Stochastic Uncoupled Dynamics and Nash Equilibrium*

Sergiu Hart† Andreu Mas-Colell‡

October 13, 2005

Abstract

In this paper we consider dynamic processes, in repeated games, that are subject to the natural informational restriction of uncoupledness. We study the almost sure convergence of play (the period-by-period behavior as well as the long-run frequency) to Nash equilibria of the one-shot stage game, and present a number of possibility and impossibility results. Basically, we show that if in addition to random experimentation some recall, or memory, is introduced, then successful search procedures that are uncoupled can be devised. In particular, to get almost sure convergence to pure Nash equilibria when these exist, it suffices to recall the last two periods of play.

Journal of Economic Literature Classification Numbers: C7, D83.

Keywords: uncoupled, Nash equilibrium, stochastic dynamics, finite recall, finite memory, finite automaton, exhaustive experimentation

*Previous versions: October 2004, May 2005. We thank Bob Aumann, Dean Foster, Fabrizio Germano, Sham Kakade, Gábor Lugosi, Yishay Mansour, Abraham Neyman, Yosi Rinott, Jeff Shamma, Benjy Weiss, Peyton Young, the associate editor, and the referees for their comments and suggestions. Part of this research was conducted during the Conference on Mathematical Foundations of Learning Theory at the Centre de Recerca Matemàtica in Barcelona, June 2004.

†Center for the Study of Rationality, Institute of Mathematics, and Department of Economics, The Hebrew University of Jerusalem, Feldman Building, Givat Ram, 91904 Jerusalem, Israel. *E-mail*: hart@huji.ac.il; *URL*: http://www.ma.huji.ac.il/hart.
Research partially supported by a grant of the Israel Academy of Sciences and Humanities.

‡Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain. *E-mail*: andreu.mas-colell@upf.edu.

1 Introduction

A dynamic process in a multi-player setup is *uncoupled* if the moves of every player do not depend on the payoff (or utility) functions of the other players. This is a natural informational requirement, which holds in most models. In Hart and Mas-Colell (2003b) we introduce this concept and show that uncoupled stationary dynamics cannot always converge to Nash equilibria, even if these exist and are unique. The setup was that of deterministic, stationary, continuous-time dynamics.

It is fairly clear that the situation may be different when *stochastic* moves are allowed, since one may then try to carry out some version of exhaustive search: keep randomizing until by pure chance a Nash equilibrium is hit, and then stop there. However, this is not so simple: play has a decentralized character, and no player can, alone, recognize a Nash equilibrium. The purpose of this paper is, precisely, to investigate to what extent Nash equilibria can be reached when considering dynamics that satisfy the restrictions of our previous paper: uncoupledness and stationarity. As we shall see, one can obtain positive results, but these will require that, in addition to the ability to perform stochastic moves of an experimental nature, the players retain some memory from the past plays.

Because we allow random moves, it is easier to place ourselves in a discrete time framework. Thus we consider the repeated play of a given stage game, under the standard assumption that each player observes the play of all players; as for payoffs, each player knows only his own payoff function. We start by studying a natural analog of the approach of our earlier paper; that is, we assume that in determining the random play at time $t+1$ the players recall only the information contained in the current play of all players at time t; i.e., past history does not matter. We call this the case of 1-*recall*. We shall then see that the result of our earlier paper is recovered: convergence of play to Nash equilibrium cannot be ensured under the hypotheses of uncoupledness, stationarity, and 1-recall (there is an exception for the case of generic two-player games with at least one pure Nash equilibrium).

Yet, the exhaustive search intuition can be substantiated if we allow for

(uncoupled and stationary) strategies with longer recall. Perhaps surprisingly, to guarantee almost sure convergence of play to pure Nash equilibria when these exist, it suffices to have 2-*recall*: to determine the play at $t+1$ the players use the information contained in the plays of all players at periods t and $t-1$. In general, when Nash equilibria may be mixed, we show that convergence of the long-run empirical distribution of play to (approximate) equilibria can be guaranteed using longer, but finite, recall. Interestingly, however, this does not suffice to obtain the almost sure convergence of the period-by-period behavior probabilities. As it turns out, we can get this too within the broader context of finite memory (i.e., finite-state automata).

In conclusion, one can view this paper as contributing to the demarcation of the border between those classes of dynamics for which convergence to Nash equilibrium can be obtained and those for which it cannot.

The paper is organized as follows. Section 2 presents the model and defines the relevant concepts. Convergence to pure Nash equilibria is studied in Section 3, and to mixed equilibria, in Section 4 (with a proof relegated to the Appendix). We conclude in Section 5 with some comments and a discussion of the related literature, especially the work of Foster and Young (2003a, 2003b).

2 The Setting

A basic static (one-shot) *game* is given in strategic (or normal) form, as follows. There are $N \geq 2$ *players*, denoted $i = 1, 2, ..., N$. Each player i has a finite set of *actions* A^i; let $A := A^1 \times A^2 \times ... \times A^N$ be the set of action combinations. The *payoff function* (or *utility function*) of player i is a real-valued function $u^i : A \to \mathbb{R}$. The set of *randomized* or *mixed* actions of player i is the probability simplex over A^i, i.e, $\Delta(A^i) = \{ x^i = (x^i(a^i))_{a^i \in A^i} : \sum_{a^i \in A^i} x^i(a^i) = 1 \text{ and } x^i(a^i) \geq 0 \text{ for all } a^i \in A^i\}$; as usual, the payoff functions u^i are multilinearly extended, so $u^i : \Delta(A^1) \times \Delta(A^2) \times ... \times \Delta(A^N) \to \mathbb{R}$.

We fix the set of players N and the action sets A^i, and identify a game by its payoff functions $U = (u^1, u^2, ..., u^N)$.

For $\varepsilon \geq 0$, a *Nash ε-equilibrium* x is an N-tuple of mixed actions $x = (x^1, x^2, ..., x^N) \in \Delta(A^1) \times \Delta(A^2) \times ... \times \Delta(A^N)$, such that x^i is an *ε-best reply* to x^{-i} for all i; i.e., $u^i(x) \geq u^i(y^i, x^{-i}) - \varepsilon$ for every $y^i \in \Delta(A^i)$ (we write $x^{-i} = (x^1, ..., x^{i-1}, x^{i+1}, ..., x^N)$ for the combination of mixed actions of all players except i). When $\varepsilon = 0$ this is a *Nash equilibrium*, and when $\varepsilon > 0$, a Nash *approximate* equilibrium.

The dynamic setup consists of a repeated play, at discrete time periods $t = 1, 2, ...$, of the static game U. Let $a^i(t) \in A^i$ denote the action of player i at time[1] t, and put $a(t) = (a^1(t), a^2(t), ..., a^N(t)) \in A$ for the combination of actions at t. We assume that there is standard monitoring: at the end of period t each player i observes everyone's realized action, i.e., $a(t)$.

A *strategy*[2] f^i of player i is a sequence of functions $(f_1^i, f_2^i, ..., f_t^i, ...)$, where, for each time t, the function f_t^i assigns a mixed action in $\Delta(A^i)$ to each history $h_{t-1} = (a(1), a(2), ..., a(t-1))$. A strategy profile is $f = (f^1, f^2, ..., f^N)$.

A strategy f^i of player i has *finite recall* if there exists a positive integer R such that only the history of the last R periods matters: for each $t > R$, the function f_t^i is of the form $f_t^i(a(t-R), a(t-R+1), ..., a(t-1))$; we call this *$R$-recall*.[3] Such a strategy is moreover *stationary* if the ("calendar") time t does not matter: $f_t^i \equiv f^i(a(t-R), a(t-R+1), ..., a(t-1))$ for all $t > R$.

