and no greater than an 100ε gain to any player from deviating.

Step 4 (Payoffs). First we consider the equilibrium payoffs. Let $\bar{\pi}'$ denote the probability that no player ever enters the punishment state and let $\bar{v}'$ be the expected payoff vector conditional on this. Let $\bar{v} = \sum_{h=1}^{L'} (\bar{L}^h/L^R) \bar{v}^h$. The payoff $\bar{v}'$ differs from $\bar{v}$ solely due to communications phases, yielding

$$|\bar{v}' - \bar{v}| \leq \frac{l^C L^C \bar{d}}{l^R L^R + l^C L^C} = \frac{\bar{d}}{1 + (l^R/l^C)(L^R/L^C)}, \tag{4.1}$$

where recall that $\bar{d}$ is the greatest difference between any payoffs. Since $|\bar{v}-v| \leqslant \varepsilon$ by construction, we conclude

$$|W_i^T(\sigma)-v_i| \leqslant \varepsilon + \bar{d}\left[\frac{1}{1+(l^R/l^C)(L^R/L^C)} + (1-\bar{\pi}')\right]. \tag{4.2}$$

This shows we must take l^R/l^C very large and $\bar{\pi}'$ close to one. We will show that the weak law of large numbers implies that $\bar{\pi}'$ close to one if l^R and l^C are both sufficiently large (relative to ε').

Step 5 (Gain to Deviating). Next, we consider how much a player might gain by deviating. During a communications stage (or in any period for that matter) the greatest per-period gain is $\bar{d}$. However, only $l^CL^C/(l^RL^R+l^CL^C)$ of the periods in the game lie in communications phases so we have that the greatest per period gain over the whole game due to deviations during communications is

$$[\text{communications}] \qquad \frac{\bar{d}}{1+(l^R/l^C)(L^R/L^C)}. \tag{4.3}$$

Next we consider review stages. Fixing a deviator, there are three possibilities: all other players are in the punishment state, all are in the review state, or they disagree on the state. Of course a player contemplating deviating may not be certain which of these is true, but since he can only benefit from this extra information, we may suppose he knows which case he faces.

If all other players are in the punishment state they will remain there regardless of player i's play. Regardless of how he plays during a punishment phase, player i gets at most $\underline{v}_i^h$. Since $|\sum (L^h/L^R)\,\underline{v}_i^h - v_i| \leqslant \varepsilon$, the largest gain over $W_i^T(\sigma)$ that player i can obtain by deviating through the punishment state is

$$\begin{aligned}[\text{punishment}] \qquad & \varepsilon + v_i - W_i^T(\sigma) \\ & \leqslant 2\varepsilon + \bar{d}\left[\frac{1}{1+(l^R/l^C)(L^R/L^C)} + (1-\bar{\pi}')\right],\end{aligned} \tag{4.4}$$

where the inequality follows from (4.2).

Next, suppose that other players disagree about the state (so in particular, $N>2$). Then player i can possibly get $\bar{d}$ per period. Let $\hat{\pi}$ be a lower bound on the probability that all opponents agree on a punishment state at the end of the subsequent communications stage (and therefore agree for the rest of the game). Then the deviating player can at best hope to get $\bar{d}$ for one stage with probability $\hat{\pi}$, and $\bar{d}$ in all periods with probability $(1-\hat{\pi})$. Since there are l' review phases, $\bar{d}$ per period for one phase is actually worth $\bar{d}/l'$ yielding

$$\text{[confusion]} \qquad \frac{\hat{\pi}\bar{d}}{l'} + (1-\hat{\pi})\,\bar{d}. \tag{4.5}$$

This leaves review stages where all opponents are in the reward state. Let $1-\pi'$ be an upper bound on the probability that player i both gains more than 2ε and no opponent enters the punishment state at the end of the stage. Then player i can at best gain 2ε in all periods, and can gain more than 2ε and remain with all opponents in the reward state next stage with probability no more than $1-\pi'$. Since at best he can gain $\bar{d}$, player i gains at most

$$\text{[reward]} \qquad 2\varepsilon + \frac{\pi'\bar{d}}{l'} + (1-\pi')\bar{d} \tag{4.6}$$

in reward stages.

