Evolution of Beliefs and the Nash Equilibrium of Normal Form Games*

Masaki Aoyagi

Department of Economics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

The paper formulates a simple two-person model of learning with pattern recognition and discusses its implications. In particular, it focuses on the asymptotic behavior of players' beliefs when the game has a mixed-strategy Nash equilibrium. *Journal of Economics Literature* Classification Numbers: C72, D83.   © 1996 Academic Press, Inc.

1. INTRODUCTION

In models of learning, it is recognized that the path of play displays some conspicuous patterns when players use simple rules in assessing their opponents' behavior. If the players themselves become aware of such patterns, they may want to utilize them in an attempt to better assess their opponents' behavior. This paper formulates a simple two-person model of learning that allows such pattern recognition and discusses its implications. In particular, it focuses on the convergence of players' beliefs to a mixed Nash equilibrium of a game.

Consider the learning process of two myopic players who repeatedly play the game of matching pennies as depicted in Fig. 1. Suppose that both players use *fictitious play* to assess their opponents' behavior. Namely, each player maximizes the current-period payoff with the assumption that the opponent's next mixed action is given by the empirical frequency of the opponent's past pure action choices.

\	H	T
H	1, −1	−1, 1
T	−1, 1	1, −1

FIGURE 1

We know from the theorem by Robinson [11][1] that the sequence of each player's beliefs converges to $0.5H + 0.5T$, which corresponds to the Nash equilibrium of the game. Figure 2 illustrates the locus of converging beliefs as a spiral around the Nash equilibrium point. When beliefs follow such a systematic path, however, the players' action choices also exhibit some regularity: For example, suppose player 1 plays H in some period. This implies that he believed that 2 was at least as likely to play H as T in that period. If he in fact observes 2 playing H, then, player 1 never switches from H to T next period since his belief has been reinforced by his current observation. Similarly, player 2 never switches from T to H having observed 1 playing H. If players become aware of such regularity in the opponent's action choices, however, will they still be willing to stick to fictitious play? It is natural to expect that player 2, for instance, tries to take advantage of (H, H) this period by playing T next period.

Our formulation incorporates this type of reasoning into learning. More specifically, players in our model form their beliefs by analyzing history in the following manner: Let a *pattern* be any finite string of action profiles. Player i looks at the history up to the last period of play, and examines each pattern that appears in the tail of the current history. For example, if the current history is given by

$$((T, T), (T, H), ..., (H, H), (H, T)),$$

player i first looks at the one-length pattern (H, T), then the two-length pattern $((H, H), (H, T))$, and so on. For each one of those patterns, player i counts the number of its occurrences in the entire history, and he *recognizes* the pattern if the number exceeds his subjective threshold level. When he does not recognize any pattern, his belief is set equal to the overall empirical frequency as in fictitious play. When he recognizes some patterns, on the other hand, he chooses *one* of them according to his subjective criterion and *uses* it in the following manner: he looks at the opponent's pure action choice immediately after each occurrence of the
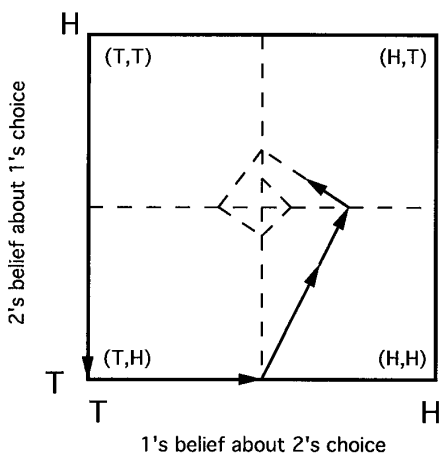
FIG. 2.   Fictitious play in matching pennies. A sample path of beliefs and the corresponding action choices (in parentheses).

pattern, and calculates the *empirical frequency conditional on the pattern*. His new belief is then set equal to a convex combination of the overall and conditional empirical frequencies. The weight used in the convex combination generally depends on history. Therefore, when we define a player's *assessment rule* to be a mapping that assigns a belief to each finite history, each assessment rule in our model is characterized by such parameters as the threshold number for the repetition of each pattern, the choice rule when more than one pattern is recognized, and the weight to be used in the convex combination given each history.

One important aspect of the current formulation is that adjustments based on patterns are not once and for all. Namely, if players adjust their beliefs based on some pattern, this adjustment itself may cause a skewed empirical distribution when conditioned on some other (possibly more complex) patterns. For example, in the game of matching pennies above, suppose that players begin to recognize only one-length patterns, and that their beliefs are set equal to the empirical frequencies conditional on those one-length patterns.[2] Suppose further that the *two*-length tail of the current history (in periods $t-1$ and t) is given by $((H, T), (H, T))$. It can be seen that player 2 will always play T following this two-length pattern as long as his assessment is based on patterns of length one: Since player 2 is using one-length patterns, his assessment for period t was given by the empirical frequency conditional on (H, T). The fact that he played T in period t indicates that he placed more weight on player 1's H conditional on (H, T).

---

[2] Namely, the convex combination between the overall and conditional frequencies is such that the latter gets the whole weight.

Since player 1 in fact played H in period t, player 2 now places even more weight on 1's playing H conditional on (H, T). Since the pattern (H, T) is to be used again in period $t+1$, player 2 must play T in period $t+1$. Although player 1 cannot detect such regularity when he uses only one-length patterns, he will become aware of it as soon as he begins to recognize two-length patterns. Similar adjustments will repeatedly take place for longer patterns as long as those patterns are recognizable for players.

In that players use empirical frequencies of some form, our model may be viewed as a generalization of fictitious play. First introduced by Brown [1], fictitious play and its generalization have been extensively studied as models of learning in recent years. Robinson [11] proves that fictitious play converges to a Nash equilibrium in any zero-sum games. Miyasawa [9] proves the same in 2×2 games. Krishna [6] and Monderer and Shapley [10] further provide the convergence result in supermodular games and potential games, respectively. Fudenberg and Kreps [2] extend the idea of fictitious play to allow players to conduct experiments in extensive-form stage-games.

Apart from the issue raised in this paper, it is also known that fictitious play has a few problems as a learning process. Shapley [12] constructs a 3×3 game with a unique mixed equilibrium such that fictitious play ends up in the limit cycle around the Nash equilibrium point (given a certain pair of initial actions).[3] Jordan [5] makes a similar observation in a three-person version of matching pennies with a unique mixed equilibrium. For a class of assessment rules that includes fictitious play,[4] he shows that any such process ends up in the limit cycle around the Nash equilibrium point given generic initial beliefs.

Another problem with fictitious play is that the joint empirical frequency may never approach a (mixed) Nash equilibrium even when the marginal frequency converges to it. Fudenberg and Kreps [3] and Jordan [5] both make this observation.[5] Fudenberg and Kreps [3] take a more rigorous interpretation of convergence to a mixed equilibrium, and propose the use of convergence in mixed actions as opposed to convergence in marginal empirical frequencies. With this notion of convergence, they show that (1) a strategy profile which is not a Nash equilibrium is unstable, and (2) for any mixed equilibrium, there exist assessment rules and players' behavioral rules that make the equilibrium stable, both under the

[3] Note that Shapley's [12] limit cycle exhibits patterns similar to the ones described above since it is generated by fictitious play.

[4] The assessment rules that (1) are functions of the overall empirical frequency only, and (2) asymptotically approximate the overall frequency at a uniform rate (over all the frequencies). Some Bayesian learning process as well as fictitious play falls in this category.

[5] See Section 6.

conditions that the assessment rules are asymptotically empirical and that players are asymptotically myopic.[6]

The paper proceeds as follows: In the next section, we develop the basic notation used in the analysis. We define the notion of stability in Section 3. Roughly, a (mixed) action profile is *stable* if players' beliefs converge to the profile with probability one. When assessment rules are *adaptive* as defined in Section 3, it is shown that a stable profile (if any) must be a Nash equilibrium profile. Since our assessment rules will turn out to be adaptive, the question then is whether a given Nash equilibrium is stable or not. Section 5 takes up this major question of the paper.

Section 4 formally describes the assessment rule proposed above. It also presents some results which indicate that the proposed rule actually recognizes patterns. In particular, it is shown (under some conditions on the parameters) that when the opponent biases his action today on the most recent history of some given length,[7] the player using the proposed assessment rule asymptotically predicts the opponent's next move.

