Stochastic Uncoupled Dynamics and Nash Equilibrium [1]

Manuscript by Daniel Balle
Seminar Algorithmic Game Theory, ETH Zurich

May 8, 2017

1 Introduction

Many adaptive dynamic games are subject to the natural requirement of *uncoupledness*, which simply restricts the strategies of all players to depend only on their own utility functions. A well-known example of such a strategy is the best-response algorithm, as players decide on their next action based purely on the maximization of their individual payoffs.

However in their previous paper, *Uncoupled Dynamics Do Not Lead to Nash Equilibrium* [2], conveniently named after the impossibility result they established, the authors showed that this simple and natural requirement was sufficient to break the guarantee of the convergence to a Nash Equilibrium. Surprisingly proving that no uncoupled strategy is up to the challenge the following straightforward two-player game is enough, which we will do shorty.

	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 simple two-player game

To address the main reason for this regrettable finding, namely the inability of players to detect equilibria individually due to their decentralized behavior under uncoupled dynamics, we will introduce the concept of *history of play*. More specifically we will study how the past can influence the players strategies and the convergence to Nash Equilibria by considering the following questions:

What if players could remember previous plays?
What if they had memories?

2 Model and Concepts

2.1 Static Setup

Let us first present the very familiar concept of a game in our setting

Definition 1. *A basic static (or one-shot) game can be given in the following strategic form:*

- $N \geq 2$ *players denoted by* $i \in \{1, 2, ..., N\}$
- *A finite set of actions* A^i *for each player* i.
  The set of action combinations is denoted as $A := A^1 \times A^2 \times ... \times A^N$
- *A payoff (or utility) function* $u^i : A \to \mathbb{R}$ *for each player* i.

We then identify a game by its payoff functions $U := (u^1, ..., u^N)$.

As mentioned previously this paper investigates convergence under stochastic dynamics and we will thus extend this definition by introducing the following concept:

Definition 2. *In a game which allows for random moves a player* i *assigns a probability* $x^i(a)$ *to each action* $a \in A^i$. *We call* x^i *the* randomized *or* mixed *action of player* i.

- *The set of all mixed actions for player* i *is* $\Delta(A^i)$
- *We denote by* $\Delta := \Delta(A^1) \times ... \times \Delta(A^N)$ *the set of randomized action combinations or* N*-tuples.*
- *The payoff function is multi-linearly extended to* $u^i : \Delta \to \mathbb{R}$

The concepts of best-replying and Nash equilibria naturally extend to this stochastic setting:

Definition 3. *We say that the randomized actions* $x^i \in \Delta(A^i)$ *is an* ϵ-best reply *to* $x^{-i} := (x^1, ..., x^{i-1}, x^{i+1}, ..., x^N)$ *if for all* $y^i \in \Delta(A^i)$*:*

$$u^i(x) \geq u^i(y^i, x^{-i}) - \epsilon$$

Definition 4. *A* Nash ϵ-equilibrium *is a randomized action combination* $\underline{x} = (\underline{x}^1, ..., \underline{x}^N) \in \Delta$ *such that each* $\underline{x}^i$ *is an* ϵ*-best reply to* $\underline{x}^{-i}$.

2.2 Dynamic Setup & History of Play

We will now consider repeated play of U at discrete time periods $t = 1, 2, ...$ and denote by:

- $a^i(t) \in A^i$ the action of player i at time t
- $a(t) = (a^1(t), ..., a^N(t)) \in A$ the action combination of all players at t

Note that the assumption of a round-robin scheme under which most games are played is dropped and all players realize an action at time period t, after which everyone observes the outcome $a(t)$.

As this notations suggests this introduces the very interesting concept of a *history of play*, i.e. players get a sense of time and progression of the game. The past is formally defined as the sequence $(a(1), ..., a(t-1))$ of actions combinations leading up to time t. Let's denote by H^{t-1} the set of all histories of length $t-1$. In the dynamic setup of the game U players can now base their decisions on information from the past.

Definition 5. *The* strategy function $f_t^i : H \to \Delta(A^i)$ *of player* i *at time period* t *assigns a mixed action* $x^i \in \Delta(A^i)$ *to each history* $(a(1), ..., a(t-1)) \in H^{t-1}$

The *strategy* of a player i is thus simply a sequence of strategy functions $f^i := (f_1^i, ..., f_t^i, ...)$. However we will restrict ourselves exclusively to *stationary* dynamics, i.e. the time t has no influence on the players decision and all strategy functions are identical $f^i \equiv f_t^i$ for all t. We will also refer to the output of these strategy functions as *behavior probabilities*, i.e. mixed actions conditioned on the history of play.