Adding these bounds, we find an upper bound on the gain to deviating

$$\text{[total]} \qquad \frac{2\bar{d}}{1+(l^R/l^C)(L^R/L^C)} + 4\varepsilon + \frac{\hat{\pi}\bar{d}}{l'} + (1-\hat{\pi})\bar{d}$$

$$+\frac{\pi'\bar{d}}{l'} + \bar{d}(1-\pi') + \bar{d}(1-\bar{\pi}'). \tag{4.7}$$

Our goal is to find l', l^R, l^C, ε' to simultaneously make (4.2) and (4.6) smaller than 100ε.

Step 6. Fix l' so that $\bar{d}/l' < \varepsilon$. Let $l^C(l^R)$ be the largest integer smaller than γl^R where

$$\frac{\bar{d}}{1+(1/\gamma)(L^R/L^C)} = \varepsilon.$$

We may then simplify (4.2) to

$$|W_i^T(\sigma) - v_i| \leqslant 2\varepsilon + (1-\bar{\pi}')\bar{d}, \tag{4.2'}$$

and (4.7) to

$$8\varepsilon + (1-\hat{\pi})\bar{d} + (1-\pi')\bar{d} + (1-\bar{\pi}')\bar{d}. \tag{4.7'}$$

Consequently, it suffices to show that we can choose l^R and ε' so that when $l^C = l^C(l^R)$, $(1-\bar{\pi}')$, $(1-\hat{\pi})$, $(1-\pi')$ are all less than or equal to $\varepsilon/\bar{d}$.

By Lemma A.1 in the Appendix, for the hth reward phase there exists ε^h and l^h such that for all $\varepsilon' \leqslant \varepsilon^h$ and $l^R \geqslant l^h$ the probability that either i gains less than or equal 2ε or at least one opponent switches to punishment at the end of the stage is at least $(1-\varepsilon/\bar{d})^{1/L^R}$. Consequently, for $\varepsilon' \leqslant \min \varepsilon^h$, $l^R \geqslant \max l^h$, we have $\pi' \geqslant 1-\varepsilon/\bar{d}$.

By Lemma A.2 in the Appendix, for the (j, k, i) that are active communications links there exist ε^{jki} and l^{jki} such that for all $\varepsilon' \leqslant \varepsilon^{jki}$ and $l^C \geqslant l^{jki}$ if player j plays $\hat{\alpha}_j^{jki}$ and players $k' \neq i$ play $\alpha_{k'}^{jki}$ the probability of k switching to punishment at the end of the phase is at least $(1-\varepsilon/\bar{d})^{1/L^C}$. Consequently, for $\varepsilon' \leqslant \min \varepsilon^{ijk}$, $l^C \geqslant \max l^{ijk}$, we have $\hat{\pi} \geqslant 1-\varepsilon/\bar{d}$. Note that $l^C \geqslant \max l^{ijk}$, provided $l^C = l^C(l^R)$ and $l^R \geqslant (1+\gamma)\, l^{ijk}/\gamma$.

Fix then $\varepsilon' = \min\{\varepsilon^h, \varepsilon^{ijk}\}$, and $l^* = \max\{l^h, (1+\gamma)\, l^{ijk}/\gamma\}$. By the weak law of large numbers, there exists an l^{**} such that for $l' \geqslant l^{**}$, $\bar{\pi}' \geqslant 1-\varepsilon/\bar{d}$. Choosing $l' = \max\{l^*, l^{**}\}$ then completes the proof. ∎

Remark. How frequently does punishment occur? As $T \to \infty$, the proof shows that the probability of punishment goes to zero. Examining the proof of Theorem 3.1, we see that infinite equilibria are constructed by splicing together a sequence of truncated equilibria. For approximate discounted equilibria, the construction repeats the same equilibrium over and over, so that the probability of punishment in each round is constant. By the zero-one law, this means punishment occurs infinitely often.

In the case of time-average equilibria, it is irrelevant to payoffs whether punishment occurs infinitely often or not. The truncated equilibria that are spliced together have decreasing probability of punishment, say $\pi_1, \pi_2, \ldots \to 0$. By the zero-one law, if $\sum_{t=1}^{\infty} \pi_t = \infty$, punishment occurs infinitely often, while if $\sum_{t=1}^{\infty} \pi_t < \infty$, punishment almost surely stops in finite time. However, we can choose a subsequence $\{\pi_t'\}$ such that $\sum_{t=1}^{\infty} \pi_t' < \infty$. The corresponding truncated equilibria when spliced together form a time-average equilibrium in which punishment almost surely ceases. On the other hand, we may form a sequence in which the tth truncated equilibrium is repeated $1/\pi_t$ times. Splicing this sequence together clearly yields a time-average equilibrium with $\sum_{t=1}^{\infty} \pi_t' = \infty$, and so punishment occurs infinite often.

