Stochastic Uncoupled Dynamics and Nash Equilibrium* (Extended Abstract)

Sergiu Hart† Andreu Mas-Colell‡

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 to Nash equilibria, and present a number of possibility and impossibility results. Basically, we show that if in addition to random moves some recall 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, bounded recall

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 (2003) 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, in addition to making random moves, any positive result will also require the players to recall pieces of the past.

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

*This version: November 2004. We thank Dean Foster, Fabrizio Germano, Sham Kakade, Gábor Lugosi, Yossi Rinott, Benji Weiss, and Peyton Young for their comments. 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`.

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 retain 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 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 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 the players at periods t and $t-1$. In general, when Nash equilibria may be mixed, we show that convergence to (approximate) equilibria can be guaranteed using longer, but finite, recall.

Finally, 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. The omitted proofs can be found in the full version of the paper.[1] 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, $\Delta(A^i) = \{\ x^i = (x^i(a^i))_{a^i \in A^i} : \sum_{a^i \in A^i} x^i(a^i) = 1$ and $x^i(a^i) \geq 0$ 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[2] 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*[3] f^i of player i is a sequence of functions $(f^i_1, f^i_2, ..., f^i_t, ...)$, where, for each time t,

[1] `http://www.ma.huji.ac.il/hart/abs/uncoupl-st.html`

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

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

the function f_t^i assigns a mixed action in $\Delta(A^i)$ to each history $(a(1), a(2), ..., a(t-1))$. 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.* 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 will 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.

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 (2003). 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 to pure Nash equilibria in all games where such equilibria exist.*

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 uncoupled, 1-recall, stationary strategy

	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

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 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

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

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 and β we essentially have the game of Jordan

	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

α

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

β

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

γ

Figure 2: A generic three-player game

(1993) (see Hart and Mas-Colell (2003, 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 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 to pure Nash equilibria in every two-player generic game where such equilibria exist.*

The proof is omitted.

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 one more period of recall suffices.

Theorem 3 *There exist uncoupled,* 2*-recall, stationary strategy mappings that guarantee almost sure convergence to pure Nash equilibria 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 again the same action);
- 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 to reach a state in S_1 in finitely many periods. Indeed:

- At each state (a', a) in S_2 all players randomize; hence there is 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 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 randomizing; hence there is 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 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. □

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 also, in a slightly more complex way, 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). The convergence will now 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 will 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 will refer to $(\Phi_t[a])_{a \in A} \in \Delta(A)$ as the *joint distribution of play*,[5] and to $\left(\Phi_t[a^i]\right)_{a^i \in A^i} \in \Delta(A^i)$ as the *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 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 $x = (x^1, x^2, ..., x^N)$ 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. The length of the recall R depends on the precision ε and the bound on payoffs M (as well as on number of players N and the number of actions $|A^i|$).

Proof. Given $\varepsilon > 0$, let K be such that[6]

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

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

[5] Also known as the "empirical (or sample) distribution of play."

[6] 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).

distribution of the last K actions of i, i.e., $z^i(a^i) := |\{K+1 \le k \le 2K : a^i_k = a^i\}/K$ 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.

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 marginal distribution of play will converge to z^i. At a state s in S_2 every player randomizes, so there is positive probability that everyone will play K-periodically, leading in $r = \max\{1 \le k \le K : a_{K+k} \ne a_k\}$ steps to S_1. At a state s in S_3, there is positive probability of reaching S_2 in $K+1$ steps: in the first step the play is some $a \ne a_{K+1}$, 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 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 *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 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 T after which the occurrence probabilities $\Pr[a(t) = a \mid h_T]$, where $h_T = (a(1), a(2), ..., a(T))$, also converge to the same Nash ε-equilibrium x:

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

for every $a \in A$.

The proof of Theorem 5 is relatively intricate and is omitted. T is the time when some ergodic set is reached. Of course, (1) follows from (3). Note that neither (4) nor its marginal implications,

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

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*

$$\Pr[a^i(t) = a^i \mid a(1), a(2), ..., a(t-1)] = f^i(a(1), a(2), ..., a(t-1))(a^i).$$

We next show that this cannot be guaranteed in general.