To study the convergence to Nash Equilibria under the influence of the past we will first introduce the concept of *recall* which limits the information player may use to decide on their next move to a finite number of play periods immediately preceding the current time period.

Definition 6. *A strategy has* R-recall *if only the last* R *action combinations matter, i.e.* f^i *is of the form* $f^i(a(t-R), ..., a(t-1))$ *for all* $t > R$.

Players have 1-recall strategies when they base their decision exclusively on the outcome of the previous action combination $a(t-1)$, as they do for example under best-response dynamics.

Naturally the strategy of each player also depends on the repeated game induced by $U = (u^1, ..., u^N)$. We call the association of a strategy profile to the payoff functions a *strategy mapping* $f(U) = (f^1(U), ..., f^N(U))$. The second restriction our dynamic setup has to satisfy is *uncoupledness* which dictates that the moves of every player do not depend on the payoff functions of other players. Thus $f^i(U) \equiv f^i(u^i)$.

3 Pure Equilibria

Now that we have formally defined our model we will study the convergence of uncoupled stationary dynamics to Nash Equilibria. We will do so by first considering the simpler scenario of Pure Nash Equilibria $\underline{a} = (\underline{a}^1, ..., \underline{a}^N)$.

We will begin by giving a generalization of the impossibility conclusion established by the authors in their previous paper *Uncoupled Dynamics Do Not Lead to Nash Equilibrium* [2]. However as a follow-up to this rather regrettable result we will then show how allowing for a longer recall past the current situation dramatically improves the outlook on convergence.

3.1 The Bad News

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

Proof. A simple proof by contradiction based on the example given during the introduction suffices to establish this first theorem. Consider thus once again the following two-player game of figure 2.

	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 2: A simple two-player game

Now suppose for sake of contradiction that there exists an uncoupled, 1-recall, stationary strategy mapping f that guarantees convergence to pure Nash Equilibria when these exists. In the game above this would be (γ, γ).

Observation 1. *In each action combination $a(t)$ at least one of the two players is best-replying.*

Based on the above observation we can first show that for such a strategy mapping the following must hold for our two-player game.

Lemma 1. *If player i is best-replying in state $a(t)$ he will play the same move at $t+1$*

Proof. Suppose without loss of generality that in state $a(t)$ player 1 is best-replying. We then create a new game $U' = (u^1, \bar{u}^2)$ by changing only the utility function u^2 of player 2 such that $\bar{u}^2(a(t)) = 2$ and $\bar{u}^2(\gamma, \gamma) = 0$. In this new game $a(t)$ is now the unique pure Nash equilibria. Since our strategy mapping converges and is 1-recall neither player will choose a different action at time $t+1$. Yet by uncoupledness the strategy of player 1 is independent of the utility function of player 2, and thus he will not move in the original game U either. □

Observation 2. *In any action combination $a(t)$ in which only player i plays γ, player i is not best-replying and thus player j is.*

From the above lemma and observation it follows that the state (γ, γ) can never be reached when starting from any other state. Indeed

1. if neither player chose γ only the player currently not best-replying might deviate to γ in the next round, and
2. in any situation in which exactly one player plays γ the opponent is best-replying and thus won't move.

This contradicts our convergence assumption. □

Remark 1. In turns out that we obtain a positive result in any two player game if we require *genericity*, i.e. uniqueness of best replies for each player. The proof of this proposition is omitted here.

[1] A sequence X_n converges **almost surely** to X if $Pr[\lim_{n\to\infty} X_n = X] = 1$

3.2 Not All Is Lost

While the previous impossibility result might seem quite discouraging the simple addition of recall yields a drastic improvement. As we will now demonstrate how having players remember only their previous two action is sufficient to guarantee convergence to a pure Nash equilibria.

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

We will now see the first of three construction proofs, all very similar in nature but increasingly more intricate.

Proof. In the first part of the following proof we will describe a strategy mapping f and then proceed to show that f satisfies all requirements listed above in the second part.

Part 1 A state is now identified as the play of the two previous periods $(a', a) := (a(t-1), a(t)) \in A \times A$. The strategy mapping f^i of each player i is then defined as follows:

- if $a' = a$ and a^i is a best reply of player i to a^{-i} then player i plays the same action a^i;
- otherwise player i picks an action $\bar{a}^i$ uniformly at random from A^i, i.e. he plays a mixed action x^i such that $x^i(a^i) = 1/|A^i|$ for all a^i.