This discussion should be contrasted with the results of Radner [21] in which punishment is infinitely often, and Fudenberg, Kreps, and Maskin [4], in which it almost surely ceases.

Appendix

Proof of Lemma 4.2. We may regard $\sum_{j \neq i} M_j + 1$ dimensional Euclidean space as having components π_j, $j \neq i$, and g_i. Define F to be the subset of this space such that there exists a mixed strategy α_i for player i with $\pi_j = \pi_j(\cdot, \alpha_i, \bar{\alpha}_i^h)$ and $g_i \leqslant g_i(\alpha_i, \bar{\alpha}_{-i}^h)$. Let $\bar{\pi}$ be the $\sum_{j \neq i} M_j$ vector with components $\pi_j(\cdot, \bar{\alpha}^h)$, and let $\bar{g}_i = g_i(\bar{\alpha}^h)$. Then $(\bar{\pi}, \bar{g}_i) \in F$ and for $\Delta > 0$, enforceability implies $(\bar{\pi}, \bar{g}_i + \Delta) \notin F$. Moreover, F is a convex polyhedral set and its extreme points correspond to pure strategies for player i.

Since F is convex polyhedral, we may characterize $(\pi, g_i) \in F$ by the linear inequalities $A(\pi - \bar{\pi}) + b(g_i - \bar{g}_i) \geqslant c$, where A is a matrix, b and c are vectors. Consider minimizing $\lambda(\pi - \bar{\pi}) - (g_i - \bar{g}_i)$ subject to this constraint. Suppose this problem has a solution equal to zero. Then $\lambda\pi \geqslant g_i + \eta$ where $\eta = \lambda\bar{\pi} - \bar{g}_i$, and $\lambda\bar{\pi} = \bar{g}_i + \eta$, the desired conclusion.

Since $(\bar{\pi}, \bar{g}_i) \in F$, a feasible solution to the primal exists, so a minimum of zero exists, if and only if the dual has a feasible plan yielding zero. The dual is to maximize ωc subject to

$$(\omega A, \omega b) = (\lambda, -1)$$
$$\omega \geqslant 0.$$

In other words, if we can find $\omega \geqslant 0$, $\omega c = 0$, $\omega b = -1$, then $\lambda = \omega A$ is the desired solution.

Recall that $(\bar{\pi}, \bar{g}_i) \in F$, so $0 \geqslant c$. Moreover, if $A(\bar{\pi} - \bar{\pi}) + b(g_i - \bar{g}_i) \geqslant c$ and $g_i' \geqslant g_i$, $A(\pi - \bar{\pi}) + b(g_i' - \bar{g}_i) \geqslant c$ implying that $0 \geqslant b$. Finally, for $\Delta > 0$, $b\Delta \ngeqslant c$. We conclude for some component p, $c_p = 0$, $b_p < 0$. Choose then $\omega_q = 0$, $q \neq p$, and $\omega_p = 1$. This is the desired solution. ∎

LEMMA A.1. *Suppose players $j \neq i$ are in the reward state and follow their equilibrium strategies in review phase h. For every $\beta > 0$ there exists an ε^h and l^h such that for $\varepsilon' \leqslant \varepsilon^h$ and $l^R \geqslant l^h$,*

$$\text{Prob}\left\{\exists j \neq i \mid \bar{\pi}_j^h(\cdot) - \pi_j(\cdot, \alpha^{-h})| \leqslant \varepsilon' \text{ and } (1/\bar{L}^h l^R) \sum_{t=1}^{L^h l^R} r_i(t) > g_i(\bar{\alpha}^h) + 2\varepsilon\right\} \leqslant \beta$$

regardless of the strategy used by player i. (Recall that $r_i(t) = r_i(z_i(t))$ is player i's realized payoff in period t.)

