



## Steady State Learning and Nash Equilibrium

Drew Fudenberg; David K. Levine

*Econometrica*, Vol. 61, No. 3 (May, 1993), 547-573.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9682%28199305%2961%3A3%3C547%3ASSLANE%3E2.0.CO%3B2-Z>

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

*Econometrica* is published by The Econometric Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/econosoc.html>.

---

*Econometrica*

©1993 The Econometric Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2003 JSTOR

## STEADY STATE LEARNING AND NASH EQUILIBRIUM

BY DREW FUDENBERG AND DAVID K. LEVINE<sup>1</sup>

We study the steady states of a system in which players learn about the strategies their opponents are playing by updating their Bayesian priors in light of their observations. Players are matched at random to play a fixed extensive-form game, and each player observes the realized actions in his own matches, but not the intended off-path play of his opponents or the realized actions in other matches. Because players are assumed to live finite lives, there are steady states in which learning continually takes place. If lifetimes are long and players are very patient, the steady state distribution of actions approximates that of a Nash equilibrium.

KEYWORDS: Learning, Nash equilibrium, multi-armed bandits.

### 1. INTRODUCTION

WE STUDY AN EXTENSIVE-FORM GAME played repeatedly by a large population of players who are matched with one another at random, as in Rosenthal (1979). As in Fudenberg and Kreps (1991), we suppose that players do not know the strategies their opponents are using, but learn about them from their past observations of the opponents' play. At the end of each period, each player observes the realized actions of his own opponents; players do not observe play in the other matches. The key feature of such learning is that the stage game may have information sets that are not reached in the course of play, and observing the opponents' realized actions does not reveal how the opponents would have played at the unreached information sets. Thus, even if the players play the stage game many times, they may continue to hold incorrect beliefs about the opponents' play unless they engage in a sufficient amount of "experimentation."

In our model, individual players play only a finite number of periods; new players enter so that the total size of the population is constant. Entering players believe that they face a fixed but unknown distribution of opponents' strategies. They have noninformative prior beliefs about this distribution which they update using Bayes rule, and their actions in each period are chosen to maximize the expected present value of their payoffs given their beliefs. Steady states always exist. When lifetimes are short, steady state play is mostly determined by the priors, as players have little chance to learn from experience. Steady states need not be Nash equilibria if lifetimes are long but players are impatient, because players may not learn their opponents' off-path play. Instead, steady states with impatient players correspond to the "self-confirming

<sup>1</sup> This paper presents the results on steady state learning from our (1990) paper "Steady State Learning and Self-Confirming Equilibrium." We thank Andreu Mas-Colell and three anonymous referees for useful comments on the previous paper, and Youngse Kim and Peter Klibanoff's careful proofreading. Financial support from National Science Foundation Grants SES 9008770 and SES 9023697, the John S. Guggenheim Foundation, and the UCLA Academic Senate is gratefully acknowledged.

equilibria” introduced in our (1990) paper.<sup>2</sup> Our main result is that if lifetimes are long and players are sufficiently patient, then expected-value maximization implies that players will choose to do enough experimentation that steady states correspond to Nash equilibria of the stage game.

To motivate these results, fix an extensive-form game of perfect recall, and suppose that each player knows its structure. In general, this knowledge is not enough for a player to decide how to play; he must also predict the play of his opponents. Nash equilibrium and its refinements describe a situation in which each player correctly predicts the play of his opponents. To understand when equilibrium analysis is justified therefore requires an explanation of how players’ predictions are formed, and when they are likely to be accurate.

One classic explanation is that an outside “mediator” suggests a strategy profile to the players, who accept it unless some player could gain by deviating. A second classic explanation is that the game is dominance solvable and it is common knowledge that players are Bayesian rational, so that introspection by the players leads them all to the same predictions. A more recent explanation, introduced by Aumann (1987) and further developed by Brandenburger and Dekel (1987) and Tan and Werlang (1988), is that predictions will be correct if they are consistent with Bayesian updating from a common prior distribution.

This paper contributes to the development of a fourth explanation, that Nash equilibrium is the result of learning from past observations. In our model, the steady state distributions of strategies played approximate those of Nash equilibria if players live a long time and are also sufficiently patient. The intuition for this is roughly the following.

If a player’s prior is not degenerate, he will learn a great deal about the results of any action he chooses many times; this is “passive learning.” If the player is patient, he will choose to invest in discovering what his best strategy is. This is accomplished by “active learning” or “experimentation,” meaning that the player will sometimes play actions that do not maximize the current period’s expected payoff given his current beliefs, so that he may learn whether his beliefs are in fact correct. Without experimentation, players can persistently maintain incorrect beliefs about their opponents’ off-path play, which is why steady states need not correspond to equilibria unless players experiment sufficiently often.

Our focus on active learning differentiates the paper from the literature on learning in rational expectations environments, as pioneered by Bray (1982). In that literature, players observe system-wide aggregates, and thus have no reason to experiment. The focus on active learning also differentiates the paper from that of Canning (1990), which in other respects is quite similar. Canning studies two-player simultaneous-move stage games, and supposes that players live forever but only remember their last  $T$  observations. He shows that when  $T$  is

<sup>2</sup> This concept is closely related to the “conjectural equilibria” of Battigalli and Guatoli (1988), the “rationalizable conjectural equilibria” of Rubinstein and Wolinsky (1990), and the “private beliefs equilibrium” of Kalai and Lehrer (1991a).

large, the steady states approximate Nash equilibria. We should also mention the work of Kalai and Lehrer (1991b) which studies Bayesian learning in a setting where the same players are matched with one another in every period. In their model, unlike Canning's, active learning is possible, but active learning is not necessary for their main result, which is that play eventually resembles that of a Nash equilibrium of the repeated game.

Our study was inspired primarily by Fudenberg and Kreps (1991), who emphasized the importance of active learning in justifying Nash equilibrium. They showed by example how passive learning without experimentation can lead to steady states that are not Nash. There are two main differences between their work and ours. First, Fudenberg and Kreps develop a model of bounded rationality, making ad-hoc assumptions about players' behavior, while we assume that players are Bayesian expected-utility maximizers. Second, Fudenberg and Kreps analyze the dynamic evolution of a system where all players become more experienced over time, and characterize the states to which the system can converge, while we analyze the steady states of a stationary system.

The steady state model has several comparative advantages. Our model can describe situations where the players' lifetimes are moderate and the inexperienced players have a substantial influence, although we do not study such situations here. Our model is mathematically more tractable, which enables us to solve for the optimal Bayesian policies. Finally, in the Fudenberg and Kreps model, players act as if they are facing a stationary environment even though the environment is not stationary. In the steady states we analyze, the players' assumption of stationarity is justified.

Both papers avoid the question of global stability, for which general results seem unlikely.<sup>3</sup> Fudenberg and Kreps do develop several notions of local stability and establish global convergence for the class of  $2 \times 2$  simultaneous-move stage games. Our model is rich enough to analyze local stability, but we should point out that such an analysis would need to consider the evolution of the system when the players' belief that the system is in a steady state is mistaken.

## 2. THE STAGE GAME

The *stage game* is an  $I + 1$ -player extensive form game of perfect recall. Player  $i = I + 1$  is nature. The game tree  $X$ , with nodes  $x \in X$ , is finite. The terminal nodes are  $z \in Z \subset X$ .<sup>4</sup> Information sets, denoted by  $h \in H$ , form a partition of  $X \setminus Z$ . The information sets where player  $i$  has the move are  $H_i \subset H$ , while  $H_{-i} \equiv H \setminus H_i$  are information sets for other players (or nature). The feasible actions at information set  $h \in H$  are denoted  $A(h)$ ;<sup>5</sup>  $A_i = \bigcup_{h \in H_i} A(h)$  is the set of all feasible actions for player  $i$ , and  $A_{-i} = \bigcup_{j \neq i} A_j$  are the feasible actions for player  $i$ 's opponents.

<sup>3</sup> See Shapley (1964) and Canning (1989).

<sup>4</sup> To avoid a trivial special case in one of our proofs, we will suppose  $\#Z > 1$ .

<sup>5</sup> We denote the actions so that  $A(h) \cap A(h') = \emptyset$  for  $h \neq h'$ .

A pure strategy for player  $i$ ,  $s_i$ , is a map from information sets in  $H_i$  to actions satisfying  $s_i(h) \in A(h)$ ;  $S_i$  is the set of all such strategies. We let  $s \in S \equiv \times_{i=1}^{I+1} S_i$  denote a pure strategy profile for all players including nature, and  $s_{-i} \in S_{-i} \equiv \times_{j \neq i} S_j$ . Each strategy profile determines a terminal node  $\zeta(s) \in Z$ . We suppose that all players know the structure of the extensive form—that is, the game tree  $X$ , information partitions  $H_i$  and action sets  $A_i$ . Hence, each player knows the space  $S$  of strategy profiles, and can compute the function  $\zeta$ . Each player  $i$  receives a payoff in the stage game that depends on the terminal node. Player  $i$ 's payoff function is denoted  $u_i: Z \rightarrow \mathbf{R}$ . We let  $U$  be the largest difference in utility levels,  $U \equiv \max_{i,z,z'} |u_i(z) - u_i(z')|$ .

Let  $\Delta(\cdot)$  denote the space of probability distributions over a set. Then a mixed strategy profile is  $\sigma \in \times_{i=1}^{I+1} \Delta(S_i)$ . For ease of exposition, we assume that nature plays a known mixed strategy  $\sigma_{I+1}^0$ .<sup>6</sup>

Let  $Z(s_i)$  be the subset of terminal nodes that are reachable when  $s_i$  is played, that is,  $z \in Z(s_i)$  if and only if for some  $s_{-i} \in S_{-i}$ ,  $z = \zeta(s)$ . Similarly, define  $X(s_i)$  to be all nodes that are reachable under  $s_i$ , not merely terminal nodes. In a similar vein, let  $H(s_i)$  be the set of all information sets that can be reached if  $s_i$  is played. In other words,  $h \in H(s_i)$  if there exists  $x \in X(s_i)$  with  $x \in h$ .

We will also need to refer to the information sets that are reached with positive probability under  $\sigma$ , denoted  $\bar{H}(\sigma)$ . Notice that if  $\sigma_{-i}$  is completely mixed, then  $\bar{H}(s_i, \sigma_{-i}) = H(s_i)$ , as every information set that is potentially reachable given  $s_i$  has positive probability.

In addition to mixed strategies, we define behavior strategies. A behavior strategy for player  $i$ ,  $\pi_i$ , is a map from information sets in  $H_i$  to probability distributions over moves, so that  $\pi_i(h) \in \Delta(A(h))$ ;  $\Pi_i$  is the set of all such strategies. As with pure strategies,  $\pi \in \Pi \equiv \times_{i=1}^{I+1} \Pi_i$ , and  $\pi_{-i} \in \Pi_{-i} \equiv \times_{j \neq i} \Pi_j$ . We also let  $\pi_i(a)$  denote the component of  $\pi_i(h)$  corresponding to the action  $a \in A(h)$ . Finally, let  $\zeta(\pi) \in \Delta(Z)$  be the probability distribution over terminal nodes induced by the behavior strategy  $\pi$ .

Since the game has perfect recall, each mixed strategy  $\sigma_i$  induces a unique equivalent behavior strategy denoted  $\pi_i(\cdot | \sigma_i)$ .<sup>7</sup> In other words,  $\pi_i(h | \sigma_i)$  is the probability distribution over actions at  $h$  induced by  $\sigma_i$ , and  $\pi_i(a | \sigma_i)$  is the probability of the action  $a \in A(h)$ .

For each node  $x$  and each player  $i$ , let  $\{a_{-i}(l, x)\}_{l=1}^L$  be the collection of all actions of players other than  $i$  (including nature) which are predecessors of  $x$ . (Note that  $L$  is equal to the length of the path to  $x$  minus the number of nodes in the path that belong to player  $i$ .) If pure strategy  $s_i$  is such that  $x \in X(s_i)$ , player  $i$  believes that the probability of reaching node  $x$  when  $s_i$  is played and

<sup>6</sup> Our main result (Theorem 5.1, about Nash equilibria) extends in a straightforward way to the case where nature's move is drawn from a fixed but unknown distribution; extending Theorem 6.1 requires a modification of the definition of self-confirming equilibrium.

<sup>7</sup> This is known as Kuhn's Theorem (Kuhn (1953)). Recent presentations of this result can be found in Fudenberg and Tirole (1991) and Kreps (1990), among other places.

the opponent's strategies are  $\pi_{-i}$  is

$$(2.1) \quad p_i(x|\pi_{-i}) = \prod_{l=1}^L \pi_{-i}(a_{-i}(l, x)).$$

Notice the convention we use: each node  $x$  is assigned a number  $p_i(x|\pi_{-i})$  which is the probability of reaching that node if any  $s_i$  is played for which  $x \in X(s_i)$ . Naturally if  $x \notin X(s_i)$  the probability of reaching  $x$  is zero, while  $\sum_{z \in Z(s_i)} p_i(z|\pi_{-i}) = 1$ . The effect of changing  $s_i$  is not on the numbers  $p_i$ , but rather the set of nodes  $X(s_i)$  that can be reached.