Strategies have to fit the game being played. We thus consider a *strategy mapping*, which, to every game (with payoff functions) U, associates a strategy profile $f(U) = (f^1(U), f^2(U), ..., f^N(U))$ for the repeated game induced by $U = (u^1, u^2, ..., u^N)$. Our basic requirement for a strategy mapping is *uncoupledness*, which says that the strategy of each player i may depend only on the i-th component u^i of U, i.e., $f^i(U) \equiv f^i(u^i)$. Thus, for any player i and time t, the strategy f_t^i has the form $f_t^i(a(1), a(2), ..., a(t-1); u^i)$. Finally, we shall say that a strategy mapping has R-recall and is stationary if, for any U, the strategies $f^i(U)$ of all players i have R-recall and are stationary.

[1] More precisely, the actual *realized* action (when randomizations are used).

[2] We use the term "strategy" for the repeated game and "action" for the one-shot game.

[3] The recall of a player consists of past realized pure actions only, not of mixed actions played (not even his own).

3 Pure Equilibria

We start by considering games that possess *pure* Nash equilibria (i.e., Nash equilibria $x = (x^1, x^2, ..., x^N)$ where each x^i is a pure action in A^i).[4] Our first result generalizes the conclusion of Hart and Mas-Colell (2003b). We show that with 1-recall — that is, if actions depend only on the current play and not on past history — we cannot hope in all generality to converge, in an uncoupled and stationary manner, to pure Nash equilibria when these exist.

Theorem 1 *There are no uncoupled, 1-recall, stationary strategy mappings that guarantee almost sure convergence of play to pure Nash equilibria of the stage game in all games where such equilibria exist.*[5]

Proof. The following two examples, the first with $N = 2$ and the second with $N = 3$, establish our result. We point out that the second example is *generic* — in the sense that the best reply is always unique — while the first is not; this will matter in the sequel.

The first example is the two-player game of Figure 1. The only pure Nash equilibrium is (γ, γ). Assume by way of contradiction that we are given an

	α	β	γ
α	1,0	0,1	1,0
β	0,1	1,0	1,0
γ	0,1	0,1	1,1

Figure 1: A non-generic two-player game

uncoupled, 1-recall, stationary strategy mapping that guarantees convergence to pure Nash equilibria when these exist. Note that at each of the nine action pairs, at least one of the two players is best-replying. Suppose the current state $a(t)$ is such that player 1 is best-replying (the argument is symmetric for player 2). We claim that player 1 will play at $t + 1$ the same action as

[4] From now on, "game" and "equilibrium" will always refer to the one-shot stage game.

[5] "Almost sure convergence of play to pure Nash equilibria" means that almost every play path consists of a pure Nash equilibrium being played from some point on.

in t (i.e., player 1 will not move). To see this consider a new game where the utility function of player 1 remains unaltered and the utility function of player 2 is changed in such a manner that the current state $a(t)$ is the only pure Nash equilibrium of the new game. It is easy to check that in our game this can always be accomplished (for example, to make (α, γ) the unique Nash equilibrium, change the payoff of player 2 in the (α, γ) and (γ, α) cells to, say, 2). The strategy mapping has 1-recall, so it must prescribe to the first player not to move in the new game (otherwise convergence to the unique pure equilibrium would be violated there). By uncoupledness, therefore, player 1 will not move in the original game either.

It follows that (γ, γ) can never be reached when starting from any other state: if neither player plays γ currently then only one player (the one who is not best-replying) may move; if only one plays γ then the other player cannot move (since in all cases it is seen that he is best-replying). This contradicts our assumption.

The second example is the three-player game of Figure 2. There are three players $i = 1, 2, 3$, and each player has three actions α, β, γ. Restricted to α

α

	α	β	γ
α	0,0,0	0,4,4	2,1,2
β	4,4,0	4,0,4	3,1,3
γ	1,2,2	1,3,3	0,0,0

β

α	β	γ
4,0,4	4,4,0	3,1,3
0,4,4	0,0,0	2,1,2
1,3,3	1,2,2	0,0,0

γ

α	β	γ
2,2,1	3,3,1	0,0,0
3,3,1	2,2,1	0,0,0
0,0,0	0,0,0	6,6,6

Figure 2: A generic three-player game

and β we essentially have the game of Jordan (1993) (see Hart and Mas-Colell (2003b, Section III)), where every player i tries to mismatch the player $i-1$ (the predecessor of player 1 is player 3): he gets 0 if he matches and 4 if he mismatches. If all three players play γ then each one gets 6. If one player plays γ and the other two do not, the player that plays γ gets 1 and the other two get 3 each if they mismatch and 2 each if they match. If two players play γ and the third one does not then each one gets 0.

The only pure Nash equilibrium of this game is (γ, γ, γ). Suppose that we start with all players playing α or β, but not all the same; for instance, (α, β, α). Then players 2 and 3 are best-replying, so only player 1 can move in the next period (this follows from uncoupledness as in the previous example). If he plays α or β then we are in exactly the same position as before (with, possibly, the role of mover taken by player 2). If he moves to γ then the action configuration is (γ, β, α), at which both players 2 and 3 are best-replying and so, again, only player 1 can move. Whatever he plays next, we are back to situations already contemplated. In summary, every configuration that can be visited will only have at most one γ, and therefore the unique pure Nash equilibrium (γ, γ, γ) will never be reached. □

Remark. In the three-player example of Figure 2, starting with (α, β, α), the empirical joint distribution of play cannot approach the distribution of a mixed Nash equilibrium, because neither (α, α, α) nor (β, β, β) will ever be visited — but these action combinations have positive probability in every mixed Nash equilibrium (there are two such equilibria: in the first each player plays $(1/2, 1/2, 0)$, and in the second each plays $(1/4, 1/4, 1/2)$).

As we noted above, the two-player example of Figure 1 is not generic. It turns out that in the case of only two players, *genericity* — in the sense that every player's best reply to pure actions is always unique — does help.

Proposition 2 *There exist uncoupled,* 1*-recall, stationary strategy mappings that guarantee almost sure convergence of play to pure Nash equilibria of the stage game in every two-player generic game where such equilibria exist.*

Proof. We define the strategy mapping of player 1 (and similarly for player 2) as follows. Let the state (i.e., the previous period's play) be $a = (a^1, a^2) \in A = A^1 \times A^2$. If a^1 is a best reply to a^2 according to u^1, then player 1 plays a^1; otherwise, player 1 randomizes uniformly[6] over all his actions in A^1.

[6]I.e., the probability of each $a^1 \in A^1$ is $1/|A^1|$. Of course, the uniform distribution may be replaced — here as well as in all other constructions in this paper — by any probability distribution with full support (i.e., such that every action has positive probability).

With these strategies, it is clear that each pure Nash equilibrium becomes an absorbing state of the resulting Markov chain.[7] Moreover, as we shall show next, from any other state $a = (a^1, a^2) \in A$ there is a positive probability of reaching a pure Nash equilibrium in at most two steps. Indeed, since a is not a Nash equilibrium, then at least one of the players, say player 2, is not best-replying. Therefore there is a positive probability that in the next period player 2 plays $\bar{a}^2$, where $(\bar{a}^1, \bar{a}^2)$ is a pure Nash equilibrium (which is assumed to exist), whereas player 1 plays the same a^1 as last period (this has probability one if player 1 was best-replying, and positive probability otherwise). The new state is thus $(a^1, \bar{a}^2)$. If $a^1 = \bar{a}^1$, we have reached the pure Nash equilibrium $\bar{a}$. If not, then player 1 is not best-replying — it is here that our genericity assumption is used: the *unique* best reply to $\bar{a}^2$ is $\bar{a}^1$ — so now there is a positive probability that in the next period player 1 will play $\bar{a}^1$ and player 2 will play $\bar{a}^2$, and thus, again, the pure Nash equilibrium $\bar{a}$ is reached.