Proof. From Lemma 4.2 there exist for $j \neq i$, m_j-vectors of weights λ_j and a scalar η such that $\sum_{j \neq i} \lambda_j \pi_j(\cdot, \alpha_i, \bar{\alpha}_{-i}^h) \geqslant g_i(\alpha_i, \bar{\alpha}_{-i}^h) + \eta$ and $\sum_{j \neq i} \lambda_j \pi_j(\cdot, \bar{\alpha}^h) = g_i(\bar{\alpha}^h) + \eta$. Consider then the random variable $x(t) = \sum_{j \neq i} \lambda_j(z_j(t))$, where $\lambda_j(z_j(t))$ is the component of $z_j(t)$ corresponding to z_j. The idea of this construction is to use a single random variable $x(t)$ to summarize the information received by all other players. Let

$$\sum_{j \neq i} \lambda_j \bar{\pi}_j^h(\cdot) = (1/\bar{L}^h l^R) \sum_{t=1}^{L^{-h} l^R} x(t) \equiv \bar{x}.$$

It is clear then that there exists ε^h such that if $\bar{x} > g_i(\bar{\alpha}^h) + \eta + \varepsilon$ then $|\bar{\pi}_j^h - \pi_j(\cdot, \bar{\alpha}^h)| > \varepsilon^h$ for some j. That is, if the sample mean $\bar{x}$ is far from its theoretical distribution under $\bar{\alpha}_j^n$ then some player $j \neq i$ will observe that his empirical information $\bar{\pi}_j^h$ is far from what it would be if player i had played $\bar{\alpha}^n$. Consequently, it suffices to show with $\bar{\rho}_i = \sum_{t=1}^{\bar{L}^h l^R} r_i(t)$ that

$$\text{Prob}\{\bar{x} \leqslant g_i(\bar{\alpha}^h) + \eta + \varepsilon \text{ and } \rho_i > g_i(\bar{\alpha}^h) + 2\varepsilon\} \leqslant 1 - \beta.$$

Fix a strategy for player i and consider

$$\tilde{x}(t) \equiv x(t) - E[x(t) \mid x(t-1), ..., r(t-1), ...].$$

These are uncorrelated random variables with zero mean and are bounded independent of the particular strategy of player i. Similar considerations apply to $\tilde{r}(t) = r(t) - E[r(t) \mid x(t-1), ..., r(t-1), ...]$. Consequently the weak law of large numbers shows that as $l^R \to \infty$, $\bar{\tilde{x}}$, $\bar{\tilde{r}} \to 0$ in probability uniformly over strategies for player i. Let $\alpha_i(t)$ be player i's mixed action at t conditional on $x(t-1), ..., r(t-1), ...$. Then

$$\begin{aligned} E[x(t) \mid x(t-1), ..., r(t-1), ...] &= \sum_{j \neq i} \lambda_j \pi_j(\cdot, \alpha_i(t), \bar{\alpha}_{-i}^h) \geqslant g_i(\alpha_i(t), \bar{\alpha}_{-i}^h) + \eta \\ &= E[r(t) \mid x(t-1), ..., r(t-1), ...] + \eta. \end{aligned}$$

Consequently $\text{Prob}\{\bar{x} - \bar{r} - \eta < -\varepsilon\} \to 0$. Since $\bar{x} \leqslant g_i(\bar{\alpha}^h) + \eta + \varepsilon$ and $\bar{r} > g_i(\bar{\alpha}^h) + 2\varepsilon$ implies $\bar{x} - \bar{r} < \eta - \varepsilon$, this gives the desired conclusion. ∎

LEMMA A.2. *Suppose player j is in the punishment state, and that (j, k, i) is an active communication phase. If all players $k' \neq i$ follow their equilibrium strategies then for every $1 > \beta$ there exists an ε^{jki} and l^{jki} such that for $\varepsilon' \leqslant \varepsilon^{jki}$ and $l^R \geqslant l^{jki}$*

$$\text{Prob}\{|\bar{\pi}_k^{jki}(\cdot) - \pi_k(\cdot, \alpha^{jki})| > \varepsilon'\} \geqslant \beta.$$

Proof. This is similar to, but simpler than Lemma A.1. Observe that the set of vectors $\pi_k = \pi_k(\cdot, \hat{\alpha}^{jki}, \alpha_i, \alpha_{-j-i}^{jki})$ for different α_i is compact convex, and by assumption does not contain $\pi_k(\cdot, \alpha^{jki})$. Consequently, we may find a m_k-vector of weights λ and a scalar $\eta > 0$ such that $\lambda \pi_k \geqslant \eta$ and

$\lambda\pi_k(\cdot, \alpha^{jki}) = 0$. Set $x(t) = \lambda(z_k(t))$. Again, $\lambda\bar{\pi}_j^{jik}(\cdot) = \bar{x}$. so there exists ε^{jki} such that if $\bar{x} > \eta/2$, $|\bar{\pi}_k^{jik} - \pi_j(\cdot, \bar{\alpha}^h)| > \varepsilon^{jki}$. Again $\tilde{x}(t) = x(t) - E[x(t) \mid x(t-1), \ldots]$ has $\bar{\tilde{x}}$ converging uniformly to zero in probability. Since $E[x(t) \mid x(t-1), \ldots] = \lambda\pi_k(\cdot, \hat{\alpha}^{jki}, \alpha_i(t), \alpha^{jki}_{-j-i}) \geqslant \eta$, we get the desired conclusion. ∎

