GAMES AND ECONOMIC BEHAVIOR 5, 320–367 (1993)

Learning Mixed Equilibria*

DREW FUDENBERG

Department of Economics, Harvard University, Cambridge, Massachusetts 02138

AND

DAVID M. KREPS

Graduate School of Business, Stanford University, Stanford, California 94305, and Department of Economics, Tel Aviv University, Tel Aviv, Israel

Received July 8, 1992

We study learning processes for finite strategic-form games, in which players use the history of past play to forecast play in the current period. In a generalization of fictitious play, we assume only that players asymptotically choose best responses to the historical frequencies of opponents' past play. This implies that if the stage-game strategies converge, the limit is a Nash equilibrium. In the basic model, plays seems unlikely to converge to a mixed-strategy equilibrium, but such convergence is natural when the stage game is perturbed in the manner of Harsanyi's purification theorem. *Journal of Economic Literature* Classification Number: C72. © 1993 Academic Press, Inc.

1. INTRODUCTION

Nash equilibrium describes a situation in which players have identical and exactly correct beliefs about the strategies each player will choose. How and when might the players come to have correct beliefs, or at least beliefs that are close enough to being correct that the outcome corresponds to a Nash equilibrium? One explanation is that the players play the game over and over and that their beliefs come to be correct as the result of learning from past play. This explanation has been explored at some length in the recent literature, in models that take a number of different forms and stress different aspects of the problems.[1]

This paper explores learning models that are in the spirit of the model or method of fictitious play (Brown, 1951; Robinson, 1951) in which players choose their strategies to maximize their current period's expected payoff on the assumption that their opponents will play each strategy with probability equal to its historical frequency. We extend the previous literature in three ways:

(1) We provide some minor extensions to the basic model of fictitious play, by generalizing the classes of rules by which players form their beliefs and use them to choose their actions.

(2) We study models in the spirit of fictitious play for games with more than two players.

(3) Most importantly, we reformulate the study of convergence to mixed-strategy equilibria. We argue that the notion of convergence used previously in the literature, that the empirical marginal distributions converge, is not an appropriate notion of what it means to play a mixed-strategy profile, and we suggest and analyze the stronger criterion of the convergence of intended behavior. We show that all Nash equilibria and only Nash equilibria are possible limit points under this mode of convergence. Finally, we investigate the *global* stability of mixed equilibria in the setting of Harsanyi's (1973) purification theorem.

Section 2 gives a general formulation of learning in a strategic-form game. This formulation, and our subsequent analysis, supposes that the same players play each other repeatedly (as opposed to a model with a large number of player 1's, player 2's, etc.) and that in each round of play, players observe the (pure) strategies chosen by their rivals.

Section 3 reviews the model of fictitious play for two players. We separate the questions addressed by fictitious play (and models of learning in general) into two groups. First, if play "settles down" or converges in some appropriate sense, what are the possible limit points? Second, is play guaranteed to converge?

With regard to the first question, recall that there are two modes of

* We are grateful to Robert Anderson, Robert Aumann, Vincent Crawford, Hugo Hopenhayn, David Levine, Marco Li Calzi, Ramon Marimon, Tom Sargent, and Josè Scheinkman for helpful comments. We thank IDEI, Toulouse, and the Institute for Advanced Studies, Tel Aviv University, for their hospitality while this research was being conducted. The financial assistance of the National Science Foundation (Grants SES 88-08204, SES 90-08770, SES 89-08402, and SES 92-08954) and the John Simon Guggenheim Foundation is gratefully acknowledged.

0899-8256/93 $5.00

[1] To provide a guide to the literature would take too long, so we remain content with a partial list of recent references: Canning (1991), Crawford and Haller (1990), Eichenberger *et al.* (1991), Ellison (1993), Fudenberg and Levine (1993), Fudenberg and Kreps (1988), Hendon *et al.* (1991), Jordan (1991), Kalai and Lehrer (1993), Kandori *et al.* (1993), Milgrom and Roberts (1990, 1991), Nyarko (1991), and Young (1993). We comment on some of these papers as we proceed, when they bear directly on our own analysis.

convergence in the literature on fictitious play. In the first, there is a finite time T such that a single strategy profile is played in every period from T on; it is easy to see that any such profile must be a Nash equilibrium. In the second mode of convergence, play cycles among different strategy profiles in such a way that the empirical frequencies of each player's choices converge to some (mixed) strategy. The corresponding strategy profile is also a Nash equilibrium; this is the traditional sense in which fictitious play is said to converge to mixed-strategy equilibria. Section 3 briefly reviews results from the literature that use these two convergence notions.

In Section 4, we generalize fictitious play by considering more general assumptions about the ways in which players construct their assessments and then choose their immediate actions. We assume throughout that players' choices of actions are *asymptotically myopic*; i.e., in the long run, players choose in a way that maximizes their immediate payoffs. This assumption requires some explanation and rationale, which we provide. As for players' assessments, if they are *adaptive* (following Milgrom and Roberts (1991)), and if intended play converges to a pure-strategy profile, the profile must be a Nash equilibrium, for any number of players. But more is required if the second form of convergence—convergence of empirical frequencies to a mixed-strategy profile—is to have only Nash equilibria as limit points. A sufficient condition is that assessment rules are *asymptotically empirical*, which means that players' assessments converge together with empirical frequencies. Moreover, this condition suffices only for two-player games.

Section 5 presents several objections to convergence of empirical frequencies as an appropriate mode of convergence for learning to play a mixed-strategy profile. In summary, these objections are: (1) Although assessments are converging (if they are asymptotically empirical), the strategies that are chosen are not; (2) in examples, correlations will be observed over time in the actions of players who choose their actions independently; and (3) because of (2), convergence in this sense for games with more than two players is problematic.

For these reasons, in Section 6 we propose a stronger mode of convergence, namely *convergence of (intended) behavior*. This raises some technical complications: When players use mixed strategies, the realized distribution of play need not equal the intended one, which makes notions of convergence inherently probabilistic. We attend to these complications and then show that only Nash equilibria are possible limit points under this mode of convergence, as long as behavior is asymptotically myopic and assessments are asymptotically empirical. Moreover, for any game and Nash equilibrium of the game, there is a model of asymptotically myopic behavior and asymptotically empirical assessments for which the

equilibrium is a limit point of intended behavior with probability as close to one as desired.[2] These results are not limited to two-player games.

The problem with this alternative mode of convergence is that, while convergence to mixed behavior is possible, it is hard to see why it should occur. The difficulty is the standard one with mixed strategies: If, based on their assessments, players choose their actions to maximize precisely their expected payoffs, then (unless their assessments are precisely those of the mixed equilibrium) their intended behavior will not converge. If players are not restricted to precise maximization, then any behavior (that puts weight on actions in the equilibrium mixture) will be satisfactory. Something outside of payoff-maximization considerations is required to lead players to the precise mixtures needed in the equilibrium. In our basic formulation, we see no natural way of doing this.

As a way around this problem, in Sections 7 and 8 we consider learning in games in which each player's payoff is subject to a sequence of i.i.d. random shocks that are observed only by that player, as suggested by Harsanyi's (1973) purification theorem. In this context, a mixture over two strategies does not correspond to mixing by a player who is indifferent, but rather to a player who (in each period) strictly prefers one of the two strategies, depending on the (period's) realization of the player's payoff perturbation. Yet from the perspective of other players, who do not know the precise value of this period's perturbation for the player, the actions of the first player are random. We show in Section 7 that in this context all of our earlier results go through without difficulty. Then, in Section 8, we specialize to the class of 2×2 games with a unique equilibrium in mixed strategies, and we show that any learning process that is close to fictitious play (in a sense to be made precise) will converge with probability one to the unique equilibrium. Here we use and adapt results from the theory of stochastic approximation (Arthur *et al.*, 1987; Kushner and Clark, 1978; and Ljung and Söderstrom, 1983).[3]

Before setting out, let us note that the results given here are only a small part of the overall story. Among other things, we are assuming that players encounter the same opponents repeatedly and yet act myopically, a fairly unsavory combination (see Section 4); they observe the full (stage-game-pure) strategies chosen by rivals in each round of play; and they play the same game over and over. We hope to return to each of these three simplifying assumptions in subsequent work.

[2] To the extent that some Nash equilibria seem unreasonable, such as those where players use weakly dominated strategies, this last result indicates that our assumptions are too weak. This is discussed as well in Section 6.

[3] In this respect, our work is similar to that of Marcet and Sargent (1989a, 1989b) on learning rational expectations equilibrium.

2. Formulation

Fix an I-player, finite, strategic-form game, hereafter referred to as the *stage game*. The players are indexed $i = 1, 2, \ldots, I$, and we let $-i$ denote the "other" players; i.e., $-i = \{1, 2, \ldots, i-1, i+1, \ldots, I\}$.[4] Let S^i, $i = 1, \ldots, I$, be the finite set of pure strategies (or actions) for player i, let $S = S^1 \times \cdots \times S^I$ be the set of pure-strategy profiles, and let u^i: $S \to R$ give player i's payoffs. In the usual fashion, let Σ^i be the mixed strategies for player i, let $\Sigma = \Sigma^1 \times \cdots \times \Sigma^I$ be the set of mixed-strategy profiles, and extend the domain of u^i from S to Σ. Also, for each i let S^{-i} denote $\Pi_{j \neq i} S^j$, and let Σ^{-i} denote the set of probability distributions over S^{-i}; for $s^i \in S^i$ and $\sigma^{-i} \in \Sigma^{-i}$, let $u^i(s^i, \sigma^{-i})$ denote i's expected utility if she chooses pure strategy s^i and her rivals act according to the (possibly correlated) distribution σ^{-i}.

Imagine that these players play the game repeatedly, at dates $t = 1, 2, \ldots$. Imagine that after each round of play, players observe the actual actions chosen by their opponents; i.e., the pure strategy that is chosen is observed. If a player chooses his action using a mixed strategy, the mixing is not observed. Then a history of play up to time t, denoted ζ_t, is a string of (pure) strategy profiles $\zeta_t = (s_1, \ldots, s_{t-1})$, where $s_{t'} \in S$ for $t' = 1, \ldots, t-1$. The set of all histories of play up to time t, or $(S)^{t-1}$, is denoted by $\mathcal{Z}_t$.[5] By convention, $\mathcal{Z}_1$ denotes the (singleton) set consisting of the null history. Also, $\mathcal{Z}$ denotes the set of all possible infinite histories; i.e., $\mathcal{Z} = (S)^\infty$, with typical element $\zeta = (s_1, s_2, \ldots)$.

The basic object of this paper is a *model of learning and behavior*, which specifies how the players behave and what they believe as time passes. A model of learning and behavior consists formally of two pieces, *behavior rules* and *assessment rules* for each player. We take these in turn.

Behavior Rules

We denote by ϕ^i the *behavior rule* that player i uses in the infinitely repeated game. That is, $\phi^i = (\phi_1^i, \phi_2^i, \ldots)$, where ϕ_t^i: $\mathcal{Z}_t \to \Sigma^i$. The notation ϕ (for a profile $(\phi^1, \ldots, \phi^I)$ of behavior rules for the players), ϕ_t (for the profile of behavior rules at date t as a function of ζ_t), and $\phi_t(\zeta_t)$ are all used.

[4] We use male pronouns for players in general, and for players numbered 2, 4, 6, etc., and lettered j and $-i$. We use female pronouns for players numbered 1, 3, etc., and for players lettered i and k.

[5] Insofar as possible, we follow the convention that subscripts refer to time and superscripts to players. When we write $(\cdot)^t$, however, we mean the usual t-fold Cartesian product of the argument within the parentheses. (We try to avoid this as much as possible.)

Fix a profile of behavior rules ϕ. Given any $t \geq 1$ and $\zeta_t \in \mathcal{Z}_t$, we can use ϕ to construct a conditional probability distribution (conditional on ζ_t) for the rest of the path of play in the usual fashion: ζ_t and ϕ give a probability distribution on s_t (the actual play at date t) via $\phi_t(\zeta_t)$. This gives us conditional probabilities on $\mathcal{Z}_{t+1}$, and with transition probabilities for s_{t+1} given by $\phi_{t+1}(\zeta_{t+1})$, we can extend the conditional distribution to $\mathcal{Z}_{t+2}$, and so on. These then give a probability distribution over the space of complete histories $\mathcal{Z}$ by the Kolmogorov extension theorem. We write $\mathbf{P}(\cdot \mid \zeta_t)$ for this conditional probability, keeping in mind that this is for a fixed profile of behavior strategies.

One part of this construction must be emphasized. Given history ζ_t, the probability that i plays s^i at time t is $\phi_t^i(\zeta_t)(s^i)$. When we construct the conditional probability distribution on ζ_t, we must specify the joint probability that 1 plays s^1, 2 plays s^2, and so on. We insist that

$$\mathbf{P}(s_t = (s^1, \ldots, s^I) \mid \zeta_t) = \phi_t^1(\zeta_t)(s^1) \times \cdots \times \phi_t^I(\zeta_t)(s^I).$$

That is, players randomize (in their behaviors) independently.

Assessments

To model the behavior rules of players, we employ some ancillary formalisms. Specifically, we want to speak of what each player assesses concerning the behavior of her rivals, at each date t and contingent on each possible history ζ_t.

For the analysis in this paper, it suffices to specify (for each player i, time t, and partial history ζ_t) what i believes her rivals will do in the round about to be played. Formally, for each t, let μ_t^i denote a function with domain $\mathcal{Z}_t$ and range Σ^{-i}, representing i's *assessment* over the possible pure-strategy profiles that her rivals will choose at date t, as a function of ζ_t. Also, we use μ^i to denote a full system of assessments or *assessment rule* for i; i.e., μ^i is a sequence $(\mu_1^i, \mu_2^i, \ldots)$. Note well that $\sigma^{-i} \in \Sigma^{-i}$ encodes more than i's marginal assessments for her rivals' behavior; i is allowed to make an assessment concerning the joint behavior of her rivals that admits correlations in their play.[6] This may at first seem troubling

[6] For example, imagine a three-player game in which player 1 has a choice between pure strategies a and b and player 2 has a choice between a' and b'. Imagine that player 3 thinks that player 1 mixes between a and b either with probabilities $\frac{3}{4}$ and $\frac{1}{4}$ or with probabilities $\frac{1}{4}$ and $\frac{3}{4}$, with the same mixing probabilities used at each date, irrespective of the history of play. That is, player 3 believes either that player 1 is using the behavior rule $\hat{\phi}^1$ that is given by $\hat{\phi}_t^1(\zeta_t)(a) \equiv \frac{3}{4}$ or that 1 uses $\check{\phi}^1$ given by $\check{\phi}_t^1(\zeta_t)(a) \equiv \frac{1}{4}$. Player 3 entertains similar beliefs about the behavior rule used by player 2; we use $\hat{\phi}^2$ and $\check{\phi}^2$ to denote the two possibilities. Moreover, player 3 believes that her rivals randomize at each date independently; i.e., if player 1 is using $\hat{\phi}^1$ and player 2 is using $\hat{\phi}^2$, then player 3 assesses that the probability of

when contrasted with the independence assumption made in the earlier construction of the probability measures **P**. There is no conflict, however. The measure **P** reflects the *objective* probability measure that governs the evolution of play, as a function of the behavior rules by the players. We do not allow players to correlate their (mixed) strategies at any date; hence **P** is constructed with independence at each date. On the other hand, the $\mu_t^i(\zeta_t)$ represent a player's *subjective* assessment of what her rivals are about to do; unless and until i knows what behavior rules her rivals are using, correlation in her assessments can reflect her strategic uncertainty.[7]

3. Fictitious Play

The model of *fictitious play* (Brown, 1951; Robinson, 1951) can be viewed as a model of learning and behavior. First we give the details of fictitious play, and then we discuss its interpretation as a model of learning and behavior.

In fictitious play, there are two players; i.e., $I = 2$. In this setting, we interpret $-i$ as "not i"; i.e., $-i = 3 - i$ for $i = 1, 2$. Otherwise, the general setting is just as in Section 2. The behavior and assessment rules are built up as follows.