In Section 5, we examine the stability of Nash equilibrium. When a Nash equilibrium is strict (and hence pure), it is readily verified that it is stable provided that initial beliefs lie in its small neighborhood. Therefore, our main interest is in the stability of a *mixed* Nash equilibrium. For this purpose, we will look at two particularly simple specifications of the choice between the overall and conditional empirical frequencies:[8] In our first specification, a player uses the conditional empirical frequency only if the discrepancy between the two frequencies exceeds some fixed level. He uses the overall frequency otherwise. We will refer to this rule as *discontinuous adjustment*. In our second specification, a player uses the conditional frequency whenever possible. We will refer to this rule as *conditional fictitious play*.

Under discontinuous adjustment, it is shown that a mixed Nash equilibrium in a wide class of games is unstable provided that players can recognize *one*-length patterns. This result is based on the following logic: for a mixed Nash equilibrium to be stable, a player cannot repeatedly switch between the overall and conditional frequencies since that implies a "jump" of his belief over the discrepancy between the two frequencies. After some point in time, hence, he must use only one of the two frequencies. If the player uses only the overall frequency, his beliefs display the same systematic movement as in regular fictitious play. This makes his action

---

[6] An assessment rule is "asymptotically empirical" if the belief is set asymptotically equal to the overall frequency. A player is "asymptotically myopic" if it asymptotically chooses a myopic best response to his belief.

[7] This assumption is purely hypothetical. In the following sections, we will analyze the process when both players use the proposed assessment rules.

[8] Namely, the weight used in the convex combination takes the form $(0, 1)$ or $(1, 0)$.

choices predictable, causing the opponent's belief to jump. On the other hand, if the player uses only the conditional frequency and if his beliefs converge, it is shown that the overall frequency must converge to the same point. This contradicts the presumption that there exists a discrepancy between the two frequencies when the players use the conditional frequency.

Under conditional fictitious play, on the other hand, it is shown that any *unique* (pure or mixed) Nash equilibrium in a *zero-sum* game is *stable*, provided that there exists an upper bound on the length of patterns that players can recognize, and that the assessment rules are *asymptotically symmetric* in the sense that the two players use patterns of the same length asymptotically.[9] The intuition behind this result is as follows: since there exists a limit on the length of recognizable patterns, we can pick some (random) time T such that after T the players use only those patterns that are used infinitely often. Furthermore, since the assessment rules are asymptotically symmetric, we can suppose that the two players use the same patterns after T. Therefore, the set of periods after T can be partitioned into subsequences of periods depending on which pattern is used. Let w be one of the patterns used infinitely often, and take a subsequence of periods in which w is used. Since the beliefs are set equal to the empirical frequency conditional on w in those periods, the corresponding subsequence of beliefs converges when regular fictitious play converges. This is true for any such subsequence, and since the Nash equilibrium is unique by assumption, every subsequence converges to the same point. It follows that the entire sequence of beliefs converges as well.

Section 6 discusses some issues that arise from our analysis. All the proofs are relegated to the Appendix.

2. THE MODEL OF REPEATED PLAY

Two players repeatedly play a normal-form game $G = (A_1 \times A_2, (\pi_1, \pi_2))$, where A_i is the *finite* set of player i's actions, and $\pi_i: A_1 \times A_2 \to \mathbf{R}$ is player i's payoff function. Let $A = A_1 \times A_2$ be the set of action profiles. We assume that each player knows the opponent's action set. Let M_i be the set of probability distributions over A_i and $M = M_1 \times M_2$. As usual, we identify M_i with the set of player i's mixed actions m_i and extend the payoff functions π_i over M. With slight abuse of notation, we use $a_i \in A_i$ to denote the degenerate mixed action that places probability one on the pure action a_i. Player j's *belief* λ_j is an element of M_i ($j \neq i$). The interpretation is that player j's belief represents his point estimate about player i's next mixed action.

[9] See Section 5 for a more accurate description.

Let $H^t = A^t$ be the set of t-length histories ($t = 0, 1, ..., \infty$) and $H = \bigcup_{t \in \mathbf{N}} H^t$ ($\mathbf{N} = \{1, 2, ...\}$) be the set of (non-null) finite histories. We identify H^0 with the singleton set $\{\phi\}$. For $h \in H^\infty$, let $h^{t, t'} = (a^t, ..., a^{t'})$ be the $(t' - t + 1)$-length partial history of h, and let $h^t = h^{1, t}$. Also, given $h, h' \in H$, let $hh' \in H$ be the history obtained by appending h' after h.

Player i's *assessment rule* is a mapping $z_i : H^0 \cup H \to M_j$. Namely, it is a function used by player i in determining his belief after each history of play. Also, player i's *behavioral rule* is a mapping $x_i : H^0 \cup H \to M_i$. It specifies which (mixed) action to take after each history. Throughout, we assume that each player's behavioral rule x_i is *myopic* in the sense that $x_i(h) \in B_i(z_i(h))$ for each $h \in H^0 \cup H$, where $B_i : M_j \to M_i$ is the best response correspondence defined by $B_i(m_j) = \arg \max_{m'_i \in M_i} \pi_i(m'_i, m_j)$.

Although assessment rules are deterministic, there is generally a stochastic element in the model since behavioral rules may entail mixed actions when players are indifferent between two or more pure actions. For this reason, we need to consider the probability space (Ω, P). Each $\omega \in \Omega$ corresponds to a sequence of particular realizations of those mixed actions, and hence corresponds to a particular realization of an infinite history. The probability distribution P over Ω is induced by the assessment rules (z_1, z_2) and the behavioral rules (x_1, x_2). In this sense, each history is a random variable, and so is player i's belief after period t, $\lambda_i^t = z_i(h^t) \in M_j$.[10]

3. POINTS OF CONVERGENCE, NASH EQUILIBRIUM, AND STABILITY

There can be at least three notions of convergence in the present framework: convergence in beliefs, in mixed actions, and in empirical frequencies. Given the myopia of players, convergence in mixed actions (to a mixed Nash equilibrium) is a much stronger notion than that in beliefs as the former requires the latter. In matching pennies, for example, players will not mix between H and T unless their beliefs attach the same weight to H and T by the opponents. Our discussion focuses mainly on convergence in beliefs. We will discuss its connection to convergence in empirical frequencies under some circumstances.

As a first step, we relate the Nash equilibrium to a point of convergence in beliefs. We begin with the property on assessment rules which stipulates that when a player does not observe some pure actions of his opponent for a long period of time, then he begins to think that those pure actions are not to be played at all.

---

[10] For the same reason, any variable with the superscript t should be regarded as a random variable in what follows.

\	L	R
U	1, 1	0, 0
D	0, 0	1, 1

<div align="center">FIGURE 3</div>

DEFINITION. An assessment rule z_i of player i is *adaptive* if the following is true: For any $K_j \subset A_j$ and $h = (a^1, a^2, \ldots) \in H^\infty$, if there exists a time $T_h < \infty$ such that $a_j^t \in K_j$ for all $t \geqslant T_h$, then $\lim_{t \to \infty} \lambda_i^t(K_j) = 1$.

The definition follows that of Milgrom and Roberts [8].[11] Note that it is a path-by-path property. Proposition 1 states that if both players' assessment rules are adaptive, then any point of convergence in beliefs corresponds to a Nash equilibrium of the game G.

PROPOSITION 1. *Suppose the assessment rules z_1 and z_2 are adaptive. If, for a (mixed) action profile $m \in M$, $\lim_{t \to \infty}(\lambda_2^t, \lambda_1^t) = m$ along some sample path, then m is a Nash equilibrium of G.*

As Milgrom and Roberts [8] and Fudenberg and Kreps [3] note, the result is rather straightforward when the limit is a pure action profile. The proposition guarantees its validity for a mixed action profile in two-person games.[12]

It is not difficult to see that the converse of Proposition 1 is not true. Namely, a sequence of beliefs may converge to a Nash equilibrium even when players attach positive weight on some unobserved actions. For example, in the game of Fig. 3, suppose $(\lambda_2^t, \lambda_1^t) = ((0.5 + t^{-1}) L + (0.5 - t^{-1}) R, (0.5 + t^{-1}) U + (0.5 - t^{-1}) D)$ and $(a_1^t, a_2^t) = (U, L)$ for $t \in \mathbf{N}$. Although adaptiveness is violated, the limit profile $m = (0.5U + 0.5D, 0.5L + 0.5R)$ is a mixed Nash equilibrium of G.