Therefore an absorbing state — i.e., a pure Nash equilibrium — must eventually be reached with probability one. □

Interestingly, if we allow for longer recall the situation changes and we can present positive results for general games. In fact, for the case where pure Nash equilibria exist the contrast is quite dramatic, since allowing just one more period of recall suffices.

Theorem 3 *There exist uncoupled, 2-recall, stationary strategy mappings that guarantee almost sure convergence of play to pure Nash equilibria of the stage game in every game where such equilibria exist.*

Proof. Let the state — i.e., the play of the previous two periods — be $(a', a) \in A \times A$. We define the strategy mapping of each player i as follows:

- if $a' = a$ (i.e., if *all* players have played exactly the same actions in the past two periods) and a^i is a best reply of player i to a^{-i} according to u^i, then player i plays a^i (i.e., he plays the same action yet again);

[7] A standard reference for Markov chains is Feller (1968, Chapter XV).

- in all other cases, player i randomizes uniformly over A^i.

To prove our result, we partition the state space $S = A \times A$ of the resulting Markov chain into four regions:

$$\begin{aligned} S_1 &:= \{(a,a) \in A \times A : a \text{ is a Nash equilibrium}\}; \\ S_2 &:= \{(a',a) \in A \times A : a' \neq a \text{ and } a \text{ is a Nash equilibrium}\}; \\ S_3 &:= \{(a',a) \in A \times A : a' \neq a \text{ and } a \text{ is not a Nash equilibrium}\}; \\ S_4 &:= \{(a,a) \in A \times A : a \text{ is not a Nash equilibrium}\}. \end{aligned}$$

Clearly, each state in S_1 is absorbing. Next, we claim that all other states are transient: there is a positive probability of reaching a state in S_1 in finitely many periods. Indeed:

- At each state (a',a) in S_2 all players randomize; hence there is a positive probability that next period they will play a — and so the next state will be (a,a), which belongs to S_1.

- At each state (a',a) in S_3 all players randomize; hence there is a positive probability that next period they will play a pure Nash equilibrium $\bar{a}$ (which exists by assumption) — and so the next state will be $(a,\bar{a})$, which belongs to S_2.

- At each state (a,a) in S_4 at least one player is not best-replying and thus is randomizing; hence there is a positive probability that the next period play will be some $a' \neq a$ — and so the next state will be (a,a'), which belongs to $S_2 \cup S_3$.

In all cases there is thus a positive probability of reaching an absorbing state in S_1 in at most three steps. Once such a state (a,a), where a is a pure Nash equilibrium, is reached (this happens eventually with probability one), the players will continue to play a every period.[8] □

[8] Aniol Llorente (personal communication, 2005) has shown that it suffices to have 1-recall for one player and 2-recall for all other players, but this cannot be further weakened (see the example of Figure 1).

Thus extremely simple strategies may nevertheless guarantee convergence to pure Nash equilibria. The strategies defined above may be viewed as a combination of search and testing. The search is a standard random search; the testing is done individually, but in a coordinated manner: the players wait until a certain "pattern" (a repetition) is observed, at which point each one applies a "rational" test (he checks whether or not he is best-replying). Finally, the pattern is self-replicating once the desired goal (a Nash equilibrium) is reached. (This structure will appear again, in a slightly more complex form, in the case of mixed equilibria; see the proofs of Proposition 4 and Theorem 5 below.)

4 Mixed Equilibria

We come next to the general case (where only the existence of mixed Nash equilibria is guaranteed). Here we consider first the long-run frequencies of play, and then the period-by-period behavior probabilities. The convergence will be to approximate equilibria. To this effect, assume that there is a bound M on payoffs; i.e., the payoff functions all satisfy $|u^i(a)| \leq M$ for all action combinations $a \in A$ and all players i.

Given a history of play, we shall denote by Φ_t the empirical frequency distribution in the first t periods: $\Phi_t[a] := |\{1 \leq \tau \leq t : a(\tau) = a\}|/t$ for each $a \in A$, and similarly $\Phi_t[a^i] := |\{1 \leq \tau \leq t : a^i(\tau) = a^i\}|/t$ for each i and $a^i \in A^i$. We shall refer to $(\Phi_t[a])_{a \in A} \in \Delta(A)$ as the *empirical joint distribution of play*,[9] and to $(\Phi_t[a^i])_{a^i \in A^i} \in \Delta(A^i)$ as the *empirical marginal distribution of play of player* i (up to time t).

Proposition 4 *For every M and $\varepsilon > 0$ there exists an integer R and an uncoupled, R-recall, stationary strategy mapping that guarantees, in every game with payoffs bounded by M, almost sure convergence of the empirical marginal distributions of play to Nash ε-equilibria; i.e., for almost every history of play there exists a Nash ε-equilibrium of the stage game $x = (x^1, x^2, ..., x^N)$*

[9] Also known as the "long-run sample distribution of play."

such that, for every player i and every action $a^i \in A^i$,

$$\lim_{t\to\infty} \Phi_t[a^i] = x^i(a^i). \tag{1}$$

Of course, different histories may lead to different ε-equilibria (i.e., x may depend on the play path). The length of the recall R depends on the precision ε and the bound on payoffs M (as well as on the number of players N and the number of actions $|A^i|$).

Proof. Given $\varepsilon > 0$, let K be such that[10]

$$\left[||x^i - y^i|| \leq \frac{1}{K} \quad \text{for all } i \right] \implies \left[|u^i(x) - u^i(y)| \leq \varepsilon \quad \text{for all } i \right] \tag{2}$$

for $x^i, y^i \in \Delta(A^i)$ and $|u^i(a)| \leq M$ for all $a \in A$. Let $\bar{y} = (\bar{y}^1, \bar{y}^2, ..., \bar{y}^N)$ be a Nash 2ε-equilibrium, such that all probabilities are multiples of $1/K$ (i.e., $K\bar{y}^i(a^i)$ is an integer for all a^i and all i). Such a $\bar{y}$ always exists: take a $1/K$-approximation of a Nash equilibrium and use (2). Given such a Nash 2ε-equilibrium $\bar{y}$, let $(\bar{a}_1, \bar{a}_2, ..., \bar{a}_K) \in A \times A \times ... \times A$ be a fixed sequence of action combinations of length K whose marginals are precisely $\bar{y}^i$ (i.e., each action a^i of each player i appears $K\bar{y}^i(a^i)$ times in the sequence $(\bar{a}^i_1, \bar{a}^i_2, ..., \bar{a}^i_K)$).

Take $R = 2K$. The construction parallels the one in the Proof of Theorem 3. A state is a history of play of length $2K$, i.e., $s = (a_1, a_2, ..., a_{2K})$ with $a_k \in A$ for all k. The state s is K-periodic if $a_{K+k} = a_k$ for all $k = 1, 2, ..., K$. Given s, for each player i we denote by $z^i \in \Delta(A^i)$ the frequency distribution of the last K actions of i, i.e., $z^i(a^i) := |\{K+1 \leq k \leq 2K : a^i_k = a^i\}/K$ for each $a^i \in A^i$; put $z = (z^1, z^2, ..., z^N)$.

We define the strategy mapping of each player i as follows:

- if the current state s is K-periodic and z^i is a 2ε-best reply to z^{-i}, then player i plays $a^i_1 = a^i_{K+1}$ (i.e., continues his K-periodic play);
- in all other cases player i randomizes uniformly over A^i.

[10] We use the maximum (ℓ^∞) norm on $\Delta(A^i)$, i.e., $||x^i - y^i|| := \max_{a^i \in A^i} |x^i(a^i) - y^i(a^i)|$; it is easy to check that $K \geq M \sum_i |A^i|/\varepsilon$ suffices for (2).