We now model the idea that a player has beliefs about his opponents' play. Let  $\mu_i$  be a probability measure over  $\Pi_{-i}$ , the set of other players' behavior strategies. Fix  $s_i$ . Then the marginal probability of a node  $x \in X(s_i)$  is

$$(2.2) \quad p_i(x|\mu_i) = \int p_i(x|\pi_{-i}) \mu_i(d\pi_{-i}).$$

This in turn gives rise to preferences

$$(2.3) \quad u_i(s_i, \mu_i) \equiv u_i(s_i, p_i(\cdot|\mu_i)) \equiv \sum_{z \in Z(s_i)} p_i(z|\mu_i) u_i(z).$$

It is important to note that even though the beliefs  $\mu_i$  are over opponents' behavior strategies, and thus reflect player  $i$ 's knowledge that his opponents choose their randomizations independently, the marginal distribution  $p_i(\cdot|\mu_i)$  over nodes can involve correlation between the opponents' play. For example, if players 2 and 3 simultaneously choose between  $U$  and  $D$ , player 1 might assign probability  $1/4$  to  $\pi_2(U) = \pi_3(U) = 1$ , and probability  $3/4$  to  $\pi_2(U) = \pi_3(U) = 1/2$ . Even though both profiles in the support of  $\mu_i$  suppose independent randomization by players 2 and 3, the marginal distribution on their joint actions is  $p(U, U) = 7/16$  and  $p(U, D) = p(D, U) = p(D, D) = 3/16$ , which is a correlated distribution. This correlation reflects a situation where player 1 believes some unobserved common factor has helped determine the play of both of his opponents. Since the opponents are in fact randomizing independently, we should expect player  $i$ 's marginal distribution to reflect this if he obtains sufficiently many observations, but until observations are accumulated, the correlation in  $p(\cdot|\mu_i)$  can persist.

Frequently  $\mu_i$  will be either a point mass at  $\pi_{-i}$ , or have a continuous density  $g_i$  over  $\pi_{-i}$ . In this case we write  $p_i(x|\pi_{-i})$ ,  $u_i(x, \pi_{-i})$ ,  $p_i(x|g_i)$ , and  $u_i(x, g_i)$  respectively.

### 3. STEADY STATES

Corresponding to each player (except nature) in the stage game is a population consisting of a continuum of players in the dynamic game. In each population, the total mass of players is one. There is a doubly infinite sequence of periods,  $\dots, -1, 0, 1, \dots$ , and each individual player lives  $T$  periods. Every period  $1/T$  new players enter the  $i$ th population, and we make the steady state

assumption that there are  $1/T$  players in each generation, with  $1/T$  players of age  $T$  exiting each period.

Every period each player  $i$  is randomly and independently matched with one player from each population  $i' \neq i$ , with the probability of meeting a player  $i'$  of age  $t$  equal to its population fraction  $1/T$ .<sup>8</sup> For example, if  $T = 2$ , each player is as likely to be matched with a "new" player as an "old" one. Each player  $i$ 's opponents are drawn independently.

Over his lifetime, each player observes the terminal nodes that are reached in the games he has played, but does not observe the outcomes in games played by others. Thus, each player will observe a sequence of private histories. The private history of player  $i$  through time  $t$  is denoted  $y_i^t = (s_i(1), z(1), \dots, s_i(t), z(t))$ . Let  $Y_i$  denote the set of all such histories of length no more than  $T$ . We let  $t(y_i)$  denote the length of a history  $y_i \in Y_i$ . New players have the null history 0, and we set  $z(0) = 0$ .

A rule for a player of the  $i$ th kind is a map  $r_i: Y_i \rightarrow S_i$  that specifies player  $i$ 's choice of pure strategy for each possible observation. (Note that if  $t(y_i) = T$ ,  $r_i(y_i)$  has no meaning because player  $i$  does not get to play at  $T + 1$ .)

Suppose for the moment that all players in population  $i$  use the same, arbitrary, deterministic rule  $r_i$ , and face a sequence of opponents whose play is a random draw from a stationary distribution. (This is in fact the case in the steady states of our model.) We will soon specialize to the case where the rules are derived from maximizing expected discounted values given prior beliefs, but it is helpful to develop the mechanics of the matching model before introducing that complication.

A steady state for given rules  $r_i$  specifies fractions  $\theta_i(y_i)$  of each population  $i$  in each experience category  $y_i$  such that after each player meets one opponent at random and updates his experience accordingly, the population fractions are unchanged. Specifically, if  $\theta_i \in \Delta(Y_i)$ , the fraction of population  $i$  playing  $s_i$  is

$$(3.1) \quad \bar{\theta}_i(s_i) = \sum_{\{y_i | r_i(y_i) = s_i\}} \theta_i(y_i).$$

Also define  $\bar{\theta}_{i+1} = \sigma_{i+1}^0$ . We may then define a map  $f: \times_{i=1}^I \Delta(Y_i) \rightarrow \times_{i=1}^I \Delta(Y_i)$  by assigning  $f[\theta]_i(y_i)$  to be the fraction of player  $i$ 's with experience  $y_i$  after randomly meeting opponents drawn from  $\theta$ . The new entrants to the population have no experience, so

$$(3.2) \quad f[\theta]_i(0) = 1/T.$$

The fraction having experience  $(y_i, r_i(y_i), z)$  is the fraction of the existing  $\theta_i(y_i)$  that met opponents playing strategies that led to  $z$ . Noting that  $\zeta_i^{-1}[s_i, \cdot](z)$  are those strategies, let  $S_i(z) = \zeta_i^{-1}[r_i(y_i), \cdot](z)$  so that

$$(3.3) \quad f[\theta]_i(y_i, r_i(y_i), z) = \theta_i(y_i) \sum_{s_{-i} \in S_i(z)} \prod_{k \neq i} \bar{\theta}_k(s_k).$$

<sup>8</sup> Boylan (1990) has shown that this deterministic system is the limit of a stochastic finite-population random-matching model as the number of players goes to infinity.

Finally, it is clear that

$$(3.4) \quad f[\theta]_i(y_i, s_i, z) = 0 \quad \text{if} \quad s_i \neq r_i(y_i).$$

DEFINITION 3.1:  $\theta \in \times_{i=1}^I \Delta(Y_i)$  is a *steady state* if  $\theta = f[\theta]$ .

To illustrate this definition, consider the game “matching pennies,” with  $S_i = S_2 = \{H, T\}$ . Suppose that  $T = 2$  and

$$(3.5) \quad \begin{aligned} r_1(0) &= H, & r_1(H, H) &= H, & r_1(H, T) &= T, \\ r_2(0) &= T, & r_2(H, T) &= T, & r_2(T, T) &= H. \end{aligned}$$

(Note that we do not need to specify  $r_1(T, \cdot)$  or  $r_2(\cdot, H)$  as such histories never occur—young player 1’s always play  $H$  and young player 2’s always play  $T$ .)

In a steady state we have:

$$(3.6) \quad \begin{cases} \theta_1(0) = 1/2, & \theta_1(H, H) = \theta_1(0)\theta_2(T, T), \\ & \theta_1(H, T) = \theta_1(0)[\theta_2(0) + \theta_2(H, T)], \\ \theta_2(0) = 1/2, & \theta_2(H, T) = \theta_2(0)[\theta_1(0) + \theta_1(H, H)], \\ & \theta_2(T, T) = \theta_2(0)\theta_1(H, T), \end{cases}$$

a system of quadratic equations.

Computation shows that  $\theta_1(0) = 1/2$ ,  $\theta_1(H, H) = 1/10$ ,  $\theta_1(H, T) = 4/10$ ,  $\theta_2(0) = 1/2$ ,  $\theta_2(H, T) = 3/10$ , and  $\theta_2(T, T) = 2/10$ , from which it follows that  $\bar{\theta}_1(H) = \theta_1(0) + \theta_1(H, H) = 6/10$  and  $\bar{\theta}_2(H) = \theta_2(T, T) = 2/10$ . Note that the average play of the player 1’s and the player 2’s corresponds to a mixed strategy of the stage game, even though all individuals in both populations use deterministic rules. (Canning (1989, 1990) makes the same point.)

THEOREM 3.1: For any rules  $r_i$  and nature’s moves  $\bar{\theta}_{I+1}$ , a steady state exists.

PROOF: For given  $\bar{\theta}_{I+1}$ ,  $f$  is a polynomial map from  $\times_{i=1}^I \Delta(Y_i)$  to itself, and so it has a fixed point. Q.E.D.

Given a steady state  $\theta \in \times_{i=1}^I \Delta(Y_i)$ , we may easily compute the population fractions  $\bar{\theta}_i \in \Delta(S_i)$  playing each strategy by (3.1). Conversely, given the steady state fractions we can calculate the experience levels recursively by

$$(3.7) \quad \begin{aligned} \theta'_i(y_i(0)) &= 1/T, \\ \theta'_i(y_i, r_i(y_i), z) &= \theta'_i(y_i) \sum_{s_{-i} \in S_i(z)} \Pi_{k \neq i} \bar{\theta}_k(s_k), \\ \theta'_i(y_i, s_i, z) &= 0 \quad \text{if} \quad s_i \neq r_i(y_i). \end{aligned}$$

If we then recalculate  $\bar{\theta}_i$  using (3.1) we have a polynomial function  $\bar{f}$  mapping  $\times_{i=1}^I \Delta(S_i)$  to itself. We may equally well characterize a steady state as a fixed



point of  $\bar{f}$ , and calculate the corresponding fixed point of  $f$  using (3.7). Since  $\Delta(S_i)$  is much smaller than  $\Delta(Y_i)$ , this is of some practical importance.

4. VALUE MAXIMIZATION GIVEN BAYES STATIONARY BELIEFS

Our interest is not in arbitrary rules  $r_i$  but in rules derived from maximizing the expected discounted sum of per-period payoffs given exogenous prior beliefs. More precisely, we suppose that each player's objective is to maximize

$$(4.1) \quad \frac{1 - \delta}{1 - \delta^T} E \sum_{t=1}^T \delta^{t-1} u_t$$

where  $u_t$  is the realized stage game payoff at  $t$  and  $0 \leq \delta < 1$ . Each population believes that it faces a fixed (time invariant) probability distribution of opponents' strategies, but is unsure what the true distribution is.

Population  $i$ 's prior beliefs are over behavior strategies. We will assume throughout that the players' beliefs about nature's moves are correct, that is, are concentrated on the true value  $\pi_{I+1}(\cdot | \sigma_{I+1}^0)$ . We will also suppose that the players' beliefs about their opponents strategies are nondoctrinaire.

DEFINITION 4.1: Beliefs  $\mu_i$  are *nondoctrinaire* if for each player  $i$  the projection of  $\mu_i$  on  $\prod_{j \neq i, I+1} \Pi_j$  is given by a strictly positive density function  $g_i$ .

It is important to emphasize that player  $i$ 's beliefs about player  $j$  correspond to the average play of all player  $j$ 's and not the play of a particular individual. As in the matching pennies example of the last section, a mixed distribution over player  $j$ 's play may be the result of different subpopulations of the player  $j$ 's playing different pure strategies.

For notational convenience, we suppose that all player  $i$ 's begin the game with the same prior beliefs  $g_i^0$ . All of our results extend immediately to the case of finitely many subpopulations of player  $i$ 's with different initial beliefs.

We let  $g_i(\cdot | z)$  denote the posterior beliefs starting with prior  $g_i$  after  $z$  is observed:

$$(4.2) \quad g_i(\pi_{-i} | z) = p_i(z | \pi_{-i}) g_i(\pi_{-i}) / p_i(z | g_i).$$

Let  $V_i^k(g_i)$  denote the maximized average discounted value (in current units) starting at  $g_i$  with  $k$  periods remaining. Bellman's equation is

$$(4.3) \quad V_i^k(g_i) = \max_{s_i \in S_i} \left[ (1 - \phi_k) u_i(s_i, g_i) + \phi_k \sum_{z \in Z(s_i)} p_i(z | g_i) V_i^{k-1}(g_i(\cdot | z)) \right]$$

where  $V_i^0(g_i) \equiv 0$ , and  $\phi_k = \delta(1 - \delta^{k-1}) / (1 - \delta^k)$ . Let  $s_i^k(g_i)$  denote a solution of this problem.