Our analysis in Section 5 focuses on the *stability* of a Nash equilibrium as defined below. Stability is a stronger concept of convergence than that used in Proposition 1 since it requires convergence with probability one as opposed to along some path.

DEFINITION. Given the assessment rules z_i and the behavioral rules x_i ($i = 1, 2$), a (mixed) action profile $m \in M$ is *stable* if $P(\lim_{t \to \infty}(\lambda_2^t, \lambda_1^t) = m) = 1$.

---

[11] The definition of "adaptiveness" by Milgrom and Roberts [8] is in fact the combination of our definition and the myopia of the players.

[12] Fudenberg and Kreps [3, Proposition 4.2] prove that the limit is a (mixed) Nash equilibrium when assessment rules are "asymptotically empirical" instead of "adaptive." Our statement is stronger since any asymptotically empirical rule is adaptive but not vice versa.

By Proposition 1, every stable point (if any) is a Nash equilibrium of the game G when the assessment rules z_1 and z_2 are both adaptive.

4. ASSESSMENT RULES WITH PATTERN RECOGNITION

We now formulate assessment rules that incorporate the idea of pattern recognition as motivated in the introduction.

To begin with, a *pattern* is defined to be any element of the set of finite histories, H. Each player is interested in the patterns that appear in the "tail" of the current history: Suppose the current history is given by $h \in H^t$ ($t \in \mathbf{N}$). For every positive integer $n \leqslant t$, players look at the pattern w of length n which coincides with the tail $h^{t-n+1,\,t}$ of h. We say that the pattern w is *recognized* by player i at h if the number of times w appears in the history h exceeds player i's threshold level.

To formalize the ideas, we first need to specify the way players count the number of appearances of some pattern $w \in H^n$ in $h \in H^t$. This number is given by

$$R(h, w) = \sum_{\tau = n}^{t} \chi_w(h^{\tau - n + 1,\,\tau}),$$

where the function $\chi_w : H^n \to \{0, 1\}$ is defined by $\chi_w(h) = 1$ if $h = w$ and $\chi_w(h) = 0$ otherwise. We denote the number of appearances of the n-length "tail pattern" $w = h^{t-n+1,\,t}$ by $r(h, n)$ ($= R(h, w)$). For instance, $r(h^{2t}, 1) = t$, $r(h^{2t}, 2) = t$ and so on if $h^{2t} = (a, b, a, b, ..., a, b, a, b)$ for some $a, b \in A$. Also, if $h^{2t} = (a, a, ..., a)$ for some $a \in A$, then $r(h^{2t}, 2)$ is $2t - 1$ rather than t. This is appropriate if the opponent bases his action on bounded memory of length two, since then his action after one (a, a) will be exactly the same as after the overlapping (a, a).[13]

The threshold levels of player i are given by $k_i(n) \in \{2, 3, ..., \infty\}$. Given those numbers, we say that player i *recognizes* a pattern at history h if

$$r(h, n) \geqslant k_i(n). \tag{1}$$

The interpretation is that the pattern w appears sufficiently many times in the history so that it begins to draw attention of player i.[14] For simplicity,

---

[13] See Proposition 2.

[14] Hence, when $k_i(n) = \infty$ for some $n \in \mathbf{N}$, player i does not recognize any patterns of length n.

we assume that the threshold level depends only on the length of patterns and not on the patterns themselves.

In general, there may be more than one n that satisfies (1) at history h. In those circumstances, player i selects the number $N_i(h)$ such that

$$N_i(h) \in \{n \in \mathbf{N} : r(h, n) \geqslant k_i(n)\} \tag{2}$$

to decide on the particular pattern that is actually *used*. Namely, player i uses $w = h^{t - N_i(h) + 1, t}$ when forming his belief at history h as described below. We generally do not impose any restriction on the choice made in (2). One possibility is that player 1 takes the *maximum* number. See Example 1 below for this point. If there exists no $n \in \mathbf{N}$ for which (1) is satisfied, then $N_i(h) = 0$.

Whenever the history $h \in H^t$ $(t \geqslant 1)$ is such that $N_i(h) > 0$, player i faces two modes of empirical frequencies: One is the overall empirical frequency $f_i^t \in M_j$ defined by

$$f_i^t(a_j) = \frac{\sum_{a_i \in A_i} R(h, a)}{t} \quad (a_j \in A_j),$$

and the other is the empirical frequency *conditional on the pattern* $g_i^t \in M_j$ defined by

$$g_i^t(a_j) = \frac{\sum_{a_i \in A_i} R(h, wa)}{r(h, n)} \quad (a_j \in A_j),$$

where $a = (a_i, a_j)$, $w = h^{t - n + 1, t}$, and $n = N_i(h)$. For convenience, define $g_i^t = f_i^t$ when $N_i(h) = 0$. In other words, the overall empirical frequency is the usual empirical frequency as used in fictitious play, while the empirical frequency conditional on the pattern represents the proportion of the opponent's pure actions observed immediately after the occurrences of the pattern in question.

EXAMPLE 1. Let $a = (a_1, a_2)$, $b = (b_1, b_2)$, $c = (c_1, c_2) \in A$. Consider the history

period	1	2	3	$\cdots$	$3t-5$	$3t-4$	$3t-3$	$3t-2$	$3t-1$	$3t$	$3t+1$
observation	a	b	c	$\cdots$	a	b	c	a	b	c	?

If $k_i(n) < \infty$ for some $n \in \mathbf{N}$, the empirical frequency conditional on the pattern, g_i^{3t}, indicates that a_j is far more likely in period $3t + 1$ than b_j or c_j. This is independent of the choice made in (2): $N_i(h^{3t}) = 1, 2, \ldots$ will all give rise to the same conditional empirical frequency.

Suppose instead that the history is given by

period	1	2	3	$\cdots$	$3t-5$	$3t-4$	$3t-3$	$3t-2$	$3t-1$	$3t$	$3t+1$
observation	b	a	b	$\cdots$	b	a	b	b	a	b	?

Suppose $k_i(1), k_i(2) < \infty$. If player i uses patterns of length 1 rather than those of length 2 in (2), he may still think neither a_j nor b_j is very unlikely in period $3t+1$. This is because the empirical frequency conditional on just b attaches the same weight on a_j and b_j (and when compared with the overall frequency, it might even suggest b_j is unlikely). On the other hand, if he uses patterns of length 2, he will find that a_j is extremely unlikely as it has never been played after (a, b). The same is true if player i refers to longer patterns. This example shows that choosing the shortest pattern in (2) sometimes leads to a very unintuitive prediction.

Players take a convex combination of f_i^t and g_i^t in order to determine their beliefs at history $h \in H^t$

$$\lambda_i^t = \beta_i(h) \, g_i^t + (1 - \beta_i(h)) f_i^t, \tag{3}$$

where $\beta_i \in [0, 1]$ is the weight which generally depends on the history h.

It is easy to verify that any assessment rule satisfying (3) is adaptive. We provide below some general properties of assessment rules described by (3). It is, however, very difficult to draw conclusions about the stability of a Nash equilibrium for general weight functions β_i. For this reason, we need to restrict the functional forms of β_i in our analysis of the stability of mixed equilibria in Section 5. Note that if $\beta_i(h) = 0$ for all h, then (3) just reduces to regular fictitious play, while if $\beta_i(h) = 1$ for all h, (3) implies *conditional fictitious play* in which players only look at the empirical frequency conditional on the pattern whenever possible. We will analyze this rule in Section 5. In the rest of this section, we present some properties of the learning rule (3) which indicate that it actually recognizes patterns.

We first show that when player i's assessment rule satisfies (3) (together with some other conditions), he will *asymptotically detect* the opponent's action choice based on finite memory. To be more precise, given $s \in \{0, 1, \ldots\}$, player j's behavioral rule $x_j: H^0 \cup H \to M_j$ has *s-bounded recall* if $x_j(h') = x_j(hh')$ for each $h' \in H^s$ and $h \in H^0 \cup H$. Namely, a (mixed) action specified in each period depends only on the on the history during the past s periods.[15] A behavioral rule has *bounded recall* if it has s-bounded recall for some $s \in \{0, 1, \ldots\}$.

---

[15] Note that our definition of bounded recall is more general than the usual definition: s-bounded recall is usually defined to be memory consisting only of the *opponent's* actions in the past s periods whereas according to the definition used here, it may include one's own past actions. See, for example, Lehrer [7].