Partition the state space S consisting of all sequences over A of length $2K$ into four regions:

$$\begin{aligned} S_1 &:= \{s \text{ is } K\text{-periodic and } z \text{ is a Nash } 2\varepsilon\text{-equilibrium}\}; \\ S_2 &:= \{s \text{ is not } K\text{-periodic and } z \text{ is a Nash } 2\varepsilon\text{-equilibrium}\}; \\ S_3 &:= \{s \text{ is not } K\text{-periodic and } z \text{ is not a Nash } 2\varepsilon\text{-equilibrium}\}; \\ S_4 &:= \{s \text{ is } K\text{-periodic and } z \text{ is not a Nash } 2\varepsilon\text{-equilibrium}\}. \end{aligned}$$

We claim that the states in S_1 are persistent and K-periodic, and all other states are transient. Indeed, once a state s in S_1 is reached, the play moves in a deterministic way through the K cyclic permutations of s, all of which have the same z — and so, for each player i, his empirical marginal distribution of play will converge to z^i. At a state s in S_2 every player randomizes, so there is a positive probability that everyone will play K-periodically, leading in $r = \max\{1 \leq k \leq K : a_{K+k} \neq a_k\}$ steps to S_1. At a state s in S_3, there is a positive probability of reaching S_2 in $K+1$ steps: in the first step the play is some $a \neq a_{K+1}$, and, in the next K steps, a sequence $(\bar{a}_1, \bar{a}_2, ..., \bar{a}_K)$ corresponding to a Nash 2ε-equilibrium. Finally, from a state in S_4 there is a positive probability of moving to a state in $S_2 \cup S_3$ in one step. □

Proposition 4 is not entirely satisfactory, because it does not imply that the empirical *joint* distributions of play converge to joint distributions induced by Nash approximate equilibria. For this to happen, the joint distribution needs to be (in the limit) the product of the marginal distributions (i.e., independence among the players' play is required). But this is not the case in the construction in the Proof of Proposition 4 above, where the players' actions become "synchronized" — rather than independent — once an absorbing cycle is reached. A more refined proof is thus needed to obtain the stronger conclusion of the following theorem on the convergence of the joint distributions.

Theorem 5 *For every M and $\varepsilon > 0$ there exists an integer R and an uncoupled, R-recall, stationary strategy mapping that guarantees, in every game with payoffs bounded by M, the almost sure convergence of the empirical joint*

distributions of play to Nash ε-equilibria; i.e., for almost every history of play there exists a Nash ε-equilibrium of the stage game $x = (x^1, x^2, ..., x^N)$ such that, for every action combination $a = (a^1, a^2, ..., a^N) \in A$,

$$\lim_{t\to\infty} \Phi_t[a] = \prod_{i=1}^{N} x^i(a^i). \tag{3}$$

Moreover, there exists an almost surely finite stopping time[11] *T after which the occurrence probabilities $\Pr[a(t) = a \,|\, h_T]$ also converge to Nash ε-equilibria; i.e., for almost every history of play and every action combination $a = (a^1, a^2, ..., a^N) \in A$,*

$$\lim_{t\to\infty} \Pr[a(t) = a \,|\, h_T] = \prod_{i=1}^{N} x^i(a^i), \tag{4}$$

where x is the same Nash ε-equilibrium of (3).

As before, x and T may depend on the history; T is the time when some ergodic set is reached. Since the proof of Theorem 5 is relatively intricate, we relegate it to the Appendix. Of course, (1) follows from (3). Note that neither (4) nor its marginal implications,

$$\lim_{t\to\infty} \Pr[a^i(t) = a^i \,|\, h_T] = x^i(a^i) \tag{5}$$

for all i, hold for the construction of Proposition 4 (again, due to periodicity).

Now (5) says that, after time T, the overall probabilities of play converge almost surely to Nash ε-equilibria. It does not say the same, however, about the actual play or *behavior probabilities* $\Pr[a^i(t) = a^i \,|\, h_{t-1}] = f^i(h_{t-1})(a^i)$ (where $h_{t-1} = (a(1), a(2), ..., a(t-1))$). We next show that this cannot be guaranteed in general when the recall is finite.

[11] I.e., T is determined by the past only: if $T = t$ for a certain play path $h = (a(1), ..., a(t), a(t+1), ...)$, then $T = t$ for any other play path $h' = (a(1), ..., a(t), a'(t+1), ...)$ that is identical to h up to and including time t. This initial segment of history $(a(1), a(2), ..., a(T))$ is denoted h_T.

Theorem 6 *For every small enough*[12] $\varepsilon > 0$, *there are no uncoupled, finite recall, stationary strategy mappings that guarantee, in every game, the almost sure convergence of the behavior probabilities to Nash ε-equilibria of the stage game.*

Proof. Choose a stage game U with a unique, completely mixed Nash equilibrium, and assume that a certain pure action combination, call it $\bar{a} \in A$, is such that $\bar{a}^1$ is the unique best reply of player 1 to $\bar{a}^{-1} = (\bar{a}^2, ..., \bar{a}^N)$. Let U' be another game where the payoff function of player 1 is the same as in U, and the payoff function of every other player $i \neq 1$ depends only on a^i and has a unique global maximum at $\bar{a}^i$. Then $\bar{a}$ is the unique Nash equilibrium of U'. Take $\varepsilon > 0$ small enough so that all Nash ε-equilibria of U are completely mixed, and moreover there exists $\rho > 0$ such that, for any two Nash ε-equilibria x and y of U and U', respectively, we have $0 < x^i(\bar{a}^i) < \rho < y^i(\bar{a}^i)$ for all players i (recall that $x^i(\bar{a}^i) < 1 = y^i(\bar{a}^i)$ when x and y are the unique Nash equilibria of U and U', respectively).

We argue by contradiction and assume that for some R there is an uncoupled, R-recall, stationary strategy mapping f for which the stated convergence does in fact obtain.

Consider now the history $\bar{\mathbf{a}} = (\bar{a}, \bar{a}, ..., \bar{a})$ of length R that consists of R repetitions of $\bar{a}$. The behavior probabilities have been assumed to converge (a.s.) to Nash ε-equilibria, which, in both games, always give positive probability to the actions $\bar{a}^i$. Hence the state $\bar{\mathbf{a}}$ has a positive probability of occurring after any large enough time T. Therefore, in particular at this state $\bar{\mathbf{a}}$, the behavior probabilites must be close to Nash ε-equilibria. Now all Nash ε-equilibria x of U satisfy $x^1(\bar{a}^1) < \rho$, so the behavior probability of player 1 at state $\bar{\mathbf{a}}$ must also satisfy this inequality, i.e., $f^1(\bar{\mathbf{a}}; u^1)(\bar{a}^1) < \rho$. But the same argument applied to U' (where player 1 has the same payoff function u^1 as in U) implies $f^1(\bar{\mathbf{a}}; u^1)(\bar{a}^1) > \rho$ (since this inequality is satisfied by all Nash ε-equilibria of U'). This contradiction proves our claim. □

The impossibility result of Theorem 6 hinges on the finite recall assumption. Finite recall signifies that the distant past is irrelevant to present be-

[12] I.e., for all $\varepsilon < \varepsilon_0$ (where ε_0 may depend on N and $(|A^i|)_{i=1}^N$).

havior (two histories that differ only in periods beyond the last R periods will generate the same mixed actions).[13] Hence, finite recall is a special, though natural, way to get the past influencing the present through a finite set of parameters. But it is not the only framework with this implication. What would happen if, while retaining the desideratum of a limited influence from the past, we were to broaden our setting by moving from a finite recall to a "finite memory" assumption? It turns out that we then obtain a positive result: the period-by-period behavior probabilities can also be made to converge almost surely.