Part 2 To prove that this strategy mapping f does indeed guarantee convergence we will partition the state space $S = A \times A$ into four regions:

$$S_1 := \{(a, a) \in S : a \text{ is a Nash equilibrium } \}$$
$$S_2 := \{(a', a) \in S : a' \neq a \text{ and } a \text{ is a Nash equilibrium } \}$$
$$S_3 := \{(a', a) \in S : a' \neq a \text{ and } a \text{ is not a Nash equilibrium } \}$$
$$S_4 := \{(a, a) \in S : a \text{ is not a Nash equilibrium } \}$$

Note that this strategy mapping then induces a Markov chain over S, as we can assign a probability p to the transitions between any two states s' and s.

Observation 3. *Each state in S_1 is absorbing.*

This observation follows directly from the fact that in any Nash equilibrium $\underline{a}$ all players are best responding and f^i thus dictates that player i plays the same action $\underline{a}^i$.

Lemma 2. *For all states $s \in S_2 \cup S_3 \cup S_4$ there is a strictly positive probability $p > 0$ to reach a state $s' \in S_1$ in finitely many periods.*

We will consider each partition space progressively:

- For all states $(a', a) \in S_2$ each player i randomly picks a new action $\bar{a}^i$ since $a' \neq a$. Thus for every player i there is a positive probability $1/|A^i|$ that he plays the same action $\bar{a}^i = a^i$ again. Since a was a Nash equilibrium the resulting state (a, a) belongs to S_1.

- For all states $(a', a) \in S_3$ each player i again randomly picks a new action $\bar{a}^i$. Thus there is positive probability that all players play a pure Nash equilibrium $\bar{a}$. The resulting sate $(a, \bar{a})$ belongs to S_2 which we have already shown to be transient.
- For all states $(a, a) \in S_4$ at least one player i is not best-replying since a would otherwise be a pure Nash equilibrium. That player i will randomly play an action $\bar{a}^i$ resulting in a state $(a, \bar{a}) \in S_2 \cup S_3$.

Therefore S is an absorbing Markov chain as every state has a positive probability of reaching a state $s \in S_1$ in at most three steps. Once a state $(a, a) \in S_1$ where a is a pure Nash equilibrium is reached the players will continue to play a every period.

□

So we have seen that even extremely simple strategies may guarantee convergence to pure Nash equilibria when we introduce the concept of recall. And surprisingly remembering only the past two periods is already sufficient. Indeed the difficulty of convergence can be mostly attributed to the inability of players to detect equilibria individually as their behavior is decentralized under uncoupled dynamics. However by remembering the past two action combinations players can now observe a pattern and then act in coordinated manner. Even if player i is currently playing a best reply a^i he will nonetheless randomly chose his next action as long as the pattern of the state (a', a) seems to indicate that a global optimum hasn't been reached yet.

Remark 2. We are exclusively studying the possibility of converge, but not the actual rate of convergence. Thus while the previous strategy mapping now yields positive results for the detection of equilibria, the randomized search aspect is less appealing.

4 Mixed Equilibria[2]

Under the existence of pure Nash equilibria we have observed very satisfying results in terms of convergence possibility. However we will now lift this very optimistic assumption and consider the general case of mixed and approximate equilibria $\underline{x} \in \Delta$

4.1 Distributions of Play

Before studying the convergence of the actual behavior probabilities $f^i(\cdot) \in \Delta(A^i)$ we will introduce the following two distributions:

Definition 7. *For a given history of play $(a(1), ..., a(t))$ we denote by Φ_t the empirical frequency distribution for each action combination $a \in A$:*

$$\Phi_t[a] := |\{1 \leq \tau \leq t : a(\tau) = a\}|/t$$

We refer to $(\Phi_t[a])_{a \in A} \in \Delta(A)$ as the joint distribution of play.

[2]section will not be presented during the talk

Definition 8. *Similarly we refer to* $(\Phi_t^i[a^i])_{a^i \in A^i} \in \Delta(A^i)$ *as the marginal distribution of player* i*, where:*

$$\Phi_t^i[a^i] := |\{1 \leq \tau \leq t : a^i(\tau) = a^i\}|/t$$

In the remainder of this section we are interested in studying the convergence of these distributions Φ to Nash ϵ-Equilibria $\underline{x} = (\underline{x}^1, ..., \underline{x}^N) \in \Delta(A)$. However it is important to distinguish between the convergence of the joint distribution of play and the convergence of the marginal distribution of each player.