Let  $g_i(\cdot|y_i)$  be the posterior density over opponents' strategies given sample  $y_i$ , and let  $p_i(\cdot|y_i)$  be the corresponding distribution on terminal nodes. It will often be convenient to abbreviate  $V_i^k(g_i(\cdot|y_i))$  as  $V_i^k(y_i)$ ,  $s_i^k(g_i(\cdot|y_i))$  as  $s_i^k(y_i)$ , and  $u_i(s_i, g_i(\cdot|y_i))$  as  $u_i(s_i|y_i)$ .

The optimal policy is  $r_i(y_i) = s_i^{T-t(y_i)}(g_i(\cdot|y_i))$ . Note that this function is independent of the true value of the steady state.<sup>9</sup> (The steady state does influence the distribution of observations and hence the distribution of actions played.) Thus by Theorem 3.1 a steady state exists.

These steady states are not very interesting if lifetimes are short. For example if  $T = 1$  the entire population consists of inexperienced players, each of whom plays a best response to his prior beliefs and then leaves the system. Our interest is in the nature of steady states in the limit as lifetimes  $T$  increase to infinity and  $\delta$  goes to one.

For  $h \in H_{-i}$  and  $a \in A(h)$ , let  $n(a|y_i)$  be the number of times the move  $a$  has been observed in the history  $y_i$ . We define  $n(x|y_i)$  and  $n(h|y_i)$  similarly, and set  $n(s_i|y_i)$  to be the number of times player  $i$  has played  $s_i$ .

Let  $\hat{\pi}_{-i}^i(\cdot|y_i)$  be the sample average of player  $i$ 's observations about his opponents' play. That is, for each  $h \in H_{-i}$  and  $a \in A(h)$ ,

$$\hat{\pi}_{-i}^i(a|y_i) = n(a|y_i)/n(h|y_i)$$

with the convention that  $0/0 = 1$ . Let  $\hat{p}_i(x|y_i)$  be the distribution on nodes induced by the sample averages  $\hat{\pi}_{-i}^i$ , that is  $\hat{p}_i(x|y_i) = p_i(x|\hat{\pi}_{-i}^i(\cdot|y_i))$ . Since  $\hat{p}_i(\cdot|y_i)$  reflects the extensive form of the game, it is not in general equal to the sample average on terminal nodes  $z \in Z(s_i)$ . For example, consider a game where if player 1 moves  $L$ , players 2 and 3 observe player 1's move and simultaneously choose  $H_2$  or  $T_2$  and  $H_3$  or  $T_3$  respectively. If the sample  $y_i$  is 4 observations at  $(L, H_2, H_3)$ , 1 observation each of  $(L, H_2, T_3)$  and  $(L, T_2, H_3)$ , and 0 observations of  $(L, T_2, T_3)$ , then

$$\hat{p}_i((L, T_2, T_3)|y_i) = (1/6)(1/6) = 1/36$$

even though there are no observations on  $(L, T_2, T_3)$  in the sample. Since player 1 is certain that players 2 and 3 randomize independently, he treats the observed correlation in the sample as a fluke.

<sup>9</sup> Several readers have asked us the following questions: Won't players update their beliefs in the course of period- $t$  play as they observe the actions of their opponents? And shouldn't they therefore deviate from the original play of  $r_i(y_i)$ ? The answers are that yes, players will update their beliefs in the course of period- $t$  play, but that the optimal plan at the beginning of the period,  $r_i(y_i)$ , already takes this revision into account. Intuitively, player  $i$  can foresee what his posterior beliefs will be at every information set, and his optimal plan will thus maximize his expected utility at each information set, conditional on that information set being reached. (Remember that  $r_i(y_i)$  is a strategy for the extensive-form game, that is, it specifies a feasible action at every information set. It does *not* specify the "same" action at every information set; indeed by definition an action that is feasible at one information set cannot be feasible at another.)

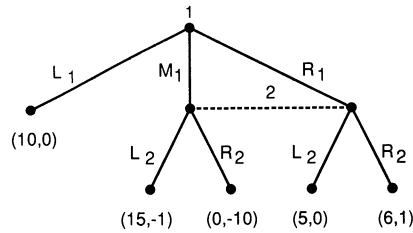


FIGURE 1

## 5. ACTIVE LEARNING AND NASH EQUILIBRIUM

Our goal is to show that if players are patient as well as long lived then steady states approximate play in a Nash equilibrium. Theorem 5.1 establishes this for the case where lifetimes  $T$  go to infinity “more quickly” than the discount factor tends to 1. We do not know whether the conclusion of the theorem holds for the other order of limits. We indicate below in the discussion of Lemma 5.2 why our approach cannot handle situations where  $\delta$  equals one.

**THEOREM 5.1:** *For any fixed nondoctrinaire priors  $g_i^0$  there is a function  $T(\delta)$  such that if  $\delta_m \rightarrow 1$  and  $T_m \geq T(\delta_m)$ , every sequence of steady states  $\bar{\theta}^m$  has an accumulation point  $\bar{\theta}$ , and every accumulation point is a Nash equilibrium.*

An accumulation point exists by compactness; the interesting part of the theorem is that the accumulation points are Nash equilibria. The idea of the proof is simply that players do enough experimentation to learn the true best responses to the steady state. The obvious argument is that if a player is very patient, and a strategy has some probability of being the best response, the player ought to try it and see. However, the simple example in Figure 1 shows a complication. Even a very patient player may find it optimal never to experiment with certain strategies. Moreover, these unused strategies need not be dominated strategies in the stage game. In the game of Figure 1, if player 1 assigns a low probability to player 2 playing  $L_2$ , his current period’s expected payoff is maximized by playing  $L_1$ . Now, if player 1 is patient, he may be willing to incur a short-run cost to obtain information about player 2’s play, but given player 1’s beliefs, the lowest-cost way of obtaining this information is by playing  $R_1$ , and player 1 may never play  $M_1$ .<sup>10</sup> Since not all experiments need be made, our proof will use a more indirect approach.

<sup>10</sup>As an aside, we note that this example also shows why our learning model does not yield results in the spirit of forward induction (Kohlberg and Mertens (1986)). Forward induction interprets all deviations from the path of play as attempts to gain in the current round. Since  $L_1$  strictly dominates  $R_1$ , forward induction argues that player 2 will believe that player 1 has played  $M_1$  whenever player 2’s information set is reached, and hence that player 2 will play  $L_2$ ; this will lead player 1 to play  $M_1$ . In contrast, in our model player 1 deviated from  $L_1$  to gain information that will help him in future rounds, and the cheapest way to do this is to play  $R_1$ . When  $R_1$  is more likely than  $M_1$ ,  $R_2$  is optimal for player 2.

Very briefly, our proof derives both upper bounds and lower bounds on the players' option values, that is, the difference between their expected payoff in the current round and their expected average present value. We argue that if  $s_i$  has positive probability in the limit, most players using  $s_i$  do not expect to learn much more about its consequences, and play  $s_i$  because it maximizes their current payoff. For these players the option value of the game is low. However, we also show that if  $s_i$  is not a best response to the steady state distribution and players are patient, then they are very unlikely to observe a sample that makes their option value small, thus obtaining a contradiction. Intuitively, if some strategy  $\hat{s}_i$  is a better response than  $s_i$  to the steady state distribution  $\bar{\theta}_{-i}$ , and player  $i$ 's beliefs assign nonnegligible probability to the opponents' strategies lying in a neighborhood of  $\bar{\theta}_{-i}$ , then player  $i$  should have a positive option value from the possibility of playing  $\hat{s}_i$ , and this option value will be large if player  $i$  is sufficiently patient. Moreover, since player  $i$ 's prior assigns positive probability to all neighborhoods of  $\bar{\theta}_{-i}$ , repeated observations from the distribution generated by  $\bar{\theta}_{-i}$  should be unlikely to substantially reduce the probability player  $i$  assigns to neighborhoods of  $\bar{\theta}_{-i}$ .

The proof of Theorem 5.1 uses five lemmas proven in Appendix B. Before proving the theorem we first discuss the lemmas. Recall that  $s_i^k(y_i)$  is the optimal choice of strategy for a player with  $k$  periods of play remaining and beliefs  $g_i(\cdot|y_i)$ .

LEMMA 5.2: *There exists a function  $\eta(n) \rightarrow 0$  as  $n \rightarrow \infty$  such that for all  $y_i$  and  $\delta$*

$$\begin{aligned}
 (5.2.1) \quad & \max_{s_i} u_i(s_i|y_i) - u_i(s_i^k(y_i)|y_i) \\
 & \leq V_i^k(y_i) - u_i(s_i^k(y_i)|y_i) \\
 & \leq [\delta/(1 - \delta)] \max_{x \in X(s_i^k(y_i))} \hat{p}(x|y_i)\eta(n(x|y_i)).
 \end{aligned}$$

The first inequality follows from the fact that, because it is possible to achieve average payoff  $\max_{s_i} u_i(s_i|y_i)$  by ignoring all subsequent observations and playing the strategy that maximizes  $u_i(s_i|y_i)$  for the remainder of the individual's lifetime,  $V_i^k(y_i) \geq \max_{s_i} u_i(s_i|y_i)$ .

To understand the second inequality, observe that  $V_i^k(y_i) - u_i(s_i^k(y_i)|y_i)$  is a measure of the option value, or anticipated capital gain, to the information generated by playing  $s_i^k(y_i)$ . The second inequality asserts that as the total number of observations on nodes that are reachable under  $s_i^k(y_i)$  becomes large, when weighted by the empirical probability  $\hat{p}$  that these nodes are reached, the option value becomes small. The idea behind this conclusion is that once a player has many observations of play at a given information set, another observation will not change his beliefs about that information set very much, as has been shown by Diaconis and Freedman (1990). (This relies on the assumption of nondoctrinaire priors.) In the context of a simple multi-armed bandit

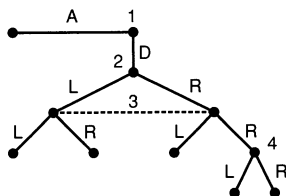


FIGURE 2

model, their result implies that the option value of each arm is bounded by a decreasing function of the number of times each arm has been played. The reason that the third expression in (5.2.1) is more complicated than this is that in our model players know the extensive form of the game. This means that they may know that some large samples are not “representative,” and for these large samples the option value may still be large.

As an example of this possibility, consider the game in Figure 2. In this game, if player 1 plays  $D$ , then players 2 and 3 simultaneously choose between  $L$  and  $R$ ; player 4 only gets to move if players 2 and 3 both play  $R$ . Now suppose that player 1 has played  $D$  200 times, and that he has observed 100 draws of  $(D, L, R)$ , and 100 draws of  $(D, R, L)$ . Then if his prior beliefs on  $\Pi_2$  and  $\Pi_3$  are uniform, his posteriors will be concentrated in the neighborhood of players 2 and 3 each playing the strategy  $(1/2L, 1/2R)$ . Given these beliefs, player 1 believes there is a substantial (about  $1/4$ ) probability that playing  $D$  will lead to player 4’s information set, so that information about player 4’s play is potentially valuable, yet player 1 has not received any observations of player 4’s play. The point is that because player 1 knows that players 2 and 3 play simultaneously, he treats the observed correlation in their play as a fluke.<sup>11</sup>

Of course, if players 2 and 3 do choose their strategies independently, the sort of spurious correlation in this example should be unlikely, so that most large samples should lead to small option values. This is established by Corollary 5.5 below.

The final important remark about Lemma 5.2 is that it places weaker and weaker bounds on the option value as  $\delta$  goes to 1, as a very patient player will have gains from learning when faced with even a small amount of uncertainty. This property is what has prevented us from extending our proof of Theorem 5.1 to limits where  $\delta$  goes to 1 faster than  $T$  goes to infinity.

Lemma 5.2 provides an upper bound on expected gains from learning about the consequences of a strategy; the next lemma provides a lower bound.

<sup>11</sup> Fudenberg and Kreps (1991, Ch. 9), develop a learning model in which players do not know the extensive form of the game. We believe that Theorem 5.1 would extend to this context, but Theorem 6.1 (on self-confirming equilibrium) would not.

Define

$$P(s_i, \Delta, y_i) \equiv \max_{s'_i} \int_{\{\pi_{-i} | u_i(s'_i, \pi_{-i}) \geq u_i(s_i, \pi_{-i}) + \Delta\}} g_i(\pi_{-i} | y_i) d\pi_{-i}$$

to be the largest posterior probability (given the sample  $y_i$ ) that any strategy yields a gain of  $\Delta$  over  $s_i$ . If this probability is large and the player is patient, the option value ought to be large.