Specifically, a strategy of player i has *finite memory* if it can be implemented by an automaton with finitely many states, such that, at each period t, its input is $a(t) \in A$, the N-tuple of actions actually played, and its output is $x^i(t+1) \in \Delta(A^i)$, the mixed action to be played next period. To facilitate comparison with finite recall, we shall measure the size of the memory by the number of elements of A it can contain; thus *R-memory* means that the memory can contain any $(a(1), a(2), ..., a(R))$ with $a(k) \in A$ for $k = 1, 2, ..., R$ (i.e., the automaton has $|A|^R$ states).

Theorem 7 *For every M and $\varepsilon > 0$ there exists an integer R and an uncoupled, R-memory, stationary strategy mapping that guarantees, in every game with payoffs bounded by M, the almost sure convergence of the behavior probabilities to Nash ε-equilibria; i.e., for almost every history of play there exists a Nash ε-equilibrium of the stage game $x = (x^1, x^2, ..., x^N)$ such that, for every action $a^i \in A^i$ of every player $i \in N$,*

$$\lim_{t\to\infty} \Pr[a^i(t) = a^i \,|\, h_{t-1}] = x^i(a^i). \tag{6}$$

Since the players randomize independently at each stage, (6) implies

$$\lim_{t\to\infty} \Pr[a(t) = a \,|\, h_{t-1}] = \prod_{i=1}^{N} x^i(a^i) \tag{7}$$

for every $a \in A$, from which it follows, by the law of large numbers, that the

[13] See Aumann and Sorin (1989) for a discussion of bounded recall.

empirical joint distributions of play also converge almost surely, i.e., (3).

Proof. We modify the construction in Proposition 4 as follows. Let $R = 2K+1$; a state is now $\tilde{s} = (a_0, a_1, a_2, ..., a_{2K})$ with $a_k \in A$ for $k = 0, 1, ..., 2K$. Let $s = (a_1, a_2, ..., a_{2K})$ be the last $2K$ coordinates of $\tilde{s}$ (so $\tilde{s} = (a_0, s)$); the frequencies z^i are still determined by the last K coordinates $a_{K+1}, ..., a_{2K}$.

There will be two "modes" of behavior. In the first mode the strategy mappings are as in the Proof of Proposition 4, except that now the recall has length $2K + 1$, and that whenever s (the play of the last $2K$ periods) is K-periodic and z^i is not a 2ε-best reply to z^{-i}, player i plays an action that is *different* from a_1^i (rather than a randomly chosen action); i.e., i "breaks" the K-periodic play. This guarantees that a K-periodic state $\tilde{s}$ (the play of the last $2K + 1$ periods) is reached *only when* z is a Nash 2ε-equilibrium. When this occurs the strategies move to the second mode, where in every period player i plays the mixed action z^i, and the state remains fixed (i.e., it is no longer updated).

Formally, we define the strategy mapping and the state-updating rule for each player i as follows. Let the state be $\tilde{s} = (a_0, a_1, ..., a_{2K}) = (a_0, s)$; then:

- **Mode I**: $\tilde{s}$ is not K-periodic.
  - If s is K-periodic and z^i is a 2ε-best reply to z^{-i}, then player i plays a_1^i (which equals a_{K+1}^i; i.e., he continues his K-periodic play).
  - If s is K-periodic and z^i is not a 2ε-best reply to z^{-i}, then player i randomizes uniformly over $A^i \backslash \{a_1^i\}$ (i.e., he "breaks" his K-periodic play).
  - If s is not K-periodic, then player i randomizes uniformly over A^i.

  In all three cases, let a be the N-tuple of actions actually played; then the new state is $\tilde{s}' = (a_1, ..., a_{2K}, a)$.

- **Mode II**: $\tilde{s}$ is K-periodic.

  Player i plays the mixed action z^i, and the new state is $\tilde{s}' = \tilde{s}$ (i.e., unchanged).

- The starting state is any $\tilde{s}$ that is not K-periodic (Mode I).

It is easy to check that once a block of size K is repeated twice, either the frequencies z constitute a Nash 2ε-equilibrium — in which case next period the cyclical play continues and we get to Mode II — or they don't — in which case the cycle is broken by at least one player (and the random search continues). Once Mode II is reached, which happens eventually a.s., the states of all players stay constant, and each player plays the corresponding frequencies forever after. □

5 Discussion and Comments

This section includes some further comments, particularly on the relevant literature.

(a) *Foster and Young*: The current paper is not the first one where, within the span of what we call uncoupled dynamics, stochastic moves and the possibility of recalling the past have been brought to bear on the formulation of dynamics leading to Nash equilibria. The pioneers were Foster and Young (2003a), followed by Foster and Young (2003b), Kakade and Foster (2003), and Germano and Lugosi (2004).

The motivation of Foster and Young and our motivation are not entirely the same. They want to push to its limits the "learning with experimentation" paradigm (which does not allow direct exhaustive search procedures that, in our terminology, are not of an uncoupled nature). We start from the uncoupledness property and try to demarcate the border between what can and what cannot be done with such dynamics.

(b) *Convergence*: Throughout this paper we have sought a strong form of convergence, namely, almost sure convergence to a point.[14] One could consider seeking weaker forms of convergence (as has been done in the related literature): almost sure convergence to the convex hull of the set of Nash ε-equilibria, or convergence in probability, or "$1-\varepsilon$ of the time being an ε-

[14]The negative results of Theorems 1 and 6 also hold for certain weaker forms of convergence.

equilibrium," and so on. Conceivably, the use of weaker forms of convergence may have a theoretical payoff in other aspects of the analysis.

(c) *Stationarity*: With stationary finite recall (or finite memory) strategies, no more than convergence to approximate equilibria can be expected. Convergence to exact equilibria requires non-stationary strategies with unbounded recall; see Germano and Lugosi (2004) for such a result.

Another issue is that non-stationarity may allow the transmission of arbitrarily large amounts of information through the time dimension (for instance, a player may signal his payoff function through his actions in the first T periods), thus effectively voiding the uncoupledness assumption.

(d) *State space*: Theorems 1 and 3 show how doubling the size of the recall (from 1 to 2) allows for a positive result. More generally, the results of this paper may be viewed as a study of convergence to equilibrium when the common state space is larger than just the action space, thus allowing, in a sense, more information to be transmitted. Shamma and Arslan (2005) have introduced procedures in the continuous-time setup (extended to discrete-time in Arslan and Shamma (2004)) that double the state space, and yield convergence to Nash equilibria for some specific classes of games. Interestingly, convergence to correlated equilibria in the continuous-time setup was also obtained with a doubled state space, consisting of the current as well as the cummulative average play; see Theorem 5.1 and Corrolary 5.2 in Hart and Mas-Colell (2003a).

(e) *Unknown game*: Suppose that the players observe, not the history of play, but only their own realized payoffs; i.e., for each player i and time t the strategy is $f_t^i\left(u^i(a(1)), u^i(a(2)), ..., u^i(a(t-1))\right)$ (in fact, the player may know nothing about the game being played but his set of actions). What results can be obtained in this case? It appears that, for any positive result, experimentation even at (apparent) Nash equilibria will be indispensable. This suggests, in particular, that the best sort of convergence to hope for, in a stationary setting, is some kind of convergence in probability as mentioned in Remark (b). On this point see Foster and Young (2003b).

(f) *Which equilibrium?*: Among multiple (approximate) equilibria, the more "mixed" an equilibrium is, the higher the probability that the strate-

gies of Section 4 will converge to it. Indeed, the probability that uniform randomizations yield in a K-block frequencies $(k_1, k_2, ..., k_r)$ is proportional to $K!/(k_1!\ k_2!\ ...\ k_r!)$, which is lowest for pure equilibria (where $k_1 = K$) and highest for[15] $p_1 = p_2 = ... = p_r = K/r$.