Definition 9. *For a given* ϵ*-Equilibria* $\underline{x} = (\underline{x}^1, ..., \underline{x}^N)$ *the* induced *joint distribution* $\bar{\Phi}$ *over every action combination* $a = (a^1, ..., a^N) \in A$ *is simply the product of the players action combinations* $\bar{\Phi}[a] = \prod_i \underline{x}^i(a^i)$*.*

Of course if the joint distribution of play converges to the Nash Equilibrium, then so do the marginal distributions. However the opposite implication is false. For this statement to hold our strategy mapping has to guarantee *independence* among the players. Consider the following example which illustrates this difference:

Example 1. Given the history of play $h = ((\alpha, \beta), (\alpha, \gamma), (\beta, \beta), (\alpha, \beta), (\alpha, \gamma))$ for some game U we get a joint distribution of play

$$\Phi_t(\alpha, \beta) = 2/5,\ \Phi_t(\alpha, \gamma) = 2/5, \Phi_t(\beta, \beta) = 1/5$$

while the marginal distributions for player 1 and 2 are respectively

$$\Phi_t^1(\alpha) = 4/5,\ \Phi_t^1(\beta) = 1/5, \Phi_t^1(\gamma) = 0/5$$
$$\Phi_t^2(\alpha) = 0/5,\ \Phi_t^2(\beta) = 3/5, \Phi_t^2(\gamma) = 2/5$$

Now suppose that these correspond exactly to the only Nash Equilibrium $\underline{x} = (\Phi^1, \Phi^2)$ of the game. Thus if our strategy mapping f is defined such that the given history of play is played repeatedly with periodicity 5 we obtain convergence of the marginal distributions of play. However the joint distribution induced by x differs from the joint distribution of play Φ:

$$\bar{\Phi}(\alpha, \beta) = 12/25,\ \bar{\Phi}(\alpha, \gamma) = 8/25,\ \bar{\Phi}(\beta, \beta) = 2/25$$

We can naturally argue that the converge of both distributions of play is preferred. Unfortunately constructing a strategy mapping which guarantees independence of play between players is slightly more complex. We will thus first direct our efforts to the convergence of the marginal distributions.

4.2 Convergence of Marginal Distributions

For the following results we will consider that there is a bound M on the payoffs, i.e. for all players i and all action combinations $a \in A$ it holds that $|u^i(a)| \leq M$.

Theorem 3. *For every* M *and* $\epsilon > 0$ *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 marginal distributions of play* Φ^i *to a Nash* ϵ*-equilibria* $\underline{x} = (\underline{x}^1, ..., \underline{x}^N)$*.*

So for every player i *and every action* $a^i \in A^i$*:*

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

Proof. As for the previous proof we will first describe the construction of a strategy mapping f and then proceed to prove that it will lead to convergence.

Part 1 Given $\epsilon > 0$ and a game U with bounded utility functions we first find some K such that there is a Nash 2ϵ-Equilibrium $\tilde{y} = (\tilde{y}^1, ..., \tilde{y}^N)$ with all probabilities being multiples of $1/K$, i.e. for all players i:

$$\forall a^i \in A^i \, : \, K\tilde{y}^i(a^i) \in \mathbb{N}$$

We can convince ourselves that we can always pick a K large enough that satisfies this requirement. Now given $\tilde{y}$ one can fix a sequence of action combinations $\tilde{a} := (\tilde{a}_1, ..., \tilde{a}_K)$ of length K such that the corresponding marginal distribution of each player i are exactly $\tilde{y}^i$, i.e.

$$\forall a^i \in A^i \, : \, \Phi^i[a^i] = \tilde{y}^i$$

Example 2. Given the following equilibrium $\tilde{y}$ over two players:

$$\tilde{y}^1(\alpha) = 2/5, \, \tilde{y}^1(\beta) = 1/5, \tilde{y}^1(\gamma) = 2/5$$
$$\tilde{y}^2(\alpha) = 1/5, \, \tilde{y}^2(\beta) = 1/5, \tilde{y}^2(\gamma) = 3/5$$

a fixed sequence of action combinations with corresponding marginals would be $\tilde{a} = ((\alpha, \alpha), (\alpha, \beta), (\beta, \gamma), (\gamma, \gamma), (\gamma, \gamma))$

The recall of our strategy mapping is then chosen to be $R = 2K$. Similar to the previous construction proof a state is now identified as a history of play of the previous $2K$ periods : $s := (a_1, a_2, ..., a_{2K})$ with $a_k \in A$.