DEFINITION. Given $\varepsilon \geqslant 0$, player j's behavioral rule $x_j: H^0 \cup H \to M_j$ is *ε-asymptotically detected* by player i if $P(\limsup_{t \to \infty} \|\lambda_i^t - x_j(h^t)\| \leqslant \varepsilon) = 1$. In particular, it is *asymptotically detected* if it is 0-asymptotically detected.

Note that the definition does not require that player i asymptotically (ε)-correctly forecast each contingent action specified by x_j. Rather it requires that player i asymptotically detect player j's behavioral rule *along the path of play*.

The following proposition states that using the assessment rule as defined in (3), player i ε-asymptotically detects any m-bounded recall behavioral rule of player j, provided that the following conditions are satisfied: (a) player i actually uses patterns of length m or longer whenever possible (Conditions (i) and (ii) in Proposition 2); (b) the threshold number for iterations $k_i(s)$ grows large when the pattern length s becomes large (Condition (iii)); (c) beliefs λ_i^t are always set within ε of the conditional empirical frequency g_i^t (Condition (iv)). Condition (b) above ensures that player i uses samples of ever increasing size over time to calculate the conditional empirical frequency g_i^t. This is necessary for player i to ignore any spurious patterns in the long-run. As a special case of (b), it is possible that player i does not use any pattern of length n' or greater for some $n' > m$, i.e., $k_i(s) = \infty$ for every $s \geqslant n'$.

PROPOSITION 2. *Let $m \in \mathbf{N}$ be an arbitrary number. Suppose player i's assessment rule satisfies* (3) *and*

   (i) $k_i(m) < \infty$;

   (ii) $N_i(h) \geqslant m$ *whenever* $r(h; m) \geqslant k_i(m)$ *in* (2);

   (iii) $\lim_{s \to \infty} k_i(s) = \infty$;

   (iv) $\beta_i^t = 1$ *if* $d_i^t = \|g_i^t - f_i^t\| > \varepsilon$ *for some* $\varepsilon \geqslant 0$.

Then any m-bounded recall behavioral rule x_j of player j is ε-asymptotically detected by player i. If, in addition, $k_i(s) < \infty$ for each $s \in \mathbf{N}$, any bounded recall behavioral rule of player j is ε-asymptotically detected by player i.

As a special case of the definition of asymptotic detection, we say that player i's assessment rule z_i is *consistent* if it asymptotically detects any (pure or mixed) 0-bounded recall behavioral rule of player j. The following is a Corollary to Proposition 2 which identifies the condition under which the assessment rule is consistent. In comparison with Proposition 2, Conditions (i) and (ii) are no longer necessary since no matter what conditional or overall frequency player i may use, he will detect x_j since it is now history-independent. Condition (iv) is also unnecessary: under Condition (iii), player i uses samples of ever increasing size over time to calculate

either the conditional frequency g_i^t or the overall frequency f_i^t. The frequencies g_i^t and f_i^t will hence be almost surely the same in the limit when x_j is history independent.

COROLLARY 2.1.   *If* $\lim_{s \to \infty} k_i(s) = \infty$, *then player i's assessment rule z_i satisfying* (3) *is consistent.*

On the other hand, when $k_i(s)$'s are uniformly bounded by some constant (i.e., when condition (iii) is not satisfied), consistency may be violated as seen in the following example.

EXAMPLE 2.   Suppose player j plays the mixed action m_j every period independently of history. Suppose $k_i(s)$ is equal to, say, 53 for all $s \in \mathbf{N}$ and $\beta_i(h) = 1$ for all $h \in H$. Suppose also that the selection criterion in (2) is to choose the maximum number always. Consistency requires that for any $\delta > 0$, there be a random time T after which player i's belief stays in the δ-neighborhood of m_j. Suppose $m_j(a_j) = 1/2$ for some $a_j \in A_j$. No matter how large T might be, there is a pattern longer than T that player i recognizes after T. When he recognizes such a pattern for the first time, his belief λ_i^t is such that $\lambda_i^t(a_j) = l/53$ for some $l = 0, 1, ..., 53$ for any $a_j \in A_j$, and hence it cannot be in the δ-neighborhood of m_j if δ is small enough. Player i hence catches spurious patterns even though the opponent's behavioral rule is history-independent.

5. STABILITY OF NASH EQUILIBRIUM

We begin with the stability property of a strict Nash equilibrium. When the game G has a strict Nash equilibrium $a \in A$, it is stable under (3) given that the initial beliefs are in the small neighborhood of a. The result is straightforward and patterns emerging in this case are rather trivial as imagined: they consist exclusively of a.

PROPOSITION 3.   *Suppose the assessment rules z_1 and z_2 satisfy* (3). *Any strict Nash equilibrium $a \in A$ of the game G is stable provided that $(z_2(h^0), z_1(h^0))$ is close enough to a.*

The stability of a mixed Nash equilibrium is far more difficult to analyze. In fact, it is no longer possible to proceed with a general weight function β_i. We will below postulate two simple forms for β_i and discuss their implications. In particular, we will look at the case where players make a zero–one choice between the two frequencies.

5.1. *Discontinuous Adjustment*

We first study the weight function β_i which depends on the distance $d_i^t = \|g_i^t - f_i^t\|$ between the overall and conditional empirical frequencies as follows: take the overall frequency if the distance is not so large, but take the conditional frequency otherwise. Formally,

$$\beta_i^t = 0 \quad \text{if} \quad d_i^t \leqslant \kappa_i \quad \text{and} \quad \beta_i^t = 1 \quad \text{if} \quad d_i^t > \kappa_i, \qquad (4)$$

where κ_i is a positive constant. Therefore, the assessment rule is discontinuous at $d_i^t = \kappa_i$. One interpretation of such an assessment rule is that players are rather cautious in using the conditional empirical frequency when it is in the small neighborhood of the overall frequency. Since such assessment rules always set beliefs within κ_i of the conditional empirical frequency as required in Proposition 2, they will κ_i-asymptotically detect the oppenent's m-bounded recall behavioral rule if (a) player i actually uses patterns of length m or longer whenever possible, and (b) player i uses sufficiently many observations to calculate the conditional empirical frequency g_i^t after sufficiently long patterns (i.e., Conditions (i)–(iii) of Proposition 2).

When the players' assessment rules are given by (3) with the weight function as specified in (4), we find that a mixed Nash equilibrium in a wide class of games is unstable, provided that $\kappa_i > 0$ is small enough.[16] Let $c_i(m_i) = \{a_i \in A_i : m_i(a_i) > 0\}$ denote the support of the mixed action $m_i \in M_i$.

PROPOSITION 4. *Suppose the assessment rules z_i ($i = 1, 2$) are discontinuous at κ_i (i.e., they satisfy (3) when the weight function is as in (4)), they recognize one length patterns (i.e., $k_i(1) < \infty$), and there exists an upperbound on the length of recognizable patterns (i.e., $k_i(s) = \infty$ for every $s \geqslant n$ for some $n \in \mathbf{N}$). When $\kappa_i > 0$ ($i = 1, 2$) are sufficiently small, no mixed Nash equilibrium $m \in M$ of the game G is stable provided that for either $i = 1$ or 2 and $j \neq i$,*

(i) *for every $a \in c_1(m_1) \times c_2(m_2)$, there exists $a_i' \in c_i(m_i)$ such that $\pi_i(a_i', a_j) < \pi_i(a)$, or*

(ii) *there exists $a_i \in c_i(m_i)$ such that for every $a_j \in c_j(m_j)$, there exists $a_i' \in c_i(m_i)$ such that $\pi_i(a_i', a_j) < \pi_i(a)$.[17]*

---

[16] Note that the rule just reduces to fictitious play when κ_i is sufficiently large.

[17] Cases (i) and (ii) do not exhaust all the possibilities of instability. Condition (i) states that for each pure action profile in the support of m, either player 1 or 2 has an inferior response within the support. Condition (ii) states that either player 1 or 2 has a pure action in the support that is not the worst response to the opponent's any pure action in the support.