LEMMA 5.3: *For all  $1 > \varepsilon > 0$  and  $\Delta > 0$ , there are  $\underline{\delta} < 1$  and  $\underline{K}$  such that for all  $\delta \in [\underline{\delta}, 1)$  and for all  $k \geq \underline{K}$*

$$(5.3.1) \quad \Delta \cdot P(s_i^k(y_i), \Delta, y_i) - \varepsilon \leq \frac{V_i^k(y_i) - u_i(s_i^k(y_i) | y_i)}{1 - \varepsilon}.$$

Now we turn to the issue of what fraction of the population has sampled frequently enough to have discovered (approximately) the true distribution of opponents' play. Stating the desired result requires some additional notation. Let  $m$  index a particular environment, characterized by a horizon  $T_m$ , a plan  $r_i^m$  for the  $i$ th kind of player, and the population fraction of opponents' actions  $\bar{\theta}_{-i}^m$ . From (3.7) we can calculate for population  $i$  the corresponding steady state fractions  $\theta_i^m$  of players with different histories. Let  $\tilde{Y}_i \subset Y_i$  be a subset of the histories for player  $i$ . We define

$$\theta_i^m\{\tilde{Y}_i\} = \sum_{y_i \in \tilde{Y}_i} \theta_i^m(y_i)$$

to be the steady state fraction of population  $i$  in category  $\tilde{Y}_i$ . First we examine the relationship between the probabilities of nodes as measured by maximum likelihood, and the true population value.

LEMMA 5.4: *For all  $\varepsilon > 0$  and functions  $\eta$  such that  $\eta(n) \rightarrow 0$  as  $n \rightarrow \infty$  there is an  $N$  such that for all  $T_m, \bar{\theta}_{-i}^m, r_i^m$ , and  $s_i$ ,*

$$(5.4.1) \quad \theta_i^m \left\{ \max_{x \in X(s_i)} \hat{p}_i(x | y_i) \eta(n(x | y_i)) > \varepsilon, \text{ and } n(s_i | y_i) > N \right\} \leq \varepsilon.$$

This asserts that few people have a large sample on the strategy  $s_i$ , few hits on a reachable node, and a high maximum likelihood estimate of hitting it. Notice the nature of the assertions: the fraction of the population that both have a large sample and an unrepresentative one is small. It need not be true that of those that have a large sample, most have a representative sample. Since the sampling rule is endogenous, we must allow the possibility that sampling continues only if the sample is unusual (e.g., keep flipping until tails have been observed).

Our next step is to combine Lemmas 5.2 and 5.4 to conclude that players are unlikely to repeatedly play a strategy solely for its option value. Intuitively, if strategy  $s_i$  has already been played many times, the player's observations  $y_i$

should have provided enough observations at the relevant nodes that the player is unlikely to learn very much by playing  $s_i$  again. As we explained earlier, the reason this conclusion only holds for most large samples, as opposed to all of them, is that since players know the extensive form of the game, they may know that their sample is “unrepresentative” of the true distribution.

COROLLARY 5.5: *For all  $\varepsilon > 0$  there exists  $N$  such that for all  $T_m$  and all  $\delta$ ,*

$$\theta_i^m \{V_i^k(y_i) - u_i(s_i^k(y_i)|y_i) > \delta\varepsilon/(1 - \delta), n(s_i^k(y_i)|y_i) > N\} \leq \varepsilon,$$

where  $k = T_m - t(y_i)$  is the number of periods remaining given history  $y_i$ .

PROOF OF COROLLARY: Observe that since Lemma 5.4 holds for each of finitely many  $s_i$ , it continues to hold (possibly for a different value of  $N$ ) when  $s_i$  is a function of  $y_i$ . Then substitute the second inequality of (5.2.1) into inequality (5.4.1). Q.E.D.

This corollary shows that even very patient players will eventually exhaust the option value of a strategy they have played many times; of course, the  $N$  required to satisfy (5.5) may grow large as  $\delta \rightarrow 1$ .

Our next lemma asserts that regardless of sample size, players are unlikely to become convinced of “wrong” beliefs. Given  $h \in H_{-i}$  and  $a \in A(h)$ , we can define  $\bar{\pi}_{-i}(a|\bar{\theta}_{-i})$  to be the conditional probability that  $a$  is chosen, given that  $h$  is reached, and  $\bar{p}_i(x|\bar{\theta}_{-i})$  to be the probability of reaching node  $x$ . Define

$$B_\varepsilon^i(\bar{\theta}_{-i}^m) \equiv \left\{ \pi_{-i} : \left\| \bar{p}_i(z|\bar{\theta}_{-i}^m) - p_i(z|\pi_{-i}) \right\| \leq \varepsilon \text{ for all } z \in Z \right\}$$

to be the beliefs  $\pi_{-i}$  that yield approximately the same distribution over terminal nodes as  $\bar{\theta}_{-i}^m$ . Let

$$Q_\varepsilon^i(\bar{\theta}_{-i}^m|y_i) = \int_{B_\varepsilon^i(\bar{\theta}_{-i}^m)} g_i(\pi_{-i}|y_i) d\pi_{-i}$$

be the corresponding posterior probability, and let  $Q_\varepsilon^i(\bar{\theta}_{-i}^m|0)$  be the prior probability. The result of Diaconis and Freedman mentioned earlier implies that along any sample path, the posterior beliefs converge to a point mass on the empirical distribution; the strong law of large numbers implies that players are unlikely to have a sample that both reaches an information set many times and gives misleading (that is, biased) information about play there. These two facts underly the following lemma. Note well the order of quantifiers here: a single  $\gamma$  can be used for all samples  $y_i$ , steady states  $\bar{\theta}^m$ , and lifetimes  $T_m$ .

LEMMA 5.6: *For all  $\varepsilon > 0$  there exists a  $\gamma > 0$  such that for all  $y_i, \bar{\theta}_{-i}^m, r_i^m$ , and  $T_m$ ,*

$$\theta_i^m \{Q_\varepsilon^i(\bar{\theta}_{-i}^m|y_i)/Q_\varepsilon^i(\bar{\theta}_{-i}^m|0) \leq \gamma\} \leq \varepsilon.$$

Our last lemma asserts that if the population fraction playing a strategy is positive, the population fraction that has played it a number of times must be sizeable as well.

LEMMA 5.7: *Let  $T_m \rightarrow \infty$  be the length of life, and  $\bar{\theta}^m$  be a subsequence of steady states that converge to  $\bar{\theta}$ , and let  $r_i^m$  be the corresponding rules. If  $\bar{\theta}_i(s_i) > 0$ ,*

$$(5.7.1) \quad \theta_i^m \{n_i(s_i|y_i) > N \text{ and } r_i^m(y_i) = s_i\} > \bar{\theta}_i^m(s_i) - (N/T_m).$$

With these lemmas in hand, we can now prove Theorem 5.1: Any limit point  $\bar{\theta}$  of steady states is a Nash equilibrium. Here is a rough sketch of the proof: If  $\bar{\theta}$  is not a Nash equilibrium, then some player  $i$  must be able to gain at least  $3\Delta > 0$  by not playing some strategy  $s_i$  that  $\bar{\theta}$  assigns positive probability. If player  $i$ 's beliefs assign probability close to 1 to a neighborhood of  $\bar{\theta}$ , he would assign a nonnegligible probability to the event that his opponents' strategies are such that he could gain at least  $\Delta$  by deviating; call this event  $E$ . Now player  $i$ 's prior beliefs are nondoctrinaire, and hence assign a nonzero probability to  $E$ , and from Lemma 5.6, for most samples player  $i$ 's posterior does not assign  $E$  a vanishingly small weight. For all such samples, Lemma 5.3 implies that player  $i$ 's expected gain from learning is not negligible provided his discount factor is sufficiently high and he has sufficiently many periods of life remaining.

However, since  $\bar{\theta}_i(s_i)$  is positive, Lemma 5.7 implies that a nonnegligible fraction of the player  $i$ 's have played  $s_i$  many times and intend to play it again. From Corollary 5.5, if  $\delta$  is large, most of these players must have negligible expected gains from learning about  $s_i$ , which contradicts the conclusion of the last paragraph that player  $i$  is unlikely to have samples that give him a negligible gain from learning.

PROOF OF THEOREM 5.1: We will first show that for each  $\Delta > 0$  there is a function  $T(\delta, \Delta)$  such that if  $\delta_m \rightarrow 1$  and  $T_m \geq T(\delta_m, \Delta)$  any accumulation point  $\bar{\theta}$  of the steady states  $\bar{\theta}^m$  has the property that no player can gain more than  $3\Delta$  by deviating from  $\bar{\theta}_i$ . That is, for all players  $i$ , all  $s_i \in \text{support}(\bar{\theta}_i)$  and all  $s'_i$ ,  $u_i(s_i, \bar{\theta}_{-i}) \geq u_i(s'_i, \bar{\theta}_{-i}) - 3\Delta$ . The existence of the desired function  $T(\delta)$  will follow by choosing  $\Delta(\delta) \rightarrow 0$  as  $\delta \rightarrow 1$ .

Thus, we fix a  $\Delta > 0$  and a sequence of positive numbers  $\epsilon_m \rightarrow 0$ , and let  $k(\epsilon_m, \Delta)$  satisfy the conditions of Lemma 5.3 for  $\epsilon = \epsilon_m$  and  $\Delta$ . Let  $N(\delta)$  satisfy the conditions of Corollary 5.5 for  $\delta$  and  $\epsilon' = (1 - \delta)^2/\delta$ , so that

$$(5.1.1) \quad \theta_i^m \{V_i^k(y_i) - u_i(s_i^k(y_i)|y_i) > (1 - \delta), n(s_i^k(y_i)|y_i) > N(\delta)\} \\ \leq (1 - \delta)^2/\delta.$$

Finally, choose  $T(\epsilon_m, \delta, \Delta) = [k(\epsilon_m, \Delta) + N(\delta)]/(1 - \delta)$ .

Extracting a subsequence if necessary, we suppose that  $\bar{\theta}$  is a limit of steady states  $\bar{\theta}^m$  for  $(\delta_m, T_m)$ , with  $\delta_m \rightarrow 1$  and  $T_m \geq T(\epsilon_m, \delta_m, \Delta)$ . We claim that no player can gain more than  $3\Delta$  in a single period by playing any  $s'_i$ . If this claim is false, then for  $m$  sufficiently large there is an  $s'_i$  and  $s_i$  with  $s_i \in \text{support}(\bar{\theta}_i)$



such that  $u_i(s'_i, \bar{\theta}^m_i) > u_i(s_i, \bar{\theta}^m_i) + 2\Delta$ . It follows that for all sufficiently large  $m$  and all sufficiently small  $\varepsilon > 0$ , any profile for  $i$ 's opponents that is within  $\varepsilon$  of  $\bar{\theta}^m_i$  gives a gain of at least  $\Delta$  from playing  $s'_i$  instead of  $s_i$ . Thus for any sample  $y_i$ , the maximized probability  $P(s_i, \Delta, y_i)$  that some deviation from  $s_i$  yields a gain of at least  $\Delta$  is at least the posterior probability  $Q_\varepsilon^i(\bar{\theta}^m_i|y_i)$  that player  $i$  assigns to profiles in the set  $B_\varepsilon^i(\bar{\theta}^m_i)$ . From Lemma 5.6, and the fact that  $\bar{\theta}_i(s_i) > 0$ , there is a  $\gamma > 0$  such that for all  $T_m$  and  $\bar{\theta}^m$ ,

$$\theta_i^m \{ Q_\varepsilon^i(\bar{\theta}^m_i|y_i) / Q_\varepsilon^i(\bar{\theta}^m_i|0) \leq \gamma \} \leq \bar{\theta}_i(s_i) / 4.$$

This shows that not too many player  $i$ 's can have observed samples that have caused them to substantially lower the probability they assign to an  $\varepsilon$ -neighborhood of the true steady state. Moreover,  $Q_\varepsilon^i(\bar{\theta}^m_i|0) > \underline{Q} > 0$  since the prior is uniformly bounded away from zero by our assumption of nondoctrinaire priors. Using this fact and our earlier observation that  $P(s_i, \Delta, y_i) \geq Q_\varepsilon^i(\bar{\theta}^m_i|y_i)$ , we have

$$(5.1.2) \quad \theta_i^m \{ P(s_i, \Delta, y_i) \leq \gamma \underline{Q} \} \leq \bar{\theta}_i(s_i) / 4, \quad \text{so that}$$

$$\theta_i^m \{ P(s_i, \Delta, y_i) > \gamma \underline{Q} \} > 1 - \bar{\theta}_i(s_i) / 4.$$

Inequality (5.1.2) gives us a lower bound on the fraction of player  $i$ 's who assign a nonnegligible probability to any strategy yielding a  $\Delta$  improvement over  $s_i$ . Our next steps are to argue that (i) since  $s_i$  has positive probability in the limit, there must be a nonnegligible proportion of the population that has played  $s_i$  many times, intends to play it again, and assigns a nonnegligible probability to any strategy improving on  $s_i$ , and (ii) there must therefore be a nonnegligible proportion of the players that plays  $s_i$  even though it has played it many times before *and* has a nonnegligible expected gain from learning. This last conclusion will then be shown to contradict (5.1.1).