(g) *Correlated equilibria*: We know that there are uncoupled strategy mappings with the property that the empirical joint distributions of play converge almost surely to the set of correlated equilibria (see Foster and Vohra (1997), Hart and Mas-Colell (2000), Hart (2005), and the book of Young (2004)). Strictly speaking, those strategies do not have finite recall, but enjoy a closely related property: they depend (in a stationary way) on a finite number of summary, and easily updatable, statistics from the past. The results of these papers differ from those of the current paper in several respects. First, the convergence there is to a set, whereas here it is to a point. Second, the convergence there is to correlated equilibria, whereas here it is to Nash equilibria. And third, the strategies there are natural, adaptive, heuristic strategies, while in this paper we are dealing with forms of exhaustive search (see (h) below). An issue for further study is to what extent the contrast can be captured by an analysis of the speeds of convergence (which appears to be faster for correlated equilibria).

(h) *Adaptation vs. experimentation*: Suppose we were to require in addition that the strategies of the players be "adaptive" in one way or another. For example, at time t player i could randomize only over actions that improve i's payoff given some sort of "expected" behavior of the other players at t, or over actions that would have yielded a better payoff if played at $t-1$, or if played every time in the past that the action at $t-1$ was played, or if played every time in the past (these last two are in the style of "regret-based" strategies; see Hart (2005) for a survey). What kind of results would then be plausible? Note that such adaptive or monotonicity-like conditions severely restrict the possibilities of "free experimentation" that drive the positive results obtained here. Indeed, even the weak requirement of never playing the currently worst action rules out convergence to Nash equilibria: for exam-

[15] For large K, an approximate comparison can be made in terms of entropies: equilibria with higher entropy are more likely to be reached than those with lower entropy.

ple, in the Jordan (1993) example where each player has two actions, this requirement leads to the best-reply dynamic, which does not converge to the unique Nash equilibrium.

Thus, returning to the issue, raised at the end of the Introduction, of distinguishing those classes of dynamics for which convergence to Nash equilibria can be obtained from those for which it cannot, "exhaustive experimentation" appears as a key ingredient in the former.

Appendix: Proof of Theorem 5

As pointed out in Section 4, the problem with our construction in the Proof of Proposition 4 is that it leads to periodic and synchronized behavior. To avoid this we introduce small random perturbations, independently for each player: once in a while, there is a positive probability of repeating the previous period's action rather than continuing the periodic play.[16] To guarantee that these perturbations do not eventually change the frequencies of play (our players cannot use any additional "notes" or "instructions"[17]), we use three repetitions rather than two and make sure that the basic periodic play can always be recognized from the R-history.

Proof of Theorem 5. We take $R = 3K$, where $K > 2$ is chosen so as to satisfy (2). Consider sequences $\mathbf{b} = (b_1, b_2, ..., b_{3K})$ of length $3K$ over an arbitrary finite set B (i.e., $b_k \in B$ for all k). We distinguish two types of such sequences:

- **Type E** ("Exact"): The sequence is K-periodic, i.e., $b_{K+k} = b_k$ for all $1 \leq k \leq 2K$. Thus $\mathbf{b}$ consists of three repetitions of the *basic* K-sequence $\mathbf{c} := (b_1, b_2, ..., b_K)$.

- **Type D** ("Delay"): The sequence is not of type E, and there is $2 \leq d \leq 3K$ such that $b_d = b_{d-1}$ and if we drop the element b_d from the sequence

[16] This kind of perturbation was suggested by Benjy Weiss. The randomness is needed to obtain (4) (one can get (3) using deterministic mixing in appropriately long blocks).

[17] As would be the case were the strategies of finite memory, as in Theorem 7.

$\mathbf{b}$ then the remaining sequence $\mathbf{b}_{-d} = (b_1, ..., b_{d-1}, b_{d+1}, ..., b_{3K})$ of length $3K-1$ is K-periodic. Again, let $\mathbf{c}$ denote the *basic* K-sequence,[18] so $\mathbf{b}_{-d}$ consists of three repetitions of $\mathbf{c}$, except for the missing last element. Think of b_d as a "delay" element.

We claim that the basic sequence of a sequence $\mathbf{b}$ of type D is uniquely defined. Indeed, assume that $\mathbf{b}_{-d}$ and $\mathbf{b}_{-d'}$ are both K-periodic, with corresponding basic sequences $\mathbf{c} = (c_1, c_2, ..., c_K)$ and $\mathbf{c}' = (c'_1, c'_2, ..., c'_K)$, and $d < d'$. If $d \geq K+1$, then the first K coordinates of $\mathbf{b}$ determine the basic sequence: $(b_1, b_2, ..., b_K) = \mathbf{c} = \mathbf{c}'$. If $d' \leq 2K$, then the last K coordinates determine this: $(b_{2K+1}, b_{2K+2}, ..., b_{3K}) = (c_K, c_1, ..., c_{K-1}) = (c'_K, c'_1, ..., c'_{K-1})$, so again $\mathbf{c} = \mathbf{c}'$. If neither of these two hold, then $d \leq K$ and $d' \geq 2K+1$. Without loss of generality assume that we took d' to be maximal such that $\mathbf{b}_{-d'}$ is K-periodic, and let $d' = 2K+r$ (where $1 \leq r \leq K$). Now $b_{d'-1} = c'_{r-1}$ and $b_{d'} = c_{r-1}$, so $c_{r-1} = c'_{r-1}$ (since $b_{d'} = b_{d'-1}$). But $c_{r-1} = b_{K+r} = c'_r$ (since $d < K+r < d'$; if $r = 1$ put $r - 1 \equiv K$), so $c'_{r-1} = c'_r$. If $d' < 3K$, this last equality implies that $b_{d'} = b_{d'+1}$, so $\mathbf{b}_{-(d'+1)}$ is also K-periodic (with the same basic sequence $\mathbf{c}'$), contradicting the maximality of d'. If $d' = 3K$, the equality becomes $c'_{K-1} = c'_K$, which implies that the sequence $\mathbf{b}$ is in fact of type E (it consists of three repetitions of $\mathbf{c}'$), again a contradiction.

Given a sequence $\mathbf{b}$ of type E or D, the frequency distribution of its basic K-sequence $\mathbf{c}$, i.e., $w \in \Delta(B)$ where $w(b) := |\{1 \leq k \leq K : c_k = b\}|/K$ for each $b \in B$, will be called the *basic frequency distribution* of $\mathbf{b}$.

To define the strategies, let $\mathbf{a} = (a_1, a_2, ..., a_{3K}) \in A \times A \times ... \times A$ be the state — a history of action combinations of length $3K$ — and put $\mathbf{a}^i := (a^i_1, a^i_2, ..., a^i_{3K})$ for the corresponding sequence of actions of player i. When $\mathbf{a}^i$ is of type E or D, we denote by $y^i \in \Delta(A^i)$ its basic frequency distribution. If for each player i the sequence $\mathbf{a}^i$ is of type E or D,[19] we shall say that the state $\mathbf{a}$ is *regular*; otherwise we shall call $\mathbf{a}$ *irregular*.

The strategy of player i is defined as follows.

[18] As we shall see immediately below, $\mathbf{c}$ is well defined (even though d need not be unique).

[19] Some players' sequences may be E, and others', D (moreover, they may have different d's); therefore $\mathbf{a}$ itself need not be E or D.

(*) If the state $\mathbf{a}$ is regular, the basic frequency y^i is a 4ε-best reply to the basic frequencies of the other players $y^{-i} = (y^j)_{j \neq i}$, and the sequence $\mathbf{a}^i$ is of type E, then with probability 1/2 play a_1^i (i.e., continue the K-periodic play), and with probability 1/2 play a_{3K}^i (i.e., introduce a "delay" period by repeating the previous period's action).