Definition 10. *A state is K-periodic if for all $k = 1, ..., K$ it holds that $a_{K+k} = a_k$, i.e. $s = (a_1, a_2, .., a_K, a_1, a_2, ..., a_K)$*

Definition 11. *Given the current state s we denote by $z^i \in \Delta(A^i)$ the frequency distribution of the last K actions of each player i :*

$$z^i(a^i) := |\{a^i_k = a^i \,|\, k \in \{K+1, .., 2K\}\}|/K$$

We write $z = (z^1, ..., z^N) \in \Delta$ the corresponding distribution combination over all players. The strategy mapping f^i of each player is then defined 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 the same action he played K and $2K$ periods earlier : $\bar{a}^i := a^i_1 = a^i_{K+1}$;
- otherwise player i picks an action $\bar{a}^i$ uniformly at random from A^i, i.e. he plays a mixed action x^i such that $x^i(a^i) = 1/|A^i|$ for all a^i.

Remark 3. For Pure Nash Equilibria K would be 1 and we obtain the same 2-recall strategy mapping as defined in the previous section.

Part 2 Our state space S defined as all sequences over A of length $2K$ can once again be partitioned into four regions:

$$S_1 := \{s \text{ is } K\text{-periodic and } z \text{ is a Nash } 2\epsilon\text{-equilibrium }\}$$
$$S_2 := \{s \text{ is not } K\text{-periodic and } z \text{ is a Nash } 2\epsilon\text{-equilibrium }\}$$
$$S_3 := \{s \text{ is not } K\text{-periodic and } z \text{ is not a Nash } 2\epsilon\text{-equilibrium }\}$$
$$S_4 := \{s \text{ is } K\text{-periodic and } z \text{ is not a Nash } 2\epsilon\text{-equilibrium }\}$$

Observation 4. *In the Markov chain over S induced by f each state in S_1 is absorbing.*

Indeed once a state $s \in S_1$ is reached in all players are 2ϵ-best replying since z is an equilibria. Therefore the K-periodicity of the state is preserved and the frequency distributions z remain unchanged.

Lemma 3. *For all states $s \in S_2 \cup S_3 \cup S_4$ there is a strictly positive probability $p > 0$ to reach a state $s' \in S_1$ in finitely many periods.*

- For every state $s \in S_2$ all players randomly pick a new action $\bar{a}^i$ since s is not K-periodic. Thus there is a positive probability that all players i play $\bar{a}^i = a^i_{K+1}$ leading to $\bar{a} = a_{K+1}$. In this scenario the frequency distribution z remains the same 2ϵ-equilibrium and the new state s' is still in S_2. There is therefore a positive probability that this behavior repeats and after at most K steps the state becomes K-periodic and we thus reached a state in S_1.

$$a_1, \overbrace{a_2, a_3, ...\mathbf{a}_{K+1}, \underbrace{a_{K+2}, ..., a_{2K}}_{z}, \bar{\mathbf{a}}}^{s}$$

- For every state $s \in S_3$ all players again randomize their action $\bar{a}^i$. Now suppose this leads to some action combination $\bar{a} \neq a_{K+1}$. This guarantees that at least K additional rounds have to be played to reach K-periodicity.

$$..., a_K, \overbrace{\underbrace{\mathbf{a}_{K+1}, ..., a_{2K}}_{K \text{ rounds}}, \underbrace{\bar{\mathbf{a}}, a'_1, .., a'_{K-1}}_{K \text{ rounds}}}^{s}$$

  There is then a positive probability that the next K rounds produce the sequence $\tilde{a}$ we defined earlier, therefore making $z = \tilde{y}$ a Nash 2ϵ-Equilibrium:

$$..., a_{K+1}, \overbrace{a_{K+2}, ..., \bar{\mathbf{a}}, \underbrace{\tilde{a}_1, ..., \tilde{a}_K}_{z=\tilde{y}}}^{s}$$

  And depending on whether this state is K-periodic or not we are either in S_1 or S_2.

- Finally every state $s \in S_4$ can produce a new state $s' \in S_3$ by simply breaking its K-periodicity with $\bar{a} \neq a_{K+1}$.

Therefore we will eventually always reach an absorbing state $s \in S_1$ and as observed earlier the frequency distribution z of the last K actions will become constant, i.e. it will remain the same each period. The marginal distribution of play of each player thus converges to z:

$$\lim_{t \to \infty} \Phi^i_t[a^i] = z^i(a^i)$$

□

This strategy mapping is simply a generalization of the construction seen for Pure Nash Equilibria. And just as before players get "synchronized" by acting in a coordinated manner once a K-periodic state with an equilibrium is reached. This breaks the independence between players required for the convergence of the joint distribution of play to the Nash Equilibrium.

4.3 Convergence of Joint Distribution

It turns out we can however extend the possibility results of convergence for marginals to joint distributions by proving the existence of a strategy mapping f which dictates independent behaviors for each player.