To carry out this program, use Lemma 5.7 and the facts that  $\bar{\theta}_i^m(s_i) \geq \bar{\theta}_i(s_i) / 2$  and  $N(\delta_m) / T_m \leq N(\delta_m) / T(\varepsilon_m, \delta_m, \Delta) \leq 1 - \delta_m$  to conclude

$$(5.1.3) \quad \theta_i^m \{ n(s_i|y_i) > N(\delta_m), s_i^k(y_i) = s_i \} > \bar{\theta}_i(s_i) / 2 - (1 - \delta_m).$$

From DeMorgan's law, the probability of the intersection of the events in (5.1.2) and (5.1.3) is at least the sum of the individual probabilities minus 1, so that

$$(5.1.4) \quad \theta_i^m \{ P(s_i, \Delta, y_i) > \gamma \underline{Q}, n(s_i|y_i) > N(\delta_m), s_i^k(y_i) = s_i \} \\ \geq \bar{\theta}_i(s_i) / 4 - (1 - \delta_m).$$

By construction,  $1 - \underline{k}(\varepsilon_m, \Delta) / T_m \geq 1 - \underline{k}(\varepsilon_m, \Delta) / T(\varepsilon_m, \delta_m, \Delta) \geq \delta_m$ , i.e., at least a  $\delta_m$  fraction of the player  $i$ 's have at least  $\underline{k}$  periods of life remaining (since each generation is of the same size). From Lemma 5.3, when  $\delta_m$  is sufficiently large, the expected gains to learning of all players who have at least  $\underline{k}$  periods

of life are bounded below by a function of  $\Delta \cdot P$ ; substituting this function into (5.1.4) and using DeMorgan's law again yields

$$(5.1.5) \quad \theta_i^m \left\{ \left[ V_i^k(y_i) - u_i(s_i^k(y_i)|y_i) \geq (\Delta\gamma\underline{Q} - \varepsilon_m)(1 - \varepsilon_m) \right], \right. \\ \left. n(s_i|y_i) > N(\delta_m), s_i^k(y_i) = s_i \right\} \geq \bar{\theta}_i(s_i)/4 - (1 - \delta_m).$$

If we now choose  $m$  large enough that  $(1 - \delta_m) < (\Delta\gamma\underline{Q} - \varepsilon_m)(1 - \varepsilon_m)$  and  $\bar{\theta}_i(s_i)/4 - (1 - \delta_m) > (1 - \delta_m)^2/\delta_m$ , we conclude that

$$\theta_i^m \left\{ \left[ V_i^k(y_i) - u_i(s_i^k(y_i)|y_i) > (1 - \delta_m) \right], \right. \\ \left. n(s_i|y_i) > N(\delta_m), s_i^k(y_i) = s_i \right\} > (1 - \delta_m)^2/\delta_m,$$

which contradicts (5.1.1) because the event in this display implies the event whose probability is bounded in (5.1.1). Q.E.D.

6. CONCLUSION

We conclude by examining the scope of Theorem 5.1. Can every Nash equilibrium of every finite game be realized as a limit of steady states as  $T \rightarrow \infty$  and  $\delta \rightarrow 1$ ? Is it really necessary for  $\delta \rightarrow 1$  to achieve Nash equilibrium in the limit as  $T \rightarrow \infty$ ?

Concerning the issue of which Nash equilibria can be obtained as limits of steady states, we do not have a complete answer. It is easy to see that limits of steady states can be mixed as well as pure strategy equilibria. Since any limit point must be a Nash equilibrium, this must be the case in any game that does not have a pure strategy equilibrium. However, whether as  $T \rightarrow \infty$  and  $\delta \rightarrow 1$  it is actually possible to attain some refinement of Nash equilibrium must await further research.

If  $T$  is not large, players do not have much data, and play may be quite arbitrary and heavily influenced by priors. What if  $T \rightarrow \infty$ , but  $\delta$  is not close to one? In this case players will have a great deal of information about those strategies they have chosen to play, but players may not have much incentive to invest in exploring many strategies. Consequently, play may fail to be Nash because untested beliefs about opponents' play off the equilibrium path are not correct.

To see that for  $\delta \rightarrow 1$  a sequence of steady states may fail to converge to Nash equilibrium, we consider an example due to Fudenberg and Kreps (1991). Consider the 3-person game shown in Figure 3. Suppose that the prior of player 1 is that player 3 will play  $R$  with very high probability ( $> 2/3$ ), while that of player 2 is that 3 will play  $L$  with very high probability. If  $\delta = 0$ , or is very small, consider a candidate for a steady state in which all player 1's always play  $A_1$  and all player 2's always play  $A_2$ . This is optimal in the first period of life given the priors and low discount factor, and as a result no information about player 3 is gained, and the proposed play is again optimal in the second period of life and so forth. Consequently, regardless of  $T$  this constitutes a steady state. On the other hand, in any Nash equilibrium either 1 must play  $D_1$  or 2 must play  $D_2$ .

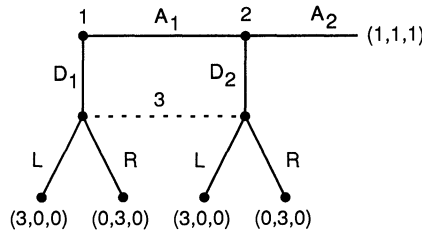


FIGURE 3

We conclude with a brief examination of the consequences of the fact that as  $T \rightarrow \infty$  players must have a great deal of information about those strategies they have chosen to play. Formally, let  $\bar{H}(\sigma)$  be the information sets reached with positive probability when the mixed strategy profile  $\sigma$  is played, let  $\bar{A}(\sigma) = \cup\{A(h)|h \in \bar{H}(\sigma)\}$  be the actions at those information sets, and let  $\bar{X}(\sigma) = \cup\{x|x \in h, h \in \bar{H}(\sigma)\}$  be the corresponding nodes. Let  $\pi_i(\cdot | \sigma_{-i})$  be the behavior strategy for  $i$ 's opponents that corresponds to  $\sigma_{-i}$ . Let us say that the beliefs  $\mu_i$  (over  $\pi_{-i}$ ) are *confirmed* for  $s_i$  and  $\sigma_{-i}$  if

$$\max_{a \in \bar{A}_{-i}(s_i, \sigma_{-i})} \int \|\pi_{-i}(a) - \pi_{-i}(a | \sigma_{-i})\| \mu_i(d\pi_{-i}) = 0,$$

that is,  $\mu_i$  puts probability one on the same play of opponents as does  $\sigma_{-i}$  at those information sets that are reached when  $(s_i, \sigma_{-i})$  is played. This captures the idea that untested beliefs about opponents' play off of the equilibrium path can be incorrect.

**THEOREM 6.1:** *For fixed nondoctrinaire priors  $g_i^0$  and  $\delta < 1$  as  $T_m \rightarrow \infty$  every sequence  $\theta^m$  of steady states has an accumulation point  $\bar{\theta}$ ; if  $\bar{\theta}_i(s_i) > 0$  there exist beliefs  $\mu_i$  that are confirmed for  $s_i$  and  $\bar{\theta}_{-i}$  and such that  $s_i$  maximizes  $u_i(\cdot | \mu_i)$ .*

**REMARK:** This notion of equilibrium is equivalent to the *self-confirming* equilibrium defined and characterized in Fudenberg and Levine (1993).

The idea of the proof should already be clear from our discussion of the proof of Theorem 5.1: Long-lived players will eventually stop experimenting and play to maximize their current expected payoff given their beliefs, and their beliefs about the payoff from any strategy they play many times are approximately correct. The formal proof is in Appendix C.

*Dept. of Economics, M.I.T., Cambridge, MA 02138, U.S.A.*

*and*

*Dept. of Economics, U.C.L.A., Los Angeles, CA 90024, U.S.A.*

*Manuscript received August, 1990; final revision received September, 1992.*

APPENDIX A

LEMMA A.1: If  $x_i, y_i \in \mathbf{R}_+$

$$\left| \prod_{i=1}^n x_i - \prod_{i=1}^n y_i \right| \leq \sum_{i=1}^n \left( \prod_{j=1}^{i-1} y_j \right) |x_i - y_i| \left( \prod_{j=i+1}^n x_j \right);$$

so if  $x_i \leq 1$

$$\left| \prod_{i=1}^n x_i - \prod_{i=1}^n y_i \right| \leq \sum_{i=1}^n \left( \prod_{j=1}^{i-1} y_j \right) |x_i - y_i|.$$

PROOF:

$$\begin{aligned} & \left| \prod_{i=1}^n x_i - \prod_{i=1}^n y_i \right| \\ &= \left| y_1 \prod_{i=2}^n x_i - \prod_{i=1}^n y_i + \prod_{i=1}^n x_i - y_1 \prod_{i=2}^n x_i \right| \\ &\leq y_1 \left| \prod_{i=2}^n x_i - \prod_{i=2}^n y_i \right| + \prod_{i=2}^n x_i |x_1 - y_1| \\ &\leq y_1 y_2 \left| \prod_{i=3}^n x_i - \prod_{i=3}^n y_i \right| + y_1 \prod_{i=3}^n x_i |x_2 - y_2| + \prod_{i=2}^n x_i |x_1 - y_1| \\ &\leq \sum_{i=1}^n \left( \prod_{j=1}^{i-1} y_j \right) |x_i - y_i| \left( \prod_{j=i+1}^n x_j \right). \end{aligned} \quad Q.E.D.$$

LEMMA A.2: Suppose that  $\lambda(\pi|y)$  is a likelihood function for  $\pi$  given a sample  $y$ , that  $g^0(\pi)$  and  $h^0(\pi)$  are two prior densities both of which are bounded above by  $\bar{g}$  and bounded below by  $\underline{g} > 0$ . If  $g(\pi|y)$  and  $h(\pi|y)$  are the corresponding posterior densities, either  $g(\pi|y) = h(\pi|y) = 0$  or

$$\frac{g(\pi|y)}{h(\pi|y)} \leq \left( \frac{\bar{g}}{\underline{g}} \right)^2.$$

PROOF: Proceeds via the calculations

$$\begin{aligned} \frac{g(\pi|y)}{h(\pi|y)} &= \frac{\lambda(\pi|y)g^0(\pi)}{\int \lambda(\pi|y)g^0(\pi) d\pi} \frac{\int \lambda(\pi|y)h^0(\pi) d\pi}{\lambda(\pi|y)h^0(\pi)} \\ &= \frac{g^0(\pi)}{h^0(\pi)} \left[ 1 + \frac{\int \lambda(\pi|y)(h^0(\pi) - g^0(\pi)) d\pi}{\int \lambda(\pi|y)g^0(\pi) d\pi} \right] \\ &\leq \left( \frac{\bar{g}}{\underline{g}} \right) \left[ 1 + \frac{[\bar{g} - \underline{g}] \int \lambda(\pi|y) d\pi}{\underline{g} \int \lambda(\pi|y) d\pi} \right] \\ &= \left( \frac{\bar{g}}{\underline{g}} \right)^2. \end{aligned} \quad Q.E.D.$$

APPENDIX B

We begin the proof of the various lemmas by demonstrating the basic fact that in large samples the posterior is uniformly close to the empirical distribution: this is used in several places below.

LEMMA B.1: For all noninformative priors  $g_i$  there is a nonincreasing function  $\eta(n) \rightarrow 0$  as  $n \rightarrow \infty$  such that for all samples  $y_i$ , information sets  $h$ , actions  $a \in A(h)$ , strategies  $s_i$ , and terminal nodes

$z \in Z(s_i)$ ,

$$(B.1.1) \quad \int \| p_i(z|\pi_{-i}) - \hat{p}_i(z|y_i) \| g_i(\pi_{-i}|y_i) d\pi_{-i} < \max_{x \in X(s_i)} \hat{p}_i(x|y_i) \eta(n(x|y_i)), \text{ and}$$

$$(B.1.2) \quad \int \| \pi_{-i}(a) - \hat{\pi}_{-i}^i(a|y_i) \| g_i(\pi_{-i}|y_i) d\pi_{-i} < \eta(n(h|y_i)).$$

REMARK: Diaconis and Freedman (1990) show that for all samples, Bayes estimates of multinomial probabilities converge to the sample average at a rate that is independent of the particular sample so long as the prior assigns strictly positive density to all distributions. One complication in our model is that even if strategy  $s_i$  has been played many times, there may be information sets  $h$  that are reachable under  $s_i$  but have not been reached in the sample. A related complication is caused by the fact that the distribution on terminal nodes generated by the sample average strategies  $\hat{\pi}$  does not equal the sample average on the terminal nodes. This explains the complicated form of the right-hand side of (B.1.1).