(A) For each player i, strategy $s^i \in S^i$, and history ζ_t, let $\kappa(\zeta_t)(s^i)$ be the number of times that i played s^i in the $t - 1$ observations that comprise ζ_t.[8]

(B) For each player i, there is an "initial weight" function η^i: $S^{-i} \to [0, \infty)$ such that $\Sigma_{s^{-i} \in S^{-i}} \eta^i(s^{-i}) > 0$.

(a, a') in any round is $(\frac{3}{4})(\frac{3}{4}) = \frac{9}{16}$. Imagine that player 3's prior belief is: 1 will use $\hat{\phi}^1$ and 2 will use $\hat{\phi}^2$ with probability .4; 1 will use $\hat{\phi}^1$ and 2 will use $\check{\phi}^2$ with probability .05; 1 will use $\check{\phi}^1$ and 2 will use $\hat{\phi}^2$ with probability .05; and 1 will use $\check{\phi}^1$ and 2 will use $\check{\phi}^2$ with probability .4. And, finally, imagine that 3 uses the sequence of observed play and Bayes' rule to update her beliefs about the joint behavior rule profile used by her rivals. With these data, we can integrate out to find 3's assessment about what 1 and 2 will do at any date t, given any history ζ_t. It is evident that although 3 believes that 1 and 2 are randomizing independently, her initial uncertainty about what behavior rules they are using and the correlation in her initial beliefs about their behavior rule profile imply that she will be making assessments about their play at each date that reflect correlation. *If* we condition on player 1 playing a at date 1, this makes it more likely that player 1 is using $\hat{\phi}^1$ than $\check{\phi}^1$, which makes it more likely that player 2 is using $\hat{\phi}^2$, which makes a' more likely. Note as well that even though player 3 believes (with probability one) that her rivals do not change their behavior from date to date as a function of what happens in the course of play, her assessments $\mu_t^3(\zeta_t)$ very much depend on ζ_t, since the history of play up to date t gives player 3 information about what behavior rule profile her rivals are in fact using.

[7] We are grateful to Bob Aumann for convincing us of how important this is.

[8] We do not bother to write κ_t^i, since the two arguments determine the length of the history and the player whose strategy is being counted.

(C) For each player i, date $t \geq 1$, history ζ_t, and strategy $s^{-i} \in S^{-i}$, we define $\eta_t^i(\zeta_t)(s^{-i}) = \eta^i(s^{-i}) + \kappa(\zeta_t)(s^{-i})$. (Note that $\eta_1^i(\zeta_1)(s^{-i}) = \eta^i(s^{-i})$ for all s^{-i}.) Then player i's assessment rule μ^i is given by normalizing the η^i; i.e.,

$$\mu_t^i(\zeta_t)(s^{-i}) = \frac{\eta_t^i(\zeta_t)(s^{-i})}{\Sigma_{s' \in S^{-i}} \eta_t^i(\zeta_t)(s')}.$$

(D) For player i, at each date t with history ζ_t, $\phi_t^i(\zeta_t)$ is a maximizer of

$$\sum_{s^{-i} \in S^{-i}} u^i(\sigma^i, s^{-i}) \mu_t^i(\zeta_t)(s^{-i}) \tag{3.1}$$

over all $\sigma^i \in \Sigma^i$.

In (D), we have not pinned down the definition of $\phi_t^i(\zeta_t)$ when there is more than one maximizer of (3.1). We do require that ϕ_t^i make a particular prescription in such cases (which, of course, can be a mixed strategy), but we do not say what it is. Formally, we would say that a model of learning and behavior is consistent with the model of fictitious play if there are initial weight functions η^i such that (C) holds as a definition of the assessment rules μ^i and (D) holds as a condition on the behavior rules ϕ^i.

We trust that most readers are familiar with the model of fictitious play, but it may help the uninitiated to give a simple example. Imagine two players who repeatedly play the strategic-form game in Fig. 1, with player 1 choosing a row and player 2 a column. We assume that the game begins with the players holding "beliefs"

$$\eta^1 = (1, 0, 4.32) \quad \text{and} \quad \eta^2 = (3, 5.7),$$

where we write these functions as vectors with the understanding that the first component of η^1 corresponds to column 1, the second component to column 2, and so on.

		Player 2		
		Column 1	Column 2	Column 3
Player 1	Row 1	5, 1	8, 4.7	2, 3
	Row 2	2, 3	2, 1	4, 2

FIG. 1. A strategic-form game.

TABLE 1

AN EXAMPLE OF FICTITIOUS PLAY

	"Beliefs" about rival	Expected payoffs	Choice of action
		Round number 1	
Player 1	(1, 0, 4.32)	(2.56, 3.62)	Row 2
Player 2	(3, 5.7)	(2.31, 2.28, 2.34)	Column 3
		Round number 2	
Player 1	(1, 0, 5.32)	(2.47, 3.68)	Row 2
Player 2	(3, 6.7)	(2.38, 2.14, 2.31)	Column 1
		Round number 3	
Player 1	(2, 0, 4.32)	(2.82, 3.45)	Row 2
Player 2	(3, 7.7)	(2.49, 1.95, 2.26)	Column 1

Refer to the first two lines of Table 1, which are labeled round number 1. The first line gives data for player 1, first her relative beliefs about what player 2 will do in the first round (i.e., the vector η^1) and next the expected payoffs she will accrue given those beliefs if she chooses row 1 and then row 2. Row 2 gives the higher payoff, and that is written down as her choice. Similarly, given 2's beliefs about what player 1 will do (the vector η^2), 2's best choice is column 3.

Move to the second two lines, labeled round number 2. Player 1's beliefs about the actions of player 2 are changed to reflect what happened in the first round. Since player 2 chose column 3 in the first round, the entry for column 3 in 1's beliefs is increased by 1. (That is, $\eta_2^1 = (1, 0, 5.32)$.) We recompute the expected payoffs to player 1 of playing either row, using these reassessed beliefs, and we see that row 2 continues to be player 1's best choice. But player 2 now finds that column 1 is optimal, when his beliefs are changed to reflect player 1's choice of row 2 in the first period. Hence in the second round, row 2 and column 1 are chosen. This gives beliefs for round number 3, and so on.

Fictitious play was not originally advanced as a model of how individuals would behave (and learn) when playing a game repeatedly; it was advanced instead as a method for computing Nash equilibria[9] or perhaps as a model of the preplay thought process of individual players. How well does it stand as a model of learning and behavior? The following two questions are raised immediately.

(1) *Is there any particular sense to how assessments are being formed?* It can be shown that the assessment rules μ^i are consistent with a Bayesian

[9] The connection will become clear in a bit.

model in which each player believes her rival is playing the same (unknown) mixed strategy in each round, independent of what came before, and where each player's prior assessment concerning this unknown behavior strategy has a Dirichlet distribution.

(2) *Is it sensible or realistic to assume that players would behave myopically, in the sense that, in each round, they choose a strategy that maximizes their immediate expected payoff, given their assessments?* Behavior that is myopic in this sense is discussed in Section 4, so for now we only note that *if* each player believes that his rival does not respond to the history of play—as posited in our answer to question (1) just preceding—then myopic behavior in this fashion is warranted.

Accepting the model as, at least, a very specific but interesting parameterization of learning and behavior, we can ask about its long-run implications. One possibility arises in the example of Fig. 1 and Table 1; if we follow this out until round 8, then in round 8 play reaches the profile row 1–column 2. Since this pair is a strict Nash equilibrium for the game, increasing the weight on row 1 in player 2's assessment and increasing the weight on column 2 in player 1's assessment only increases the optimality of column 2 and row 1, respectively. Thus play "gets stuck" at this pure-strategy Nash equilibrium. In general,

PROPOSITION 3.0. *In any history generated by fictitious play, if a strategy profile that is a strict Nash equilibrium is played, then all subsequent play will be that strategy profile.*

Or, speaking very loosely, strict Nash equilibria are absorbing for play according to the model of fictitious play. A related observation is the following.

PROPOSITION 3.1. *Suppose that in some history generated by fictitious play, a particular pure-strategy profile is played for all but a finite number of periods. Then that strategy profile must be a Nash equilibrium.*

We refrain from giving the proof here; this is an easy corollary to Proposition 4.1, which is proved later.

Thus we see one possibility; play might "stick" at some pure-strategy profile. If so, this profile must be a Nash equilibrium. (It goes almost without saying that judicious choice of the initial weight functions will allow fictitious play to stick at any strict equilibrium. Depending on how ties are broken, this is true as well of any equilibrium, even those in weakly dominated strategies.)

Proposition 3.1 implies that fictitious play cannot converge to a single pure-strategy profile in games that have no pure-strategy equilibria. Moreover, even in games that do have pure-strategy equilibria, fictitious play may fail to lock on to a single pure-strategy profile. For example, take the game in Fig. 1, and change the entry 4.7 in row 1–column 2 to a 4.

Note that row 1–column 2 is still a strict Nash equilibrium. Begin fictitious play with the same initial weight vector as before, and it turns out that column 2 will never be played. Instead, play "cycles" around the best response cycle row 1–column 1 to row 1–column 3 to row 2–column 3 to row 2–column 1, where "cycles" is put in quotes because the periods of the cycles increase through time. In the limit, however, the relative frequencies of play of the various strategies converge. That is, player 1 plays row 1 one-third of the time in the limit, and player 2 plays column 1 two-fifths of the time and column 3 three-fifths of the time. Hence the players' beliefs about how each other will be playing converge to the corresponding mixed strategies. It is straightforward to see that these mixed strategies constitute a mixed Nash equilibrium. More generally, we have

PROPOSITION 3.2. *Suppose that in some history generated by fictitious play, the empirical frequencies of pure-strategy choices converge to some (mixed) strategy profile. Then that strategy profile is a Nash equilibrium.*

The proof is omitted for now; this is a corollary of Proposition 4.2. Note that this proposition implies Proposition 3.1 as a special case.

It is natural to ask whether, in every game and for every set of initial conditions, convergence at least in the sense of Proposition 3.2 will take place under fictitious play. There are entirely trivial reasons why convergence may fail, connected with the way in which ties (among optimal strategy choices) are broken. However, if due care is taken in dealing with ties, then it is known that convergence in this sense is ensured for zero-sum games (Robinson, 1951) and two-by-two games (Miyasawa, 1961). However, for general games convergence is not ensured; the first (nontrivial) example is given in Shapley (1964).

4. EXTENSIONS OF FICTITIOUS PLAY

One problem with the model of fictitious play is its very rigid, ad hoc specification. Assessments are formed according to the empirical frequencies of past play (up to the initially given weight vectors), and actions are chosen to maximize precisely immediate expected payoffs. Neither part of this specification is essential to the results given above; we can obtain similar results for a broad class of models of learning and behavior. In this section, we present some results of this sort.

Myopic Behavior

DEFINITIONS. Given an assessment rule $\mu^i = (\mu_1^i, \mu_2^i, \ldots)$ for player i, we say that the behavior rule $\phi^i = (\phi_1^i, \phi_2^i, \ldots)$ for i is *myopic* relative

to μ^i if, for every t and ζ_t, $\phi_t^i(\zeta_t)$ maximizes i's immediate expected payoff, given assessment $\mu_t^i(\zeta_t)$. That is, $u^i(\phi_t^i(\zeta_t), \mu_t^i(\zeta_t)) = \max_{s^i \in S^i} u^i(s^i, \mu_t^i(\zeta_t))$.

The behavior rule ϕ^i is *asymptotically myopic* relative to μ^i if for some sequence of strictly positive numbers $\{\varepsilon_t\}$ with limit zero, for every t and ζ_t, $\phi_t^i(\zeta_t)$ comes within ε_t of maximizing i's immediate expected payoff, given assessment $\mu_t^i(\zeta_t)$. That is, $u^i(\phi_t^i(\zeta_t), \mu_t^i(\zeta_t)) + \varepsilon_t \geq \max_{s^i \in S^i} u^i(s^i, \mu_t^i(\zeta_t))$.

The behavior rule ϕ^i is *strongly asymptotically myopic* relative to μ^i if for some sequence of strictly positive numbers $\{\varepsilon_t\}$ with limit zero, for every t and ζ_t, every $\bar{s}^i$ in the support of $\phi_t^i(\zeta_t)$ comes within ε_t of maximizing i's immediate expected payoff, given assessment $\mu_t^i(\zeta_t)$. That is, $u^i(\bar{s}^i, \mu_t^i(\zeta_t)) + \varepsilon_t \geq \max_{s^i \in S^i} u^i(s^i, \mu_t^i(\zeta_t))$ for all $\bar{s}^i$ in the support of $\phi_t^i(\zeta_t)$.

Note that in asymptotically myopic behavior, the player can use slightly suboptimal pure strategies with large probability, or he can use grossly suboptimal pure strategies with small probability, or both, as long as the "average" suboptimality, averaged according to the probabilities with which the pure strategies are played, is small enough. In strong asymptotic myopia, grossly suboptimal pure strategies cannot be used at all.

We work throughout with models of learning and behavior for which behavior is at least asymptotically myopic with respect to the assessment rules. Even this less restrictive assumption has one feature that is potentially troublesome: It implicitly supposes that players do not try (asymptotically) to influence the future play of their opponents. To see this, consider the game in Fig. 2, and imagine that player 2 selects actions according to the model of fictitious play. In this game, row 2 is dominant for player 1, and so if player 1's behavior is asymptotically myopic for any assessment rule, she will play row 1 eventually. Since player 2 uses the assessment and behavior rules of fictitious play, he will eventually choose column 1; play converges to the pure-strategy equilibrium, row 2–column 1. But if player 1 does not behave asymptotically myopically and instead chooses row 1 each time, then player 2 would eventually choose column 2. If player 1 discounts her payoffs with a discount factor close to one, this gives her a higher overall payoff. The point is a simple one. As long as player 2 is playing according to the model of fictitious play, player 1 can

		Player 2	
		Column 1	Column 2
Player 1	Row 1	1, 0	3, 2
	Row 2	2, 1	4, 0

FIG. 2. A strategic-form game illustrating the possibility of Stackelberg leadership.

exploit this and manipulate 2's beliefs in order to receive her "Stackelberg leader" outcome (cf. Fudenberg and Levine, 1989).

In light of this example, our assumption of asymptotically myopic behavior requires some defense and explanation. We defend the assumption with stories that combine two justifications in varying proportions: First, even if a player's possible influence on an opponents' future play is large, the player may discount the future sufficiently that the effects are unimportant.

Second, even if the players are relatively patient, they may believe that their current action will have little, if any, effect on what will happen in the future. Suppose, for example, that player i believes that her rivals choose actions in each period according to some fixed but unknown (and possibly mixed) strategy profile, which is not influenced by the actions of other players. Moreover, because i learns her rivals' actual play at each date regardless of what i chooses to do, i's immediate choice of action will not affect what i learns, and thus (as long as i's behavior is subsequently myopic) it will not affect i's own subsequent actions.[10] Weakening this slightly, if i believes that her rivals will be playing a fixed strategy asymptotically, then asymptotically myopic behavior (for the same reasons) is warranted.

We are not very happy with either of these two justifications on its own. In order to permit learning to take place, play must be repeated "frequently," more frequently than would be suggested by a substantial discount rate, except for extraordinarily impatient players. And the story that players regard their rivals as playing fixed strategies repeatedly suffers from internal inconsistency; Why should a player imagine that his rivals are so different from himself? A belief that one's rivals will settle down to repeated play of a single strategy profile (justifying asymptotic myopia) is more palatable, especially when each player, in consequence, settles down to repeated play of a single strategy. But even in this more palatable story, each player is (effectively) assuming that his rivals settle down more quickly than the player does himself.

More convincing justifications of myopia can be given by enriching our story and combining the two justifications. Rather than thinking of a small group of players who interact repeatedly, we think of situations in which there are a large number of (potential) players who interact in small groups.

[10] Suppose instead that i's choice of action in round t affects the information she receives about the strategy choices of her rivals in that period. (This would be natural, for example, if we imagined that the stage game is an extensive-form game, and players only observe the outcome of each round of play.) Then i's choice of action today might affect *her own* subsequent actions; and she might choose to invest in information today by taking an action that is (myopically) suboptimal but that may generate useful information for guiding future choices.

Imagine that we have 5000 players 1, 5000 players 2, and so on, that repeated meetings between any small group of players are rare, and that whenever a player meets some group of rivals, he is unaware of how these rivals acted in the past. To be more precise, imagine that one of the following three stories holds.

Story 1. At each date t, one group of players is selected to play the game. (That is, one from each of the 5000 players of each type is selected to play.) They do so, and their actions are revealed to all the potential players. Those who play at date t are then returned to the pool of potential players, and another group is chosen at random for date $t + 1$.

Story 2. At each date t there is a random matching of all the players, so that each player is assigned to a group with whom the game is played. At the end of the period, it is reported to all how the entire population played. (That is, at the end of the period, it is announced that 20% of the 1's chose row 1, and so on.) The play of any particular player is never revealed.

Story 3. At each date t there is a random matching of the players, and each group plays the game. Each player recalls at date t what happened in the previous encounters in which he was involved, without knowing anything about the identity or experiences of his current rivals.

In each of these stories, myopic behavior seems "sensible," for reasons that mix to varying degrees the two basic justifications given above. In the first story, mainly the first justification is at work. Although the game is played relatively frequently, any single individual plays very infrequently, and at any reasonable discount rate, immediate payoff considerations will dominate any long-run considerations. In the second and third stories, it is more a matter of each player believing (now, with good reason) that his own immediate actions will have little impact on how his future rivals will behave. In story 2, this is because each player may believe that how he behaves will have little influence on the reported aggregate distribution; in story 3, this is because each player attaches low probability to the possibility that his current rivals will be future rivals any time soon, or even that future rivals will indirectly be affected by the player's own immediate play through an effect on the player's immediate rivals, who then (through some chain of individuals) affects future rivals. These stories make myopic behavior more plausible intuitively, although whether this plausible intuition has some firm, formal basis remains a question worth exploring.

Adaptive Assessments

Once we assume that behavior is (asymptotically) myopic, the next step is to specify the assessment rules μ^i used by the players and, in particular,

how these assessments are revised as the players observe the actions of others. In the models we consider, players believe that, at least asymptotically, the past choices of opponents are to some extent representative of future choices. A fairly weak property that captures this idea is suggested by Milgrom and Roberts (1991).[11]

DEFINITION. The assessment rule μ^i is *adaptive* if for every $\varepsilon > 0$ and for every t, there is some $T(\varepsilon, t)$ such that for all $t' > T(\varepsilon, t)$ and histories $\zeta_{t'}$, $\mu^i_{t'}(\zeta_{t'})$ puts probability no more than ε on the set of pure strategies by i's opponents that were not played at all between times t and t' (according to $\zeta_{t'}$).

In words, the definition says that i puts very little weight on strategies by her rivals that have not been played for a long (enough) time. The class of adaptive assessment rules is very broad, including, for example, assessments that take a weighted average of the history of past plays by one's opponent, as long as the weight put on any initial segment of history can be made small by making t sufficiently large. Four examples of adaptive assessment rules are: (a) assess that one's rivals will play in period t whatever was played in $t - 1$ (the assessments that go with Cournotian dynamics); (b) assess that one's rivals will play according to an exponentially weighted average of past plays; (c) assess that one's rivals are equally likely to play any action that has been played at least 1% of the time, with zero probability for all other actions; and (d) assess according to the scheme of fictitious play (where all previous observations are equally weighted).

While the class of adaptive assessment rules is broad, there are arguments that restricting attention to this class is too restrictive. See, for example, the discussion in Milgrom and Roberts (1991, p. 89ff) concerning sophisticated learning.

Convergence to a Pure-Strategy Profile

We are now in a position to generalize Proposition 3.1. Fix assessment rules μ^i and behavior rules ϕ^i for our two players. To state the result, we require a definition.

DEFINITION. The infinite history $\zeta = (s_1, s_2, \ldots)$ is said to be *compatible* with behavior rules ϕ if for each $t = 1, 2, \ldots$ and for $i = 1, \ldots, I$, the action s^i_t is in the support of $\phi^i_t(\zeta_t)$.

That is, ζ is something that could be observed with positive probability over all finite time horizons for players who behave according to the behavior rules that are given.

[11] Milgrom and Roberts define adaptive *behavior* as opposed to *assessments*, but it will be apparent that our formal definition is just a gloss on theirs.

PROPOSITION 4.1.[12] *Let ζ be an infinite history $(s_1, s_2, \ldots)$ such that for some $s_* \in S$ and for some T, $s_t = s_*$ for all $t \geq T$. If ζ is compatible with adaptive assessment rules μ^i and behavior rules ϕ^i that are strongly asymptotically myopic relative to the assessment rules, then s_* must be a Nash equilibrium of the stage game.*

Proof. Normalize the payoffs in the game so that the range of payoffs for each player is no greater than one. Suppose ζ_t is the partial history $(s_1, s_2, \ldots, s_t)$ of ζ. Maintain the hypotheses of the proposition; ζ is compatible with the ϕ^i, and ζ's components are eventually s_*. Suppose that s_* is not a Nash equilibrium. Then (without loss of generality) player 1 has a better response to s^{-1}_* and s^1_*. Let $\bar{s}^1$ be 1's better response, and set

$$\varepsilon = \frac{u^1(\bar{s}^1, s^{-1}_*) - u^1(s^1_*, s^{-1}_*)}{4} > 0.$$

Because the μ^i are adaptive and because history eventually settles on repeated play of s_*, we can find a T sufficiently large so that for all $t > T$, the probability assessment of player 1, $\mu^1_t(\zeta_t)$, puts probability at least $1 - \varepsilon$ on the play of s^{-1}_*. Thus the expected payoff to 1 for all $t > T$ from playing s^1_* (against 1's assessment of her rivals' strategy choices) is at least $4\varepsilon(1 - \varepsilon) - \varepsilon = 3\varepsilon - 4\varepsilon^2 > \varepsilon$ worse than 1's payoff from playing $\bar{s}^1$.[13] Thus s^1_* is more than ε suboptimal against s^{-1}_* for all $t > T$, which implies that 1's behavior rule is not strongly asymptotically myopic, a contradiction. ■

Note that the proposition assumes that behavior rules are strongly asymptotically myopic. If we deleted the modifier strongly, the result would be false as stated. Consider, for example, repeated play of the prisoners' dilemma, and behavior where, at period t, each player chooses to cooperate with probability $1/t$ and to defect with probability $(t - 1)/t$. Consider the infinite history where each player cooperates in each period. This history is compatible with the behavior rules. And (for any assessment rules) the behavior rules are asymptotically myopic, because they involve playing a suboptimal strategy with vanishing probability. But the strategy profile that is "repeated" in each period is not a Nash equilibrium. The problem of course is that compatibility only requires that each finite history

[12] Compare with Milgrom and Roberts (1991, Theorem 3[ii]).

[13] This is computed as the probability assessed that -1 plays s^{-1}_*, at least $1 - \varepsilon$, times $4\varepsilon = u^1(\bar{s}^1, s^{-1}_*) - u^1(s^1_*, s^{-1}_*)$, less the probability that -1 plays anything other than s^{-1}_*, which is no more than ε, times the maximum possible difference in payoffs playing $\bar{s}^1$ and s^1_*, which by the normalization is one.

has positive probability for the behavior rules. To obtain a result in the spirit of Proposition 4.1, but with asymptotic myopia instead of strong asymptotic myopia, we must either be more careful about how we make histories consistent with the given behavior rules or study not the actual history of play but the intended strategies of the players. We provide one result along these lines at the end of Section 6.

Convergence to Mixed Strategies in Empirical Frequencies for $I = 2$

Next we proceed to generalizations of fictitious play and convergence in the second sense of Section 3, where we look for convergence of the empirical frequencies of observations to some (possibly mixed) strategy profile.

For the remainder of this section, assume that the game has two players only; i.e., $I = 2$. We take up the case of more than two players in Section 5.

Let $\bar{\sigma}(\zeta_t)$: $S \to [0, \infty)$ give the vector of proportions of strategy profiles in S along the partial history ζ_t; i.e., $\bar{\sigma}(\zeta_t)(s)$ gives the number of times s was played in periods 1 through $t-1$, divided by $t-1$. We write $\bar{\sigma}^i(\zeta_t)$ for the marginal frequency distribution on S^i induced by $\bar{\sigma}(\zeta_t)$; i.e. $\bar{\sigma}^i(\zeta_t)(s^i) = \Sigma_{s^{-i} \in S^{-i}} \bar{\sigma}(\zeta_t)(s^i, s^{-i})$. Then, in the spirit of Proposition 3.2, we look for conditions on assessment and behavior rules that guarantee,

Suppose ζ is an infinite history $(s_1, s_2, \ldots)$ such that for some $\sigma_* \in \Sigma$,

$$\lim_{t \to \infty} \bar{\sigma}^i(\zeta_t) = \sigma_*^i,$$

for $i = 1, 2$. Then σ_* is a Nash equilibrium of the stage game.[14]

To get this result, it is insufficient that behavior be strongly asymptotically myopic with respect to adaptive assessment rules. Consider, for example, the game *matching pennies,* and suppose that at dates $t \geq 4$, the two players assess equal probabilities for any strategy by their rival that has occurred at least 10% of the time in the past; until date 4, they assess equal probabilities for the two strategies. As for behavior, players behave myopically optimally in all instances, with the following specification if the asssessment leaves the player indifferent: If t is divisible by 3, then play "heads"; otherwise play "tails." What happens is that the sequence of plays is tails, tails, heads, tails, tails, heads, ..., for both players, and each always assesses equal probabilities for his rival's two strategies.

[14] Please note carefully, this is not quite the same as asking that $\lim_t \bar{\sigma}(\zeta_t) = \sigma_*$. We ask only that the marginal frequencies converge and not the joint frequency distribution. There is a lot behind this observation, to which we return in the next section.

Empirical frequencies converge to $(\frac{1}{3}, \frac{2}{3})$ for each, which (of course) is not a Nash equilibrium of the stage game.[15]

The difficulty, it should be clear, comes from the fairly weak requirements of being an adaptive assessment rule. When only one strategy choice by rivals is eventually observed, adaptive assessment rules converge together with the (degenerate) empirical frequencies of observations. But when rivals use more than one (pure) strategy with nonvanishing frequency (or even with vanishing frequency that vanishes sufficiently slowly), adaptive decision rules can assign probability to that strategy that is unrelated to its limiting empirical frequency. To obtain the result that we seek, we must sharpen considerably the criterion imposed on assessment rules. The simplest and most direct criterion that works runs as follows.

DEFINITION. The assessment rule μ^i is *asymptotically empirical* if for every $\zeta \in \mathfrak{Z}$,

$$\lim_{t \to \infty} \|\mu_t^i(\zeta_t) - \bar{\sigma}^{-i}(\zeta_t)\| = 0,$$

where the ζ_t are subhistories of the fixed ζ.[16]

It is easy to see that any asymptotically empirical assessment rule is adaptive, that there are adaptive assessment rules that are not asymptotically empirical, and that the assessment rule in the model of fictitious play is asymptotically empirical.

Is it reasonable to insist that assessment rules are asymptotically empirical? This property is natural if one's picture of a rival's dynamic behavior is that the rival is playing some (unknown) strategy repeatedly, or even if one supposes that one's rival will converge to repeated, independent play of some (unknown) strategy. But if you think that your rival's strategy may shift repeatedly through time, then some assessment scheme that puts nonvanishing weight on more recent observations (which is precluded if assessments are asymptotically empirical) would be more reasonable.

PROPOSITION 4.2. *Let ζ be an infinite history $(s_1, s_2, \ldots)$ such that for some $\sigma_* \in \Sigma$,*

$$\lim_{t \to \infty} \bar{\sigma}^i(\zeta_t) = \sigma_*^i,$$

[15] At the cost of complicating the description of the assessment and behavior rules, we can modify this example so that the two players eventually play the mixed strategies $(\frac{1}{3}, \frac{2}{3})$ at all dates; nonconvergence of their intended strategies is not the issue.

[16] Whenever we are dealing with finite dimensional vectors as here, $\|\cdot\|$ denotes the sup norm.

for $i = 1, 2$. If ζ is compatible with asymptotically empirical assessment rules μ^i and behavior rules ϕ^i that are strongly asymptotically myopic relative to the assessment rules, then σ_* is a Nash equilibrium of the stage game.

The proof resembles the proof of Proposition 4.1 with the following amendments. First, since the assessment rules μ^i are asymptotically empirical, the assessments of player i (given by μ^i) at the partial histories ζ_t converge to the mixed strategy σ_*^{-i}. If σ_*^i is not a best response to σ_*^{-i}, then there is some pure strategy $\tilde{s}^i$ for player i that is strictly better against σ_*^{-i} than is some $\hat{s}^i$ in the support of σ_*^i. By a standard argument, for some $\varepsilon > 0$ and sufficiently large T, $\hat{s}^i$ will be worse against $\mu^i(\zeta_t)$ than is $\tilde{s}^i$ by more than ε, for all $t > T$. Thus $\hat{s}^i$ will not be played eventually (by asymptotic myopia). But this would contradict $\hat{s}^i$ being in the support of the limiting frequencies of i's strategy choices.[17]

5. Objections to Convergence of the Empirical Distributions as a Convergence Criterion

Notwithstanding the results of the previous section, the convergence criterion employed fails to capture what we want for a model of "learning to play mixed strategies." Our objections begin with the obvious observation that in examples such as fictitious play, players are (almost) never playing mixed strategies. They are instead jumping from one pure strategy to another, (typically) in cycles of ever-increasing length, so behavior is not converging.

The rebuttal to this is that while behavior is not converging, beliefs are. Mixed equilibria are sometimes interpreted as equilibria in beliefs; each side believes the other to be acting in a manner that makes the first (nearly) indifferent among several actions. Under this interpretation, the convergence criterion used in the previous section is fairly natural *if* players ignore the cycles in their own and their opponent's play.

However, these cycles can lead to phenomena so striking that we do not believe they would be ignored. Consider, for example, a symmetric *battle of the sexes* as depicted in Fig. 3. Imagine play of this game using the precise method of fictitious play, where each player begins with the

[17] The conclusion of Proposition 4.2 does not require the full power of asymptotically empirical assessment rules; e.g., the conclusion still holds for assessment rules that do not approach the empirical frequencies along histories where the empirical frequencies do not converge. More concretely, suppose that μ^i reports the empirical frequencies of $-i$'s choices over the most recent $\alpha\%$ of history, for α strictly between 0 and 100. This assessment rule is not asymptotically empirical per our formal definition, but it is empirical enough so that Proposition 4.2 holds.

		Player 2: Column 1	Player 2: Column 2
Player 1	Row 1	0, 0	2, 1
	Row 2	1, 2	0, 0

FIG. 3. The battle of the sexes.

relative beliefs vector $(1, \sqrt{2})$. The symmetry of the situation implies that if player 1 chooses top in the first round, 2 will choose left, and vice versa. In fact, with the numbers we are given, top-left will be played, and each player's relative beliefs going into the second round will be given by $(2, \sqrt{2})$. The symmetry again implies play of either top-left or bottom-right, and so on, inductively.[18] From general results about fictitious play, we know that empirical frequencies will converge to the Nash equilibrium probabilities $(\frac{2}{3}, \frac{1}{3})$. But this will be realized with perfect correlation in the two players' choices: Top-left will be played two-thirds of the time, and bottom-right one-third. Players will get zero round after round, there will be perfect correlation in their actions, and yet, according to the theory, they will persist in believing that they are "converging" to the mixed Nash equilibrium.

Moreover, this example shows that Proposition 4.2 will run into difficulties for the case $I > 2$. Imagine a three-player game, in which the actions of player 3, from the perspective of players 1 and 2, are irrelevant. Players 1 and 2 simply play the battle of the sexes against each other in each round. Player 3, to choose an optimal strategy, must forecast the joint actions of her rivals; for the sake of definiteness, suppose her optimal action is *tic* if she believes that they will play to a main-diagonal cell with probability $\frac{2}{3}$ or more, and her optimal action is *tac* otherwise.

What should 3 conclude, asymptotically, if 1 and 2 act in accordance with the particular model of fictitious play given above? Should she conclude that their actions are perfectly correlated, always playing along the main diagonal, hence *tic* is optimal? Or should she conclude that 1 will play top two-thirds of the time, 2 will play left two-thirds of the time, and hence top-left has asymptotic probability four-ninths, bottom-right has probability one-ninth, and thus *tac* is optimal?

There are (at least) two different ways we could proceed, depending on how we extend the definition of asymptotically empirical assessments. One possible definition is precisely the definition given before, interpreting

[18] Because the payoffs are rational and the relative weights have an irrational ratio, there will never be a tie; each player will have a unique best response at all times.

$\overline{\sigma}^{-i}(\zeta_t)$ as the marginal frequency distribution along ζ_t of profiles from S^{-i}. Under this definition, i's assessment (asymptotically) reflects any correlations that are observed empirically in the play of her rivals. The example shows how this definition permits convergence (under fictitious play) to non-Nash (correlated) assessments, so that Proposition 4.2 fails.[19] An alternative definition supposes that players asymptotically assess independent play by their rivals, regardless of the empirical frequencies.[20] Then we obtain Proposition 4.2 for $I > 2$. However, this seems to us to be somewhat unnatural; if there is correlation asymptotically, we feel that it is unnatural to assume that players ignore it.

Moreover, if it is unnatural for player 3 to ignore correlation in the choices of players 1 and 2, then isn't it equally unnatural for player 1 to ignore correlation in her choice of strategy and that of player 2? If so, then the example indicates that even for two-player games, asymptotic empiricism as formulated may be dubious.

All these objections (past our first and most basic objection) are grounded in the battle-of-the-sexes example; if that example is nongeneric, perhaps these objections have less force. In fact, it can be shown that the example is nongeneric for 2×2 games: In a 2×2 game, for generically chosen payoffs, the actions of the two players (under the model of fictitious play) will be asymptotically uncorrelated. However, we conjecture that robust examples of asymptotic correlation can be found in larger games. The basis for this conjecture is the game rock–scissors–paper. Fictitious play in this game must converge to the unique Nash equilibrium $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, since the game is zero sum. And, for most initial weight vectors, this happens while (asymptotically) avoiding the three cells along the main diagonal. (Each of the other cells has asymptotic frequency $\frac{1}{6}$.) We conjecture that these properties hold for a neighborhood of games around rock–scissors–paper, although we are unable to prove either convergence to the Nash equilibrium frequencies (since most games in a neighborhood will be nonzero sum) nor are we sure of the asymptotic frequencies of the cells. Robust examples can be created easily, though, if we move from the strictures of exact fictitious play.

Because we cannot verify our conjecture, we do not leave our secondary objections based on asymptotic correlation in empirical frequencies neat and tidy. Nonetheless, in our view the first objection—that this mode of convergence does not correspond to learning to play mixed strategies—suffices to motivate research into stronger modes of convergence. With this motivation, then, we proceed.

[19] One can repair the proposition in this case by restricting to histories ζ where the empirical joint frequencies are the products (in the limit) of the empirical marginal frequencies. But this repair seems a bit cheesy.

[20] One way to formalize this is to define, for each t, ζ^t, and $s \in S$, $\overline{\eta}(\zeta_t)(s) = \Pi_{i \in I} \overline{\sigma}^i(\zeta_t)(s^i)$. That is, $\overline{\eta}(\zeta_t)$ gives the "frequency distribution" obtained by using the marginal frequencies $\overline{\sigma}^i$ and forcing independence. Let $\overline{\eta}^{-i}(\zeta_t)$ give the S^{-i} marginal distribution of $\overline{\eta}(\zeta_t)$. Then asymptotic empiricism in this second sense is the condition $\lim_t \|\mu^i(\zeta_t) - \overline{\eta}^{-i}(\zeta_t)\| = 0$ along every history ζ.

6. Convergence of Behavior Strategies

Rather than look for convergence of empirical frequencies (and hence assessments about the actions of others), we look for convergence of the behavior strategies employed by players. That is, we study convergence (in t) of $\phi_t^i(\zeta_t)$ to some $\sigma_*^i \in \Sigma^i$, for each player i.

Because we wish to consider games with $I > 2$, we must first specify how we will adapt the definition of asymptotically empirical assessments. We proceed in the easiest fashion, by using the definition precisely as it was given earlier, but interpreting $-i$ as the set of i's rivals. That is, i's assessments asymptotically exhibit any correlation that is observed empirically in the choices of her rivals.

Two problems surface immediately. First, if $\phi_t^i(\zeta_t)$ is meant to converge to a mixed strategy, then player i must be willing to play one of several pure strategies. In the model of fictitious play, we insisted that players choose only myopic best replies, computed on the basis of their assessments of the actions of their rivals. How likely is it that a player, based on some history of play, would assess for his rival precisely the mixed strategy that makes him (the first player) indifferent? If this is unlikely, how can we ever have players willfully randomizing?

It is here that the asymptotic parts of asymptotic myopia and asymptotic empiricism come into play. We do not insist in general that players play only myopic best responses; they can play slightly suboptimal responses, as long as the degree of suboptimality vanishes as time (t) passes. So if assessments converge to the equilibrium mixed strategies quickly enough relative to the rate at which the allowable suboptimality vanishes, we can sustain mixed strategies even if assessments do not match precisely the equilibrium mixtures. At the same time, we do not insist that players' assessments are precisely empirical; if the empirical frequencies of play converge to some equilibrium mixed strategy, then players' beliefs can sit at precisely that limit mixed strategy, justifying the play of mixed strategies even if behavior is precisely myopic.

This means that the divergence from precisely myopic behavior and precisely empirical beliefs that we allow carry a lot of power in our story, at least insofar as convergence to mixed strategies is concerned. As we shall see, we don't require both at once. That is, our results obtain with asymptotic myopia and beliefs that are precisely empirical, or with behavior that is precisely myopic and beliefs that are asymptotically empirical.

But we need one or the other, if we are to hope for convergence of behavior to mixed-strategy profiles.

The Formal "Nonconvergence" Criterion: Unstable Strategy Profiles

The second problem that is raised is that statements of convergence in terms of behavior strategies must be probabilistic statements. To see what is at issue here, imagine playing the matching pennies game repeatedly. Suppose that along some history ζ, the empirical frequencies of the two rows and the two columns approach $(\frac{1}{2}, \frac{1}{2})$, but the behavior rules converge to the mixed strategies $(\frac{1}{3}, \frac{2}{3})$. This, you may object, is very unlikely. How could players' behavior strategies be converging to $(\frac{1}{3}, \frac{2}{3})$ and at the same time empirical frequencies are approaching $(\frac{1}{2}, \frac{1}{2})$? Unlikely is just the right word. There is nothing that prevents this—any history ζ is compatible with behavior rules that have players mixing strictly in each round—but by the strong law of large numbers, this history belongs to an event of probability zero. If behavior strategies are converging to $(\frac{1}{3}, \frac{2}{3})$, then the strong law of large numbers says that empirical frequencies will converge to $(\frac{1}{3}, \frac{2}{3})$ with probability one. Given asymptotically empirical assessment rules, this would rule out players continuing to play anything close to the $(\frac{1}{3}, \frac{2}{3})$ strategies.

Accordingly, when giving results in the spirit of Propositions 4.1 and 4.2, we give results of the following form: If σ_* is not a Nash equilibrium, then for every initial condition there is probability zero that behavior will remain forever in a small-enough neighborhood of σ_*. The formalities run as follows.

Fix a set of behavior rules ϕ^i (which will be accompanied by assessment rules μ^i, although for the time being only the behavior rules are needed). Recall that $\mathbf{P}(\cdot|\zeta_t)$ represents the objective conditional probability distribution on the space $\mathfrak{X}$ created by starting at ζ_t and using the behavior rules thereafter.

DEFINITION. A strategy profile $\sigma_* \in \Sigma$ is *unstable* if there exists some $\varepsilon > 0$ such that $\mathbf{P}(\|\phi_{t'}(\zeta_{t'}) - \sigma_*\| < \varepsilon \text{ for all } t' \geq t \mid \zeta_t) = 0$ for all t and ζ_t.

Note that ε here is independent of the starting conditions ζ_t.

PROPOSITION 6.1. *Fix behavior rules ϕ^i that are asymptotically myopic relative to some asymptotically empirical assessment rules μ^i. Then every strategy profile σ_* that is not a Nash equilibrium is unstable.*

Note that in this proposition, behavior rules are required to be (only) asymptotically myopic relative to the assessment rules. Compare with Propositions 4.1 and 4.2, in which strong asymptotic myopia was assumed. We return to this point at the end of this section.

Although the details of the proof of this proposition are tedious, the idea is fairly simple. *If* behavior lies forever in a small neighborhood of the strategy profile σ_*, then empirical frequencies will eventually lie in a small neighborhood too, and thus so will the players' assessments. If the strategy profile is not a Nash equilibrium, then (eventually) some player will want to move far away from the strategy profile, a contradiction to the supposition.

The key technical result is an application of the strong law of large numbers, which we state in the form of a lemma. The proof of this lemma is given in the Appendix.

LEMMA 6.2. *Let $\{x_t; t = 1, 2, \ldots\}$ be a sequence of random variables on a probability space with range some finite set A. Fix a probability distribution π on A and an $\varepsilon > 0$, and let Λ be the (measurable) subset of the probability space consisting of all sample points such that for $t = 1, 2, \ldots$, the distribution of each x_t conditional on $\{x_1, \ldots, x_{t-1}\}$, denoted $\pi_t(\cdot|x_1, \ldots, x_{t-1})$, satisfies*

$$\max_{a \in A} |\pi_t(a \mid x_1, \ldots, x_{t-1}) - \pi(a)| < \varepsilon.$$

Let $\tau_t(a)$ be the random variable $\Sigma_{t'=1}^{t} 1_a(x_{t'})$; that is, $\tau_t(a)$ is the number of times that $x_{t'} = a$ for $t' = 1, \ldots, t$. Then

$$\limsup_{t \to \infty} \frac{\tau_t(a)}{t} \leq \pi(a) + \varepsilon \qquad \text{and} \qquad \liminf_{t \to \infty} \frac{\tau_t(a)}{t} \geq \pi(a) - \varepsilon$$

for all $a \in A$, almost surely conditional on Λ.[21]

Proof of Proposition 6.1. Suppose that σ_* is a strategy profile that is not a Nash equilibrium. Then there is some player i and a pure strategy $\tilde{s}^i$ such that $\tilde{s}^i$ is strictly better against σ_*^{-i} than is σ_*^i. Since $u^i(\sigma^i, \sigma^{-i})$ is continuous in both arguments, we can find an $\varepsilon > 0$ so that for all σ^{-i} that are within $I\varepsilon$ of σ_*^{-i} and σ^i that are within ε of σ_*^i, both in the sup norm (on Σ^{-i} and Σ^i, respectively),

$$u^i(\sigma^i, \sigma^{-i}) + \varepsilon < u^i(\tilde{s}^i, \sigma^{-i}).$$

(Interpret σ^{-i} here as any element of Σ^{-i}; i.e., σ^{-i} need not be based on independent play by i's rivals. But σ_*^{-i} is composed of independent choices by i's rivals according to the components of the strategy profile σ_*.)

We claim that for this ε and for all asymptotically empirical behavior

[21] If Λ has zero prior probability, the lemma is taken to be vacuous.

rules ϕ and assessment rules μ that are asymptotically myopic relative to these assessment rules,

$$\mathbf{P}(\|\phi_{t'}(\zeta_{t'}) - \sigma_*\| < \varepsilon \text{ for all } t' \geq t \mid \zeta_t) = 0.$$

To see why, suppose to the contrary that for some ζ_t, this probability is strictly positive. We proceed to derive a contradiction.

From the lemma, we know that on the set Λ of positive probability (conditional on ζ_t), where $\|\phi_{t'}(\zeta_{t'}) - \sigma_*\| < \varepsilon$ for all $t' \geq t$, the limits inferior and superior of the empirical frequency distribution $\overline{\sigma}^{-i}(\zeta_{t'})$ almost surely lie within $(I - 1)\varepsilon$ of σ_*^{-i}. (If $|\phi^j_{t'}(\zeta_{t'})(s^j) - \sigma^j_*(s^j)| < \varepsilon$ for all s^j, then for any $\hat{s}^{-i} = (\hat{s}^j)_{j \neq i}$, $|\Pi_{j \neq i}\, \phi^j_{t'}(\zeta_{t'})(\hat{s}^j) - \Pi_{j \neq i}\, \sigma^j_*(\hat{s}^j)| < (I - 1)\varepsilon$.) Since assessments are asymptotically empirical, along every infinite history in Λ there is a T such that for all $t' > T$, the assessments of player i lie within $I\varepsilon$ of σ_*^{-i}. But then asymptotic myopia implies that along every $\zeta \in \Lambda$, $\phi^i_{t'}(\zeta_t)$ will eventually be more than ε away from σ^i_*, which contradicts the definition of Λ. ■

Two remarks about the proof are in order.

(1) Note that the neighborhood of σ_* that is used in the proof is independent of the behavior and assessment rules that are assumed to be given; the value of ε depends only on the strategy profile σ_* and the extent to which it is not a Nash equilibrium.

(2) The full strength of asymptotic empiricism is not required for this proof. What is essential is that *if behavior lies forever in some small neighborhood of a strategy, then assessments come to lie in another small neighborhood of that strategy*. We used the strong law of large numbers to show that the empirical frequencies would lie in a small neighborhood, and then we were able to enlist the asymptotically empirical character of assessment rules. But suppose the assessment rule took the following form: $\mu^i_t(\zeta_t)$ is asymptotically equal to the empirical frequency of observed strategy choices by one's rivals over the most recent $\sqrt{t}$ periods. Since the length of this segment of history grows without bound, we can still enlist the strong law of large numbers to come to the desired conclusion. Or suppose $\mu^i_t(\zeta_t)$ is a weighted average of past observations, with the greatest weight on the most recent observation. As long as that greatest weight falls off to zero fast enough as t goes to infinity, we can enlist a variation of the strong law of large numbers and obtain the desired result. (Of course, this rules out exponential moving averages, where the weight on the most recent observation does not vanish at all.)

We stick to asymptotic empiricism for the remainder of this paper, since it is expositionally the easiest thing to deal with. But you should note that it is a bit more restrictive than we actually need.

Locally Stable Strategy Profiles

Proposition 6.1 shows that every strategy profile that is not a Nash equilibrium is unstable. We know by example that there are some strategy profiles that are not unstable. It is natural to wonder whether any Nash equilibria are unstable. The answer is no; no Nash equilibrium profile is unstable.

To avoid double negatives, we make the following definition.

DEFINITION. A strategy profile σ_* is *locally stable* if there exists some asymptotically empirical assessment rules and behavior rules that are asymptotically myopic with respect to the assessment rules such that for every $\varepsilon > 0$, we can find some t and $\zeta_t \in \mathfrak{Z}_t$ such that

$$\mathbf{P}(\lim_{t' \to \infty} \phi_{t'}(\zeta_{t'}) = \sigma_* \mid \zeta_t) > 1 - \varepsilon.$$

We do not insist that behavior converges to the target strategy with probability one, but only that the probability can be made arbitrarily close to one (for a fixed model of behavior and assessment rules) for *some* choice of initial conditions. We couldn't have a probability one statement as long as the target strategy is not pure (except for degenerate cases), since as long as players use mixed strategies, there is positive probability of a very long run of "bad luck," which would lead players away from the target strategy.

PROPOSITION 6.3. *Every Nash equilibrium profile σ_* is locally stable.*

The proof is left for the Appendix, but some remarks are in order here. In the definition of local stability, we allow assessments to be asymptotically empirical and behavior to be asymptotically myopic. In fact, we can obtain local stability of any Nash equilibrium with either (a) a model with asymptotically empirical assessments and precisely myopic behavior or (b) a model with precisely empirical assessments and asymptotically myopic behavior. We provide details of the first sort of construction in the Appendix, but the idea behind each is easily given.

For the first construction, we suppose that players maintain precisely the equilibrium beliefs unless and until sufficient evidence against this accumulates. Of course, as long as players maintain equilibrium beliefs, the equilibrium strategies are among their myopic best responses, and we suppose that this is how they play. The work then is in showing that if players play according to the equilibrium strategies, we can rig things so that there is arbitrarily small probability that "sufficient evidence" against these assessments will arise.

For the second sort of construction, we would assume that players maintain precisely empirical beliefs, but that they stick to playing the

targeted equilibrium strategies unless and until the "cost" of doing so becomes too large. The work here is in showing that we can rig things so that there is arbitrarily small probability that they will ever have to abandon the equilibrium strategies.

In both constructions, the trick is to note that as long as play is according to equilibrium strategies, by the strong law of large numbers, empirical frequencies will converge to the equilibrium frequencies. Thus in the first case there is asymptotically no need to abandon the assessments that opponents use the equilibrium strategies, and in the second case, the cost of keeping to the equilibrium strategies (computed vis à vis the empirical frequencies) vanishes asymptotically.

In both constructions, players use precisely the equilibrium strategy for no positive reason at all. This suffers from the usual problem of implausibility of a mixed Nash equilibrium; the precise mixing probabilities have nothing to do with one's own payoffs, but are chosen to make one's rivals indifferent. Thus, although the constructions show that convergence to a mixed-strategy equilibrium is possible, neither one convinces us that it would in fact happen, except perhaps for players who have been trained in game theory and therefore know how they are "expected" to act. Put simply, mixed-strategy equilibria as a positive prediction are hard (for us) to defend in the stark environment of this chapter. Instead, we prefer to follow Harsanyi (1973) in interpreting mixed-strategy equilibria as a shorthand description of pure-strategy equilibria in games where parameters of the game (such as the players' payoffs) are subject to small random perturbations, which are private information. Sections 7 and 8 consider learning in this context and argue that mixed-strategy outcomes are indeed plausible when interpreted in this way.

Asymptotically Myopic Behavior and Compatible Histories

Before moving to this development, we have one piece of pending business to take care of.

In both Propositions 4.1 and 4.2, we assumed that the players' behavior rules were strongly asymptotically myopic relative to their assessment rules, whereas in Proposition 6.1, we assumed that behavior was (only) asymptotically myopic. We earlier indicated why the results in Section 4 would not work with asymptotically myopic behavior; viz., compatibility of a history and a profile of behavior rules is (too) weak. To obtain results in the spirit of Section 4 without strong asymptotic myopia, we must work with the sort of probabilistic convergence criteria used in this section.

Because we do not assume (in Propositions 4.1 and 4.2) that strategies converge, to avoid problems of correlation we must restrict attention to the case of two players. Also, because strategies are not assumed to converge, we cannot use the definitions of unstable and locally stable strategy profiles given above. Instead, for fixed behavior and assessment rules, we make the following definition.

DEFINITION. A strategy profile $\sigma_* \in \Sigma$ is *unstable in empirical frequencies* if there exists some $\varepsilon > 0$ such that for all t and ζ_t, $\boldsymbol{P}(\|\overline{\sigma}^i(\zeta_{t'}) - \sigma^i_*\| > \varepsilon \text{ for all } t' \geq t \text{ and } i = 1, 2 \mid \zeta_t) = 0$.

Note that this captures some of the spirit of Proposition 4.2, in that this notion of stability asks for empirical frequencies to remain close to a target profile σ_*. At the same time, it is a probabilistic statement about the likelihood of this event.

PROPOSITION 6.4. *In two-player games, for asymptotically empirical assessment rules μ^i and behavior rules ϕ^i that are asymptotically myopic with respect to the assessment rules, every strategy profile σ_* that is not a Nash equilibrium is unstable in empirical frequencies.*

We omit the proof. The idea is that if empirical (marginal) frequencies lie close to σ_*, then so must beliefs. But if beliefs are close to σ_*, and σ_* is not a Nash equilibrium, then for some i and some pure strategy s^i in the support of σ^i_*, s^i will be used with vanishing probability. And then (by the strong law of large numbers) the frequency of s^i must fall to zero, which contradicts the hypothesis that empirical frequencies stay close to σ^i_* (which puts positive weight on s^i).

7. LEARNING IN GAMES WITH RANDOMLY PERTURBED PAYOFFS

Although Proposition 6.3 shows that all Nash equilibria are locally stable, including equilibria in mixed strategies, we have suggested that convergence of intended behavior to a mixed-strategy profile in the standard model seems implausible, as it requires that players use just the right mixed strategy whenever they are indifferent, and it is not apparent why they should or would choose to do so. Of course, this apparent drawback of mixed-strategy equilibria is not special to our learning-theoretic approach, but arises whenever mixed strategies are considered.

In response to this problem, Harsanyi (1973) proposed that the mixed-strategy equilibria of a game could be interpreted as pure-strategy equilibria of a related game of incomplete information, in which each player's payoff is randomly perturbed by a stochastic shock, which is private information. If the distribution of payoff perturbations is absolutely continuous with respect to Lebesgue measure, then the strategy of each player would be (essentially) pure, because (given the strategy of others) each player would have a strict best response for almost all of his payoffs. But from the perspective of his opponents, the actions of the player would be random, because the opponents do not know the value of the player's

payoff perturbation. For example, a mixed strategy placing probability $\frac{1}{3}$ on one pure strategy and $\frac{2}{3}$ on another corresponds to a situation in which the payoff shocks and opponents' strategy are such that the player has a strict preference for the first strategy when his payoff perturbation comes from a set with probability $\frac{1}{3}$ and a strict preference for the second when his payoff perturbation comes from the complementary set. Harysanyi showed that for generic strategic-form payoffs, every mixed-strategy equilibrium can be "purified" by any small, sufficiently well-behaved payoff perturbation.

In the spirit of Harsanyi's work, this section extends our assumptions on behavior and notions of convergence to games in which the players' payoffs are subject to an i.i.d. sequence of payoff perturbations. As we show, this allows us to construct a more satisfactory model of learning to play a mixed-strategy equilibrium. The concluding section then examines the question of global convergence to a mixed equilibrium in 2×2 games.

The Model

Consider I players $i = 1, \ldots, I$ playing a strategic-form game at times $t = 1, 2, \ldots$, where the action spaces A^i are the same in each period, but the payoffs are subject to random and privately observed shocks. Specifically, the payoff to player i from the action profile $a = (a^i, a^{-i})$ in period t is $u_t^i(a) = v^i(a) + e_t^i(a^i)$. This is the *augmented* or *perturbed* version of the *underlying game*, which is the game where the payoff functions are simply the v^i. We call $e_t^i = (e_t^i(a^i))_{a^i \in A^i}$ the *date-t perturbation* of player i's payoffs. We assume that for each i the $\{e_t^i;\ t = 1, 2, \ldots\}$ are independent and identically distributed and that the perturbations of different players are independent. We denote the probability distribution of each e_t^i by ρ^i, and we denote its support, which we suppose is compact, by $E^i \subset R^{A^i}$.

In the stage game for period t, each player i observes the shock e_t^i to her own payoffs, but does not observe the shocks to her opponents' payoff functions. Hence in the stage game, a pure strategy for player i is a map from E^i to A^i. (We do not need to consider mixed strategies in the augmented game.)

To model learning in the repeated game, we suppose that at date t, player i knows the sequence of (pure) action profiles that have occurred in the past, the shocks to her own payoffs, and also her current payoff perturbation e_t^i; player i does not learn the past payoff shocks of her opponents. We adopt the following notation:

(a) We use $A = \Pi_{i=1}^I A^i$ to denote the *space of action profiles*, with typical element a. Profiles of actions by all players except i are denoted by $a^{-i} \in A^{-i}$. *Probability distributions over actions* by player i are denoted by $\alpha^i \in \mathcal{A}^i$, and probability distributions over A^{-i} (which can reflect correlations in the actions of i's opponents) are denoted by $\alpha^{-i} \in \mathcal{A}^{-i}$.

(b) *Histories of actions up to time t* are denoted by $\zeta_t \in \mathcal{Z}_t$; i.e., $\zeta_t = (a_1, \ldots, a_{t-1}) \in (A)^{t-1} = \mathcal{Z}_t$. *Complete histories of actions* are denoted by $\zeta \in \mathcal{Z}$.

(c) We write $\bar{\alpha}_t(\zeta_t)$ to denote the *empirical disribution of action profiles* up to time t along the history ζ_t, and we use $\bar{\alpha}_t^{-i}(\zeta_t)$ to denote the *empirical distribution of action profiles by i's opponents* up to time t along the history ζ_t.

(d) In addition to ζ_t, at time t player i knows her own history of payoff perturbations up to and including time t, or $(e_1^i, \ldots, e_t^i)$. We use $\xi_t^i \in \mathcal{X}_t^i$ to denote the vector of all this information; i.e., ξ_t^i looks like $(\zeta_t, (e_1^i, \ldots, e_t^i))$. Dropping the subscript t, ξ^i denotes a complete history for player i of action profiles by all players at all dates and all of i's payoff perturbations; dropping the superscript i, as in ξ_t or ξ, denotes (respectively) a time t history or a complete history of plays, and all the players' payoff perturbations.

(e) A *behavior rule* for player i is denoted by $\phi^i = (\phi_1^i, \phi_2^1, \ldots)$, where $\phi_1^i: \mathcal{X}_t^i \to A^i$.[22]

(f) An *assessment rule* for player i is denoted by $\mu^i = (\mu_1^i, \mu_2^i, \ldots)$, where $\mu_t^i: \mathcal{Z}_t \to \mathcal{A}^{-i}$.

All of this is a straightforward extension of our earlier model. In particular, i's assessment rule gives her predictions how her opponents will play at each date, based on history so far. In this regard, note that the domain of μ_t^i is $\mathcal{Z}_t$ and not $\mathcal{X}_t^i$. We assume that players other than i never observe i's payoff perturbations, so it seems sensible that i would not have assessments of the actions of her opponents depending on her payoff perturbations. Nonetheless, at the cost of some notational complexity, we could assume that i's assessments at date t depend on all of ξ_t^i, as long as asymptotic empiricism is properly defined.

Given behavior rules for all the players and the exogenous probability distributions on payoff perturbations, we can construct the induced conditional probability distribution, conditional on ξ_t, on the space $\mathcal{X}$. We use $\mathbf{P}(\cdot\,|\xi_t)$ to denote this probability distribution.

DEFINITION. For augmented games:

(a) The assessment rule μ^i *is asymptotically empirical* if

$$\lim_{t\to\infty} \|\mu_t^i(\zeta_t) - \bar{\alpha}_t^{-i}(\zeta_t)\| = 0$$

for every $\zeta \in \mathcal{Z}$.

(b) The behavior rule ϕ^i is *asymptotically myopic relative* to μ^i if for

[22] Measurability of the behavior rules is always assumed.

some sequence of nonnegative numbers $\{\varepsilon_t\}$ converging to zero, i's choice of action at every ξ_t^i is at most ε_t suboptimal against $\mu_t^i(\zeta_t)$.[23,24]

Nash Equilibria of the Augmented Game

Before examining learning in the context of this model, we review the structure of Nash equilibria in the augmented (stage) game.

A Nash equilibrium of the augmented game is, as usual, a strategy profile such that each player's chosen strategy $s^i(\cdot)\,(: E^i \to A^i)$ maximizes her expected payoff given the strategies of her opponents, or equivalently that for ρ^i-almost every e^i, $a^i = s^i(e^i)$ maximizes the expectation (over e^{-i}) of $v^i(a^i, s^{-i}(e^{-i})) + e^i(a^i)$.

Assumption 7.1. For each i, the distribution ρ^i is absolutely continuous with respect to Lebesgue measure on R^{A^i}.

This assumption simplifies the analysis, because it implies that for any distribution of the opponents' actions, i has a strict preference for one of her actions at ρ^i-almost every e^i.

LEMMA 7.2. *For every $\alpha^{-i} \in \mathcal{A}^{-i}$, the set of e^i for which*

$$\arg\max_{a^i \in A^i} \Big[\sum_{a^{-i} \in A^{-i}} [v(a^i, a^{-i}) + e^i(a^i)]\alpha^{-i}(a^{-i}) \Big]$$

is a singleton has measure one under ρ^i.

We omit the proof, which is based on the observation that the complement of this set lies in a finite union of lower-dimensional hyperplanes. Note that this is true whether α^{-i} reflects independent or correlated play by i's rivals.

For each e^i and α^{-i}, let $b^i(e^i, \alpha^{-i})$ specify some best response for i to α^{-i} when her payoff perturbation is e^i, and let $\beta^i(\alpha^{-i})$ be the distribution that b^i induces on player i's actions:

$$\beta^i(\alpha^{-i})(a^i) = \rho^i\{e^i \in E^i : b^i(e^i, \alpha^{-i}) = a^i\}.$$

(Lemma 7.2 shows that β^i is well defined, since for every α^{-i}, b^i is uniquely determined for ρ^i-almost every e^i.)

It is straightforward to prove the following technical result.

[23] Suboptimality here is measured *given* the period t payoff perturbation. We believe that nothing of interest changes with a weaker definition in which suboptimality is measured averaging over e_t^i, but the proofs are somewhat more involved.

[24] Throughout, we are loose in our notation, taking as understood things such as: in (a), ζ_t denotes the date-t subhistory of the fixed ζ; and in (b), ζ_t is the actions-profile part of ξ_t^i.

LEMMA 7.3. *The function β^i is continuous.*

For each i and strategy profile s^i, let $\pi^i(s^i)$ denote the distribution on A^i induced by s^i; i.e., $\pi^i(s^i)(a^i) = \rho^i\{e^i \in E^i : s^i(e^i) = a^i\}$. For s^{-i}, a profile of strategies for i's opponents (that is, $s^{-i} = (s^j)_{j \neq i}$, where s^j: $E^j \to A^j$), let $\pi^{-i}(s^{-i})$ denote the distribution on A^{-i} induced if i's rivals use the strategies in s^{-i}. Note that $\pi^{-i}(s^{-i}) \in \mathcal{A}^{-i}$ is a product measure (since the various e^j are independent of each other).

LEMMA 7.4. *(a) A strategy profile s is a Nash equilibrium of the stage game if and only if, for each player i and ρ^i-almost all e^i, $s^i(e^i) = b^i(e^i, \pi^{-i}(s^{-i}))$.*

(b) If $(\alpha^1, \ldots, \alpha^I) \in \mathcal{A}^1 \times \ldots \times \mathcal{A}^I$ satisfies $\beta^i(\alpha^{-i}) = \alpha^i$ for all i (where it is understood that α^{-i} is the product measure on $\mathcal{A}^{-i}$ whose margins are the various α^j for $j \neq i$), then every strategy profile s such that $s^i(e^i) = b^i(e^i, \alpha^{-i})$ for all i and ρ^i-almost every e^i is a Nash equilibrium.

This is largely a matter of marshalling definitions, hence the proof is omitted. This lemma shows that to analyze Nash equilibria, it suffices to work with the induced marginal distributions over actions, which motivates the following definition.

DEFINITION. The vector of marginal distributions $\alpha = (\alpha^1, \ldots, \alpha^I) \in \mathcal{A}^1 \times \ldots \times \mathcal{A}^I$ is a *Nash distribution* if $\beta^i(\alpha^{-i}) = \alpha^i$ for all i.

Local Stability

As one would expect, our results about the relationship between Nash equilibrium and local stability carry over to the context of augmented games. Since all that each player observes about the others, and all that matters for a player's decisions, are the actions chosen, we define stability and stability of behavior rules ϕ in terms of the induced distributions on actions $\pi^i(\phi_t^i(\xi_t))$.

DEFINITION. A profile $\alpha_* \in \mathcal{A}^1 \times \ldots \times \mathcal{A}^I$ is *unstable* if there exists some $\varepsilon > 0$ such that for all t and ξ_t,

$$\mathbf{P}(\|\pi^i(\phi_{t'}^i(\xi_{t'}^i)) - \alpha_*^i\| < \varepsilon \text{ for all } t' \geq t \text{ and } i \mid \xi_t) = 0.$$

PROPOSITION 7.5. *Fix asymptotically empirical assessment rules μ^i and behavior rules that are asymptotically myopic relative to the μ^i. Then if α_* is not a Nash distribution, α_* is unstable.*

This proposition does not follow immediately from our earlier results (in particular, from Proposition 6.1 applied to the augmented game), because in Proposition 6.1, it is assumed that each player sees the full strategy of his opponents in each round of play. In this setting, each player sees

only the actions chosen by his opponents; he sees neither the payoff perturbation that leads to that choice of action nor the actions that would have been chosen for other payoff perturbations.

The proof of Proposition 7.5 is left to the Appendix.

The next step in parallel with Section 6 is to note that every Nash equilibrium is locally stable for some asymptotically empirical assessments and asymptotically myopic behavior rules. This can be most easily shown by adapting the first construction in the proof of Proposition 6.3, in which players believe that the distribution over their rivals' actions corresponds to the equilibrium unless and until they receive sufficient evidence otherwise. With these beliefs, the arbitrary nature of the players' behavior rules is eliminated; rather than just happening to mix in the way the equilibrium prescribes, the players have a strict preference for the behavior they choose. Of course, the players' beliefs are still cooked to favor the equilibrium, so we do not yet have a really satisfactory explanation of how players might learn to play a mixed equilibrium. The final section provides such an explanation for a special class of two-player games.

8. Global Convergence in a Class of 2 × 2 Games

This section shows by example how learning in augmented games can lead to a mixed equilibrium even when the assumed behavior and assessment rules do not build in an arbitrary predilection for equilibrium play. To this end, we restrict attention to behavior and assessments that take precisely the form of fictitious play, as specified in (A) through (D) of Section 3. Moreover, we do more than show that convergence to a mixed equilibrium can occur even when the equilibrium is not artificially built in the behavior rules: In the games we consider, play converges to the (augmented version) of the mixed equilibrium with probability one, regardless of the initial beliefs of the players.

We do not aim for very general results here. Rather, we content ourselves with the special case of 2 × 2 games that (before being augmented) have a unique Nash equilibrium, which moreover is completely mixed. At the end of the section, we speculate about possible extensions that would provide a sufficient condition for local stability of mixed equilibria under fictitious play in other augmented games. We suspect, however, that convergence cannot be guaranteed for general augmented games; we conjecture that an augmented version of Shapley's example will provide the desired counterexample, but we have not verified this.

The remainder of this section discusses the following result, the proof of which is given in the Appendix.

Proposition 8.1. *Take any 2 × 2 game that has a unique, completely mixed Nash equilibrium, and consider any augmentation that satisfies Assumption 8.3 given below. If behavior rules and assessments are as in the model of fictitious play, the induced marginal distributions on actions converge, with probability one, to the unique Nash distributions of the augmented game.*

Previous results about the global convergence of behavior in learning processes have focused on games that are solvable by iterated strict dominance (Moulin, 1984; Milgrom and Roberts, 1990; Borgers and Janssen, 1991; Guesnerie, 1992). In contrast, the augmented games we consider are not dominance solvable.

Preliminaries

Fix a 2 × 2 augmented game with expected payoffs (v^1, v^2) and payoff perturbation vectors e^1 and e^2. Write the action sets for each player $A = \{1, 2\}$, so that, for example, $v^2(1, 2)$ is 2's payoff if he chooses column 2 and player 1 chooses row 1. (Player 1's choice of row is listed first.) Assume that the game (v^1, v^2) has a unique Nash equilibrium that is completely mixed.

Because (v^1, v^2) has a unique Nash equilibrium that is completely mixed, (v^1, v^2) has a strict best-response cycle. Rearrange rows, if necessary, so that this best-response cycle is counterclockwise. That is, $v^1(1, 1) < v^1(2, 1)$, $v^2(2, 1) < v^2(2, 2)$, $v^1(2, 2) < v^1(1, 2)$, and $v^2(1, 2) < v^2(1, 1)$.

Let $F^1(z)$ be the probability that, on any given date, $(e^1(2) - e^1(1))/(v^1(1, 2) - v^1(2, 2)) \leq z$. This probability is derived from the distribution function ρ^1 in the obvious fashion. Note that F^1 is continuous on R^1.

Let $F^2(z)$ be the probability that, on any given date, $(e^2(1) - e^2(2))/(v^2(2, 2) - v^2(2, 1)) \leq z$. Note that F^2 is continuous.

We want to compute $\beta^1(\alpha^2)$, the marginal probability that 1 plays row 1 if she assesses probability α^2 that 2 plays column 1. If she plays row 1, her expected payoff is $\alpha^2 v^1(1, 1) + (1 - \alpha^2)v^1(1, 2) + e^1(1)$, while row 2 nets for her $\alpha^2 v^1(2, 1) + (1 - \alpha^2)v^1(2, 2) + e^1(2)$. Simple algebra shows that the former is greater if

$$\frac{e^1(2) - e^1(1)}{v^1(1, 2) - v^1(2, 2)} < 1 - \alpha^2\left(1 + \frac{v^1(2, 1) - v^1(1, 1)}{v^1(1, 2) - v^1(2, 2)}\right).$$

Define

$$x = 1 + \frac{v^1(2, 1) - v^1(1, 1)}{v^1(1, 2) - v^1(2, 2)}.$$

Then row 1 is chosen whenever $(e^1(2) - e^1(1))/(v^1(1, 2) - v^1(2, 2)) < 1 - x\alpha^2$. This gives

$$\beta^1(\alpha^2) = F^1(1 - x\alpha^2).$$

Note that $x > 1$ and that β^1 is a nonincreasing function of α^2.

A similar computation shows that if we define

$$y = 1 + \frac{v^2(1, 1) - v^2(1, 2)}{v^2(2, 2) - v^2(2, 1)},$$

then

$$\beta^2(\alpha^1) = 1 - F^2(1 - y\alpha^1).$$

Note that $y > 1$ and that β^2 is a nondecreasing function of α^1.

LEMMA 8.2. *Fix any* 2×2 *game with a unique Nash equilibrium that is strictly mixed. Then every augmented version of the game that satisfies Assumption 7.1 has unique Nash distributions and thus has Nash equilibrium strategies that are essentially unique.*

Proof. Nash distributions are pairs (α^1_*, α^2_*) where $\beta^1(\alpha^{-i}_*) = \alpha^i_*$ for $i = 1, 2$. To show existence of a solution to these two equations, note that $(\alpha^1, \alpha^2) \mapsto (\beta^1(\alpha^2), \beta^2(\alpha^1))$ is a continuous function mapping the unit square into itself, and use Brouwer's fixed point theorem. To show uniqueness, suppose that (α^1_*, α^2_*) and $(\tilde{\alpha}^1_*, \tilde{\alpha}^2_*)$ are two Nash distributions. Without loss of generality, assume $\alpha^1_* \neq \tilde{\alpha}^1_*$ and, in fact, $\alpha^1_* > \tilde{\alpha}^1_*$. Because β^2 is nondecreasing, this implies that $\alpha^2_* \geq \tilde{\alpha}^2_*$. And because β^1 is nonincreasing, this implies that $\alpha^1_* \leq \tilde{\alpha}^1_*$, a contradiction. ∎

We hereafter denote the probabilities of row 1 and column 1 in the unique Nash distributions by α^1_* and α^2_*. In general, it is not the case that α^1_* and α^2_* are both strictly between zero and one, even if the original (unaugmented) game has a unique, completely mixed equilibrium. However, the Nash distribution probabilities are strictly between zero and one if the supports of the perturbations are sufficiently small.

Assumption 8.3. For $i = 1, 2$, the density function F^i is uniformly bounded on its support. The density function for F^1 is strictly positive on some neighborhood of $1 - x\alpha^2_*$, and the density function for F^2 is strictly positive on some neighborhood of $1 - y\alpha^1_*$.

This assumption is stated in somewhat implicit form, since it involves the density functions of the distribution functions F^1 and F^2. Restating it in terms of the original distribution functions ρ^1 and ρ^2 of the perturbation vectors is tedious but not difficult. A set of sufficient conditions for this assumption is that the supports of the perturbation vectors E^1 and E^2 are connected and small enough that the equilibrium marginal distributions

are strictly between zero and one, and the density functions for ρ^1 and ρ^2 are bounded and strictly positive on the interiors of their supports.

The Intuition for the Proof of Proposition 8.1

The proof of Proposition 8.1 is fairly involved, and it is easy to get lost in the details. While we reserve those details for the Appendix, we sketch here the intuition behind the proof.

We fix behavior and assessment rules where behavior is myopic with respect to the assessments, and the assessment rules conform precisely to the model of fictitious play, as given in Section 3. We write α^i_t for the probability assessed by $-i$ at date t that player i will play her first action. This is a random variable, depending on the history of play up to date t. We show that $\lim_{t\to\infty}(\alpha^1_t, \alpha^2_t) = (\alpha^1_*, \alpha^2_*)$ with probability one; since behavior is myopic and myopic behavior is continuous in beliefs, this implies that behavior converges to the Nash equilibrium strategies.

For notational simplicity, we suppose that the players' assessments equal the empirical distributions at all dates $t \geq 2$, which corresponds to the case of initial weights identically equal to zero in the fictitious play model. (It will become clear that allowing for nonzero initial weights does not alter the analysis.) In this case

$$\alpha^i_{t+1} = \begin{cases} (t\alpha^i_t + 1)/(t + 1), & \text{if } i \text{ plays action 1 in round } t, \\ t\alpha^i_t/(t + 1), & \text{if } i \text{ plays action 2 in round } t, \end{cases}$$

which is more conveniently written as

$$\alpha^i_{t+1} - \alpha^i_t = \begin{cases} (1 - \alpha^i_t)/(t + 1), & \text{if } i \text{ plays action 1 in round } t, \\ -\alpha^i_t/(t + 1), & \text{if } i \text{ plays action 2 in round } t. \end{cases}$$

The key thing to note here is that the size of the changes of the α^i vanishes asymptotically, at rate $O(1/t)$.

The theory of stochastic approximation (Arthur *et al.*, 1987; Kushner and Clark, 1978; Ljung and Söderström, 1983) shows when such asymptotically vanishing stochastic variations can be ignored, i.e., when the successive random variables will almost surely evolve according to the evolution of their expected values. Starting from some value of (α^1_t, α^2_t), compute the (conditional) expected values $(\alpha^1_{t+1}, \alpha^2_{t+1})$, then use these to compute expected values of $(\alpha^1_{t+2}, \alpha^2_{t+2})$, and so on. If for every starting value of (α^1_t, α^2_t), this "successive expected values" sequence converges to (α^1_*, α^2_*), and if certain regularity conditions are met, then the random process will almost surely approach (α^1_*, α^2_*).

So fix some (α^1_t, α^2_t) and compute the conditional expected values of

$(\alpha_{t+1}^1, \alpha_{t+1}^2)$. Or, since it makes matters a bit more transparent, let us compute $\mathbf{E}[\alpha_{t+1}^i - \alpha_t^i | \xi_t]$ for $i = 1, 2$, where $\mathbf{E}[\cdot|\xi_t]$ denotes expectation taken with respect to $\mathbf{P}(\cdot|\xi_t)$.

Given α_t^1, the probability that player 2 will play his first strategy is $\beta^2(\alpha_t^1) = 1 - F^2(1 - y\alpha_t^1)$, so the expectation of the difference between α_{t+1}^2 and α_t^2 is

$$\frac{1 - \alpha_t^2}{t+1}(1 - F^2(1 - y\alpha_t^1)) + \frac{-\alpha_t^2}{t+1}F^2(1 - y\alpha_t^1).$$

This simplifies to

$$\mathbf{E}[\alpha_{t+1}^2 - \alpha_t^2 | \xi_t] = \frac{1}{t+1}(1 - F^2(1 - y\alpha_t^1) - \alpha_t^2).$$

A similar calculation gives

$$\mathbf{E}[\alpha_{t+1}^1 - \alpha_t^1 | \xi_t] = \frac{1}{t+1}(F^1(1 - x\alpha_t^2) - \alpha_t^1).$$

The fact that the step size in these difference equations is going to zero suggests that the evolution of the successive expected values approximates that in the related differential equation system

$$\frac{d\alpha_t^1}{dt} = F^1(1 - x\alpha_t^2) - \alpha_t^1 \quad \text{and} \quad \frac{d\alpha_t^2}{dt} = 1 - F^2(1 - y\alpha_t^1) - \alpha_t^2, \qquad (\heartsuit)$$

where we reinterpret the α^i's as the successive expected values, and the time index has been changed: Because the amount of change between time t and $t + 1$ in the differential equations is independent of t, more and more steps of the difference equations are compressed per unit of time of the differential equation system as t becomes larger.

The trajectories of this system of differential equations are most easily studied by comparing them with those of the system

$$\frac{d\alpha_t^1}{dt} = F^1(1 - x\alpha_t^2) - \alpha_*^1 \quad \text{and} \quad \frac{d\alpha_t^2}{dt} = 1 - F^2(1 - y\alpha_t^1) - \alpha_*^2.$$

The second system has closed, convex orbits around the point (α_*^1, α_*^2), and that relative to the second system, the first always points strictly inward. (See Fig. 4.) Thus the first system spirals in toward (α_*^1, α_*^2). This suggests that the successive expected value sequence approaches (α_*^1, α_*^2) from any starting point, and then the methods of the theory of stochastic approximation will yield the almost sure convergence that we desire.

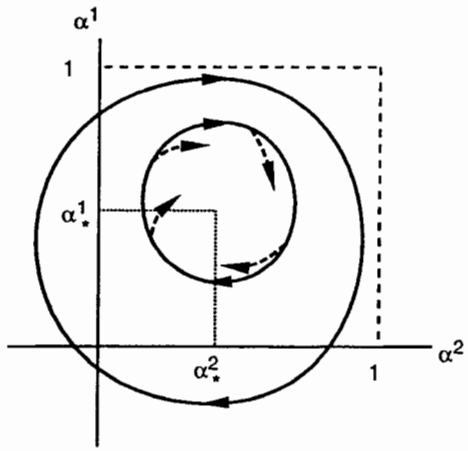

FIG. 4. Dynamics of expected beliefs. The solid curves represent trajectories of the second system of differential equations given in the text. These closed orbits give the level sets of the Lyapunov function L. (Note that these level sets are not restricted to stay inside the unit square.) The dashed arrows show the trajectories of the first system of differential equations, the system that describes the dynamics of expected beliefs. (These do stay within the unit square.) Since the dashed arrows always point inward relative to the closed orbits of the second system, the first system gives trajectories that spiral in toward (α_*^1, α_*^2).

In relating this intuition to the proof we give in the Appendix, there are two things to watch for. First, we use the closed orbit trajectories of the second system of differential equations as level curves for a Lyapunov function. Second, we derive parts of the theory of stochastic approximation that we need, because the Lyapunov function we construct is a bit less regular than is required for the general results as they are stated in the literature.[25]

Extensions of Proposition 8.1

Proposition 8.1 gives a relatively restricted global convergence result. It is restricted in that the behavior and assessment rules that are permitted are quite specific; behavior must be precisely myopic with respect to assessments that are formed according to the model of fictitious play. It is further restricted in that it considers only 2×2 games, and then only 2×2 games in which the unaugmented game has a single equilibrium

[25] Specifically, our Lyapunov function is not twice continuously differentiable in general.

that is completely mixed, and then only those for which the augmentation satisfies Assumption 8.3.

Thinking first about the behavior and assessment rules, it is clear that extensions are possible. (Indeed, extensions along these lines are suggested by Arthur *et al.* (1987).) Assessments can be asymptotically empirical and behavior asymptotically myopic, as long as the "rates of convergence" to fictitious play and myopic behavior are sufficiently fast. Specifically, the dynamic process being studied is determined at each stage by the two possible conditional values of $\alpha_{t+1}^i - \alpha_t^i$ (two for $i = 1$ and two more for $i = 2$) and the probabilities of those values. The probabilities are determined by the behavior rule (which determines which action a player chooses, thus which action his rival observes); the possible values of $\alpha_{t+1}^i - \alpha_t^i$ are determined by the assessment rule. As the detailed proof indicates, we can tolerate changes that, in terms of these differences and probabilities, contribute differences that are uniformly $O(1/t^2)$ or smaller. One can control the probabilities by imposing a rate of convergence test on the sequence $\{\varepsilon_t\}$ that governs the asymptotic part of asymptotically myopic behavior. But for the differences $\alpha_{t+1}^i - \alpha_t^i$, a bit more delicacy is called for.

One's first instinct might be to impose a uniform rate of convergence to empirical assessments; the natural condition would seem to be that

$$|\alpha_t^i - \bar{\alpha}^i(\zeta_t)| = o(1/t^2), \qquad (\spadesuit)$$

where $\bar{\alpha}^i(\zeta_t)$ is the fraction of the time that i has played her first strategy. But this is stronger than is needed. We do not need to know that α_t^i is close to empirical frequencies, but only that if α_t^i is fairly far from empirical frequencies, α_{t+1}^i is going to be about the same distance away (from the new empirical frequency) in the same direction. Specifically, we need to know that

$$\alpha_{t+1}^i - \alpha_t^i = \begin{cases} (1 - \alpha_t^i)/(t+1) + o(1/t^2), & \text{if } i \text{ plays action 1 in round } t, \\ -\alpha_t^i/(t+1) + o(1/t^2), & \text{if } i \text{ plays action 2 in round } t. \end{cases} \qquad (\clubsuit)$$

In this regard, note that for fictitious play beliefs with nonzero initial weight vectors, ($\clubsuit$) holds (so our proof extends to this case) even though ($\spadesuit$) fails.

Extending our results to other classes of 2×2 games or beyond 2×2 games (and to games with more than two players) seems to offer greater challenges. We conjecture that results on local stability can be derived, along the following line. Take an augmented game and any equilibrium

distribution for that game. Write down the continuous-time dynamics for the expected values of the empirical frequencies, as in Eq. ($\blacktriangledown$).[26] If this system is locally stable at the equilibrium values by the usual eigenvalue test, then the equilibrium will be locally stable in the sense of this paper for fictitious-play learning dynamics. (Compare with Arthur *et al.* (1987).) It is interesting to speculate whether the continuous-time system that goes with the Shapley (1964) counterexample is unstable at the equilibrium. If so, then its instability under fictitious play (for augmentations) might follow.

Regardless of these conjectures, we hope that the limited results we have managed to derive here indicate that, with Harsanyi's notion of purification, it is plausible that in some cases players would learn to play a mixed-strategy equilibrium.

Appendix: Proofs Omitted in the Text

Proof of Lemma 6.2. Fix any $a \in A$. Construct a "standard" probability space $\{\Omega, F, P\}$, where $\Omega = [0, 1]^{\{1,2,\ldots\}}$ and, writing ω_t as the tth component of ω, the sequence $\{\omega_t\}$ is an independent sequence of random variables, each uniformly distributed on the unit interval. Enumerate A as $\{a_1, a_2, \ldots, a_N\}$ (where N is the cardinality of A), with $a_1 = a$. Now define random variables y_t on this standard probability space as follows: For $t = 1$, $y_1(\omega) = a_n$ for that index n such that

$$\sum_{m=1}^{n-1} \pi_1(a_m) < \omega_1 \leq \sum_{m=1}^{n} \pi_1(a_m).$$

Then, inductively in t, let $y_t(\omega) = a_n$ for that index n such that

$$\sum_{m=1}^{n-1} \pi_t(a_m | y_1(\omega), \ldots, y_{t-1}(\omega)) < \omega_t \leq \sum_{m=1}^{n} \pi_t(a_m | y_1(\omega), \ldots, y_{t-1}(\omega)).$$

This construction uses the uniformly distributed random variables to construct a sequence of random variables whose joint distribution is identical to the joint distribution of the original sequence $\{x_t\}$. Accordingly, we can define Λ in terms of the constructed probability space, and if we prove that the stated bounds on the limits superior and inferior of $\tau_t(a)/t$ hold almost surely conditional on Λ for each a taken one at a time, then they hold almost surely on Λ for all the (finitely many) a's simultaneously, which then gives the desired result.

But showing that the two bounds hold on Λ for the y_t sequence and the fixed a is easy. Because we set $a_1 = a$, the set of points $\omega \in \Lambda$ for which $y_t(\omega) = a$ contains the set $\{\omega : \omega_t \leq \pi(a) - \varepsilon\}$ and is contained within the set $\{\omega : \omega_t \leq \pi(a) + \varepsilon\}$. This is so because $\pi(a)\ \varepsilon \leq \pi_t(a | y_1, \ldots, y_{t-1}) \leq \pi(a) + \varepsilon$ for all t, for points in Λ by definition; then compare with how

[26] This is conceptually straightforward, but it is *not* a simple exercise in practice, which is one reason we offer conjectures instead of results.

we determine those ω for which $y_t(\omega) = a$. For $r \in [0, 1]$, let $\nu_t(r, \omega)$ be the number of times that $\omega_{t'} \leq r$ for $t' = 1, \ldots, t$. Then the estimate

$$\nu_t(\pi(a) - \varepsilon, \omega) \leq \tau_t(a, \omega) \leq \nu_t(\pi(a) + \varepsilon, \omega) \qquad \text{for } \omega \in \Lambda$$

follows from the asserted set inclusions. By the strong law of large numbers,

$$\lim_{t\to\infty} \frac{\nu_t(r, \omega)}{t} = r$$

with probability one for each r individually, so this holds with probability one both for $r = \pi(a) - \varepsilon$ and $r = \pi(a) + \varepsilon$. This, combined with the previous bounds on $\pi_t(a)$ on Λ, gives precisely the desired result. ■

In Proposition 6.3, we are required to show that the probability of a certain type of event can be made as close to one as desired. The proof is simplified by the following lemma, which shows that it suffices to make the probability of this type of event strictly positive.

LEMMA A.1. *Suppose that for a strategy profile σ_* there exists some asymptotically empirical assessment rules and some behavior rules that are asymptotically myopic with respect to those rules, such that for some t and $\zeta_t \in \mathfrak{Z}_t$,*

$$\mathbf{P}\left(\lim_{t'\to\infty} \phi_{t'}(\zeta_{t'}) = \sigma_* | \zeta_t\right) > 0.$$

Then, σ_* is locally stable.

To prove the lemma, let Λ be the event $\{\lim_{t'\to\infty}\phi_{t'}(\zeta_{t'}) = \sigma_*\}$. By the usual arguments, this is measurable with respect to the σ-field generated by the $\{\zeta_1, \zeta_2, \ldots\}$. Thus by Paul Levy's zero-or-one law (Chung, 1974, p. 341), the probability of Λ conditional on $\zeta_{t'}$ approaches the indicator function of Λ as t' approaches infinity.[27] Since Λ has positive probability conditional on ζ_t, for some (positive probability) $\zeta_{t'}$ that are continuations of ζ_t, the conditional probability of Λ can be made as close to one as desired, which gives the result.

Proof of Proposition 6.3. As noted in the discussion following the statement of this proposition, we can sharpen the result in either of two ways; we can require that behavior is precisely myopic and assessments are asymptotically empirical, or we can require that assessments are precisely empirical and behavior is asymptotically myopic. We give a construction for the first of these sharpenings here; the other is left to the reader's ingenuity (or see earlier working paper versions of this paper).

Fix the Nash equilibrium profile σ_*. By virtue of the lemma, we only need to find myopic behavior rules and asymptotically empirical assessment rules, and a t and ζ_t such that

$$\mathbf{P}\left(\lim_{t'\to\infty} \phi_{t'}(\zeta_{t'}) = \sigma_* | \zeta_t\right) > 0.$$

[27] If you have trouble squaring this assertion with what you find in Chung (1974), recall that $\zeta_{t'}$ "contains" ζ_t for $t < t'$, so conditioning on $\zeta_{t'}$ is the same as conditioning on $\{\zeta_1, \ldots, \zeta_{t'}\}$.

The behavior rules are easily defined: For each player i, date t, and history ζ_t such that the player i's assessment $\mu_t^i(\zeta_t)$ is that her rivals are playing according to σ_*^{-i}, player i uses the strategy σ_*^i. If her assessment is anything else (including any form of correlation in the play of her rivals), her behavior can be assigned arbitrarily, as long as it is myopic.

To construct the assessment rules, create a probability space on which is defined a sequence of random strategy profiles $\{s_1, s_2, \ldots\}$, where each s_t is independently and identically distributed according to σ_*.

Let

$$\delta(\zeta_t) = \max_{i=1,\ldots,I} [\|\bar{\sigma}^{-i}(\zeta_t) - \sigma_*^{-i}\|],$$

(where σ_*^{-i} is taken to be the joint distribution on S^{-i} with margins given by the σ_*^j and with independence determining all joint probabilities). That is, $\delta(\zeta_t)$ is the difference (in the sup norm) between the empirical frequencies and the target strategy. By the strong law of large numbers, with probability one the empirical frequencies of strategies and joint strategy profiles all converge to the corresponding probabilities under σ_*. So for $n = 1, 2, \ldots$, let L_n be a positive integer sufficiently large so that the event

$$\{\delta(\zeta_t) \leq 1/n \text{ for all } t > L_n\}$$

has probability at least equal to $(2^{n+1} - 1)/2^{n+1}$. Assume that $L_{n+1} > L_n$. For each $t = 1, 2, \ldots$, let $n(t) = 0$ if $t < L_1$, let $n(t) = 1$ if $L_1 \leq t < L_2$, and so on. Note that $\lim_{t\to\infty} n(t) = \infty$.

We claim that by construction, the event

$$\{\delta(\zeta_t) \leq 1/n(t),\ t = 1, 2, \ldots\}$$

has probability at least one-half. To see why, note that this event can be written as

$$\bigcap_{m=0}^{\infty} \{\delta(\zeta_t) \leq 1/n \text{ for all } t \text{ such that } n(t) = n\},$$

and note that the probability of the events in this intersection are 1 for $n = 0$, $\frac{3}{4}$ (or more) for $n = 1$, $\frac{7}{8}$ (or more) for $n = 2$, and so on. Apply De Morgan's law to see that the complement of this event is the union of events, the first of which has probability 0, the second probability no more than $\frac{1}{4}$, the third probability no more than $\frac{1}{8}$, and so on. Hence the probability of this union is $\frac{1}{2}$ at most, and the probability of its complement—the event we are interested in—is at least $\frac{1}{2}$.

Now we give the assessment rules. For any history ζ_t, let

$$\mu_t^i(\zeta_t) = \begin{cases} \sigma_*^{-i}, & \text{if } \|\bar{\sigma}^{-i}(\zeta_t) - \sigma_*^{-i}\| < 1/n(t), \\ \bar{\sigma}^{-i}(\zeta_t), & \text{otherwise.} \end{cases}$$

That is, the i believes that her rivals are playing (independently) according to σ_*^{-i} unless and until the accumulated evidence against this hypothesis becomes severe (as measured

by $1/n(t)$), at which point the player reverts to empirical beliefs. These assessment rules are clearly asymptotically empirical.

With these behavior and assessment rules, the players begin with σ_*, and the sequence $\{L_n\}$ has been constructed precisely so that, with probability $\frac{1}{2}$ or more, the players continue to use σ_* forever. To see this, note that if $\delta(\zeta_{t'}) \leq 1/n(t')$ for $t' = 1, 2, \ldots, t$, then $\phi_t = \sigma_*$. Thus the probability of the event $\{\delta(\zeta_t) \leq 1/n(t), t = 1, 2, \ldots\}$ under the measure induced by the behavior rules ϕ is precisely the same as the probability of this event under the probability distribution where all strategy profiles are drawn independently and identically according to the distribution σ_*. By construction, this event has probability at least $\frac{1}{2}$ under the i.i.d.-generated measure, so it has the same probability (at least $\frac{1}{2}$) under the measure generated by the behavior rules ϕ. And, of course,

$$\{\delta(\zeta_t) \leq 1/n(t), t = 1, 2, \ldots\} \subseteq \{\phi_t(\zeta_t) = \sigma_*, t = 1, 2, \ldots\}.$$

This completes the proof.

The proof of Proposition 7.5 uses the following lemma, which shows that as time passes, asymptotically myopic behavior converges to myopic behavior, in the sense that the induced distributions over outcomes become close.

LEMMA A.2. *For any $\delta > 0$ there exists an $\varepsilon > 0$ such that, for any beliefs α^{-i} and for any s^i that ε-maximizes player i's payoff against α^{-i} for ρ^i-almost every e^i,*

$$\|\pi^i(s^i) - \beta^i(\sigma^{-i})\| \leq \delta.$$

Proof of Lemma A.2. We show that for any δ there exists an ε such that for all α^{-i}, the set of e^i for which player i has more than one ε-best response has measure (under ρ^i) no greater than δ. To see this, fix some α^{-i} and note that the set of e^i that makes player i indifferent between any two given actions lies on a lower-dimensional hyperplane, and there are (at most) $(\#A^i)^2$ such hyperplanes, where $\#A^i$ denotes the cardinality of A^i. Put a "sleeve" of diameter ε around each of these hyperplanes, and the set of e^i where player i has multiple ε-best responses is contained in the union of these sleeves. In the compact set E^i, there is a uniform (in α^i) upper bound on the Lebesgue measure of the union of $(\#A^i)^2$ sleeves of this sort, and this uniform upper bound goes to zero as ε goes to zero. Because ρ^i is absolutely continuous with respect to Lesbesgue measure, there is thus a uniform upper bound on the ρ^i-measure of these sets, going to zero as ε goes to zero, which establishes the result. ■

Proof of Proposition 7.5. Fix an α_* that is not a Nash distribution. Without loss of generality, suppose that $\beta^1(\alpha_*^{-1}) \neq \alpha_*^1$. Let $\delta = \|\beta^1(\alpha_*^{-1}) - \alpha_*^1\|$. Since β^1 is continuous, we can find ε' sufficiently small that $\|\beta^1(\alpha^{-1}) - \beta^1(\alpha_*^1)\| \leq \delta/4$ for all α^{-1} within $2\varepsilon'$ of α_*^{-1}.

We show that profile α_* is unstable for $\varepsilon = \min(\varepsilon'/I, \delta/4)$. Suppose not, so that for some history ξ_t,

$$\mathbf{P}(\|\pi^i(\phi^i_{t'}(\xi^i_{t'})) - \alpha^i_*\| < \varepsilon \text{ for all } t' \geq t \text{ and for } i = 1, \ldots, I | \xi_t) > 0.$$

Lemma 6.2 then implies that (almost surely on this event of positive probability) the empirical marginal frequencies of actions eventually lie within ε of the α^i_*. Then because assessments are asymptotically empirical, we conclude that (almost surely on this event), $\|\mu^1_{t'}(\zeta_{t'}) - \alpha_*^{-1}\| \leq I\varepsilon \leq \varepsilon'$. This in turn implies that the distribution on actions $\beta^1(\mu^1_{t'}(\zeta_{t'}))$ induced by the myopic best response to $\mu^1_{t'}(\zeta_{t'})$ is within $\delta/4$ of $\beta^1(\alpha_*^{-1})$.

From Lemma A.2, there is one ε'' such that the set of ε''-best responses to $\mu^1_{t'}(\zeta_{t'})$ is within $\delta/4$ of $\beta^1(\mu^1_{t'}(\zeta_{t'}))$ for any $\zeta_{t'}$. Let t' be large enough that the suboptimization allowed for by player 1's decision rule is less than this ε''. The triangle inequality then implies that the marginal distribution over player 1's actions is within $\delta/2$ of $\beta^1(\alpha_*^{-1})$, and hence at least $\delta/2 \geq \varepsilon$ away from α_*^1, which contradicts the hypothesis.

Proof of Proposition 8.1. For $(\alpha^1, \alpha^2) \in [0, 1] \times [0, 1]$, let $L(\alpha^1, \alpha^2)$ be the function defined by

$$L(\alpha^1, \alpha^2) = \int_{\alpha_*^1}^{\alpha^1} [1 - F^2(1 - y\beta^1) - \alpha_*^2]\, d\beta^1 - \int_{\alpha_*^2}^{\alpha^2} [F^1(1 - x\beta^2) - \alpha_*^1]d\beta^2.$$

The function L (a mnemonic for Lyapunov) has the following properties:

(a) $L(\alpha_*^1, \alpha_*^2) = 0$.

(b) If $(\alpha^1, \alpha^2) \neq (\alpha_*^1, \alpha_*^2)$,then $L(\alpha^1, \alpha^2) > 0$.

(c) L is continuously differentiable with gradient vector

$$\nabla L = (1 - F^2(1 - y\alpha^1) - \alpha_*^2, \alpha_*^1 - F^1(1 - x\alpha^2)).$$

This gradient vector is zero at (α_*^1, α_*^2) and nonzero everywhere else.

(d) L is convex. In fact, the level curves of L are trajectories of

$$\frac{d\alpha_t^1}{dt} = F^1(1 - x\alpha_t^2) - \alpha_*^1 \quad \text{and} \quad \frac{d\alpha_t^2}{dt} = 1 - F^2(1 - y\alpha_t^1) - \alpha_*^2,$$

which gives closed trajectories that wind around the point (α_*^1, α_*^2).

Property (a) is immediate from the definition of L. For property (b), note first that $1 - F^2(1 - y\alpha_*^1) = \alpha_*^2$ and $F^1(1 - x\alpha_*^2) = \alpha_*^1$. Then enlist Assumption 8.3 to note that $1 - F^2(1 - y\alpha^1)$ is nondecreasing, and it is strictly increasing in a neighborhood of α_*^1, and $F^1(1 - x\alpha^2)$ is nonincreasing, and it is strictly decreasing in a neighborhood of α_*^2. Property (b) then follows. Property (c) is virtually a matter of definitions, together with the properties of $1 - F^2(1 - y\alpha^1)$ and $F^1(1 - x\alpha^2)$ just noted. For part (d), compute the Hessian of L to show convexity; the rest is an exercise in integration.

Note that L is separable; i.e., $L(\alpha^1, \alpha^2) = L^1(\alpha^1) - L^2(\alpha^2)$, where we define

$$L^1(\alpha^1) = \int_{\alpha_*^1}^{\alpha^1} (1 - F^2(1 - y\beta^1) - \alpha_*^2]d\beta^1,$$

and

$$L^2(\alpha^2) = \int_{\alpha_*^2}^{\alpha^2} [F^1(1 - x\beta^2) - \alpha_*^1]d\beta^2.$$

Now fix assessment rules according to the model of fictitious play, and suppose that behavior is precisely myopic with respect to these assessments.

For every ζ^t, $t > 1$, define

$$L_t(\zeta_t) = L(\alpha_t^1, \alpha_t^2) = L^1(\alpha_t^1) + L^2(\alpha_t^2).$$

In words, $L_t(\zeta_t)$ is the value of the function L at the vector of assessments by the two players. We have proved the theorem if we prove that $\lim_{t\to\infty} L_t(\zeta_t) = 0$ with probability one, since this implies that the beliefs are converging (with probability one) to α_*^1 and α_*^2; hence (by earlier analysis and myopic behavior) the marginal distributions over actions are converging to those values.

The first step in proving that $\lim_t L_t(\zeta_t) \to 0$ is to derive an upper bound for $\mathbf{E}[L_{t+1}(\zeta_{t+1}) - L_t(\zeta_t)|\zeta_t]$. Specifically, we show that there exists a nonpositive continuous function ι on the unit square with $\iota(\alpha^1, \alpha^2) < 0$ for $(\alpha^1, \alpha^2) \neq (\alpha_*^1, \alpha_*^2)$, such that if K is the uniform bound on the density of the two perturbation scalars,

$$\mathbf{E}[L_{t+1}(\zeta_{t+1}) - L_t(\zeta_t)|\xi_t] \leq \frac{\iota(\alpha_t^1, \alpha_t^2)}{t+1} + \frac{K(x+y)}{2(t+1)^2}. \qquad (\bigstar)$$

We obtain this bound by looking at the two (separable) pieces of $L_t(\zeta_t)$. That is, we write

$$\mathbf{E}[L_{t+1}(\zeta_{t+1}) - L_t(\zeta_t)|\xi_t] = \mathbf{E}[L^1(\alpha_{t+1}^1) - L^1(\alpha_t^1)|\xi_t] - \mathbf{E}[L^2(\alpha_{t+1}^2) - L^2(\alpha_t^2)|\xi_t],$$

and we begin with estimates of each of the two expectations on the right-hand side.

Consider first the term involving L^1. Given α_t^1, we have that α_{t+1}^1 will equal $(t\alpha_t^1 + 1)/(t+1)$ if player 1 plays row 1 in round t, and α_{t+1}^1 will equal $t\alpha_t^1/(t+1)$ if player 1 plays row 2. Since 1's beliefs about 2's actions are given by α_t^2, player 1 will play row 1 in round t with probability $F^1(1 - x\alpha_t^2)$. Now

$$L^1\left(\frac{t\alpha_t^1 + 1}{t+1}\right) - L^1(\alpha_t^1) = \int_{\alpha_t^1}^{(t\alpha_t^1+1)/(t+1)} [1 - F^2(1 - y\beta^1) - \alpha_*^2]\mathrm{d}\beta^1,$$

which is bounded above by

$$\left[\frac{t\alpha_t^1 + 1}{t+1} - \alpha_t^1\right][1 - F^2(1 - y\alpha_t^1) - \alpha_*^2] + \frac{1}{2}Ky\left[\frac{t\alpha_t^1 + 1}{t+1} - \alpha_t^1\right]^2,$$

where K is the uniform bound on the density functions of F^1 and F^2.[28] Simplifying, this is

$$\left[\frac{1 - \alpha_t^1}{t+1}\right][1 - F^2(1 - y\alpha_t^1) - \alpha_*^2] + \frac{1}{2}Ky\left[\frac{1 - \alpha_t^1}{t+1}\right]^2.$$

[28] This is the length of the interval of integration, times the value of the integrand at $\beta^1 = \alpha_t^1$, plus a bound on the integral of the difference between the true integrand and the integrand at α_t^1. This latter bound comes from noting that $\beta \mapsto F^2(1 - y\beta)$ is Lipschitz continuous with Lipschitz constant Ky.

Similarly,

$$L^1\left(\frac{t\alpha_t^1}{t+1}\right) - L^1(\alpha_t^1) \leq \left[\frac{-\alpha_t^1}{t+1}\right][1 - F^2(1 - y\alpha_t^1) - \alpha_*^2] + \frac{1}{2}Ky\left[\frac{\alpha_t^1}{t+1}\right]^2.$$

Hence $\mathbf{E}[L^1(\alpha_{t+1}^1) - L^1(\alpha_t^1)|\xi_t]$ is bounded above by

$$\left[\frac{1 - F^2 - \alpha_*^2}{t+1}\right][(1 - \alpha_t^1)F^1 - \alpha_t^1(1 - F^1)] + \frac{Ky}{2(t+1)^2}$$
$$= \left[\frac{1 - F^2 - \alpha_*^2}{t+1}\right][F^1 - \alpha_t^1] + \frac{Ky}{2(t+1)^2},$$

where we suppress the arguments (respectively $1 - x\alpha_t^2$ and $1 - y\alpha_t^1$) of F^1 and F^2.

Similar calculations show that $\mathbf{E}[L^2(\alpha_{t+1}^2) - L^2(\alpha_t^2)|\xi_t]$ is bounded below by

$$\left[\frac{F^1 - \alpha_*^1}{t+1}\right][1 - F^2 - \alpha_t^2] - \frac{Kx}{2(t+1)^2}.$$

Thus

$$\mathbf{E}[L_{t+1}(\zeta_{t+1}) - L_t(\zeta_t)|\xi_t] = \mathbf{E}[L(\alpha_{t+1}^1, \alpha_{t+1}^2) - L(\alpha_t^1, \alpha_t^2)|\xi_t]$$
$$= \mathbf{E}[L^1(\alpha_{t+1}^1) - L^1(\alpha_t^1)|\xi_t] - \mathbf{E}[L^2(\alpha_{t+1}^2) - L^2(\alpha_t^2)|\xi_t]$$

is bounded above by

$$\left[\frac{1 - F^2 - \alpha_*^2}{t+1}\right][F^1 - \alpha_t^1] - \left[\frac{F^1 - \alpha_*^1}{t+1}\right][1 - F^2 - \alpha_t^2] + \frac{K(x+y)}{2(t+1)^2}.$$

Write $[F^1 - \alpha_t^1]$ as $[F^1 - \alpha_*^1 + \alpha_*^1 - \alpha_t^1]$ and write $[1 - F^2 - \alpha_t^2]$ as $[1 - F^2 - \alpha_*^2 + \alpha_*^2 - \alpha_t^2]$, substitute these two expressions into the upper bound, and simplify. This gives

$$\mathbf{E}[L_{t+1}(\zeta_{t+1}) - L_t(\zeta_t)|\xi_t] \leq \left[\frac{1 - F^2 - \alpha_*^2}{t+1}\right][\alpha_*^1 - \alpha_t^1] - \left[\frac{F^1 - \alpha_*^1}{t+1}\right][\alpha_*^2 - \alpha_t^2] + \frac{K(x+y)}{2(t+1)^2}.$$

Define

$$\iota(\alpha^1, \alpha^2) = [1 - F^2(1 - y\alpha^1) - \alpha_*^2][\alpha_*^1 - \alpha^1] - [F^1(1 - x\alpha^2) - \alpha_*^1][\alpha_*^2 - \alpha^2].$$

The signs of $1 - F^2 - \alpha_*^2$ and $\alpha_*^1 - \alpha^1$ are opposite, and the signs of $F^1 - \alpha_*^1$ and $\alpha_*^2 - \alpha^2$ are the same, so that ι is a nonpositive function. Moreover, if $\alpha^1 \neq \alpha_*^1$, then $1 - F^2 - \alpha_*^2 \neq 0$, and if $\alpha^2 \neq \alpha_*^2$, then $F^1 - \alpha_*^1 \neq 0$, so $\iota(\alpha^1, \alpha^2) < 0$ for $(\alpha^1, \alpha^2) \neq (\alpha_*^1, \alpha_*^2)$. Putting everything together, this gives the bound ($\bigstar$).

Next, for each t and ζ_t, let

$$\psi_t(\zeta_t) = \mathbf{E}[L_t(\zeta_t) - L_{t-1}(\zeta_{t-1})|\xi_{t-1}],$$

and

$$\hat{L}_t(\zeta_t) = L_t(\zeta_t) - \sum_{t'=1}^{t} \max\{\psi_{t'}(\zeta_{t'}), 0\}.$$

By construction, $\{\hat{L}_t(\zeta_t)\}$ forms a supermartingale over the information sequence $\{\zeta_t\}$. By the bound (★),

$$\sum_{t=1}^{\infty} \max\{\psi_t(\zeta_t), 0\} \leq \sum_{t=1}^{\infty} \frac{K(x+y)}{2(t+1)^2} < \infty,$$

so $\{\hat{L}_t\}$ is a bounded supermartingale, and hence has a limit almost surely. As an immediate consequence, $L_t(\zeta_t)$ itself has a limit almost surely.

Finally, it is not possible that, with positive probability, this limit is a value greater than zero. To see this, suppose that along some history ζ, $\lim_{t\to\infty} L_t(\zeta_t) > 0$. From the construction of the function ι given previously, it is easy to show that, in this case, $\lim\sup_{t\to\infty} \iota(\alpha_t^1, \alpha_t^2) < 0$, so that for all $t > T$ for some large T, $\iota(\alpha_t^1, \alpha_t^2) < \lambda < 0$. Increase T if necessary so that $T > K(x + y)/|\lambda|$. Then it is a matter of algebra to show that $\psi_t(\zeta_t) < \lambda/(2(t+1))$ for all $t > T$. Accordingly, if we define $\bar{L}_t(\zeta_t)$ as $L_t(\zeta_t) - \Sigma_{t'=1}^{t} \psi_{t'}(\zeta_{t'})$, we know that $\lim_{t\to\infty} \bar{L}_t(\zeta_t) = \infty$ for those ζ where $L_t(\zeta_t)$ has nonzero limit.

But $\{\bar{L}_t(\zeta_t)\}$ is a martingale that is bounded below. If $L_t(\zeta_t)$ has nonzero limit with positive probability, $\bar{L}_t(\zeta_t)$ has limit ∞ with positive probability. In this case, $\lim_{t\to\infty} \mathbf{E}[\bar{L}_t(\zeta_t)] = \infty$, which contradicts the fact that $\{\bar{L}_t(\zeta_t)\}$ is a martingale. ■

REFERENCES

ARTHUR, W. B., ERMOLIEV, Y. M., AND KANIOVSKI, Y. M. (1987). "Limit Theorems for Proportions of Balls in a Generalized Urn Scheme," Mimeo, IIASA.

BORGERS, T., AND JANSSEN, M. (1991). "On the Dominance Solvability of Large Cournot Games," Mimeo, University College, London.

BROWN, G. W. (1951). "Iterative Solutions of Games by Fictitious Play," in *Activity Analysis of Production and Allocation* (T. C. Koopmans, Ed.). New York: Wiley.

CANNING, D. (1991). "Social Equilibrium," Mimeo, Cambridge University.

CHUNG, K.-L. (1974). *A Course in Probability Theory*, 2nd ed. New York: Academic Press.

CRAWFORD, V., AND HALLER, H. (1990). "Learning How to Cooperate: Optimal Play in Repeated Coordination Games," *Econometrica* **58,** 571–596.

EICHENBERGER, J., HALLER, H., AND MILNE, F. (1991). "Naive Bayesian Learning in 2 × 2 Matrix Games," *J. Econ. Behav. Org.*, in press.

ELLISON, G. (1993). "Learning, Local Interaction, and Coordination," *Econometrica*, in press.

FUDENBERG, D., AND KREPS, D. M. (1988). "A Theory of Learning, Experimentation, and Equilibrium in Games," Mimeo, Stanford University.

FUDENBERG, D., AND LEVINE, D. K. (1989). "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica* **57,** 759–778.

FUDENBERG, D., AND LEVINE, D. K. (1993). "Steady State Learning and Nash Equilibrium," *Econometrica* **61,** 547–574.

GUESNERIE, R. (1992). "An Exploration of the Eductive Justification of the Rational Expectations Hypothesis," *Amer. Econ. Rev.* **82,** 1254–1278.

HARSANYI, J. (1973). "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points," *Int. J. Game Theory* **2,** 1–23.

HENDON, E., JACOBSEN, H., AND SLOTH, B. (1991). "Fictitious Play in Extensive Form Games and Sequential Equilibrium," Mimeo, University of Copenhagen.

JORDAN, J. (1991). "Bayesian Learning in Normal Form Games," *Games Econ. Behav.* **3,** 60–81.

KALAI, E., AND LEHRER, E. (1993). "Rational Learning Leads to Nash Equilibrium," mimeo, Northwestern University. [*Econometrica*, in press]

KANDORI, M., MAILATH, G., AND ROB, R. (1993). "Learning, Mutation, and Long-Run Equilibria in Games," *Econometrica* **61,** 29–56.

KUSHNER, H., AND CLARK, D. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. New York: Springer-Verlag.

LJUNG, L., AND SÖDERSTRÖM, T. (1983). *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press.

MARCET, A., AND SARGENT, T. J. (1989a). "Convergence of Least-Squares Learning Mechanisms in Self-Referential Linear Stochastic Models," *J. Econ. Theory* **48,** 337–368.

MARCET, A., AND SARGENT, T. J. (1989b). "Convergence of Least-Squares in Environments with Hidden State Variables and Private Information," *J. Polit. Econ.* **97,** 1306–1322.

MILGROM, P., AND ROBERTS, J. (1990). "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica* **58,** 1255–1278.

MILGROM, P., AND ROBERTS, J. (1991). "Adaptive and Sophisticated Learning in Repeated Normal Form Games," *Games Econ. Behav.* **3,** 82–100.

MIYASAWA, K. (1961). "On the Convergence of the Learning Process in a 2 × 2 Non-Zero Sum Two-Person Game," Mimeo, Princeton University.

MOULIN, H. (1984). "Dominance Solvability and Cournot Stability," *Math. Soc. Sci.* **7,** 83–103.

NYARKO, Y. (1991). "Learning and Agreeing to Disagree Without Common Priors," Mimeo, New York University.

ROBINSON, J. (1951). "An Iterative Method of Solving a Game," *Ann. Math.* **54,** 296–301.

SHAPLEY, L. (1964). "Some Topics in Two-Person Games", in *Advances in Game Theory* (M. Dresher, L. S. Shapley, and A. W. Tucker, Eds.). Princeton, NJ: Princeton Univ. Press.

YOUNG, H. P. (1993). "The Evolution of Conventions," *Econometrica* **61,** 57–84.