Theorem 4. *For every M and $\epsilon > 0$ 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 distribution of play Φ to the induced distribution of a Nash ϵ-equilibrium $\underline{x} = (\underline{x}^1, ..., \underline{x}^N)$.*

$$\lim_{t\to\infty} \Phi_t[a^i] = \prod_i \underline{x}^i(a^i)$$

In addition the occurrence probability of a given action combination at time t also converges to the same Nash ϵ-equilibrium:

$$\lim_{t\to\infty} Pr[a(t) = a] = \prod_i \underline{x}^i(a^i)$$

In other words, after a sufficient period of time the probability of encountering some action combination a coincides exactly with the overall probability of playing a under a Nash Equilibria $\underline{x}$, which by independence is the product of each players mixed action $\underline{x}^i(a^i)$. Note that this however was not the case for the strategy mapping described in the proof of theorem 3, because of the synchronized behavior which led to a periodic action combination sequence. Let's illustrate this quickly by looking at our familiar example:

Example 3. Because of the periodic play dictated by f the sequence $h = ((\alpha,\beta),(\alpha,\gamma),(\beta,\beta),(\alpha,\beta),(\alpha,\gamma))$ will get repeated indefinitely. As seen earlier the marginals then converge to the Nash Equilibrium x :

$$x^1(\alpha) = 4/5,\ x^1(\beta) = 1/5, x^1(\gamma) = 0/5$$
$$x^2(\alpha) = 0/5,\ x^2(\beta) = 3/5, x^2(\gamma) = 2/5$$

Yet the action combination (β,γ) which has probability $x^1(\beta)\cdot x^2(\gamma) = 2/25$ of being played in our Nash Equilibrium has zero occurrence probability $Pr[a(t) = (\beta,\gamma)] = 0$, i.e. it will never happen.

Proof. We will present an abbreviated version of the proof as it is quite long and intricate. The main idea behind the construction of this strategy mapping is to introduce small *random perturbations* every now and then for every player independently. We thus avoid periodicity and synchronization between the players and guarantee independent behavior of play.

Part 1 For a given $\epsilon > 0$ and a game U with bounded utility functions we once again find some K as in the previous proof. The recall of our strategy

mapping is then chosen to be $R = 3K$. The intuition behind this assignment is that players should still be able to recognize some basic periodic play even with the introduction of random errors.

A state is then defined as a history of play $s := (a_1, ..., a_{3K}) \in A \times ... \times A$ of length $3K$, and we write $s^i := (a_1^i, .., a_{3K}^i) \in A^i \times ... \times A^i$ the corresponding action sequence of player i. We then formally identify these random perturbations as repeated or "delayed" actions.

Definition 12. *Given a sequences of actions $s^i := (a_1^i, ..., a_{3K}^i)$ of length $3K$ of some player i we can sometimes identify it as one of the following two special types:*

- **Type E** *("Exact"): the sequence is K-periodic, i.e. $a_{K+k}^i = a_k^i$ forall $k \in \{1, ..., 2K\}$ and consists of a repeated* basic *sequence $c^i := (a_1^i, ..., a_K^i)$.*
- **Type D** *("Delayed"): there is some delay $a_d^i = a_{d-1}^i$ such that if a_d^i were to be dropped the remaining sequence s_{-d}^i would be K-periodic and again consisting of a repeated* basic *sequence c^i.*

For every sequence s^i of type E or D we define the frequency distribution of actions in its basic sequence [3] $c^i := (c_1^i, ..., c_K^i)$, called the *basic frequency distribution* $y^i \in \Delta(a^i)$, as follows:

$$y^i(a^i) := |\{c_k^i = a^i \,|\, k \in \{1, .., K\}\}|/K$$

We write $y = (y^1, ..., y^N) \in \Delta$ the basic distribution combination over all players.

Remark 4. This will look familiar to the more attentive readers as it draws parallels to the frequency distribution z^i of a player i over his last K actions from the previous proof - which we will also use further down.

We say that a state s is *regular* if for all players i the corresponding action sequence s^i is either of type E or D, and *irregular* otherwise. The strategy mapping f^i of each player i is then defined as follows :

- if the current state s is regular, the basic frequency distribution y^i is a 4ϵ-best reply to y^{-i} and s^i is of type **E**, then
  - with probability $1/2$ player i continues his K-periodicity by playing $\bar{a}^i := a_1^i = a_{K+1}^i = a_{2K+1}^i$,
  - and with probability $1/2$ he produces a **delay** by repeating his previous action $\bar{a}^i := a_{3K}^i$.
- if the current state s is regular, the basic frequency distribution y^i is a 4ϵ-best reply to y^{-i} and s^i is of type **D**, then player i also continues his K-periodic play by playing either a_K^i or a_{K+1}^i depending on where the delay is.
- otherwise player i picks an action $\bar{a}^i$ uniformly at random from A^i

[3] the basic sequence c^i of any sequence s^i is uniquely defined (proof omitted here).