PROOF: Fix a strategy  $s_i$  and terminal node  $z \in Z(s_i)$ . Let  $\{a_{-i}(l, z)\}_{l=1}^L$  be the actions by other players (including nature) that lead to  $z$ , and let  $x(1)$  through  $x(L)$  be the nodes where those actions are taken. It follows from (2.1) and Lemma A.1 in Appendix A that

$$(B.1.3) \quad \| p_i(z|\pi_{-i}) - \hat{p}_i(z|y_i) \| \leq \sum_{l=1}^L \hat{p}_i(x(l)|y_i) \| \pi_{-i}(a_{-i}(l, z)) - \hat{\pi}_{-i}^i(a_{-i}(l, z)|y_i) \|.$$

For each  $h \in H_{-i}$  and  $\varepsilon < 1/\#A(h)$  let  $B_\varepsilon^h$  be the sphere in  $\Delta(A(h))$  of radius  $\varepsilon$  in the sup norm centered at  $\hat{\pi}_{-i}^i(h)$ , and let  $\Pi^{-h}$  be the set of behavior strategies for information sets other than  $h$ . Since  $|\pi_{-i}(a) - \hat{\pi}_{-i}^i(a|y_i)| \leq \varepsilon$  on the set  $B_\varepsilon^h \times \Pi^{-h}$  for all  $a \in A(h)$ ,

$$(B.1.4) \quad \int \| \pi_{-i}(a) - \hat{\pi}_{-i}^i(a|y_i) \| g_i(\pi_{-i}|y_i) d\pi_{-i} \leq \varepsilon + \int_{\sim B_\varepsilon^h \times \Pi^{-h}} g_i(\pi_{-i}|y_i) d\pi_{-i}.$$

Suppose first that  $\tilde{g}_i^0$  is a nondoctinaire prior for which  $\pi_{-i}(h)$  is independent of  $\pi_{-i}(h')$ ,  $h \neq h'$ . Then the corresponding posterior  $\tilde{g}_i$  consists of a product  $\tilde{g}_i = \prod_h \tilde{g}_i^h$  where  $\tilde{g}_i^h$  is a multinomial. Diaconis and Freedman (1990) show that for a multinomial, if  $\varepsilon < 1/\#A(h)$  the posterior odds ratio

$$\int_{\sim B_\varepsilon^h} \tilde{g}_i^h(\pi_{-i}(h)|y_i) d\pi_{-i}(h) / \int_{\sim B_\varepsilon^h} \tilde{g}_i^h(\pi_{-i}(h)|y_i) d\pi_{-i}(h)$$

goes to zero as the sample size  $n_i(h|y_i) \rightarrow \infty$  at an exponential rate that is independent of the particular sample. Since the multinomials are independent, the posterior odds ratio studied by Diaconis and Freedman equals

$$\frac{\int_{\sim B_\varepsilon^h \times \Pi^{-h}} \tilde{g}_i(\pi_{-i}|y_i) d\pi_{-i}}{\int_{B_\varepsilon^h \times \Pi^{-h}} \tilde{g}_i(\pi_{-i}|y_i) d\pi_{-i}} \equiv L_i(B_\varepsilon^h|y_i).$$

Now consider an arbitrary strictly positive prior  $g_i^0$ . Since both  $g_i^0$  and  $\tilde{g}_i^0$  are bounded above and below, we can conclude from Lemma A.2 that there is a constant  $k > 1$  such that for all  $y_i$  and  $\pi_i$ ,

$$\tilde{g}_i(\pi_{-i}|y_i)/k \leq g_i(\pi_{-i}|y_i) \leq k\tilde{g}_i(\pi_{-i}|y_i).$$

Thus,

$$\frac{\int_{\sim B_\varepsilon^h \times \Pi^{-h}} g_i(\pi_{-i}|y_i) d\pi_i}{\int_{B_\varepsilon^h \times \Pi^{-h}} g_i(\pi_{-i}|y_i) d\pi_{-i}} \leq k^2 L_i(B_\varepsilon^h|y_i),$$

and hence goes to zero at the same exponential rate. In particular, for all  $\varepsilon > 0$  there is an  $\eta^\varepsilon(n) \rightarrow 0$  such that (B.1.4) is less than  $\varepsilon + \eta^\varepsilon(n(h|y_i))/L$ . Choose  $N_i$  so that  $\eta^{1/t}(n) \leq 1/t$  for  $n \geq N_i$ . Then  $\eta(n) \equiv \eta^{1/t}(n)$  for  $N_1 + \dots + N_{i-1} \leq n < N_1 + N_2 + \dots + N_i$  satisfies  $\eta(n) \rightarrow 0$  and (B.1.4) less than  $\eta(n(h|y_i))/L$ . We conclude from (B.1.4) that (B.1.2) is valid.

Now we combine (B.1.2) with (B.1.3) to obtain (B.1.1): Let  $h(l)$  be the information set containing  $x(l)$ , and note that  $n(h(l)|y_i) \geq n(x(l)|y_i)$ . Then

$$\begin{aligned}
 \text{(B.1.5)} \quad & \int \| p_i(z|\pi_{-i}) - \hat{p}_i(z|y_i) \| g_i(\pi_{-i}|y_i) d\pi_{-i} \\
 & \leq \sum_{l=1}^L \hat{p}_i(x(l)|y_i) \eta(n(h(l)|y_i)/L) \\
 & \leq \max_{x \in X(s_i)} \hat{p}_i(x|y_i) \eta(n(x|y_i)). \qquad \text{Q.E.D.}
 \end{aligned}$$

LEMMA 5.2: *There exists a function  $\eta(n) \rightarrow 0$  such that for all  $y_i$  and  $\delta$*

$$\begin{aligned}
 \text{(5.2.1)} \quad & \max_{s_i} u_i(s_i|y_i) - u_i(s_i^k(y_i)|y_i) \\
 & \leq V_i^k(y_i) - u_i(s_i^k(y_i)|y_i) \\
 & \leq [\delta/(1-\delta)] \max_{x \in X(s_i^k(y_i))} \hat{p}(x|y_i) \eta(n(x|y_i)).
 \end{aligned}$$

PROOF: The first inequality follows from the fact that  $u_i(s_i|y_i) \leq V_i^k(y_i)$  since playing  $s_i$  for the rest of the game is feasible and yields  $u_i(s_i|y_i)$  in expected value each period.

To demonstrate the second inequality, observe that because the  $V_i^k$ 's are average payoffs  $V_i^{k-1}(y_i) \leq V_i^k(y_i)$ . It follows from the Bellman equation (4.3) that

$$\begin{aligned}
 (1 - \phi_k) V_i^k(y_i) & \leq (1 - \phi_k) u_i(s_i^k(y_i)|y_i) \\
 & \quad + \phi_k \sum_{z \in Z(s_i)} p_i(z) [V_i^{k-1}(y_i, z) - V_i^{k-1}(y_i)],
 \end{aligned}$$

or since  $\phi_k \leq \delta$ ,

$$V_i^k(y_i) \leq u_i(s_i^k(y_i)|y_i) + \delta/(1-\delta) \sum_{z \in Z(s_i)} p_i(z) [V_i^{k-1}(y_i, z) - V_i^{k-1}(y_i)].$$

The final sum represents the expected value of new information. Note that  $V_i^k(y_i) = V_i^k(g_i(\cdot|y_i))$ ; that is, the value function depends only on the current beliefs about  $\pi_{-i}$ . Consequently, the value of new information should be small if the expected change in beliefs is small. To make this idea precise, we introduce the  $l_1$  norm on densities  $\|g_i - g_i'\|_1 = \int \|g_i(\pi_{-i}) - g_i'(\pi_{-i})\| d\pi_{-i}$ . It is standard that  $V_i^k(g_i)$  is Lipschitz in  $g_i$  with Lipschitz constant  $U$  equal to the largest difference in any two payoffs. Consequently

$$\begin{aligned}
 & \sum_{z \in Z(s_i)} p_i(z) [V_i^{k-1}(y_i, a) - V_i^{k-1}(y_i)] \\
 & \leq U \sum_{z \in Z(s_i)} p_i(z) \|g_i(\cdot|(y_i, z)) - g_i(\cdot|y_i)\|_1.
 \end{aligned}$$

We may then calculate

$$\begin{aligned}
 & p_i(z|y_i) \|g_i(\cdot|(y_i, z)) - g_i(\cdot|y_i)\|_1 \\
 & = p_i(z|y_i) \int \| [p_i(z|\pi_{-i}) g_i(\pi_{-i}|y_i) / p_i(z|y_i)] - g_i(\pi_{-i}|y_i) \| d\pi_{-i} \\
 & = \int \| p_i(z|\pi_{-i}) - p_i(z|y_i) \| g_i(\pi_{-i}|y_i) d\pi_{-i} \\
 & \leq \int \| p_i(z|\pi_{-i}) - \hat{p}_i(z|y_i) \| g_i(\pi_{-i}|y_i) d\pi_{-i} \\
 & \quad + \int \| \hat{p}_i(z|y_i) - p_i(z|y_i) \| g_i(\pi_{-i}|y_i) d\pi_{-i} \\
 & \leq 2 \int \| p_i(z|\pi_{-i}) - \hat{p}_i(z|y_i) \| g_i(\pi_{-i}|y_i) d\pi_{-i},
 \end{aligned}$$

where the last inequality follows from

$$p_i(z|y_i) = \int p_i(z|\pi_{-i}) g_i(\pi_{-i}|y_i) d\pi_{-i}.$$

We can now take  $\eta$  to be the function whose existence is proved in Lemma B.1, multiplied by the constant  $2U(\#Z)$ . *Q.E.D.*

LEMMA 5.3: For all  $1 > \varepsilon > 0$  and  $\Delta > 0$ , there are  $\underline{\delta} < 1$  and  $\underline{K}$  such that for all  $\delta \in [\underline{\delta}, 1)$  and for all  $k \geq \underline{K}$

$$(5.3.1) \quad \Delta \cdot P(s_i^k(y_i), \Delta, y_i) - \varepsilon \leq \frac{V_i^k(y_i) - u_i(s_i^k(y_i)|y_i)}{1 - \varepsilon}.$$

PROOF: Let  $\bar{U}$  be the largest absolute payoff. From classical statistics, for all  $\varepsilon$  and  $\Delta$  there exists a  $t$  and a  $t$ -period test procedure for the hypothesis  $\pi_{-i} \in \{\pi_{-i} | u_i(s_i, \pi_{-i}) \geq u_i(s_i^k(y_i), \pi_{-i}) + \Delta\}$  such that the type I and type II errors are less than  $\varepsilon/2\bar{U}$  for all steady states  $\bar{\theta}_{-i}$ . Consider the policy of first running this test, and then playing  $s_i$  for the remaining  $k - t$  periods if the hypothesis is accepted and playing  $s_i^k(y_i)$  otherwise. For notational convenience, let  $\bar{\phi}(t, k) = (\delta^t - \delta^k)/(1 - \delta^k)$  be the weight placed on the last  $k - t$  periods. Then the hypothesis-testing policy yields a utility of at least

$$(5.3.2) \quad - (1 - \bar{\phi}(t, k))\bar{U} + \bar{\phi}(t, k) [u_i(s_i^k(y_i), y_i) + (1 - \varepsilon/2\bar{U})\Delta \cdot P(s_i^k(y_i), \Delta, y_i) - \varepsilon/2] \leq V_i^k(y_i)$$

since the hypothesis-testing value cannot exceed the value of the optimal policy.

Rearranging terms and using  $\bar{\phi}(t, k) \leq 1$ ,  $1 - \varepsilon \leq 1$ , we find that

$$(5.3.3) \quad \Delta P(s_i^k(y_i), \Delta, y_i) \leq 1 / [ \bar{\phi}(t, k)(1 - \varepsilon/2\bar{U}) ] \cdot [ (1 - \bar{\phi}(t, k))\bar{U} - \bar{\phi}(t, k)u_i(s_i^k(y_i), y_i) + \varepsilon/2 + V_i^k(y_i) ] \leq \frac{(1 - \bar{\phi}(t, k))\bar{U} + \varepsilon/2}{\bar{\phi}(t, k)(1 - \varepsilon/2\bar{U})} + \frac{1}{1 - (\varepsilon/2\bar{U})} [V_i^k(y_i) - u_i(s_i^k(y_i)|y_i)].$$

Since as  $\delta \rightarrow 1$ ,  $\bar{\phi}(t, k) \rightarrow 1$ , we can choose  $\underline{\delta}$  and  $\underline{K}$  such that for each  $\delta \in [\underline{\delta}, 1)$  the conclusion of the lemma follows. *Q.E.D.*

We turn now to sampling theory: what fraction of the population can have such a badly biased sample that maximum likelihood estimation yields a poor estimate of the true steady state values?