Theorem 6 *For every small enough*[7] $\varepsilon > 0$, *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.*

The proof is omitted.

5 Discussion and Comments

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

(a) *Convergence*: We emphasize that throughout this paper we have sought a very strong form of convergence, namely, almost sure convergence to a point.[8] 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-equilibrium," and so on. Conceivably, the use of weaker forms of convergence may have a theoretical payoff in other aspects of the analysis. Comments on some of these trade-offs are included in the next two remarks.

(b) *Behavior probabilities*: Theorem 6 has established that we cannot hope to obtain in all generality the almost sure convergence of the actual play (or behavior) probabilities.[9] Can this be obtained for some of the weaker notions of Remark (a), and if so, then at what cost (in terms, for example, of the type of convergence obtained for the sample distribution of play)? The construction of Theorem 5 does not yield good results in this respect: the joint distributions of play, and also the occurrence probabilities, converge in a strong sense, but there is no convergence, even "in probability," of the behavior probabilities. The results of Foster and Young (2003a) are of direct relevance to this point.

(c) *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 $f_t^i(u^i(a(1)), u^i(a(2)), ..., u^i(a(t-1))$ (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 is some kind of convergence in probability as mentioned in Remark (a). On this point see Foster and Young (2003b).

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

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

[9] This convergence is obtained for pure equilibria — see the constructions of Proposition 2 and Theorem 3.

(d) *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.

In terms of results, some of the differences between Foster and Young's work and ours have already been mentioned in previous remarks: we use a stronger notion of convergence, but do not obtain the convergence of the behavior strategies, nor do we handle the "unknown game" situation of Remark (c).

(e) *Finiteness*: In this paper we "make the past finite" by means of a finite recall assumption. This is standard but it is not the only way to do so. One could allow for a finite number of updatable statistics, or for carrying a finite number of "instructions" from the past, or for finite automata. It would be of interest to determine to what extent our conclusions generalize to these environments.

(f) *Adaptive strategies*: 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 (2004) for a survey). Would the positive results of this paper still obtain? Note that such adaptive or monotonicity-like conditions severely restrict the possibilities of "free experimentation" that drive the positive results obtained here.

(g) *Correlated equilibria*: We know that there are uncoupled strategy mappings with the property that the joint distributions of play converge almost surely to the set of correlated equilibria (see Foster and Vohra (1997), Hart and Mas-Colell (2000), Hart (2004), 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; see Remark (e). 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. An issue for further study is to what extent the contrast could be captured by an analysis of the speeds of convergence.

References

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

Foster, D. P. and R. V. Vohra (1997), "Calibrated Learning and Correlated Equilibrium," *Games and Economic Behavior* 21, 40–55.

Foster, D. P. and H. P. Young (2003a),"Learning, Hypothesis Testing, and Nash Equilibrium," *Games and Economic Behavior* 45, 73–96.

Foster, D. P. and H. P. Young (2003b), "Regret Testing: A Simple Payoff-Based Procedure for Learning Nash Equilibrium," University of Pennsylvania and Johns Hopkins University (mimeo).

Germano, F. and G. Lugosi (2004), "Global Nash Convergence of Foster and Young's Regret Testing," Universitat Pompeu Fabra (mimeo).

Hart, S. (2004), "Adaptive Heuristics," The Hebrew University of Jerusalem, Center for the Study of Rationality DP-372 (mimeo); *Econometrica* (forthcoming); http://www.ma.huji.ac.il/hart/abs/heurist.html

Hart, S. and A. Mas-Colell (2000), "A Simple Adaptive Procedure Leading to Correlated Equilibrium," *Econometrica* 68, 1127–1150.

Hart, S. and A. Mas-Colell (2003), "Uncoupled Dynamics Do Not Lead to Nash Equilibrium," *American Economic Review* 93, 1830–1836.

Kakade, S. M. and D. P. Foster (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.

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