Part 2 Using both z and y our state space S defined as all sequences over A of length $3K$ can once again be partitioned into four regions:

$$S_1 := \{s \text{ is regular and } y \text{ is a Nash } 4\epsilon\text{-equilibrium }\}$$
$$S_2 := \{s \text{ is irregular and } z \text{ is a Nash } 2\epsilon\text{-equilibrium }\}$$
$$S_3 := \{s \text{ is regular and } y \text{ is not a Nash } 4\epsilon\text{-equilibrium }\}$$
$$S_4 := \{s \text{ is irregular and } z \text{ is a Nash } 2\epsilon\text{-equilibrium }\}$$

In a fashion which should be quite familiar by now we would then use the following two claims to complete our proof:

Claim 1. *All states $s \in S_1$ are ergodic, i.e. there is a nonzero probability of exiting the state and the probability of an eventual return to it is 1.*

Claim 2. *All states s in $S_2 \cup S_3 \cup S_4$ have a positive probability of reaching a state $s' \in S_1$ in finitely many steps.*

□

5 Behavior Probabilities

The previous section established the convergence of the occurrence probability $Pr[a(t) = a]$ to Nash ϵ-Equilibria, which also implies the convergence of the marginal occurrence probabilities:

$$\lim_{t\to\infty} Pr[a^i(t) = a^i] = \underline{x}^i(a^i)$$

These were however unconditioned on the history of play. We will thus try to extend our possibility results one step further and study the convergence of the actual *behavior probabilities* as assigned by the strategy functions $f^i : H \to \Delta(A^i)$ of each player i:

$$Pr[a^i(t) = a^i | a(1), ..., a(t-1)] \equiv f^i(a(t-R), ..., a(t-1))(a^i)$$

Under finite R-recall of course these probabilities are conditioned on only the previous R actions. Unfortunately our series of encouraging results seem to come to an end here as we will show that the convergence to the behavior probabilities cannot be guaranteed in general.

Theorem 5. *For every small enough $\epsilon > 0$, there are no uncoupled, finite recall, stationary strategy mappings f that guarantee in every game, the almost sure convergence of the behavior probabilities to Nash ϵ-equilibria.*

Proof. The silver lining of this rather regrettable impossibility result is that at least a simple proof by contradiction is a nice change of pace given the three previous construction proofs. Thus consider the following two games:

Observation 5. *Game U has a unique, completely mixed Nash equilibrium for $\underline{x}^1 = \underline{x}^2 = (0.5, 0.5)$ while $\underline{a} = (\alpha, \alpha)$ is the unique pure Nash equilibrium of game U'.*

	α	β
α	$1,0$	$0,1$
β	$0,1$	$1,0$

(a) Game U

	α	β
α	$1,1$	$0,0$
β	$0,1$	$1,0$

(b) Game U'

For sake of contradiction assume that there is an uncoupled, R-recall, stationary strategy mappings f that guarantee in every game, the almost sure convergence of the behavior probabilities. Therefore in both games the action α of either player will be assigned a positive probability $x^i(\alpha) > 0$. And hence the state $s = (\underline{a}, ..., \underline{a})$ of length R will have positive probability of occurring after a significant time T in game U and U'.

Once in this state s the behavior probabilities $f^i(s)$ should be close to the the actual Nash Equilibria $\underline{x}$ and $\underline{a}$ of respectively U and U'.

Observation 6. *The utility function u^1 of player 1 is the same in both games.*

However since f is an uncoupled strategy mapping, the above observation implies that the behavior probability $f^1(s)(\alpha)$ of player 1 should then be identical in both games. This contradicts our assumption that f converges to the Nash Equilibria. □

6 Memory - Beyond Recall

In light of this unfortunate impossibility result we naturally ask ourselves if we can obtain convergence by giving our players even more abilities while still guaranteeing uncoupled strategies.

Finite recall implied that the distant past was irrelevant and only the last R periods were taken into consideration when deciding on the next mixed action. However it is not unreasonable to believe that certain periods might be much more impactful than others, and thus even if their occurrence is far beyond the horizon of our recall, may still strongly influence a players decisions - very much like human behavior.

We can thus imagine lifting this continuity restriction while still maintaining the desire to limit a players recollection of the past. We are therefore effectively introducing the concept of *memory*.

Definition 13. *A player's strategy f^i has finite R-memory if it can be implemented by a finite-state automaton in $|A|^R$ states.*

More specifically at every period t the input to the "strategy" automaton would be the action combination $a(t) \in A$ and its output would be the mixed action $x^i \in \Delta(A^i)$ to be played during period $t+1$. And indeed allowing players to have "memories" leads to a positive result.

Theorem 6. *For every M and $\epsilon > 0$ 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 a Nash ϵ-equilibrium $\underline{x} = (\underline{x}^1, ..., \underline{x}^N)$:*

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

Proof. Define K as in the proof for theorem 3 for the convergence of the marginal distributions of play. Our strategy mapping is then chosen to have $R := 2K + 1$ memory. Thus a state for some player i is a sequence $\tilde{s}^i = (a_0, a_1, ...a_{2K})$ of any arbitrary action combinations $a_k \in A$ which the player decides to remember, not necessarily the previous $2K + 1$ periods of play. We could also call these his *memories*. Our strategy mapping will be constructed in such a way that all players share the same memories. All players will thus always be in the same state $\tilde{s}^i = \tilde{s}$. We then denote by s the most recent $2K$ memories of $\tilde{s}$, i.e. $\tilde{s} = (a_0, s)$. And similarly to previous proofs we identify by $z^i \in \Delta(A^i)$ the frequency distribution of actions from player i over the K most recent memories in $\tilde{s}$:

$$z^i(a^i) := |\{a_k^i = a^i \,|\, k \in \{K+1, .., 2K\}\}|/K$$

Depending on the state $\tilde{s}$ all players currently share, the strategy mapping f^i, and in particular the state-updating rule for the associated automaton is then defined as follows:

- $\tilde{s}$ is not K-periodic (**Mode I**)
  - if s is K-periodic and z^i is a 2ϵ-best reply to z^{-i}, then player i plays the same action he played K and $2K$ periods earlier : $\bar{a}^i := a_1^i = a_{K+1}^i$, i.e. *he continues his periodic play.*
  - if s is K-periodic and z^i is not a 2ϵ-best reply to z^{-i}, then player i randomizes $\bar{a}^i$ uniformly over $A^i \setminus \{a_1^i\}$, i.e. *he breaks his periodic play.*
  - if s is not K-periodic player i picks an action $\bar{a}^i$ uniformly at random from A^i.

  After having observed all realized actions $\bar{a}$ the new state of f^i becomes $\bar{s} = (s, \bar{a}) = (a_1, ..., a_{2K}, \bar{a})$.
- $\tilde{s}$ is K-periodic (**Mode II**)
  Player i plays the mixed action z^i and the state remains the same $\bar{s} = \tilde{s}$.
- All automata are initialized with some state $\tilde{s}$ which is not K-periodic.

Observation 7. *Once the strategy mapping reaches Mode II the player will play the same mixed action z^i forever.*

Therefore all players should only enter the second mode, i.e. have a K-periodic state $\tilde{s}$ once $z = (z^1, ..., z^N)$ corresponds to the Nash equilibria $\underline{x}$. However for $\tilde{s}$ to become K-periodic, first s has to be K-periodic.

$$\tilde{s} = a_0, \overbrace{\underbrace{a_1, ..., a_K}_{K \text{ rounds}}, \underbrace{a_{K+1}, .., a_{2K}}_{K \text{ rounds}}}^{s}$$

Given that all players will update their state to be $\bar{s} = (s, \bar{a}) = (a_1, ..., a_{2K}, \bar{a})$, K-periodicity can only be achieved if $\bar{a} = a_1$, i.e. all players simultaneously continued their periodic play of s. Yet each player will only display this behavior if his z^i was a 2ϵ-best reply to z^{-i}. Otherwise at least one player will have purposely broken the periodic play. Therefore all players will enter Mode II only if they are all best-replying, thus making z a Nash Equilibrium. □

Remark 5. So all players will remember the last action combinations until the Nash equilibria is reached, at which point this becomes the only memory they hold on to.

7 Closing

Throughout this paper we have now explored the influence of the past on the decisions of players and how it affects the convergence to Nash Equilibria by addressing the inability of players to detect them individually due to their decentralized behavior. While surprisingly a simple 2-recall was sufficient to guarantee the convergence to pure equilibria with a random search and discovery strategy, we had to give players the ability to remember arbitrary plays to extend this result to mixed equilibria.

References

[1] Sergiu Hart, Andreu Mas-Colell. *Stochastic Uncoupled Dynamics and Nash Equilibrium*. Journal of Economic Literature (2004)

[2] Sergiu Hart, Andreu Mas-Colell. *Uncoupled Dynamics Do Not Lead to Nash Equilibrium*. The American Economic Review (2003)