LEMMA B.2: For all  $\varepsilon > 0$  and  $\eta(n) \rightarrow 0$  as  $n \rightarrow \infty$ , there is an  $N$  for all  $\bar{\theta}_{-i}^m$ ,  $r_i^m$ ,  $s_i$ , and  $T_m$  such that if  $h \in H_{-i}$ ,  $a \in A(h)$ , then

$$(B.2.1) \quad \theta_i^m \{ | \hat{\pi}_{-i}^t(a|y_i) - \bar{\pi}_{-i}(a|\bar{\theta}_{-i}^m) | > \varepsilon, \text{ and } n(h|y_i) > N \} \leq \varepsilon.$$

If  $x \in X(s_i)$ ,

$$(B.2.2) \quad \theta_i^m \{ n(x|y_i) \leq [ \bar{p}_i(x|\bar{\theta}_{-i}^m) - \varepsilon ] n(s_i|y_i), \text{ and } n(s_i|y_i) > N \} \leq \varepsilon,$$

$$(B.2.3) \quad \theta_i^m \left\{ \max_x \| \hat{p}_i(x|y_i) - \bar{p}_i(x|\bar{\theta}_{-i}^m) \| > \varepsilon, \text{ and } n(s_i|y_i) > N \right\} \leq \varepsilon, \text{ and}$$

$$(B.2.4) \quad \theta_i^m \left\{ \max_x \bar{p}_i(x|\bar{\theta}_{-i}^m) \eta(n(x|y_i)) > \varepsilon, \text{ and } n(s_i|y_i) > N \right\} \leq \varepsilon.$$

The first statement says that the population fraction having both a badly biased sample on  $a \in A(h)$  and a large number of observations on  $h$  is uniformly small regardless of  $\bar{\theta}_{-i}^m$ . The second says that the population fraction that has both played a strategy making  $x$  likely many times, and

has seen  $x$  only rarely is uniformly small regardless of  $\bar{\theta}_{-i}^m$ . Notice what is asserted: the fraction both with a large sample and a biased sample is small. It need not be true that of those who have a large sample most have an unbiased sample.

PROOF OF LEMMA B.2: Since each period's actions are i.i.d., we can model the distribution on opponents' actions faced by player  $i$  by a probability space whose elements are what he will see the  $k$ th time he arrives at information set  $h$ .<sup>12</sup> The Glivenko-Cantelli Theorem shows that the empirical distribution at each information set  $h$  converges to the distribution induced by  $\bar{\theta}_{-i}^m$  as the number of observations  $n(h|y_i) \rightarrow \infty$ , at a rate that holds uniformly over all values of  $\bar{\theta}_{-i}^m$  and the finite number of information sets. It remains to explain why the rate is uniform over all "sampling rules"  $r_i(y_i)$ . This follows from the fact that the desired inequalities hold *even* if player  $i$  is informed of the entire sample path of each information set before choosing his rule: Through such anticipatory sampling he may be able to ensure that most of the long samples are biased, but since there are few paths where the empirical distribution is not close to the theoretical one after  $N$  samples, there is no sampling rule for which the probability of long and biased samples is large.

A similar argument yields (B.2.2), except that now we use a probability space in which the elements are what player  $i$  will see the  $k$ th time he plays strategy  $s_i$ .

Next we turn to (B.2.3). By Lemma A.1 if  $\{a_{-i}(l, x)\}_{l=1}^L$  are the moves by other players or nature and  $x(1), \dots, x(L)$  are the nodes leading to  $x$

$$(B.2.5) \quad \begin{aligned} & \|\hat{p}_i(x|y_i) - \bar{p}_i(x|\bar{\theta}_{-i}^m)\| \\ & \leq \sum_{l=1}^L \bar{p}_i(x(l)|\bar{\theta}_{-i}^m) \|\hat{\pi}_{-i}^i(a_{-i}(l, x)|y_i) - \bar{\pi}_{-i}(a_{-i}(l, x)|\bar{\theta}_{-i}^m)\| \\ & \leq \sum_{l=1}^L \min \{ \bar{p}_i(x(l)|\bar{\theta}_{-i}^m), \|\hat{\pi}_{-i}^i(a_{-i}(l, x)|y_i) - \bar{\pi}_{-i}(a_{-i}(l, x)|\bar{\theta}_{-i}^m)\| \} \end{aligned}$$

where the last step follows from  $\bar{p}_i, \hat{\pi}_{-i}^i, \bar{\pi}_{-i} \leq 1$ . From (B.2.5) it follows that if

$$(B.2.6) \quad \begin{aligned} & \max_{x \in X(s_i)} \|\hat{p}_i(x|y_i) - \bar{p}_i(x|\bar{\theta}_{-i}^m)\| > \varepsilon \text{ occurs,} \\ & \min \{ \bar{p}_i(x|\bar{\theta}_{-i}^m), \|\hat{\pi}_{-i}^i(a|y_i) - \bar{\pi}_{-i}(a|\bar{\theta}_{-i}^m)\| \} > \varepsilon/L \end{aligned}$$

for some  $x \in X(s_i)$  and  $a \in A(h)$  with  $x \in h$ . Consequently, if we can show how to find  $N$  for each such  $x$  and  $a$  so that

$$(B.2.7) \quad \theta_i^m \{ \bar{p}_i(x|\bar{\theta}_{-i}^m) > \varepsilon/L, \|\hat{\pi}_{-i}^i(a|y_i) - \bar{\pi}_{-i}(a|\bar{\theta}_{-i}^m)\| > 2\varepsilon/3L, n(s_i|y_i) > N \} \leq \varepsilon,$$

then (B.2.3) will follow from (B.2.6).

Choose  $N_1$  so that (B.2.1) holds for  $2\varepsilon/3L$ , and choose  $N$  so that (B.2.2) holds for  $\varepsilon/3L$  and so that  $N > 3LN_1/2\varepsilon$ . If  $x \in h$ ,  $n(h|y_i) > n(x|y_i)$ . Suppose  $\bar{p}_i(x|\bar{\theta}_{-i}^m) > \varepsilon/L$ . We now use (B.2.2), recalling that  $\varepsilon$  has been replaced with  $\varepsilon/3L$ . We may substitute  $n(h|y_i)$  for  $n(x|y_i)$  and  $\varepsilon/L$  for  $\bar{p}_i(x|\bar{\theta}_{-i}^m)$  while maintaining the final inequality. Also (B.2.2) presumes  $n(s_i|y_i) > N$ ; combining with  $N > 3LN_1/2\varepsilon$ , we get

$$(B.2.8) \quad \theta_i^m \{ n(h|y_i) \leq N_1 \text{ and } n(s_i|y_i) > N \} \leq \varepsilon/3L.$$

On the other hand by (B.2.1)

$$(B.2.9) \quad \theta_i^m \{ |\hat{\pi}_{-i}^i(a|y_i) - \bar{\pi}_{-i}^m(a|\bar{\theta}_{-i}^m)| > 2\varepsilon/3L \text{ and } n(h|y_i) > N_1 \} \leq 2\varepsilon/3L.$$

Combining the events in (B.2.8) and (B.2.9) we get

$$(B.2.10) \quad \theta_i^m \{ |\hat{\pi}_{-i}^i(a|y_i) - \bar{\pi}_{-i}^m(a|\bar{\theta}_{-i}^m)| > 2\varepsilon/3L, n(s_i|y_i) > N \} \leq \varepsilon/L \leq \varepsilon$$

whenever  $\bar{p}_i(x|\bar{\theta}_{-i}^m) > \varepsilon/L$ , which proves (B.2.7), and consequently (B.2.3).

<sup>12</sup> The more obvious probability space would have elements corresponding to what player  $i$  would see if he plays  $s_i$  in period  $t$  for each date  $t$  of his life. Then, for each sample path, the realized terminal node the player sees the first time he plays  $s_i$  depends on the period in which the strategy is first used. Our alternative generates the same probability distribution over observations.



To show (B.2.4) we proceed as in (B.2.5): If  $\bar{p}_i(x|\bar{\theta}_{-i}^m)\bar{\eta} \leq \varepsilon/2$  we are done. Let  $N_1$  be large enough that  $\eta(n) < \varepsilon/2$  for  $n \geq N_1$ ; then choose  $N$  so that (B.2.2) holds for  $\varepsilon' < \varepsilon/2\bar{\eta}$  such that  $(\varepsilon/2\bar{\eta} - \varepsilon')N \geq N_1$ . Q.E.D.

Our next step is to argue that the players are unlikely to have a large but inaccurate sample, so that they are unlikely to be confident of an incorrect forecast. We wish this to be true uniformly over the population fractions  $\theta_i^m$ , which will follow from the uniform version of the strong law of large numbers, that is, the Glivenko-Cantelli Theorem.

Recall that  $\theta_i^m(\bar{Y}_i)$  is the steady state fraction of population  $i$  whose histories  $y_i$  lie in  $\bar{Y}_i$ . Because our aggregate system is deterministic,  $\theta_i^m(\bar{Y}_i)$  is equal to the expected frequency with which the "old" (age  $T_m$ ) players encounter events in  $\bar{Y}_i$ . In particular, for a set  $\bar{Y}_i$  that consists of all subhistories of a set of terminal histories (i.e., histories of length  $T_m$ ) and for a particular terminal history  $y_i^m$ , define  $J(y_i^m, \bar{Y}_i)$  to be the number of times that a subhistory of  $y_i^m$  lies in  $\bar{Y}_i$ . Then we have

$$(B.4) \quad \theta_i^m(\bar{Y}_i) = \sum_{y_i^m \in \bar{Y}_i} J(y_i^m, \bar{Y}_i) \theta_i^m(y_i^m),$$

so that to bound the population fractions, it suffices to bound the probabilities of the corresponding length- $T_m$  histories.

Our goal is to relate  $\theta_i^m$  representing the population fractions with each experience to the fractions  $\bar{p}_i$  that determine the probability distributions over observations.

LEMMA 5.4: For all  $\varepsilon > 0$  and  $\eta(n) \rightarrow 0$  as  $n \rightarrow \infty$  there is an  $N$  such that for all  $T_m, \bar{\theta}_{-i}^m, r_i^m$ , and  $s_i$ ,

$$\theta_i^m \left\{ \max_{x \in X(s_i)} \hat{p}_i(x|y_i) \eta(n(x|y_i)) > \varepsilon, \text{ and } n(s_i|y_i) > N \right\} \leq \varepsilon.$$

PROOF: Letting  $\bar{\eta} = \sup_n \eta(n)$  by (B.2.3) we may choose  $N$  large enough that

$$\theta_i^m \left\{ \max_{x \in X(s_i)} \|\hat{p}_i(x|y_i) - \bar{p}_i(x|\bar{\theta}_{-i}^m)\| \bar{\eta} > \varepsilon/2 \text{ and } n(s_i|y_i) > N \right\} < \varepsilon/2$$

so that the conclusion follows from (B.2.4). Q.E.D.