References

1. D. Abreu, D. Pearce, and E. Stacchetti, Toward a theory of discounted repeated games with imperfect monitoring, *Econometrica* **58** (1990), 1041–1064.
2. D. Abreu, D. Pearce, and E. Stacchetti, Optimal cartel monitoring with imperfect information, *J. Econ. Theory* **39** (1986), 251–269.
3. J. Friedman, A non-cooperative equilibrium for supergames, *Rev. Econ. Stud.* **38** (1972), 1–12.
4. D. Fudenberg, D. Kreps, and E. Maskin, Repeated games with long run and short run players, *Rev. Econ. Stud.* **57** (1990), 555–574.
5. D. Fudenberg, D. Levine, and E. Maskin, The folk theorem with imperfect public information, mimeo, 1990.
6. D. Fudenberg and E. Maskin, Folk theorems for repeated games with discounting and incomplete information, *Econometrica* **54** (1986), 533–554.
7. D. Fudenberg and E. Maskin, Discounted repeated games with one-sided moral hazard, mimeo, MIT.
8. D. Fudenberg and E. Maskin, The dispensability of public randomizations in discounted repeated games, *J. Econ. Theory*, in press.
9. D. Fudenberg and D. Levine, Subgame perfect equilibria of finite- and infinite-horizon games, *J. Econ. Theory* **31** (1983), 251–263.
10. E. Green and R. Porter, Noncooperative collusion under imperfect information, *Econometrica* **52** (1984), 87–100.
11. D. Kreps and R. Wilson, Sequential equilibrium, *Econometrica* **50** (1982), 863–894.
12. E. Lehrer, Two-player repeated games with non-observable actions and observable payoffs, mimeo, Northwestern, 1986.
13. E. Lehrer, Extensions of the folk theorem to games with imperfect monitoring, mimeo, Northwestern, 1988.
14. E. Lehrer, "Internal Correlation in Repeated Games," CMSEMS Discussion Paper No. 800, Northwestern, 1988.
15. E. Lehrer, Nash equilibria of n-player repeated games with semi-standard information, mimeo, Northwestern, 1988.
16. E. Lehrer, Two players' repeated games with non-observable actions, mimeo, Northwestern, 1988.
17. E. Lehrer, Lower equilibrium payoffs in two-players repeated games with nonobservable actions, *Int. J. Game Theory*, in press.
18. E. Lehrer and D. Monderer, Discounting versus averaging in dynamic programming, mimeo, Northwestern, 1989.
19. E. Lehrer and S. Sorin, A uniform tauberian theorem in dynamic programming, mimeo, Northwestern, 1989.
20. R. Porter, Optimal cartel trigger-price strategies, *J. Econ. Theory* **29** (1983), 313–338.
21. R. Radner, Monitoring cooperative agreements in a repeated principal-agent relationship, *Econometrica* **49** (1981), 1127–1148.

22. R. Radner, Repeated principal agent games with discounting, *Econometrica* **53** (1985), 1173–1198.
23. R. Radner, Repeated partnership games with imperfect monitoring and no discounting, *Rev. Econ. Stud.* **53** (1986), 43–58.
24. R. Radner, R. Myerson, and E. Maskin, An example of a repeated partnership game with discounting and with uniformly inefficient equilibria, *Rev. Econ. Stud.* **53** (1986), 59–70.
25. A. Rubinstein and M. Yaari, Repeated insurance contracts and moral hazard, *J. Econ. Theory* **30** (1983), 74–97.
26. S. Sorin, Asymptotic properties of a non-zero sum stochastic game, *Int. J. Game Theory* **15** (1988), 101–107.
27. S. Sorin, Supergames, mimeo, CORE, 1986.
28. G. Stigler, The economics of information, *J. Polit. Econ.* **69** (1961), 213–225.
29. R. J. Aumann and L. Shapley, Long term competition: A game theoretic analysis, mimeo, 1976.
30. A. Rubinstein, Equilibrium in supergames with the overtaking criterion, *J. Econ. Theory* **27** (1979), 1–9.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium