האוניברסיטה העברית בירושלים

THE HEBREW UNIVERSITY OF JERUSALEM

COMPLETELY UNCOUPLED DYNAMICS AND NASH EQUILIBRIA

By

YAKOV BABICHENKO

Discussion Paper # 529 January 2010

מרכז לחקר הרציונליות

CENTER FOR THE STUDY OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Completely Uncoupled Dynamics and Nash Equilibria*

Yakov Babichenko†

December 20, 2009

Abstract

A *completely uncoupled* dynamic is a repeated play of a game, where each period every player knows only his action set and the history of his own past actions and payoffs.

One main result is that there exist *no* completely uncoupled dynamics with finite memory that lead to pure Nash equilibria (PNE) in almost all games possessing pure Nash equilibria. By "leading to PNE" we mean that the frequency of time periods at which some PNE is played converges to 1 almost surely.

Another main result is that this is not the case when PNE is replaced by "Nash ε-equilibria": we exhibit a completely uncoupled dynamic with finite memory such that from some time on a Nash ε-equilibrium is played almost surely.

*This paper is part of the Ph.D. research of the author at the Hebrew University of Jerusalem. The author wishes to thank his supervisor, Sergiu Hart, for his support and guidance, and Itai Arieli and Ron Peretz for useful comments and discussions.

†Center for the Study of Rationality and Department of Mathematics, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel. E-mail: yak@math.huji.ac.il

1 Introduction

An *uncoupled dynamic,* introduced by Hart and Mas-Colell [8] and [9], is the repeated play of a one-shot game, where the strategy of each player does not depend on the payoff functions of the other players. The problem of convergence of uncoupled dynamics to equilibrium has been studied to some extent. There are several reasonable uncoupled dynamics that lead the play to Nash equilibrium, as, for example, the hypothesis testing of Foster and Young [3] and the public learning process introduced by Foster and Kakade [3].

Another class of dynamics that can lead to Nash equilibrium is exhaustive search. The idea is to go through all the possible actions, whether in some order (deterministic exhaustive search) or randomly (probabilistic exhaustive search), until the players reach a state where they all are sure what the Nash equilibrium is, and they play it from then on. Hart and Mas-Colell [9] showed that convergence to Nash equilibria (pure or ε-equilibria) can be guaranteed by using finite-memory exhaustive search strategies.

Negative results (i.e., the impossibility of convergence to Nash equilibria) for uncoupled dynamics have been studied by Hart and Mas-Colell for continuous-time dynamics [8] and in for discrete-time dynamics [9] (see also Foster and Young [4] and Young [10]).

A completely uncoupled dynamic[1] is the repeated play of a game, where the strategy of each player depends only on his *own* past payoffs. The assumption is that a player knows neither his payoff function nor the actions played by other players; indeed, he doesn't even know how many players there are in the game. There are several completely uncoupled dynamics

[1] This concept is also called *radically uncoupled dynamic*, or *payoff-based dynamic*.

that lead to Nash equilibrium, including the following:

- Arslan, Marden, Shamma, and Young [1] deal with *acyclic games*, and show several completely uncoupled strategies that lead to Nash equilibrium in this class of games.

- Regret-testing strategy (Foster and Young [6], Germano and Lugosi [7]) is based on the idea that each player chooses some mixed action, plays it $1-\varepsilon$ of the time in a block of size k, and this receives block an average payoff of u. In ε of the time the player tests whether the mixed action that he plays is "good." A good action is in which $u+\tau$ is greater than a payoff that he would have gotten had he played any pure action constantly, where τ is the *tolerance level*. The test is based on the player's own payoffs only (and thus there is complete uncoupledness). If his action is indeed good, the player sticks with the same mixed action; otherwise, he chooses a new one randomly. If we assume that the player can record in memory any payoff in the game using finite memory, then regret-testing is a finite-memory strategy.[2] Foster and Young [6] showed that Nash ε-equilibria will be played with frequency of at least $1-\varepsilon$. in every two-person game, if both players use the regret-testing strategy. Germano and Lugosi [7] generalized the regret-testing strategy to multi-player games, but restricted the games to generic ones.

- Interactive trial-and-error learning (Young [11]) is based on classifying each player's situation into one of four modes: content, discontent, watchful, and hopeful. For each mode there is an appropriate behavior

[2] The regret testing is finite-recall dynamic, that is even stronger than finite-memory (see [9]).

of a player, for example, when a player is content he plays the same action with high probability. Transitions between modes are caused by the player's own payoffs only (and thus there is complete uncoupledness). If we assume that a player can record payoffs in finite memory then the interactive trial-and-error learning is a finite-memory strategy. Young [11] shows that this strategy guarantees that a pure Nash equilibria will be played with frequency of at least $1-\varepsilon$ in every generic multi-person game with such an equilibrium.

Assume that f_ε is a finite-memory strategy, such that for every $\varepsilon > 0$, f_ε guarantees convergence to some solution in $1 - \varepsilon$ of the time. In addition, assume that the solution is reached with frequency $1 - \varepsilon$ with probability $1 - p(\varepsilon)$ after $t(\varepsilon)$ steps, where $p(\varepsilon) \underset{\varepsilon \to 0}{\to} 0$. Let $\varepsilon_n \underset{n \to \infty}{\to} 0$; then a player can play the strategy f_{ε_1} for $t(\varepsilon_1)$ steps; afterwards he plays the strategy f_{ε_2} for $t(\varepsilon_2)$ steps, and so on. This procedure, called *annealing,* will guarantee convergence to the solution with limit frequency 1; but it is no longer a finite-memory strategy.

Using annealing on regret-testing or trial-and-error learning, we can construct infinite-memory strategies that guarantee convergence to mixed Nash or pure Nash equilibria (respectively) with limit frequency 1.

We can see that there is a gap between the uncoupled dynamics and the completely uncoupled dynamics when we restrict the strategies to finite memory:

1. The convergence to either pure Nash equilibria or Nash ε-equilibria can be guaranteed for general games in the uncoupled case, whereas in the completely uncoupled case only strategies that guarantee convergence for generic games are known. A natural question is whether we can guarantee

convergence for general games.

2. Convergence almost all the time to either pure Nash equilibria or Nash ε-equilibria can be guaranteed by finite-memory strategies in the uncoupled case, whereas in the completely uncoupled case only strategies that guarantee the convergence in $1-\varepsilon$ of the time are known. A natural question is whether we can improve the convergence to almost all the time.

In this paper we answer the above questions for finite-memory strategies. In particular, our goal is to formulate the minimal set of necessary conditions that guarantee convergence to Nash equilibria in a completely uncoupled model. In trying to answer questions 1 and 2 as formulated above, we show that for finite-memory completely uncoupled dynamics the following is true:

For pure Nash equilibria:

1. The assumption of "generic game" is necessary for convergence to pure Nash equilibria even for limit frequency $1-\varepsilon$ (Theorem 11).

2. Convergence to pure Nash equilibria (with limit frequency 1) *cannot* be guaranteed in generic games (Corollary 3).

For Nash ε-equilibria:

1. For generic games the convergence to Nash ε-equilibrium with limit frequency 1 *can* be guaranteed (Theorem 13), but it *cannot* be guaranteed for general games (Theorem 15).

2. Convergence to Nash ε-equilibria with limit frequency 1 *can* be guaranteed in generic games (Theorem 12), but it *cannot* be guaranteed for general games.

We can summarize these results in two tables:

	Established Results	Optimal Results
Pure Nash	$\left\{\begin{array}{c}\text{frequency } 1-\varepsilon\\ \text{and}\\ \text{generic games}\end{array}\right\}$ [3]	$\left\{\begin{array}{c}\text{frequency } 1-\varepsilon\\ \text{and}\\ \text{generic games}\end{array}\right\}$ [4]
ε-Nash	$\left\{\begin{array}{c}\text{frequency } 1-\varepsilon\\ \text{and}\\ \text{generic games}\end{array}\right\}$ [5]	$\left\{\begin{array}{c}\text{frequency } 1\\ \text{and}\\ \text{generic games}\end{array}\right\}$ [6] or $\left\{\begin{array}{c}\text{frequency } 1-\varepsilon\\ \text{and}\\ \text{general games}\end{array}\right\}$

In particular it follows that trial-and-error learning is optimal (for pure Nash equilibria), while regret-testing is not optimal (for Nash ε-equilibria). Indeed, in the latter case we can improve the type of convergence to limit frequency 1, or we can improve the generality of the solution to general games. But we cannot improve both of them simultaneously.

Another way to summarize the results of the paper is indicated in the following table, which presents those conditions where convergence to Nash equilibrium can be guaranteed and those where it cannot.

The notations of the table are as follows:

P- pure Nash equilibrium.

A- approximated Nash equilibrium (ε-equilibrium).

[3] Learning by trial and error.

[4] Theorem 11 and Corollary 3.

[5] Regret-testing.

[6] Theorems 12, 13, and 15.

*- completely uncoupled strategies with additional information of player's index (see Section 4.1).

$\times$- impossible.

$\surd$- possible.

	General Games		Generic Games	
Frequency 1	P: $\times$	P*: $\times$	P: $\times$	P*: $\surd$
	A: $\times$	A*: $\surd$	A: $\surd$	A*: $\surd$
Frequency $1-\varepsilon$	P: $\times$	P*: $\times$	P: $\surd$	P*: $\surd$
	A: $\surd$	A*: $\surd$	A: $\surd$	A*: $\surd$

In Section 2 we introduce the model and notations. In Section 3 we give an impossibility theorem for Nash equilibrium (pure or mixed) in generic games. In Section 4 we present the positive results for pure Nash equilibrium in generic games. Section 5, concerns pure Nash equilibrium in general games. In Section 6 we present the results for Nash ε-eqilibrium. Proofs are relegated to Section 7. Section 8 concludes.

2 The Model

In this section we mainly present our notations and define the objects and concepts that are used in the paper. Some of them are standard, but we recall them for the convenience of the reader.

2.1 The Game

A basic static (one-shot) game Γ is given in strategic form as follows. There are $n \geq 2$ players, denoted by $i = 1, 2, ..., n$. $N = \{1, 2, ..., n\}$ is the set of all the players. C is a countable set of all the possible actions of the players.

Each player i has a finite set of pure actions $A^i = \{a_1^i, a_2^i, ..., a_{m^i}^i\} \subset C$ where $|A^i| \geq 2$; let $A := A^1 \times A^2 \times ... \times A^n$ be the set of action profiles, which will be called the *action set* for short. Let $\mathcal{B}$, be the set of all action sets for a single player, and $\mathcal{A}$ the set of all the action profile sets.

The *payoff function* (or utility function) of player i is a real-valued function $u^i : A \to \mathbb{R}$. $u = (u^1, u^2, ..., u^n)$ is the payoff functions profile which will be called a payoff function for short. For convenience we would like to identify a game Γ with its payoff function u.

The set of mixed actions of player i is $\Delta(A^i)$, that is the probability simplex over A^i.

The payoff functions u^i are multilinearly extended from A to $\Delta(A)$:

$$u^i : \Delta(A^1) \times \Delta(A^2) \times ... \times \Delta(A^n) \to \mathbb{R}$$

For a given A let U_A^i be the set of all the payoff functions of player i (bounded by M), and $U_A = U_A^1 \times U_A^2 \times ... \times U_A^n$ the set of all the payoff function profiles; i.e., U_A is the set of all games with action set A. Let $\mathcal{U}$ be the set of all the games (with every possible action set)

$$\mathcal{U} = \bigcup_{A \in \mathcal{A}} U_A$$

and put $\mathcal{U}_n$ for the set of all the games with n players.

The actions of all the players except player i is $a^{-i} = (a^1, ..., a^{i-1}, a^{i+1}, ..., a^n)$. Similarly, $A^{-i} = A^1 \times ... \times A^{i-1} \times A^{i+1} \times ... \times A^n$ is the set of actions of all the players except player i.

An action $a^i \in A^i$ will be called a *best reply to* a^{-i} if $u^i(a^i, a^{-i}) \geq u^i(\overline{a}^i, a^{-i})$ for every $\overline{a}^i \in A^i$. A *pure Nash equilibrium* is an action profile $a = (a^1, a^2, ..., a^n) \in A$, such that a^i is a best reply to a^{-i} for all i. The set of all pure Nash equilibria is PNE. Similarly, for a mixed actions profile

$x = (x^1, x^2, ..., x^n)$, $x^i \in \Delta(A^i)$, we say that x^i is an *ε-best reply to* x^{-i} if $u^i(x^i, x^{-i}) \geq u^i(y^i, x^{-i}) - \varepsilon$ for every $y^i \in \Delta(A^i)$. $x = (x^1, x^2, ..., x^n)$ is *Nash ε-equilibrium,* if for every $i \in N$, x^i is an ε-best reply to x^{-i}.

2.2 The Dynamic Setup

The dynamic setup consists of the repeated play, at discrete times $t = 1, 2, ...$, of the static game Γ. Let $a^i(t) \in A^i$ denotes the action of player i at time t, and put $a(t) = (a^1(t), a^2(t), ..., a^n(t)) \in A$ for the combination of actions at time t.

We assume that at the end of time t each player i observes the action that he played $a^i(t)$ and his own payoff $u^i(a(t))$. At the time t player i knows his previous acts and payoffs $o^i(t) = ((a^i(t'))_{t'=1}^t, (u^i(a(t'))_{t'=1}^t)$, which will be called the *observations sequence of the player.* Let O_{A^i} be the set of all the possible observations sequences of a player with action set A^i.

The *history of play* in a game with action profile set A is $h(t) = (a(1), a(2), ..., a(t))$, where $a(t') \in A$ for every $t' \leq t$. Let $H_{t,A}$ be the set of all the histories of play of t steps, and $H_A^* := \bigcup\limits_{t=0}^{\infty} H_{t,A}$.

2.3 Strategy Mappings

A *completely uncoupled strategy of a player with actions set B*, is a Borel measurable[7] mapping $f_B : O_B^* \to \Delta(B)$ that assigns a mixed action for every possible observations sequence of the player. Denote by F_B the set of all the completely uncoupled strategies of player with actions set B. $\mathcal{F}$ is the set of all the completely uncoupled strategies (for all the actions sets).

[7] Borel measurability is a technical assumption that is needed to be able to talk about some measures that the mapping induces.

A strategy f is called a *finite-memory strategy* if it can be implemented by a finite automaton (for formal definition of finite automata strategies, see [2]). An equivalent way to think about finite-memory strategies is as follows: The strategy has a finite number of memory bits that determine (uniquely) the action of the player in the next step. In each step the played action and the received action change the memory state.

Before the game starts the knowledge of each player is his own action set, so we define a *completely uncoupled strategy mapping* $\varphi : \mathcal{B} \to \mathcal{F}$ to be a mapping that assigns a completely uncoupled strategy $\varphi(A^i) = f_{A^i} \in F_{A^i}$ for every actions set $A^i \in \mathcal{B}$. For every given strategy mapping φ in a game with action profile set $A = A^1 \times A^2 \times ... \times A^n$, the strategies of the players will be $(f_{A^1}, f_{A^2}, ..., f_{A^n})$. Let $f = (f_{A^1}, f_{A^2}, ..., f_{A^n})$ denote the *strategy profile.* The strategy profile defines a probabilistic play of the game.

A history of play $h(t)$ will be called *realizable* by a strategy profile f if after t steps of play, according to the strategy profile f, the probability that the history will be $h(t)$ is positive.

2.4 Types of convergences

Given a strategy profile f that induces a probabilistic play of the game, we will say that f *leads to* PNE/Nash ε-equilibria if for $t \to \infty$ the frequency of times when a pure Nash/ Nash ε-equilibrium is played, converges to 1, with probability 1. Similarly, we will say that f *leads to* PNE/Nash ε-equilibria $1 - \varepsilon$ of the time if for $t \to \infty$ the frequency of times when a pure Nash/Nash ε-equilibrium is played, is larger than $1 - \varepsilon$, with probability 1.

A strategy mapping φ *leads to* PNE/Nash ε-equilibria if $f = (\varphi(A^i))_{i=1}^n$ leads to PNE/ Nash ε-equilibria in every game.

2.5 Genericity

For every game u with n players and a set of actions A, we can consider u^i as an element of $[-M, M]^{|A|}$, and u as an element of $[-M, M]^{n|A|}$. Therefore, we can define Lebesgue measure $\lambda(\Omega)$ of game set Ω as a measure in $\mathbb{R}^{n|A|}$ [8]. In the same way we define the measure of a set $\Omega \subset U^i$ or $\Omega \subset U^{-i}$.

We will say that a certain property is valid *in almost every game with action profile set* A, if the property holds for all games with action profile set A except a subset of games with measure 0. We will say that a certain property is valid *in almost every game* if for every $A \in \mathcal{A}$ the property is valid in almost every game with action profile set A.

Similarly a property is valid *in all games except a set of games with a measure smaller than* ε, if it is true for every action set $A \in \mathcal{A}$.

3 Impossibility Result for Nash Equilibria in Generic Games

The following negative results shows that using finite-memory strategies the convergence of Nash equilibria (pure or mixed) with limit frequency 1 cannot be guaranteed even in generic games.

Theorem 1 *Let $A = A^1 \times A^2 \times ... \times A^n$ be an action profile set such that $A^1 = A^2$. Then there is no completely uncoupled mapping into finite-memory strategies leading to a pure Nash equilibrium in almost every game with action profile set A that possesses at least one pure Nash equilibrium, and also in almost every game with action profile set A^{-1} that possesses at least one pure Nash equilibrium.*

[8] Below measure will be understood as Lebesgue measure.

Theorem 2 *Let $A = A^1 \times A^2 \times ... \times A^n$ be an action profile set such that $A^1 = A^2$. Then there is no completely uncoupled mapping into finite-memory strategies leading to a Nash equilibria in almost every game with action profile set A, and also in almost every game with action profile set A^{-1}.*

In the proof of the theorems we construct a set of games with a single Nash equilibrium that is pure, where the strategy mapping fails. So the same proof holds for both Theorems 1 and 2.

From these theorems there immediately follows:

Corollary 3 *There is no completely uncoupled mapping into finite-memory strategies that leads to a pure Nash equilibria in almost every game where such an equilibrium exists.*

Corollary 4 *There is no completely uncoupled mapping into finite-memory strategies that leads to a Nash equilibria in almost every game.*

This theorem shows that in terms of the type of the convergence, the trail and error learning strategy (see [11]) achieves the optimal result; i.e., it leads to PNE in $1 - \varepsilon$ of the time.

4 Possibility Results for Pure Nash Equilibrium in Generic Games

If one would like to formulate a positive statement about convergence to PNE in the case of finite-memory strategies, then the strategies should be able to record (remember) some relevant parameters of the game. So we restrict the class of the games to the following:

First, the finite-memory strategies should be able to remember the whole game, so we assume:

(1) *The number of players is bounded by a constant P.*

(2) *The number of actions of every player is bounded by a constant T.*

Second, we would like the strategies to be able to record any payoff from the past. Here two problems can arise. The first is that the payoff could be too big, bigger than the memory size. To avoid it we assume:

(3) *The payoffs of the players are bounded from above by some constant M.*

The second problem is the impossibility of finite-memory strategy to record a general real number in a finite-memory, so there could be a situation where the strategy cannot distinguish between two different payoffs. To avoid this problem we assume that for some small constant $\delta > 0$,

(4) *The payoffs of a player differ from one another by at least δ.*

If (4) is valid then the strategy can record just the first $1 - \log_{10} \delta$ digits after the decimal point, and still distinguish between two different payoffs.

We consider P, T, M as given constants.

Definition 5 *For every $\delta > 0$ we define the class of games $\mathcal{D}_\delta$ to be all the games that satisfy (1)-(4).*

So in $\mathcal{D}_\delta$ a finite-memory strategy is able to record payoffs.

The following lemma shows that for δ small enough, $\mathcal{D}_\delta$ is very close to $\mathcal{U}$.

Lemma 6 *For every $\varepsilon > 0$ there exists $\delta > 0$ such that $\mathcal{D}_\delta$ is all games except a set of games with a measure smaller than ε.*

Proof. For every action set A let C_A be the set of all the games that are not in $\mathcal{D}_\delta$.

$$C_A = \{u \in \mathcal{U} : \exists i \in N, \exists a \neq a', a, a' \in A \text{ such that } |u^i(a) - u^i(a')| < \delta\}$$

For every player $i \in N$, and for every pair of actions $a, a' \in A$,$a \neq a'$, let $E_{\delta,i,a,a'} = \{u \in \mathcal{U} : |u^i(a) - u^i(a')| < \delta\}$. For every $u \in E_{\delta,i,a,a'}$ all the payoffs of all the players except player i at the action a' could be any number in the segment $(-M, M)$, whereas the payoff of player i at the action a' is in the segment $(u^i(a) - \delta, u^i(a) + \delta)$. So there exists a constant $K = K(M, P, T)$ such that $\lambda(E_{\delta,i,a,a'}) \leq K\delta$.

Clearly,

$$C_A = \bigcup_{i \in N,\ a,a' \in A\ a \neq a'} E_{\delta,i,a,a'}$$

and therefore there exists another constant $L = L(M, P, T)$ such that

$$\lambda(C_A) \leq \sum_{i \in N,\ a,a' \in A\ a \neq a'} \lambda(E_{\delta,i,a,a'}) \leq LK\delta$$

K and L are independent of A, and so for every $\varepsilon > 0$ there exists small enough $\delta > 0$ such that $LK\delta < \varepsilon$. For this δ holds $\lambda((DP_{A,\delta})^C) < \varepsilon$ for every A, as required. ∎

Remark 7 *All the positive results for finite-memory strategies in the paper, will be proved for games in the class $\mathcal{D}_\delta$. By Lemma 6 the positive result is proved for all the games except a set of games with measure ε (for every $\varepsilon > 0$).*

Before we formulate the positive results let us note that the negative result of Theorem 1 and Corollary 3 remains true if we consider just games in the class $\mathcal{D}_\delta$.

Proposition 8 *There is no completely uncoupled mapping into finite-memory strategies that leads to a PNE in almost every game in $\mathcal{D}_\delta$ where such an equilibrium exists.*

The proof is identical to that of Theorem 1, which remains true for any positive measure set of games, particularly for $\mathcal{D}_\delta$.

In Theorem 1, we considered games with action profile set $A = A^1 \times A^2 \times ... \times A^n$, such that $A^1 = A^2$. Clearly, the theorem remains valid for action profile set A with any two equal action sets ($A^i = A^j$ for $i \neq j$). We would like to show that if we restrict the class of games to those with different action sets of the players, then the convergence to PNE can be guaranteed.

Theorem 9 *For every $\delta > 0$, there exists a completely uncoupled mapping into finite-memory strategies that leads to a pure Nash equilibrium in every game in $\mathcal{D}_\delta$ with different action sets ,where such an equilibrium exists.*

As was mentioned in Remark 7, the theorem proves that for every $\varepsilon > 0$ the strategy mapping leads to PNE in all games with different action sets, except for a set of games of measure ε.

4.1 Complete uncoupledness with additional information

There exists a basic information of a strategy (a player) before the game starts, namely the domain of the strategy mapping. Beyond the basic information of the players was their action set only. Let us consider the case where there is some additional information for every player. Such information results in the model described above: we must now allow for strategies of players who are dependent not only on the action set.

Let K be the set of all the possible values of information. For example, if the information is the index number of the player, then $K = \mathbb{N}$. Let $\alpha = (\alpha^1, \alpha^2, ..., \alpha^n) : \mathcal{U}_n \to K^n$ be the *information function*, where $\alpha^i(u)$ is the information of player i. In the example above $\alpha^i(u) = i$. An information is called *uncoupled information* if the information is not about the payoff functions of the other players, or formally, $\alpha^i(u_1) = \alpha^i(u_2)$ for every two games $u_1 = (u^i, u^{-i})$ and $u_2 = (u^i, \overline{u}^{-i})$.

Up to now the strategy mapping has been $\varphi : \mathcal{B} \to \mathcal{F}$. Now we want the strategy mapping to be from $\mathcal{B} \times K$ to $\mathcal{F}$. Let $\varphi : \mathcal{B} \times K \to \mathcal{F}$ be a *completely uncoupled strategy mapping with additional information* α. Given a strategy mapping φ, for every game u with action profile set $A = A^1 \times A^2 \times ... \times A^n$, the strategies of the players will be $(\varphi(A^1, \alpha^1(u)), \varphi(A^2, \alpha^2(u)), ..., \varphi(A^n, \alpha^n(u)))$.

In the following theorem we show two examples of additional information (one may say reasonable information), that guarantees the convergence to a PNE.

Theorem 10 *If:*

1) for every player $i \in N$, α^i is the index of the player, or

2) for every player $i \in N$, α^i is the total number of players in the game,

then for every $\varepsilon > 0$ there exists a completely uncoupled mapping with additional information α into finite-memory strategies that leads to a pure Nash equilibrium in every game in $\mathcal{D}_\delta$ where such an equilibrium exists.

By Remark 7, for every $\varepsilon > 0$, this theorem guarantees to lead to PNE in all games except a set of games of measure ε.

5 Pure Nash Equilibria and General Games

The case of non-generic games is less interesting because of the following strong negative statement.:

Theorem 11 *For every $\varepsilon < \frac{1}{2}$ there is no completely uncoupled strategy mapping with additional uncoupled information, that leads to a pure Nash equilibrium in $1 - \varepsilon$ of the time, in every game with more than two players, where such an equilibrium exists.*

The theorem claims that even if we require the weaker type of convergence in $1 - \varepsilon$ of the time, and even if we allow some uncoupled additional information for the players, convergence to PNE still cannot be guaranteed. This result holds for both finite and infinite-memory strategies.

This theorem shows that in the terms of the generality of the solution, the interactive trail and error learning (see [11]) achieves the optimal result: convergence in almost all games.

6 Nash ε-equilibrium

We consider separately the cases of generic games and general games.

6.1 Generic games

Unlike the case of pure Nash equilibrium (see Preposition 7), the following theorem claims that the convergence to Nash ε-equilibria can be guaranteed by finite-memory strategies on the class of games $\mathcal{D}_\delta$.

Theorem 12 *For every $\varepsilon > 0$ and $\delta > 0$, there exists a completely uncoupled mapping into finite-memory strategies that leads to a Nash ε-equilibrium* *in every game in $\mathcal{D}_\delta$.*

By Remark 7 this theorem proves that for every $\varepsilon > 0$ there exists strategies that lead to Nash ε-equilibrium in all the games except set of games with measure ε.

This theorem shows that the suggested strategie like regret-testing (see [6] and [7]), do not achieve optimal results: the convergence in regret-testing is with limit frequency $1-\varepsilon$ of the time, although it is possible to guarantee convergence with limit frequency 1.

6.2 General games

We start with the impossibility result that says that we cannot guarantee convergence with limit frequency 1:

Theorem 13 *For $\varepsilon \leq \frac{1}{8}$, there is no completely uncoupled mapping into finite-memory strategies that leads to Nash ε-equilibria in every game.*

Now we would like to show that convergence in $1 - \varepsilon$ of the time can be guaranteed in general games. We show this result in two steps. First, we formally prove the result for a wide class of games $\mathcal{I}_\delta$, which will be defined below. Then we extend this result to general games. The proof of the generalization is tedious, so we will present just a sketch of it.

Given $\delta > 0$, we say that player i *influences* player j if there exists $a^{-i} \in A^{-i}$ and $a^i, b^i \in A^i$ such that $|u^j(a^i, a^{-i}) - u^j(b^i, a^{-i})| > \delta$.

Let $\mathcal{I}_\delta$ be the set of all games such that the number of players, number of actions of every player, and the payoffs are bounded by constants P, T, M, correspondingly (as in D_δ), and for every two players $i \neq j$, player i influences player j.

Proposition 14 *For every $\varepsilon > 0$ and $\delta > 0$, there exists a completely uncoupled mapping into finite-memory strategies that leads to a Nash ε-equilibrium in $1-\varepsilon$ of the time in every game in $\mathcal{I}_\delta$.*

This can be extended to

Theorem 15 *For every $\varepsilon > 0$, there exists a completely uncoupled mapping into finite-memory strategies that leads to a Nash ε-equilibrium in $1-\varepsilon$ of the time in every game.*

7 Proofs

Proof of Theorems 1 and 2. Assume, by contradiction, that such a strategy mapping φ exists. Note that $A^{-1} = A^{-2}$; therefore φ leads to a pure Nash equilibrium in every generic game with action profile sets A, A^{-1} and A^{-2}. By using the fact that φ leads to pure Nash equilibrium in generic games with action profile sets A^{-1} and A^{-2}, we will prove that there exists a set of payoff functions $P \subset U_A$ with a positive measure such that:

(i) for every $\Gamma \in P$, Γ has a unique pure Nash equilibrium;

(ii) the strategies $(f_{A^i})_{i=1}^n$ do not lead to it.

That will complete the proof.

Let $SPN_{A^{-1}} \subset U_{A^{-1}}$ be the set of the games with payoffs bounded by a constant $L > 0$ and with a unique Nash equilibrium that is pure. Then $\lambda(SPN_{A^{-1}}) > 0$; see Lemma 16. For every $v \in SPN_{A^{-1}}$ let $b(v)$ be the corresponding unique pure Nash equilibrium. The strategies $(f_{m^i})_{i=2}^n$ of players $2, 3, ..., n$ lead to $b(v)$ for almost every game v. Let $S \subset SPN_{A^{-1}}$ be the set of games for which $(f_{A^i})_{i=2}^n$ leads to $b(v)$; then $\lambda(S) = \lambda(SPN_{A^{-1}}) > 0$. By Lemma 17 there exists t and history $h \in H^*_{A^{-1}}$, realizable by $(f_{A^i})_{i=2}^n$, such that from the time $t+1$ and on, the players play $b(v)$ with probability 1. For every $h \in H^*_{A^{-1}}$ let T_h be the subset of all the games $v \in U_{A^{-1}}$ such that:

- h is realizable by $(f_{m^i})_{i=2}^n$.

- if h is played, then from the time $t+1$ and on, the players play some action $b \in A^{-1}$ with probability 1.

Then

$$\bigcup_{h \in H^*_{A^{-1}}} T_h \supset S$$

For every $i \in N$ f_{A^i} is Borel measurable, so T_h is a measurable set. There is a countable number of histories $h \in H^*_{A^{-1}}$ and S has positive measure; therefore, there exists $\overline{h}$ of size t such that $T_{\overline{h}}$ has a positive measure. Denote it by $R := T_{\overline{h}}$. The action that played from the time t and on is denoted by $b := (\overline{a^2}, \overline{a^3}, ..., \overline{a^n})$.

One should note that $A^1 = A^2$, so $f_{A^1} = f_{A^2}$; i.e., players 1 and 2 have the same strategy. So, by same considerations, for the action profile set A^{-2} the history $\overline{h} \in H^*_{A^{-2}}$ [9] and the subset of $\mathbb{R}^{(n-1)|A|}$: $R := T_{\overline{h}} \subset U_{A^{-2}}$ [10] setisfy:

- $\overline{h}$ is realizable by $(f_{m^i})_{i=1, i\neq 2}^n$

- if $\overline{h}$ is played, then from the time $t+1$ and on the players play the action $(\overline{a^1}, \overline{a^3}, \overline{a^4}, ..., \overline{a^n}) \in A^{-2}$ with probability 1.

Let us introduce a simplifying notation. For every $u \in U_A$ and for every

[9] $A^1 = A^2 \Longrightarrow A^{-1} = A^{-2} \Longrightarrow \mathcal{H}^*_{A^{-1}} = \mathcal{H}^*_{A^{-2}}$ therefore we can think of $\overline{h}$, which is the history of play in the game with action set A^{-1}, as the history of play in the game with action set A^{-2}.

[10] Every game in $U_{A^{-1}}$ is a game in $U_{A^{-2}}$, because $A^1 = A^2$, so we can think of the action of player 2 as the action of player 1.

subset of actions $B \subset A$, let $u|_B$ be the payoff function, defined only on the subset B.

Here we define our $P \subset U_A$ as the set of all the games with payoff function $u = (u^1, u^2, ..., u^n)$ such that on the "diagonal" $a^1 = a^2$ the payoffs $u^{-1}|_{\{a \in A | a^1 = a^2\}}$ and $u^{-2}|_{\{a \in A | a^1 = a^2\}}$ are some payoffs of the subset R. Off the diagonal we want the payoffs of all the players to be better than on the diagonal. Furthermore, we want a_2^1 to be a dominant [11] action for player 1, and a_1^i to be the dominant action of player $i \neq 1$.

Formally the n-player game u generates an $(n-1)$-player game $\widetilde{u}^{-1}$ as follows:

Player 2 chooses the action for both himself and player 1 (i.e., for every $i \neq 1$ $\widetilde{u}^{-1}(a^2, a^3, ..., a^n) := u^i(a^2, a^2, a^3, ..., a^n)$).

We define $P \subset U$ to be the set of all the payoff functions $u = (u^1, u^2, ..., u^n) \in U$ such that:

(a) $\widetilde{u}^{-1}(a^2, a^3, ..., a^n), \widetilde{u}^{-2}(a^1, a^3, a^4, ..., a^n) \in R$ for every action $a = (a^1, a^2, ..., a^n) \in A$ such that $a^1 = a^2$.

(b) $L+1 < u^i(a) \leq L+2$ for every action a such that $a^1 \neq a^2$, $i \neq 1$ and $a^i \neq a_1^i$.

(c) $L+1 < u^i(a) \leq L+2$ for every action a such that $a^1 \neq a^2$, $i = 1$ and $a^i \neq a_2^i$.

(d) $L+3 < u^i(a) \leq L+4$ for every action a such that $a^1 \neq a^2$, $i \neq 1$ and $a^i = a_1^i$.

(e) $L+3 < u^i(a) \leq L+4$ for every action a such that $a^1 \neq a^2$, $i = 1$ and $a^i = a_2^i$.

To make this construction clearer we provide the follwing example:

[11] Note that the action is not a dominant action in the game, but just dominant for actions out of the diagonal.

Let $A^1 = A^2 = \{a_1, a_2\}, A^3 = \{a^3_1, a^3_2\}$. For simplicity assume $L = 10$.

Let

$R =$

	a^3_1	a^3_2
a_1	$[0,1],[1,2]$	$[2,3],[3,4]$
a_2	$[4,5],[5,6]$	$[6,7],[7,8]$

R is the set of all games such that the payoffs of the row player satisfy: $0 \leq u^r(a_1, a^3_1) \leq 1, ..., 6 \leq u^r(a_2, a^3_2) \leq 7$ and the payoffs of the column player (player 3) satisfy: $1 \leq u^3(a_1, a^3_1) \leq 2, ..., 7 \leq u^3(a_2, a^3_2) \leq 8$.

So P is the following set of games:

a^3_1

	a^2_1	a^2_2
a^1_1	$[0,1],[0,1],[1,2]$	$[11,12],[11,12],[13,14]^*$
a^1_2	$[13,14],[13,14],[13,14]^*$	$[4,5],[4,5][5,6]$

a^3_2

	a^2_1	a^2_1
a^1_1	$[2,3],[2,3],[3,4]$	$[11,12],[11,12],[11,12]$
a^1_2	$[13,14],[13,14],[11,12]$	$[6,7],[6,7],[7,8]$

This is a simple example where the set R can be represented as a cartesian product of intervals $[0,1] \times [1,2] \times [2,3] \times ... \times [7,8]$ as an element in $\mathbb{R}^8$; in general cases it may not occur. Never the less, we define the set P by the cartesian product of the nodes on the diagonal (that are defined by R) and the nodes out of the diagonal (that are intervals).

Now we continue the proof. We will show that P satisfies the following:

1. $\lambda(P) > 0$.
2. Every game $u \in P$ has a pure Nash equilibrium.[12]

[12] This is needed for the proof of the nonexistence of strategies leading to pure Nash equilibria (because the condition is that the strategies lead to PNE in games where such an equilibrium exists). For the proof of mixed Nash equilibrium it is not needed.

3. For every game $u \in P$ there is a positive probability that the strategies $f_{A^1}, f_{A^2}, ..., f_{A^n}$, will not lead to Nash equilibrium.

This completes the proof.

Proof of 1. For every payoff function u that satisfies (a)-(e), conditions (b)-(e) restrict the payoffs out of the diagonal to be in some segment with length (or measure) 1. Let λ' be the Lebesgue measure on the space $\mathbb{R}^{n|A^{-1}|}$, which is the space of the diagonal actions. So

$$\lambda(P) = \underbrace{\lambda'(P|_{\{a\in A|a^1=a^2\}})}_{\text{on the diagonal}} \cdot \underbrace{1^{n(|A|-|A^{-1}|)}}_{\text{out of the diagonal } a^1\neq a^2} = \lambda'(P|_{\{a\in A|a^1=a^2\}}) \quad (1)$$

Let $B := \{b = (u^1, u^2, u^{-\{1,2\}})|(u^1, u^{-\{1,2\}}) \in R \text{ and } (u^2, u^{-\{1,2\}}) \in R\}$. One can see that $B = P|_{\{a\in A|a^1=a^2\}}$. By Lemma 18 with $k = |A^{-1}|$, $l = (n-2)|A^{-1}|$, $C = R$ we have $\lambda(P|_{\{a\in A|a^1=a^2\}}) > 0$. Therefore by (1) $\lambda(P) > 0$.

Proof of 2. For $a = (a_2^1, a_1^2, a_1^3, ..., a_1^n)$ the payoffs of the players are in the segment $[L+3, L+4]$. Deviation of a single player will reduce his payoff to less than $L+2$.

Proof of 3. The histories $\overline{h} \in H^*_{A^{-1}}$ and $\overline{h} \in H^*_{A^{-2}}$ are the same histories. Denote them by $(c(t'), c(t'), a^3(t'), ..., a^n(t'))_{t'=1}^t$ where $c(t')$ is the action of player 1/ player 2, if $\overline{h}$ is an element of $H^*_{A^{-2}}$/ $H^*_{A^{-1}}$ respectively. Define $\widetilde{h} \in H^*_A$ by $\widetilde{h} := (c(t'), c(t'), a^3(t'), a^4(t'), ..., a^n(t'))_{t'=1}^t$.

At the beginning there is a positive probability that at the first step $(c(1), c(1), a^3(1), a^4(1), ..., a^n(1))$ will be played, because the first step in $\overline{h}$ occurs with positive probability. If at time $(t'-1) = 1, 2, ..., t-1$ the history of play is the first $t'-1$ steps in $\widetilde{h}$, then all the played actions are on the diagonal ($\overline{a^1} = \overline{a^2}$) where their payoffs are from the set of games R. Therefore the observations sequence of every player at step $t'-1$ is exactly the same as if $\overline{h}$ occurred. Therefore $(c(t'), c(t'), a^3(t'), a^4(t'), ..., a^n(t'))$ will be played with a positive probability at step t'. Therefore by induction $\widetilde{h}$ is realizable by the strategies $(f_{A^i})_{i=1}^n$.

So the history $\widetilde{h}$ will occur with a positive probability. As a result, the action $(\overline{a^1}, \overline{a^2}, \overline{a^3}, ..., \overline{a^n})$ (note that $\overline{a^1} = \overline{a^2}$) will be played with probability 1, at all the steps $t+1$ and onward. But $(\overline{a^1}, \overline{a^2}, \overline{a^3}, ..., \overline{a^n})$ is not a Nash equilibrium in the game u (because deviation of player 1 increase his payoff above $L+1$). Hence the strategies do not lead to a Nash equilibrium. ■

Lemma 16 *For every $L > 0$ and every action set A, the subset $SPN_A \subset U_A$ of the games with a single Nash equilibrium that is pure, has a positive measure.*

Proof. Consider the following n-player game:

-$u^i(a) = \frac{L}{2}$ for every a such that $a^i = a^i_1$.

-$u^i(a) = 0$ for every a such that $a^i \neq a^i_1$.

The action $a^i = 1$ is a dominant strategy for every player i, so the game has a single Nash equilibrium $(a^1_1, a^2_1, ..., a^n_1)$ and it is pure. Also every $\frac{L}{6}$-perturbation of this game has the same single pure Nash equilibrium. Since the environment of size $\frac{L}{6}$ of every game has a positive measure, we found a subset of games with a positive measure as required. ■

Lemma 17 *Let $f = (f^1, ..., f^n)$ be a strategy profile, where every f^i is a finite-memory strategy that guarantees almost sure convergence of the play to Nash equilibria (pure or mixed) in every game. Then for a game with a single Nash equilibrium $a = (a^1, ..., a^n)$ that is pure, there exists a history $H(t) = (a(1), a(2), ..., a(t))$, realizable by f, such that from time $t+1$ and on, the players play the Nash equilibrium a with probability 1.*

Proof. Let Λ^i be the set of all the possible memory states of player i; i.e., all the states of the strategy automaton. Let $\Lambda = \Lambda^1 \times \Lambda^2 \times ... \times \Lambda^n$.

The strategies $f^1, ..., f^n$ induce a Markov process on the finite Markov chain Λ. Let $\Omega \subset \Lambda$ be the minimal reachable invariant set; then once it is reached the Markov process stays there. In addition, for every state $\omega \in \Omega$ the played action is a (with probability one), because otherwise there is some other action played with a frequency that does not converge to zero. There exist a time t and a path $(\alpha_1, \alpha_2, ..., \alpha_t)$ $\alpha_i \in \Lambda$, which is realized with a positive probability, such that $\alpha_t \in \Omega$. Let h be the history of play on the path $(\alpha_1, \alpha_2, ..., \alpha_t)$; then h is realizable by the strategy profile f, and from time $t+1$ and on, the players play the Nash equilibrium a with probability 1. ■

Lemma 18 *For every $k, l \in \mathbb{N}$ and for every set $C \subset \mathbb{R}^{k+l}$ with a positive measure $\lambda(C) > 0$, the set*

$$B := \{b = (x, y, z) \in \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^l | (x, z) \in C \text{ and } (y, z) \in C\} \subset \mathbb{R}^{2k+l}$$

has a positive measure.

Proof. Let $\mathbf{1}_C, \mathbf{1}_B$ be the characteristic functions of C, B. $\mathbf{1}_B(x, y, z) = \mathbf{1}_C(x, z)\mathbf{1}_C(y, z)$ by the definition of B. For every $z \in \mathbb{R}^l$ let $g(z) := \lambda(\{x \in \mathbb{R}^k | (x, z) \in C\})$. By Fubini's Theorem,

$$0 < \lambda(C) = \int_{\mathbb{R}^l} \left(\int_{\mathbb{R}^k} \mathbf{1}_C(x, z) dx \right) dz = \int_{\mathbb{R}^l} g(z) dz \tag{2}$$

also by Fubini's Theorem,

$$\begin{aligned}\lambda(B) &= \int\limits_{\mathbb{R}^l}(\int\limits_{\mathbb{R}^k}(\int\limits_{\mathbb{R}^k}\mathbf{1}_B(x,y,z)dx)dy)dz = \\ &= \int\limits_{\mathbb{R}^l}(\int\limits_{\mathbb{R}^k}(\int\limits_{\mathbb{R}^k}\mathbf{1}_C(x,z)\mathbf{1}_C(y,z)dx)dy)dz = \\ &= \int\limits_{\mathbb{R}^l}(\int\limits_{\mathbb{R}^k}\mathbf{1}_C(x,z)dx)(\int\limits_{\mathbb{R}^k}\mathbf{1}_C(y,z)dy)dz = \int\limits_{\mathbb{R}^l}g(z)g(z)dz\end{aligned} \tag{3}$$

By (2) $\int\limits_{\mathbb{R}^l}g(z)dz > 0$, therefore by (3) $\lambda(B) = \int\limits_{\mathbb{R}^l}g^2(z)dz > 0$. ∎

Proof of Theorem 9. The set C of all possible actions for all the players is countable; therefore the set $\mathcal{B}$ of all finite subsets of C is also countable. So there is an injective function $\gamma : \mathcal{B} \to \mathbb{N}$. So for every game $\Gamma \in DA$ the numbers $\gamma(A^1), \gamma(A^2), ..., \gamma(A^n)$ are different.

As described in Definition (3), in the class $\mathcal{D}_\delta$ the strategies are able to record observations in their finite-memory, such that every two different observations are recordd differently.

Let φ be a mapping that assigns the strategy $f(\gamma(A^i))$ for every actions set $A^i \in \mathcal{B}$. The strategy $f_{A^i}(l)$ for $l \in \mathbb{N}$ is defined below.

We start from formal descriptions of strategy $f_{A^i}(l)$, and we will explain later why the strategies $f(\gamma(A^1)), f(\gamma(A^2)), ..., f(\gamma(A^n))$ lead to a pure Nash equilibrium.

The strategy $f_{A^i}(l)$ is composed of five main steps.

Step 1 is called the **identification of index**:

Substep 1.1: The player plays a^i_1, and remembers his payoff $u^i(a^1_1, a^2_1, ..., a^n_1)$, denoted $u^i(1)$ for short.

The player moves to the next substep 1.2.

Substep 1.2.k.1, for $k \in \mathbb{N}$: If $l = k$ (i.e., $\gamma(A^i) = k$) the player plays a^i_1. Otherwise he plays a^i_2. In each case he remembers the payoff.

Substep 1.2.k.2: If $l > k$ the player plays a_2^i. Otherwise he plays a_1^i. If his payoff is $u^i(1)$ he evaluates his index j by

$$j = \#\{k | k < l \text{ and the player gets at step1.2.}k\text{.1 different payoff from } u^i(1)\}+1$$

and continues to step 2. If his payoff is not $u^i(1)$, he continues to substep 1.2.$k+1$.1.

Step 2 is called the **identification of action profile set**:

Below we will show that the indexes of different players are different, so the player with index j will be called player j.

Substep 2.k for $k \neq j$:

Substep 2.k.1: The player plays a_1^i. If his payoff is $u^i(1)$ he concludes that there are $k-1$ players and he continues to step 3. Otherwise he continues to substep 2.k.2.

Substep 2.k.l, for $l \geq 2$: The player plays a_1^i. If his payoff is $u^i(1)$ he concludes that player k has l actions and he continues to substep 2.$k+1$. Otherwise he continues to substep 2.k.$l+1$.

Substep 2.j:

The player plays his actions by the following order: $a_2^i, a_3^i, ..., a_{m^i}^i, a_1^i$ and then continues to substep 2.$j+1$.

Step 3 is called the **identification of payoff function** step:

Below we will show that after step 2, the player knows the profile action set A.

The player goes through all the actions $a \in A$ in lexicographic order (the order is defined by their indexes j). For every $a \in A$ the player plays his action a^i and remembers the payoff.

Step 4 is called the **finding a pure Nash equilibrium**:

The player knows his payoff function from step 3.

Reviewing all the actions $a = (a^i, a^{-i})$ in lexicographic order, the player plays a^i_1 if a^i is a best reply to a^{-i}; otherwise he plays a^1_2. If his payoff is $u^j(1)$ then he remembers a^i as the pure Nash equilibrium action and moves to step 5. Otherwise he goes to the next action in the lexicographic order.

Step 5 is called **playing the pure Nash equilibrium**, which is just the repeated play of his action in the pure Nash equilibrium.

Each substep can be implemented by finite automata, so the strategies are finite automata strategies.

Let's explain why the strategies $f(\gamma(A^1)), f(\gamma(A^2)), ..., f(\gamma(A^n))$ lead to a pure Nash equilibrium.