LEMMA 5.6: For all  $\varepsilon > 0$  there exists a  $\gamma > 0$  such that for all  $y_i, \bar{\theta}_{-i}^m, r_i^m$ , and  $T_m$

$$\theta_i^m \left\{ Q_\varepsilon^i(\bar{\theta}_{-i}^m|y_i) / Q_\varepsilon^i(\bar{\theta}_{-i}^m|0) \leq \gamma \right\} \leq \varepsilon.$$

PROOF: Fix  $g_i^0$  and  $\varepsilon$  and let  $B \equiv B_\varepsilon^i(\bar{\theta}_{-i}^m)$ . We must find  $\gamma$  so that regardless of  $y_i, \bar{\theta}_{-i}^m, r_i^m$ , and  $T_m$

$$\theta_i^m \left\{ \frac{\int_B g_i(\pi_{-i}|y_i) d\pi_{-i}}{\int_B g_i^0(\pi_{-i}) d\pi_{-i}} \leq \gamma \right\} \leq \varepsilon.$$

Since  $g_i^0$  is bounded away from 0, and  $\varepsilon$  is fixed, it suffices to find a  $\gamma'$  so that

$$\theta_i^m \left\{ \int_B g_i(\pi_{-i}|y_i) d\pi_{-i} \leq \gamma' \right\} \leq \varepsilon.$$

Recall that  $B = \{\pi_{-i} | \max_z \|\bar{p}_i(z|\bar{\theta}_{-i}^m) - p_i(z|\pi_{-i})\| \leq \varepsilon\}$ . By (B.1.1) of Lemma B.1, Lemma 5.4, and  $\#Z > 1$ , we may find an  $N$  so that

$$\theta_i^m \left\{ \int \|\bar{p}_i(z|\bar{\theta}_{-i}^m) - p_i(z|\pi_{-i})\| g_i(\pi_{-i}|y_i) d\pi_{-i} > 2\varepsilon\#Z/3, n(s_i|y_i) \geq N \right\} \leq \varepsilon/\#Z.$$

Making use of the fact that  $\Pr(\max_z X_z > x) \leq \#Z \max \Pr(X_z > x)$ , it follows that

$$\theta_i^m \left\{ \int \max_z \|\bar{p}_i(z|\bar{\theta}_{-i}^m) - p_i(z|\pi_{-i})\| g_i(\pi_{-i}|y_i) d\pi_{-i} > 2\varepsilon/3, n(s_i|y_i) \geq N \right\} \leq \varepsilon.$$

Now  $I \equiv \int \max_z \|\bar{p}_i(z|\bar{\theta}_{-i}) - p_i(z|\pi_{-i})\| g_i(\pi_{-i}|y_i) d\pi_{-i}$  may be written as the sum of integrals over  $B$  and  $\sim B$ , so in particular  $I > 2\varepsilon/3$  if

$$\varepsilon \int_{\sim B} g_i(\pi_{-i}|y_i) d\pi_{-i} > 2\varepsilon/3.$$

Since  $\int_B g_i(\pi_{-i}|y_i) d\pi_{-i} = 1$ ,  $I > 2\varepsilon/3$  also follows from  $\int_B g_i(\pi_{-i}|y_i) d\pi_{-i} \leq 1/3$ . We conclude

$$\theta_i^m \left\{ \int_B g_i(\pi_{-i}|y_i) d\pi_{-i} \leq 1/3, n(s_i|y_i) \geq N \right\} \leq \varepsilon.$$

If we take  $\gamma' = \min\{1/3, \min_{\{y_i | n(s_i|y_i) < N\}} \int_B g_i(\pi_{-i}|y_i) d\pi_{-i}\}$ ,

$$\theta_i^m \left\{ \int_B g_i(\pi_{-i}|y_i) d\pi_{-i} \leq \gamma' \right\} \leq \varepsilon. \tag{Q.E.D.}$$

LEMMA 5.7: Let  $T_m \rightarrow \infty$  be a sequence of lifetimes, and  $\bar{\theta}^m$  be a subsequence of steady states that converge to  $\bar{\theta}$ , and let  $r_i^m$  be the corresponding rules. If  $\bar{\theta}_i(s_i) > 0$ , then

$$(5.7.1) \quad \theta_i^m \{ n(s_i|y_i) > N \text{ and } r_i^m(y_i) = s_i \} > \bar{\theta}_i^m(s_i) - (N/T_m).$$

PROOF: Since  $\bar{\theta}_i(s_i) > 0$  there exists an  $\varepsilon > 0$  and  $\bar{m}$  such that  $\bar{\theta}_i^m(s_i) \geq 2\varepsilon$  for all  $m \geq \bar{m}$ . Now fix  $N$ . For any history  $y_i$ , there are at most  $N$  subhistories  $y'_i$  for which  $r_i^m(y'_i) = s_i$ , and  $n(s_i|y_i) \leq N$ . Since  $\theta_i^m(y_i) \leq 1/T_m$ , equation (B.4) shows that

$$(5.7.2) \quad \theta_i^m \{ n(s_i|y_i) \leq N \text{ and } r_i^m(y_i) = s_i \} < N/T_m.$$

Since  $\bar{\theta}_i^m(s_i)$  is the sum of the fractions playing  $s_i$  with  $n(s_i|y_i) \leq N$  and those playing  $s_i$  with  $n(s_i|y_i) > N$ , (5.7.1) follows. Q.E.D.

The following lemma is used in Appendix C:

LEMMA B.3: For all  $\varepsilon > 0$  there is an  $N$  such that, for all  $T_m$ ,  $\bar{\theta}_i^m$ ,  $r_i^m$ , and  $s_i$ , if  $h \in H_{-i}$  and  $a \in A(h)$ ,

$$\theta_i^m \left\{ \int \|\pi_{-i}(a) - \pi_{-i}(a|\bar{\theta}_{-i})\| g_i(\pi_{-i}|y_i) d\pi_{-i} > \varepsilon \text{ and } n(h|y_i) > N \right\} \leq \varepsilon.$$

PROOF: This combines B.2.1 from Lemma B.2 with B.1.2 from Lemma B.1. Q.E.D.

### APPENDIX C

THEOREM 6.1: For fixed nondoctrinaire priors  $g_i^0$  and  $\delta < 1$  as  $T_m \rightarrow \infty$  every sequence  $\theta^m$  of steady states has an accumulation point  $\bar{\theta}$ ; if  $\bar{\theta}_i(s_i) > 0$  there exist beliefs  $\mu_i$  that are confirmed for  $s_i$  and  $\bar{\theta}_{-i}$  and such that  $s_i$  maximizes  $u_i(\cdot|\mu_i)$ .

PROOF OF THEOREM 6.1: Let  $\bar{\theta}^m$  be a subsequence of steady states that converge to  $\bar{\theta}$ , and let  $r_i^m$  be the corresponding optimal rules. Suppose  $\bar{\theta}_i(s_i) > 0$ .

We will say that  $s_i$  is a  $\varepsilon$ -static best response to marginal beliefs  $p_i$  if, for all  $s'_i$ ,

$$u_i(s_i, p_i) + \varepsilon \geq u_i(s'_i, p_i).$$

Fix  $\varepsilon$ . Fix  $\eta(n)$  so that Lemma 5.2 holds and let  $N$  be large enough that Lemma 5.4 holds as well. By Lemma 5.2 and the fact that  $r_i^m$  is optimal

$$(6.1.1) \quad \theta_i^m \{ r_i^m(y_i) = s_i, s_i \text{ is not a } 2 \cdot \#Z(s_i)U\varepsilon/(1-\delta)\text{-static best response to } p_i(\cdot|y_i), \text{ and } n(s_i|y_i) > N \} \leq \varepsilon.$$

Let  $\bar{X}_{-i}(s_i)$  be the nodes hit (in the limit) with positive probability when  $s_i$  is played:  $\bar{X}_{-i}(s_i) = \{\bar{X}(s_i, \bar{\theta}_{-i})\}$ . Let  $\bar{p} \equiv \min_{s_i, x \in \bar{X}_{-i}(s_i)} \bar{p}(x|\bar{\theta}_{-i})$ . We may assume  $m$  is large enough that  $\bar{p}(x|\bar{\theta}_{-i}^m) \geq \bar{p}/2$  for  $x \in \bar{X}_{-i}(s_i)$  and  $\|\pi_{-i}(a|\bar{\theta}_{-i}^m) - \pi_{-i}(a|\bar{\theta}_{-i})\| \leq \varepsilon$ . Note that for  $x \in h$ ,  $n(h|y_i) \geq n(x|y_i)$ . Consequently, by (5.4.1) of Lemma 5.4, for any  $\varepsilon$  we may find an  $N$  so that

$$(6.1.2) \quad \theta_i^m \left\{ \max_{h \in H_{-i} \cap \bar{H}(s_i, \bar{\theta}_{-i})} \eta(n(h|y_i)) > 2\varepsilon/\bar{p}, \text{ and } n(s_i|y_i) > N \right\} \leq \varepsilon.$$

We may take  $N$  large enough to satisfy the conclusion of Lemma B.3.

$$(6.1.3) \quad \theta_i^m \left\{ \max_{a \in \bar{A}_{-i}(s_i)} \int \|\pi_{-i}(a) - \pi_{-i}(a|\bar{\theta}_{-i})\| g_i(\pi_{-i}|y_i) d\pi_{-i} > 3\varepsilon, \right. \\ \left. \text{and } n(s_i|y_i) > N \right\} \leq 2\varepsilon.$$

Since  $\bar{\theta}_i(s_i) > 0$ , choosing  $m$  large enough that  $N/T_m < \varepsilon'/2$  and  $(\bar{\theta}_i(s_i) - \theta_i^m(s_i)) < \varepsilon'/2$ , it follows from Lemma 5.7 that

$$(6.1.4) \quad \theta_i^m \{n(s_i|y_i) > N \text{ and } r_i^m(y_i) = s_i\} > \varepsilon'.$$

Combining (6.1.1) through (6.1.4) yields

$$(6.1.5) \quad \theta_i^m \left\{ r_i^m(y_i) = s_i, s_i \text{ is a static } 2\#ZU\varepsilon/(1-\delta) \text{ best-response to } p_i(\cdot|y_i), \right. \\ \left. \max_{a \in \bar{A}_{-i}(s_i)} \int \|\pi_{-i}(a) - \pi_{-i}(a|\bar{\theta}_{-i})\| g_i(\pi_{-i}|y) d\pi_{-i} \leq 3\varepsilon \text{ and } n(s_i|y_i) > N \right\} \\ \geq \varepsilon' - 3\varepsilon.$$

If we take  $3\varepsilon < \varepsilon'$ , we conclude that there is a  $y_i^\varepsilon$  such that  $s_i$  is a  $2\#Z(s_i)U\varepsilon/(1-\delta)$  best response to  $p_i(\cdot|g_i(\cdot|y_i^\varepsilon))$  with

$$\max_{a \in \bar{A}_{-i}(s_i)} \int \|\pi_{-i}(a) - \pi_{-i}(a|\bar{\theta}_{-i})\| g_i(\bar{\pi}_{-i}|y_i^\varepsilon) d\pi_{-i} \leq 3\varepsilon.$$

Taking  $\varepsilon \rightarrow 0$ , we see that as a measure on  $\pi_{-i}$ ,  $g_i(\cdot|y_i^\varepsilon)$  has a weak limit point  $\mu_i$ . Then  $s_i$  is a best response to  $p_i(\cdot|\mu_i)$ , that is, maximizes  $u_i(\cdot, \mu_i)$ , and

$$\max_{a \in \bar{A}_{-i}(s_i)} \int \|\pi_{-i}(a) - \pi_{-i}(a|\bar{\theta}_{-i})\| \mu_i(d\pi_{-i}) = 0. \quad Q.E.D.$$

### REFERENCES

- AUMANN, R. (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1-18.
- BATTIGALLI, P., AND P. GUATOLI (1988): "Conjectural Equilibrium," mimeo, Bocconi University.
- BOYLAN, R. (1990): "Continuous Approximation of Dynamical Systems with Randomly Matched Individuals," mimeo, Washington University, St. Louis.
- BRANDENBURGER, A., AND E. DEKEL (1987): "Rationalizability and Correlated Equilibrium," *Econometrica*, 55, 1391-1402.
- BRAY, M. M. (1982): "Learning, Estimation and the Stability of Rational Expectations," *Journal of Economic Theory*, 26, 318-339.
- CANNING, D. (1989): "Convergence to Equilibrium in a Sequence of Games With Learning," mimeo, Cambridge University.
- (1990): "Social Equilibrium," mimeo, Cambridge University.
- DIACONIS, P., AND D. FREEDMAN (1990): "On the Uniform Consistency of Bayes Estimates for Multinomial Probabilities," *The Annals of Statistics*, 18, 1317-1327.
- FUDENBERG, D., AND D. M. KREPS (1991): "Learning and Equilibrium in Games," CORE Lecture

- Series, Belgium.
- FUDENBERG, D. AND D. K. LEVINE (1990): "Steady State Learning and Self-Confirming Equilibrium," mimeo.
- (1993): "Self Confirming Equilibrium," *Econometrica*, 61, 523–545.
- FUDENBERG, D., AND J. TIROLE (1991): *Game Theory*. Cambridge, MA.: MIT Press.
- KALAI, E., AND E. LEHRER (1991a): "Private Beliefs Equilibrium," mimeo, Northwestern.
- (1991b): "Rational Learning Leads to Nash Equilibrium," mimeo, Northwestern.
- KOHLBERG, E., AND J. F. MERTENS (1986): "On the Strategic Stability of Equilibrium," *Econometrica*, 54, 1003–1034.
- KREPS, D. M. (1990): *A Course in Microeconomic Theory*. Princeton, N.J.: Princeton University Press.
- KUHN, H. (1953): "Extensive Games and the Problem of Information," *Annals of Mathematics Studies*, 28. Princeton, N.J.: Princeton University Press.
- ROSENTHAL, R. (1979): "Sequences of Games With Varying Opponents," *Econometrica*, 47, 1353–1366.
- RUBINSTEIN, A., AND A. WOLINSKY (1990): "Rationalizable Conjectural Equilibrium Between Nash and Rationalizability," mimeo, Northwestern.
- SHAPLEY, L. (1964): "Some Topics in Two-Person Games," *Annals of Mathematical Studies*, 52. Reprinted in *Advances in Game Theory*, ed. by M. Descher, L. Shapley, A. W. Tucker. Princeton, N.J.: Princeton University Press, 1–28.
- TAN, T., AND S. WERLANG (1988): "The Bayesian Foundations of Solution Concepts of Games," *Journal of Economic Theory*, 45, 370–392.