\	Rock	Paper	Scissors
Rock	0, 0	−1, 1	1, −1
Paper	1, −1	0, 0	−1, 1
Scissors	−1, 1	1, −1	0, 0

FIGURE 4

The intuition behind this result is as follows: Condition (i) or (ii) in the proposition implies that action a_i' is never played after $a = (a_i, a_j)$ as long as the players use the overall empirical frequency for assessment. Therefore, it is impossible for the players to use only the overall frequency for a long time since such a systematic movement will make the discrepancy between the two frequencies f_j^t and g_j^t significant. However, if player j switches from f_j^t to g_j^t when the discrepancy is large, his belief will "jump" by a significant amount. The only possibility then is that the players exclusively use the conditional frequencies after some time on. However, if their beliefs converge in this case, then it is shown that the overall empirical frequency must also converge to the same point. Thus, the players must return to the overall frequency, which is a contradiction.

When the game G is 2×2 with no pure Nash equilibrium, Condition (i) of Proposition 4 is satisfied. The game of "Rock–Paper–Scissors" in Fig. 4 is another example which satisfies Condition (i) of the proposition. (In its unique Nash equilibrium, each player plays each pure action with probability $1/3$.)

5.2. Conditional Fictitious Play

We next consider the weight function β_i constantly equal to one:

$$\beta_i(h) = 1 \qquad \text{for all} \quad h \in H. \tag{5}$$

As previously noted, (5) implies that player i uses the conditional empirical frequency whenever possible. Unlike in the case of discontinuous adjustment, conditional fictitious play as defined in (5) implies the stability of any *unique* equilibrium in any *zero-sum* games. In order to obtain the stability result, however, we further need a condition which guarantees that the level of pattern recognition by the two players is asymptotically the same. More specifically, we assume that the two assessment rules are *asymptotically symmetric* in the following sense:

$$P(\lim_{t \to \infty} N_1^t = \lim_{t \to \infty} N_2^t) = 1. \tag{6}$$

The condition (6) requires two things: (a) the length of patterns used by player i converges to some limit ($i = 1, 2$) along almost every sample path,[18] and (b) those limits are the same for the two players with probability one. With the upper bound of the length of recognizable patterns, an alternative interpretation of (6) is that the set of patterns used infinitely often is the same for the two players, and consists of the patterns of the same length along almost every sample path. This condition will be satisfied if, for example, the maximum length of recognizable patterns is the same for both players and they asymptotically choose the longest patterns in (2). In general, Proposition 5 is true for any games for which the following statement is true about regular fictitious play:

Fictitious play with any initial weight converges to a Nash equilibrium.

$$(\#)$$

Here "initial weight" refers to the hypothetical empirical frequencies that the players have prior to the game. For example, if the initial weight is 2/3 for H and 1/3 for T in matching pennies, the belief is updated to (3/4, 1/4) when the first "real" observation is H. Gale [4] demonstrates that ($\#$) is in fact true for zero-sum games.[19] We hence obtain the following proposition.

PROPOSITION 5. *Suppose that the assessment rules z_1 and z_2 are asymptotically symmetric* (6), *that they always use the conditional frequency* (*i.e., they satisfy* (3) *with the weight function as in* (5)), *and that there exists an upper-bound on the length of recognizable patterns* (*i.e., $k_i(s) = \infty$ for each $s \geqslant n$ for some $n \in \mathbf{N}$*). *If the game G is zero-sum with a unique* (*pure or mixed*) *Nash equilibrium $m \in M$, then m is stable.*

The intuition behind this result is as follows: Given the upper bound on the length of patterns that players can recognize, consider the (random) set W of patterns that are used infinitely often, and the random time T such that the players only use patterns in W after T. By (6), W is common for the two players, and it consists of patterns of the same length along (almost) every sample path. This implies that the set of periods after T can be partitioned into subsequences depending on the patterns used by the players. Take any pattern w in W and consider the subsequence of periods

---

[18] Note that the limit does not have to be constant across sample paths.

[19] See Lemma 7.4 on p. 253, which proves the convergence of fictitious play with arbitrary initial weight in the symmetrization of any zero-sum game. The correspondence between fictitious play with positive initial weight in a zero-sum game and that in its symmetrization can be established along the line of the discussion on p. 251–252. Although we suspect that ($\#$) is valid for other classes of games in which fictitious play with zero initial weight converges, we have not confirmed this.

in which w is used. In this subsequence, players set their beliefs equal to the empirical frequency conditional on w. Since the action profiles following w may be different from those generated by fictitious play before time T, the beliefs in the subsequence may still be different from those in fictitious play with zero initial weight even after T. It is clear, however, that they instead correspond to the beliefs generated by fictitious play with some positive initial weight. The convergence of this subsequence of beliefs is hence guaranteed by ($\#$). Since the Nash equilibrium is unique, every subsequence converges to the same point, ensuring the convergence of the entire sequence.

To understand one role of (6), consider the following example. Let a, b, $c \in A$ and suppose that player i asymptotically uses pattern ac when the two-length tail of the current history is ac, but uses pattern c when it is bc. This violates (6) since N_i^t does not converge, and hence stability is no longer guaranteed (at least) by the argument above: while the beliefs after bc depend on the empirical frequencies conditional on patterns ac and bc since they both end with c, the action profiles after ac are determined quite independently of those after bc. Hence the beliefs after bc will behave very differently from those in any fictitious play.

The following corollary concerns the empirical frequency of the process. It readily follows from Proposition 5 and Proposition A1 in the Appendix which states that, when there exists an upper-bound on the length of patterns that player i can recognize, the overall frequency f_i^t converges to m_j if the conditional frequency g_i^t converges to m_j.

COROLLARY 5.1. *Suppose the assessment rules z_1 and z_2 satisfy the conditions of Proposition 5. If the game G is zero-sum with a unique Nash equilibrium $m \in M$, then the marginal empirical frequencies converge to the Nash equilibrium, i.e., $P(\lim_{t \to \infty} g_i^t = m_j) = P(\lim_{t \to \infty} f_i^t = m_j) = 1$ ($i = 1, 2$, $j \neq i$).*

6. DISCUSSIONS

The paper raises a few open questions.

First, we continue our discussion on the stability of a mixed equilibrium under conditional fictitious play from the last section. We provide some examples below which indicate that convergence is unlikely when the two players asymptotically use patterns of different length (i.e., (6) is violated).

EXAMPLE 3.   Consider the game of matching pennies in Fig. 1. Suppose player 1's assessment rule z_1 is characterized by

(i) $k_1(1) < \infty$, $k_1(s) = \infty$ for $s \geqslant 2$,

(ii) $\beta_1(h) = 1$ for all $h \in H^0 \cup H$,

and player 2's assessment rule is characterized by

(iii) $k_2(s) = \infty$ for $s \geqslant 1$.

Namely, player 1 uses conditional fictitious play based only on one-length patterns, while player 2 recognizes no patterns and follows regular fictitious play. Note that Condition (6) is violated. We will show that player 1's belief never converges to the Nash equilibrium value $0.5H + 0.5T$.

Note first that player 2 always plays H (resp. T) after the combination (T, H) (resp. (H, T)). (This is based on the discussion in the Introduction.) When player 1 recognizes one-length patterns, hence, his belief will be such that

$$\lambda_1^t(H) = \begin{cases} 0 & \text{if} \quad a^t = (H, T); \\ 1 & \text{if} \quad a^t = (T, H). \end{cases}$$

It follows that 1's belief does not converge when the combination (H, T) or (T, H) appears infinitely often. If it is to converge, hence, the history must consist only of (H, H) and (T, T) after some time on.

Recalling again the discussion in the introduction, we next note that player 1 always plays H (resp. T) after the two-length pattern $((H, H), (H, H))$ (resp. $((T, T), (T, T))$). This implies that (H, H) and (T, T) must alternate every other period (asymptotically) if they are the only combinations to appear in the history after some time on as required in the previous step. In this case, however, player 1's beliefs cannot converge as they will be asymptotically approximated by

$$\lambda_1^t(H) = \begin{cases} 0 & \text{if} \quad a^t = (H, H); \\ 1 & \text{if} \quad a^t = (T, T). \end{cases}$$

We have therefore excluded the possibility that 1's beliefs converge.

Although Example 3 deals with a rather special case where player 2 recognizes no patterns, we expect that 1's beliefs do not converge in matching pennies whenever he asymptotically uses longer patterns than 2. Figure 5 presents one such evidence. It describes the result of a simulation of conditional fictitious play in matching pennies when player 1 recognizes one- and two-length patterns, whereas player 2 recognizes only one-length patterns.[20] Figure 5a depicts the movement of beliefs in periods 1 through

---

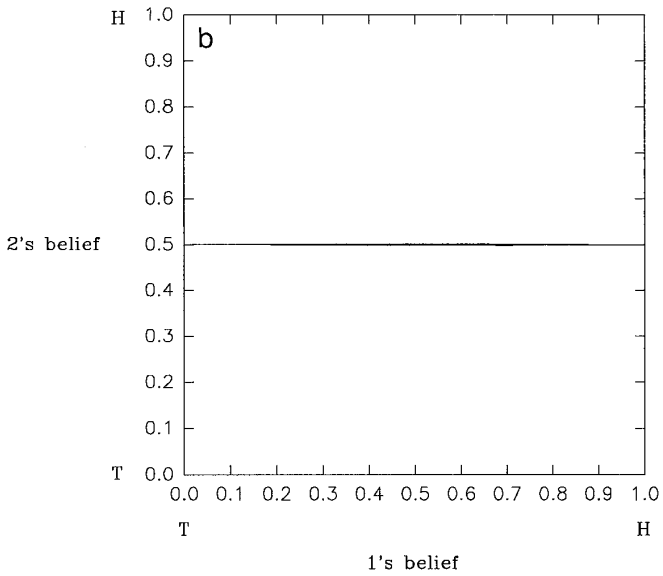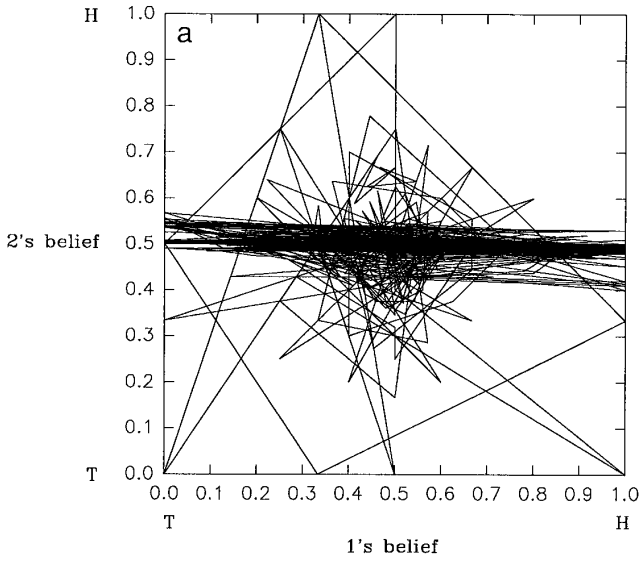[20] Player 1 (resp. 2) plays H (resp. T) in case of ties.

FIGURE 5

10,000 whereas Fig. 5b extracts periods 9,001 through 10,000. As seen, while player 2's beliefs are concentrated around 0.5 in the last 1000 periods, player 1's beliefs still oscillate between 0 and 1 in those periods. These observations lead us to the following conjecture:

When the two players asymptotically use patterns of different length (i.e., (6) is violated) a mixed Nash equilibrium is unstable in a wide class of games including matching pennies. In particular, when player 1 asymptotically uses longer patterns than player 2, although player 2's beliefs converge to a Nash equilibrium value, player 1's beliefs diverge.

The second open question concerns the discontinuity in the stability property of a mixed equilibrium between two different formulations: In discontinuous adjustment, a mixed Nash equilibrium is unstable under the conditions of Proposition 4 for any small $\kappa_i > 0$ in (4). When $\kappa_i = 0$, on the other hand, we have conditional fictitious play under which a unique mixed Nash equilibrium is stable under the conditions of Proposition 5. For example, take the game of matching pennies. Since this game satisfies the conditions of both Propositions 4 and 5, its Nash equilibrium is unstable for any $\kappa_i > 0$ but stable for $\kappa_i = 0$ ($i = 1, 2$). A possible resolution of this discontinuity may be to let $\kappa_i > 0$ depend on the length of history. If κ_i approaches zero as $t \to \infty$, we may obtain convergence even under discontinuous adjustment since the size of jumps of beliefs may be smaller in the distant future.[21]

Another important question concerns the behavior of beliefs when there is no upper bound on the length of patterns that a player can recognize. Suppose players use conditional fictitious play. Example 2 shows that conditional fictitious play does not generally satisfy consistency when $k_i(s)$'s are uniformly bounded by some constant. This suggests that it is rather hopeless to obtain the stability of a mixed Nash equilibrium when there is such a bound. Therefore, we need to have $k_i(s)$ grow indefinitely as $s \to \infty$ for convergence to take place. It is not clear, however, whether this condition suffices or not. The main difficulty in proving convergence is to locate the belief when a new pattern begins to be used. Namely, whenever a new pattern of length s begins to be used, the corresponding belief is set equal to one of the points on the lattice whose coarseness is given by $1/(k_i(s) - 1)$. If this point is always the closest one to m_j, then the belief will also get sufficiently close to m_j as $k_i(s)$ grows indefinitely. However, we do not know if this is the case. We provide below a rather weak result which states that given any small neighborhood of a Nash equilibrium, if $k_i(s)$'s grow rapidly enough as $s \to \infty$, beliefs fall in the neighborhood infinitely often.

---

[21] However, we have not been able to prove this.

PROPOSITION 6. *Suppose that the assessment rules z_1 and z_2 are asymptotically symmetric* (6), *that they always use the conditional frequency* (*i.e., they satisfy* (3) *with the weight function as in* (5)), *and that the behavioral rules x_1 and x_2 are both pure. Suppose further that the game G is zero-sum with a unique* (*pure or mixed*) *Nash equilibrium $m \in M$. Then for any $\varepsilon > 0$, there exists a function $q: \mathbf{N} \to \mathbf{N}$ such that if $k_1(s) = k_2(s) \geqslant q(s)$ for all $s \in \mathbf{N}$, then $P(\liminf_{t \to \infty} \|\lambda_j^t - m_j\| < \varepsilon) = 1$.*

The logic behind this result is as follows: Since we would have convergence if $k_i(s) = \infty$ for all $s \geqslant n$, beliefs will come into the ε-neighborhood of m if we take the time until the patterns of length $n + 1$ begin to be used long enough. This can be done by taking $k_i(n+1)$ large enough. This is true for any $n \in \mathbf{N}$, hence follows the result. We need the players' behavioral rules to be pure to have a deterministic function q which sets the lower-bound on the rate of increase for the parameters k_i's. It may seem contradictory at first that while convergence is obtained without any condition on the rate of increase of k_i in Proposition 2 (and Corollary 2.1), only something weaker is established in Proposition 6 even with such a condition. The difference, however, derives from the fact that while the opponent's actions are stationary (conditional on history of some length) in Proposition 2, they are not in Proposition 6.

Finally, as mentioned in the Introduction, one problematic feature of fictitious play is that even when the marginal frequency of each player's actions converges to a Nash equilibrium, the joint frequency of action profiles may converge to something else.[22] Learning with pattern recognition is not free from such a problem. To see this, consider the pure coordination game in Fig. 3. Suppose $\lambda_1^0(L) > 1/2$ and $\lambda_2^0(D) > 1/2$. Assume also that the players' assessment rules are *symmetric* in the sense that the two players always use patterns of the same length, and suppose further that their behavioral rules specify the same action as in the previous period in case of ties. It is not difficult to see that when players do not recognize patterns, the overall joint frequency converges to $0.5(U, R) + 0.5(D, L)$. Even when they recognize patterns, we can check that they cannot get out of this correlation. Hence, even if the overall marginal frequency converges to $1/2$ for each player, the overall joint frequency is always a convex combination of (U, R) and (D, L), which does not correspond to the Nash equilibrium profile. Note, however, that players in our model take correlation into account while in fictitious play they do not. The realized history exhibits correlation only because the two players happen to make adjustments based on patterns with the same timing. The problem is hence in part attributed to the symmetry of the assessment rules.[23] Therefore, this

---

[22] See, for example, Fudenberg and Kreps [3] and Jordan [5].

[23] The example is also not robust against slight disturbance in the payoffs.

type of problem would be generally avoided if we require the assessment rules to be only asymptotically symmetric (6), although in which case this particular mixed equilibrium is unlikely to be stable.

APPENDIX

Proof of Proposition 1. Fix $\omega \in \Omega$ such that $\lim_{t \to \infty} (\lambda_2^t(\omega), \lambda_1^t(\omega)) = m$ and let $C = \limsup_{t \to \infty} c_i(B_i(\lambda_i^t(\omega)))$, where for $Q \subset M_i$, $c_i(Q) = \bigcap_{m_i \in Q} \{a_i \in A_i : m_i(a_i) > 0\}$ is the support of Q. Note that $C \subset A_i$ is the (nonempty) set of i's pure actions that are in the support of i's mixed actions infinitely often along the path. As player j only observes i's pure actions in the set C from some time on, $\lim_{t \to \infty} \lambda_j^t(C) = 1$ by adaptiveness. Since λ_j^t converges to m_i, it must be the case that $m_i(C) = 1$.

We next show that $C \subset c_i(B_i(m_j))$. Take $a_i \in C$. By definition of C, we can take a subsequence of $\langle \lambda_i^t \rangle$ such that $a_i \in c_i(B_i(\lambda_i^{t_n}))$ and hence $a_i \in B_i(\lambda_i^{t_n})$ for $n \in \mathbf{N}$. Since the best response correspondence B_i is upper hemi-continuous, it follows that $a_i \in B_i(m_j)$, or $a_i \in c_i(B_i(m_j))$. Thus, $C \subset c_i(B_i(m_j))$ and hence $m_i(C) \leqslant m_i[c_i(B_i(m_j))]$. Since $m_i(C) = 1$ from the previous step, it follows that $m_i[c_i(B_i(m_j))] = 1$. Since the set $B_i(m_j)$ is convex, we conclude that $m_i \in B(m_j)$. ∎

Proof of Proposition 2. Note by (iii) that for each $l \in \mathbf{N}$, there exists a random time $T_l \in \mathbf{N}$ such that for every $t \geqslant T_l$, the conditional frequency g_i^t at history h^t is based on at least l observations. (This is obviously true if $k_i(s) = \infty$ for all $s > n'$ for some n' since in that case all the patterns used are length n' or smaller and there are only finitely many of them. On the other hand, when $k_i(s) < \infty$ for each $s \in \mathbf{N}$, new patterns may be used after an arbitrarily long period of time. But since there exist only finitely many patterns of any given length or less, the patterns that are used for the first time after a very long time must be very long. Since $k_i(s) \to \infty$ as $s \to \infty$ by (iii), the empirical frequencies conditional on long patterns are based on a large number of observations.)

Let $W \subset H^m$ be the (random) set of m-length histories that appear infinitely often. Take any m-length history $c \in H^m$. When $c \in W$, let $t_n \in \mathbf{N}$ be the (random) time of the nth occurrence of $c \in C$, i.e., $h^{t_n - m + 1, \, t_n} = c$ ($n \in \mathbf{N}$). Since $N_i^{t_n} \geqslant m$ when $n \geqslant k_i(m)$ by (ii), the conditional empirical frequencies $g_i^{t_n}$ at h^{t_n} ($n \geqslant k_i(m)$) always reflect the realizations of i.i.d. random draws from $x_j(c) \in M_j$ even if those empirical frequencies may be conditioned on different patterns. It then follows from the observation in the preceding paragraph that for any $l \in \mathbf{N}$, if $n \in \mathbf{N}$ is such that $t_n \geqslant T_l$ and

$n \geqslant k_i(m)$, then $g_i^{t_n}$ is based on at least l draws from $x_j(c)$. We can hence apply the strong law of large numbers to obtain

$$P(\lim_{n \to \infty} g_n^{t_n} = x_j(x) \mid c \in W) = 1 \tag{a1}$$

when $P(c \in W) > 0$. On the other hand, note that $\|\lambda_i^{t_n} - x_j(c)\| \leqslant (1 - \beta_i^{t_n}) d_i^{t_n} + \|g_i^{t_n} - x_j(x)\|$ if $d_i^{t_n} \leqslant \varepsilon$, and (iv), $\|\lambda_i^{t_n} - x_j(c)\| = \|g_i^{t_n} - x_j(c)\|$ if $d_i^{t_n} > \varepsilon$. In either case, we have $\|\lambda_i^{t_n} - x_j(c)\| \leqslant \varepsilon + \|g_i^{t_n} - x_j(c)\|$. Taking the limsup of the both sides and comparing the result with (a1), we find that

$$P(\limsup_{n \to \infty} \|\lambda_i^{t_n} - x_j(c)\| \leqslant \varepsilon \mid c \in W) = 1. \tag{a2}$$

Since (a2) is true for any $c \in H^m$ such that $P(c \in W) > 0$, it follows that $P(\limsup_{t \to \infty} \|\lambda_i^t - x_j(h^t)\| \leqslant \varepsilon) = 1$. ∎

Proof of Corollary 2.1. Since $k_i(s) \to \infty$ as $s \to \infty$, by the same discussion as in the Proof of Proposition 2, given any $l \in \mathbf{N}$, there exists a random time $T_l \in \mathbf{N}$ such that for every $t \geqslant T_l$, the conditional frequency g_i^t at history h^t is based on at least l observations. Thus, against a 0-bounded recall behavioral rule x_j, we have by the strong law of large numbers

$$\lim_{t \to \infty} g_i^t = \lim_{t \to \infty} f_i^t = x_j(h^0)$$

with probability one. Therefore, $P(\lim_{t \to \infty} \lambda_i^t = x_j(h^0)) = 1$ as desired. ∎

Proof of Proposition 3. Let $a = (a_1, a_2) \in A$ be a strict Nash equilibrium of the game G. By the continuity of π_i, there exists $\delta > 0$ such that $\|(\lambda_2, \lambda_1) - a\| < \delta$ implies $B_i(\lambda_i) = \{a_i\}$ $(i = 1, 2)$. Take $(\lambda_1^0, \lambda_2^0)$ so that $\|(\lambda_2^0, \lambda_1^0) - a\| < \delta$. Then $(a_1^1, a_2^1) = a$. Suppose $(a_1^t, a_2^t) = a$ for $t = 1, ..., T$. Then $g_i^T(a_j) = f_i^T(a_j) = 1$ and hence $\lambda_i^T(a_j) = 1$ $(i = 1, 2, j \neq i)$. It follows that $(a_1^{T+1}, a_2^{T+1}) = a$ and hence that $\lambda_i^{T+1}(a_j) = 1$. This completes the proof. ∎

Proof of Proposition 4. We first present the following result which will be used in the Proof of Proposition 4. It relates convergence in beliefs to convergence in empirical frequencies when there is an upper bound on the length of the patterns that a player can recognize.

PROPOSITION A1. *Suppose the assessment rule z_i satisfies* (3) *when $k_i(s) = \infty$ for every $s \geqslant n$ for some $n \in \mathbf{N}$. If $\lim_{t \to \infty} \lambda_i^t = m_j \in M_j$ and $\lim_{t \to \infty} \beta_i^t = 1$ along some sample path, then $\lim_{t \to \infty} f_i^t = \lim_{t \to \infty} g_i^t = m_j$ along the path.*

Namely, when there is such a bound, whenever a sequence of beliefs as described in (3) converges to some point, the empirical frequency also converges to the same point, provided that only the conditional frequency is used asymptotically.

Proof of Proposition A1.   Fix such an $\omega \in \Omega$. Since $\lim_{t \to \infty} \beta_i^t(\omega) = 1$, it immediately follows that $\lim_{t \to \infty} \|\lambda_i^t - g_i^t\| = 0$ and hence that $\lim_{t \to \infty} g_i^t = m_j$. Let $W \subset \bigcup_{s < n} H^s$ be the (finite) set of patterns that are used by player i infinitely often along $h = h(\omega) \in H^\infty$. There exists time $T \in \mathbf{N}$ such that player i only uses patterns belonging to W after T.

For $t > T$, let

$$\hat{f}_i^t(a_j) = \frac{\sum_{a_i \in A_i} R(h^{T+1, t}, a)}{t - T} \quad (a_j \in A_j),$$

and

$$\hat{g}_i^t(c)(a_j) = \frac{\sum_{a_i \in A_i} \left[R(h^t, ca) - R(h^T, ca) \right]}{R(h^t, c) - R(h^{T-1}, c)} \quad (a_j \in A_j,\, c \in W), \quad (a3)$$

when $R(h^t, c) - R(h^{T-1}, c) > 0$ $(a = (a_1, a_2))$. Namely, $\hat{f}_i^t$ is the overall empirical frequency in the history $h^{T+1, t}$, and $\hat{g}_i^t(c)$ is the (appropriately defined) empirical frequency conditional on the pattern c in $h^{T+1, t}$. Also let $\hat{g}_i^t = \hat{g}_i^t(c)$ if $c \in W$ is used at history h^t. It is clear that $\lim_{t \to \infty} \|\hat{f}_i^t - f_i^t\| = 0$ and $\lim_{t \to \infty} \|\hat{g}_i^t - g_i^t\| = 0$.