(**) If the state $\mathbf{a}$ is regular, the basic frequency y^i is a 4ε-best reply to the basic frequencies of the other players $y^{-i} = (y^j)_{j \neq i}$, and the sequence $\mathbf{a}^i$ is of type D, then play the last element of the basic sequence $\mathbf{c}$, which is a_K^i if $d > K$ and a_{K+1}^i if $d \leq K$ (i.e., continue the K-periodic play).

(***) In all other cases randomize uniformly over A^i.

As in the Proof of Proposition 4, given a state $\mathbf{a}$, for each player i let $z^i \in \Delta(A^i)$ denote the frequency distribution of $(a_{2K+1}^i, a_{2K+2}^i, ..., a_{3K}^i)$, the last K actions of i, and put $z = (z^1, z^2, ..., z^N)$; also, $y = (y^1, y^2, ..., y^N)$ is the N-tuple of the basic frequency distributions. We partition the state space S, which consists of all sequences $\mathbf{a}$ of length $3K$ over A, into four regions:

$$\begin{aligned} S_1 &:= \{\mathbf{a} \text{ is regular and } y \text{ is a Nash } 4\varepsilon\text{-equilibrium}\}; \\ S_2 &:= \{\mathbf{a} \text{ is irregular and } z \text{ is a Nash } 2\varepsilon\text{-equilibrium}\}; \\ S_3 &:= \{\mathbf{a} \text{ is regular and } y \text{ is not a Nash } 4\varepsilon\text{-equilibrium}\}; \\ S_4 &:= \{\mathbf{a} \text{ is irregular and } z \text{ is not a Nash } 2\varepsilon\text{-equilibrium}\}. \end{aligned}$$

We analyze each region in turn.

Claim 1. *All states in S_1 are ergodic.*[20]

Proof. Let $\mathbf{a} \in S_1$. For each player i, let $\mathbf{c}^i = (c_1^i, c_2^i, ..., c_K^i)$ be the basic sequence of $\mathbf{a}^i$, with y^i the corresponding basic frequency distribution. The strategies are such that the sequence of i in the next period is also of type E or D (by (*) and (**)), with basic sequence that is the cyclical permutation of $\mathbf{c}^i$ by one step $(c_2, ..., c_K, c_1)$, except when $\mathbf{a}^i$ is of type D with $d = 2$, in

[20] I.e., aperiodic and persistent; see Feller (1968, Sections XV.4-6) for these and the other Markov chain concepts that we use below.

which case it remains unchanged. Therefore the basic frequency distribution y^i does not change, and the new state is also in S_1. Hence S_1 is a closed set, and once it is reached the conditions of regularity and 4ε-best-replying for each player will always continue to be automatically satisfied; thus each player's play in S_1 depends only on whether his *own* sequence is of type E or D (again, see (*) and (**)). Therefore in the region S_1 the play becomes independent among the players, and the Markov chain restricted to S_1 is the product of N independent Markov chains, one for each player. Specifically, the state space S_1^i of the Markov chain of player i consists of all sequences of length $3K$ over A^i that are of type E or D, and the transition probabilities are defined as in (*) and (**), according to whether the sequence is of type E or D, respectively.

We thus analyze each i separately. Let $\mathbf{a}^i$ be a $3K$-sequence in S_1^i with basic sequence $\mathbf{c}^i$. The closure of $\mathbf{a}^i$ (i.e., the minimal closed set containing $\mathbf{a}^i$) consists of all $\tilde{\mathbf{a}}^i$ in S_1^i whose basic sequence is one of the K cyclical permutations of $\mathbf{c}^i$: any such state can be reached from any other in finitely many steps (for instance, it takes at most $3K-1$ steps to get to a sequence of type E, then at most $K-1$ steps to the appropriate cyclical permutation, and then another $3K$ steps to introduce a delay and wait until it reaches the desired place in the sequence).

Next, the states in S_1^i are aperiodic. Indeed, if the basic sequence $\mathbf{c}^i$ is constant (i.e., $\mathbf{c}^i = (\hat{a}^i, \hat{a}^i, ..., \hat{a}^i)$ for some $\hat{a}^i \in A^i$), then the constant sequence $(\hat{a}^i, \hat{a}^i, ..., \hat{a}^i)$ of length $3K$ is an absorbing state (since the next play of i will always be $\hat{a}^i$ by (*)), and thus aperiodic. If $\mathbf{c}^i$ is not constant, then assume without loss of generality that $c_1^i \neq c_K^i$ (if not, take an appropriate cyclical permutation of $\mathbf{c}^i$, which keeps us in the same minimal closed set). Let $\mathbf{a}^i$ be the sequence of type E that consists of three repetitions of $\mathbf{c}^i$. Starting at $\mathbf{a}^i$, there is a positive probability that $\mathbf{a}^i$ is reached again in K steps, by always making the first choice in (*) (i.e., playing K-periodically, with no delays). However, there is also a positive probability of returning to $\mathbf{a}^i$ in $3K+1$ steps, by always making the first choice in (*), *except* for the initial choice which introduces a delay (after $3K$ additional steps the delay coordinate is no longer part of the state and we return to the original

sequence[21] $\mathbf{a}^i$). But K and $3K+1$ are relatively prime, so the state $\mathbf{a}^i$ is aperiodic. Therefore, every minimal closed set contains an aperiodic state, and all states are aperiodic.

Returning to the original Markov chain (over N-tuples of actions), the product of what we have shown is that, to each combination of basic sequences $(\mathbf{c}^1, \mathbf{c}^2, ..., \mathbf{c}^N)$ whose frequency distributions constitute a Nash 4ε-equilibrium, there corresponds an ergodic set[22] consisting of all states with basic sequences that are, for each i, some cyclic permutation of $\mathbf{c}^i$. The set S_1 is precisely the union of all these ergodic sets. □

The next three claims show that all states outside S_1 are transient: from any such state there is a positive probability of reaching S_1 in finitely many steps.

Claim 2. *Starting from any state in S_2, there is a positive probability that a state in S_1 is reached in at most $2K$ steps.*

Proof. Let $\mathbf{a} \in S_2$. Since at state $\mathbf{a}$ case (***) applies to every player i, there is a positive probability that i will play a^i_{2K+1} (i.e., play K-periodically). If the new state $\mathbf{a}'$ is regular, then for each i the frequency distribution y^i of the resulting basic sequence either equals the frequency distribution z^i of the last K periods, or differs from it by $1/K$ (in the maximum norm). But z^i is a 2ε-best reply to z^{-i}, which implies that y^i is a 4ε-best reply to y^{-i} by (2) — and so $\mathbf{a}' \in S_1$. If the new state $\mathbf{a}'$ is irregular, then $\mathbf{a}' \in S_2$. Again, at $\mathbf{a}'$ there is a positive probability that every player will play K-periodically. Continuing in this way, we must at some point reach a regular state — since after $2K$ such steps the sequence of each player is surely of type E — a state that is therefore in S_1. □

Claim 3. *Starting from any state in S_3, there is a positive probability that a state in S_2 is reached in at most $5K+1$ steps.*

[21] The condition $c^i_1 \neq c^i_K$ is needed in order for the delay action, c^i_K, to be different from the K-periodic action, c^i_1.

[22] I.e., a minimal closed and aperiodic set.

Proof. Let $\mathbf{a} \in S_3$. There is a positive probability that every player will continue to play his basic sequence K-periodically, with no delays (this has probability 1/2 in (*), 1 in (**), and $1/|A^i$ in (***)). After at most $3K+1$ steps, we get a sequence of type E for every player (since all the original delay actions are no longer part of the state). During these steps the basic frequencies y^i did not change, so there is still one player, say player 1, such that y^1 is not a 4ε-best reply to y^{-1}. So case (***) applies to player 1, and thus there is a positive probability that he will next play an action $\hat{a}^1$ that satisfies $1 - y^1(\hat{a}^1) > 1/K$ (for instance, let $\hat{a}^1 \in A^1$ have minimal frequency in y^1, then $y^1(\hat{a}^1) \leq 1/|A^1| \leq 1/2$ and so[23] $1 - y^1(\hat{a}^1) \geq 1/2 > 1/K$). The sequence of every other player $i \neq 1$ is of type E, so with positive probability i plays a^i_{2K+1} (this has probability 1/2 if (*) and $1/|A^i|$ if (***)), and thus the sequence of i remains of type E. With positive probability this continues for K periods for all players $i \neq 1$. As for player 1, note that y^{-1} does not change (since all other players $i \neq 1$ play K-periodically and so their y^i does not change). If at any point during these K steps the sequence of player 1 turns out to be of type E or D, then it contains at least two repetitions of the original basic sequence (recall that we started with three repetitions, and we have made at most K steps), so the basic frequency is still y^1; but y^1 is *not* a 4ε-best reply to y^{-1}, so we are in case (***). If the sequence is not of type E or D then of course we are in case (***) — so case (***) always applies during these K steps, and there is a positive probability that player 1 will always play $\hat{a}^1$.

But after K periods the sequence of player 1 is for sure neither of type E nor D: the frequency of $\hat{a}^1$ in the last K periods equals 1, and in the middle K periods it equals $y^1(\hat{a}^1)$, and these differ by more than $1/K$ (whereas in a sequence of type E or D, any two blocks of length K may differ in frequencies by at most $1/K$). Now these two K-blocks remain part of the state for K more periods, during which the sequence of player 1 can thus be neither E nor D, and so the state is irregular. Hence case (***) applies to all players during these K periods, and there is a positive probability that each player

[23]This is where $K > 2$ is used. Note that $|A^1| \geq 2$, since otherwise player 1 would always be best-replying.

i plays a K-sequence whose frequency is $\bar{y}^i$, where $(\bar{y}^1, \bar{y}^2, ..., \bar{y}^N)$ is a Nash 2ε-equilibrium (see the beginning of the Proof of Proposition 4). So, finally, after at most $(3K+1)+K+K$ steps, a state in S_2 is reached. □

Claim 4. *Starting from any state in S_4, there is a positive probability that a state in $S_1 \cup S_2 \cup S_3$ is reached in at most K steps.*

Proof. At $\mathbf{a} \in S_4$, case (***) applies to every player. There is therefore a positive probability that each player i plays according to a K-sequence with frequency $\bar{y}^i$, where again $(\bar{y}^1, \bar{y}^2, ..., \bar{y}^N)$ is a fixed Nash 2ε-equilibrium. Continue in this way until either a regular state (in S_1 or S_3) is reached, or, if not, then after at most K steps the state is in S_2. □

Combining the four claims implies that almost surely one of the ergodic sets, all of which are subsets of S_1, will eventually be reached. Let T denote the period when this happens — so T is an almost surely finite stopping time — and let $q_T \in S_1$ be the reached ergodic state. It remains to prove (3) and (4).

Let $Q \subset S_1$ be an ergodic set. As we saw in the Proof of Claim 1, all states in Q have the same basic frequency distributions $(y^1, y^2, ..., y^N)$. The independence among players in S_1 implies that $Q = Q^1 \times Q^2 \times ... \times Q^N$, where $Q^i \subset S_1^i$ is an ergodic set for the Markov chain of player i. For each $\mathbf{a}^i \in Q^i$, let $x^i_{\mathbf{a}^i} \in \Delta(A^i)$ be the frequency distribution of all $3K$ coordinates of $\mathbf{a}^i$, then $||x^i_{\mathbf{a}^i} - y^i|| \leq 1/(3K)$ (they may differ when the sequence contains a delay). Let μ^i be the unique invariant probability measure on Q^i; then the average frequency distribution $x^i := \sum_{\mathbf{a}^i \in Q^i} \mu^i(\mathbf{a}^i)\, x^i_{\mathbf{a}^i} \in \Delta(A^i)$ also satisfies $||x^i - y^i|| \leq 1/(3K)$. But $(y^1, y^2, ..., y^N)$ is a Nash 4ε-equilibrium, and so $(x^1, x^2, ..., x^N)$ is a Nash 6ε-equilibrium (by (2)).

Once the ergodic set Q^i has been reached, i.e., $q^i_T \in Q^i$, the probability of occurrence of each state $\mathbf{a}^i = (a^i_1, a^i_2, ..., a^i_{3K})$ in Q^i converges to its invariant probability (see Feller (1968, Section XV.7)):

$$\lim_{t\to\infty} \Pr[a^i(t+k) = a^i_k \text{ for } k = 1, 2, ..., 3K \,|\, q^i_T \in Q^i] = \mu^i(\mathbf{a}^i).$$

Projecting on the k-th coordinate yields, for every $a^i \in A^i$,

$$\lim_{t\to\infty} \Pr[a^i(t) = a^i \mid q_T^i \in Q^i] = \sum_{\mathbf{a}^i \in Q^i : a_k^i = a^i} \mu^i(\mathbf{a}^i),$$

so, in particular, the limit on the left-hand side exists. Averaging over $k = 1, 2, ..., 3K$ yields on the right-hand side $\sum_{\mathbf{a}^i \in Q^i} \mu^i(\mathbf{a}^i)\, x^i_{\mathbf{a}^i}(a^i)$, which equals $x^i(a^i)$; this proves (5), from which (4) follows by independence.

A similar argument applies to the limit of the long-run frequencies Φ_t. This completes the proof of Theorem 5. □

References

Arslan, G., Shamma J. S., 2004. Distributed convergence to Nash equilibria with local utility measurements. *43rd IEEE Conference on Decision and Control*, December 2004.

Aumann, R. J., Sorin, S., 1989. Cooperation and bounded recall. *Games and Economic Behavior* 1, 5–39.

Feller, W., 1968. *An Introduction to Probability and Its Applications*, Volume I, Third Edition. J. Wiley, New York.

Foster, D. P., Vohra, R. V., 1997. Calibrated learning and correlated equilibrium. *Games and Economic Behavior* 21, 40–55.

Foster, D. P., Young, H. P., 2003a. Learning, hypothesis testing, and Nash equilibrium. *Games and Economic Behavior* 45, 73–96.

Foster, D. P., Young, H. P., 2003b. Regret testing: A simple payoff-based procedure for learning Nash equilibrium. University of Pennsylvania and Johns Hopkins University (mimeo).

Germano, F., Lugosi, G., 2004. Global Nash convergence of Foster and Young's regret testing. Universitat Pompeu Fabra (mimeo).

Hart, S., 2005. Adaptive heuristics. *Econometrica* 73, 1401–1430.

Hart, S., Mas-Colell, A., 2000. A simple adaptive procedure leading to correlated equilibrium. *Econometrica* 68, 1127–1150.

Hart, S., Mas-Colell, A., 2003a. Regret-based dynamics. *Games and Economic Behavior* 45, 375–394.

Hart, S., Mas-Colell, A., 2003b. Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review* 93, 1830–1836.

Kakade, S. M., Foster, D. P., 2003. Deterministic calibration and Nash equilibrium. University of Pennsylvania (mimeo).

Jordan, J., 1993. Three problems in learning mixed-strategy Nash equilibria. *Games and Economic Behavior* 5, 368–386.

Shamma, J. S., Arslan, G., 2005. Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Transactions on Automatic Control* 50, 312–327.

Young, H. P., 2004. *Strategic Learning and Its Limits*. Oxford University Press, Oxford.