Note that we can rewrite $\hat{f}_i^t$ as

$$\hat{f}_i^t(a_j) = \frac{\sum_{c \in W} \sum_{a_i \in A_i} \left[R(h^t, ca) - R(h^T, ca) \right]}{\sum_{c \in W} \left[R(h^t, c) - R(h^{T-1}, c) \right]} \quad (a4)$$

Comparing (a3) and (a4), we find that

$$\min_{c \in W} \hat{g}_i^t(c)(a_j) \leqslant \hat{f}_i^t(a_j) \leqslant \max_{c \in W} \hat{g}_i^t(c)(a_j). \quad (a5)$$

Fix $\varepsilon > 0$ arbitrarily. Take $T_1 > T$ such that if $t \geqslant T_1$, $\hat{g}_i^t(c)$ is well defined for all $c \in W$, $\beta_i^t > 1 - \varepsilon$, $\|\hat{f}_i^t - f_i^t\| < \varepsilon$, $\|\hat{g}_i^t - g_i^t\| < \varepsilon$, and $\|g_i^t - m_j\| < \varepsilon$.

For $t > T_1$, it follows from (a5) that for each $a_j \in A_j$,

$$
\begin{aligned}
f_i^t(a_j) &\leqslant \hat{f}_i^t(a_j) + \varepsilon \leqslant \beta_i^t \max_{c \in W} \hat{g}_i^t(c)(a_j) + (1 - \beta_i^t) \hat{f}_i^t(a_j) + \varepsilon \\
&\leqslant \sup_{s > T_1} \hat{g}_i^s(a_j) + \varepsilon \hat{f}_i^t(a_j) + \varepsilon \\
&\leqslant \sup_{s > T_1} g_i^s(a_j) + 3\varepsilon \\
&\leqslant m_j + 4\varepsilon.
\end{aligned}
$$

Similarly, we can show that $f_i^t(a_j) \geqslant m_j - 4\varepsilon$ for $t > T_1$. Since ε is arbitrary, we have $\lim_{t \to \infty} f_i^t = m_j$. ∎

We now return to the proof of Proposition 4.

Fix $\varepsilon > 0$ such that $\varepsilon < \min\{\kappa_1, \kappa_2\}/3$. Given this ε, let $T_\varepsilon \in \mathbf{N} \cup \{\infty\}$ be the (random) time such that $T_\varepsilon > 1/\varepsilon$ and $\sup_{t > T_\varepsilon} \|(\lambda_2^t, \lambda_1^t) - m\| < \varepsilon$ for the first time. We show below $T_\varepsilon(\omega) = \infty$ for any $\omega \in \Omega$ and hence conclude that m is unstable.

Suppose on the contrary that there exists $\omega \in \Omega$ for which $T = T_\varepsilon(\omega) < \infty$. We have three possible cases to consider:

(a) $d_i^t > \kappa_i$ for all $t > T$ for either $i = 1$ or 2.

This implies $\beta_i^t = 1$ for all $t > T$. Thus $\lim_{t \to \infty} \lambda_i^t = \lim_{t \to \infty} f_i^t = m_j$ and hence $\lim_{t \to \infty} d_i^t = 0$ by Proposition A1. This is a contradiction.

(b) $d_i^t > \kappa_i$ and $d_i^{t-1} \leqslant \kappa_i$ for some $t > T + 1$ for either $i = 1$ or 2.

Then $\|\lambda_i^t - \lambda_i^{t-1}\| = \|\beta_i^t g_i^t + (1 - \beta_i^t) f_i^t - f_i^{t-1}\| \geqslant d_i^t - \|f_i^t - f_i^{t-1}\| \geqslant \kappa_i - 1/t > 2\varepsilon$, which is a contradiction.

(c) $d_i^t \leqslant \kappa_i$ for all $t > T'$ for some $T' \geqslant T$ for both $i = 1$ and 2. In this case, note that $\lambda_i^t = f_i^t$ for $t > T'$ ($i = 1, 2$). We proceed case by case.

Condition (i): Note there exists a pure action profile $a \in A$ that appears infinitely often along h. Let t_n be the time of the nth occurrence of a along h ($n \in \mathbf{N}$). After period T', player j never observes action a_i' played after a.

Condition (ii): Since a_i as specified must be played infinitely often, let t_n be the time of the nth occurrence of a_i along h ($n \in \mathbf{N}$). After period T', player j never observes action a_i' played after (a_i, a_j) for any $a_j \in A_j$.

In either case, $\lim_{n \to \infty} g_j^{t_n}(a_i') = 0$. This implies $d_j^{t_n} > \kappa_j$ for n large enough if $\kappa_j > 0$ is such that $\kappa_j < \min_{a_i \in c_i(m_i)} m_i(a_i)$. We hence have a contradiction. ∎

Proof of Proposition 5. Let $W \subset \bigcup_{s < n} H^s$ be the random set of patterns that are used by player i infinitely often. Under (6), we have $W_1 = W_2 = W$ with probability one since the two players eventually use the same patterns almost surely. Also W consists of patterns of the same (random) length l almost surely. We henceforth restrict our attention to this event. Let $T < \infty$ be the random time such that every pattern used after T is in W. For any sample path $h \in H^\infty$, note that the sequence of periods $T + 1, T + 2, \ldots$ can be partitioned into subsequences depending on which l-length pattern in W is used at h^t ($t \geqslant T + 1$). Take an arbitrary pattern $c \in \bigcup_{s < n} H^s$. When $c \in W$, let t_n be the random time when the pattern c is used for the nth time along h after T. Hence $(t_n)_{n \in \mathbf{N}}$ forms one of the subsequences described

above. Since m is stable under fictitious play with arbitrary initial weight, it follows that $P(\lim_{n \to \infty} g_i^{t_n} = m_j \mid c \in W) = 1$ when $P(c \in W) > 0$ ($i = 1, 2, j \neq i$). Since $\lambda_i^{t_n} = g_i^{t_n}$ for each $n \in \mathbf{N}$, $P(\lim_{n \to \infty} \lambda_i^{t_n} = m_j \mid c \in W) = 1$. Since this is true for any pattern c such that $P(c \in W) > 0$, we have $P(\lim_{t \to \infty} \lambda_i^t = m_j) = 1$. ∎

REFERENCES

1. G. W. Brown, Iterative solution of games by fictitious play, *in* "Activity Analysis of Production and Allocation," Wiley, New York, 1951.
2. D. Fudenberg and D. M. Kreps, A theory of learning, experimentation, and equilibria in games, mimeo., MIT and Stanford University, 1988.
3. D. Fudenberg and D. M. Kreps, Learning mixed equilibria, *Games Econ. Behav.* **5** (1993), 320–367.
4. D. Gale, "The Theory of Linear Economic Models," University of Chicago Press, Chicago, IL, 1989.
5. J. S. Jordan, Three problems in learning mixed-strategy nash equilibria, *Games Econ. Behav.* **5** (1993), 368–386.
6. V. Krishna, Learning in games with strategic complementarities, mimeo., Harvard University, 1991.
7. E. Lehrer, Repeated games with stationary bounded recall strategies, *J. Econ. Theory* **46** (1988), 130–144.
8. P. Milgrom and J. Roberts, Adaptive and sophisticated learning in normal form games, *Games Econ. Behav.* **3** (1991), 82–100.
9. K. Miyasawa, On the convergence of the learning process in a 2×2 non-zero-sum two-person game, mimeo., Princeton University, 1961.
10. D. Monderer and L. Shapley, Potential games, mimeo, UCLA and Technion, 1992.
11. J. Robinson, An iterative method of solving a game, *Ann. Math.* **54** (1951), 296–301.
12. L. S. Shapley, Some topics in two-person games, *in* "Advances in Game Theory" (M. Dresher, L. S. Shapley, and A. W. Tucker, Eds.), pp. 1–28, Princeton University Press, Princeton, NJ, 1964.