At step 1: The players will go through all the natural numbers $k = 1, 2, ..., \max_{i\in N}\gamma(A^i)$. For every number k the players will know at step 1.2.k.1 whether there exists a player i with $\gamma(A^i) = k$, or not. When they get to $k = \max_{i\in N}\gamma(A^i)$, at step 1.2.$k$.2 they will know that there is no player i such that $\gamma(A^i) > k$. $\{\gamma(A^i)\}_{i=1}^n$ are different, so the indexes

$$j(i) = \#\{k \in N|\gamma(A^k) < \gamma(A^i)\} + 1$$

are also different. In addition $\{j(1), j(2), ..., j(n)\} = \{1, 2, ..., n\}$.

At step 2: First player i with index $j(i) = 1$ will play his actions $a^i_2, a^i_3, ..., a^i_{m^i}, a^i_1$. When he finishes this process, all the players will know it, because their payoff will be $u^i(1)$. Hence the players will know the number of actions of player 1. After that, the same will happen with player i'whose index is $j(i') = 2$, and so on up to player n. When it is the turn of player $n + 1$, the players will see that player $n + 1$ has only one action, which indicates that player $n + 1$ does not exist. At the end of step 2, the players will know the total number of players, and the number of actions of every player.

At step 3: The players will play all possible actions $a \in A$, in lexicographic order, and so they will know their utility function.

At step 4: The players will review all possible actions $a \in A$, until there is an action $a \in A$ in which their payoffs are $u^i(1)$. That is all the players make a best-reply action at the action profile a. Therefore, a is a pure Nash equilibrium, and at step 5 the players will play it all the time.

Given a game with different actions sets of the players and with a pure Nash equilibrium, all the players will go through all the steps simultaneously, and eventually they will get to step 5, where they will play a Nash equilibrium all the time, so the frequency of times that the players play a Nash equilibrium converges to 1. ■

Proof of Theorem 10. As in the proof of Theorem 9, the players can record every payoff in every game $\Gamma \in \mathcal{D}_\delta$ in a finite automaton.

We shall now prove that each one of the two conditions is sufficient.

Condition 1: If all the players have some different indexes from $\{1, 2, ..., n\}$, then by Theorem 9, the chain of steps

"identification of action profile set" $\rightarrow$ "identification of payoff function" $\rightarrow$ "finding of pure Nash equilibrium" $\rightarrow$ "playing the pure Nash equilibrium"

guarantees convergence to a pure Nash equilibrium.

By the above assumption, the players know their index i, so these four steps will lead to a pure Nash equilibrium.

Condition 2: The strategy mapping of every player i can depend on n, and so we define the strategy of player i as follows:

First, the player plays the following step called **random identification of action profile set of n players:**

The player uniformly randomizes a natural number $1 \leq c^i \leq n$. Afterwards he plays the step "identification of action profile set" as if his index were $j = c^i$. If at the end of this step he concludes that there are n players, he remembers this action set and finishes the step. Otherwise he repeats the step.

Afterwards the player continues the step chain "identification of payoff function" $\rightarrow$ "finding of pure Nash equilibrium" $\rightarrow$ "playing the Nash equilibrium".

When each player has used this strategy, the following happens:

If $(c^1, c^2, ..., c^n)$ is a permutation of $(1, 2, ..., n)$, then after the "identification of action profile set" all the players will conclude that there are n players, and finally find a pure Nash equilibrium.

If at least two of the players have randomized the same number, let j be the smallest number such that $j \notin \{c^1, c^2, ..., c^n\}$. At the end of step 2 all the players will conclude that there are $j - 1$ players, and they will randomize their numbers again.

In every randomization the players randomize a permutation with probability $\frac{n!}{n^n}$, so eventually they will randomize a permutation and will reach a pure Nash equilibrium.

Because the nuber of options in the randomization is finite, the strategy is a finite automaton strategy. ■

Proof of Theorem 11. Consider the following two three-player games:

$$\Gamma_1 : \underbrace{\begin{array}{|c|c|c|} \hline & a^2_1 & a^2_2 \\ \hline a^1_1 & 1,1,1^* & 1,1,1^* \\ \hline a^1_2 & 1,1,1^* & 1,1,1^* \\ \hline \end{array}}_{a^3_1} \quad \underbrace{\begin{array}{|c|c|c|} \hline & a^2_1 & a^2_2 \\ \hline a^1_1 & 1,0,1 & 0,1,1 \\ \hline a^1_2 & 0,1,1 & 1,0,1 \\ \hline \end{array}}_{a^3_2}$$

$$\Gamma_2 : \underbrace{\begin{array}{|c|c|c|}\hline & a_1^2 & a_2^2 \\ \hline a_1^1 & 1,0,1 & 0,1,1 \\ \hline a_2^1 & 0,1,1 & 1,0,1 \\ \hline\end{array}}_{a_1^3} \quad \underbrace{\begin{array}{|c|c|c|}\hline & a_1^2 & a_2^2 \\ \hline a_1^1 & 1,1,1^* & 1,1,1^* \\ \hline a_2^1 & 1,1,1^* & 1,1,1^* \\ \hline\end{array}}_{a_2^3}$$

The pure Nash equilibria of Γ_1 are $\{(i,j,1)\}_{i,j=1}^{2}$. The pure Nash equilibria of Γ_2 are $\{(i,j,2)\}_{i,j=1}^{2}$.

The strategy of player 3, is the same in both games, and the histories of player 3 are independent of the actions of players 1 and 2. So player 3 does not play one of the actions a_1^3 or a_2^3 with limit frequency larger than $\frac{1}{2}$ with probability 1. Denote this action by a_i^3 $(i = 1,2)$. So at the game Γ_i the strategies will not lead to a pure Nash equilibrium.

For a different number of actions or a different number of players, one can easily construct a similar example in which convergnce to pure Nash equilibrium cannot be guaranteed. ■

Proof of Theorem 12. Let $\nu = \nu(\varepsilon)$ be a number small enough (for example, one can take $\nu = \frac{\varepsilon}{4M}$), such that for every game there exists a Nash $\frac{\varepsilon}{2}$-equilibrium with mixed actions that are integer multiplications of ν. Such a ν exists, because every game has a Nash equilibrium, and we can approximate it by integer multiplications of ν. Now we can make a discretization of $\Delta(A^i)$ for all i, taking only the actions that are integer multiplications of ν. Denote this discretization by $\widetilde{\Delta}(A^i)$ which is a finite set. So $\widetilde{\Delta}(A) = \widetilde{\Delta}(A^1) \times \widetilde{\Delta}(A^2) \times ... \times \widetilde{\Delta}(A^n)$ is also a finite set, and we can define a lexicographic order on $\widetilde{\Delta}(A)$.

We shall call the new step **finding an Nash $\frac{\varepsilon}{2}$-equilibrium** that will be useful bellow.

The player reviews all the mixed actions in $\widetilde{\Delta}(A)$ in lexicographic order. For every action $x = (x^i, x^{-i}) \in \widetilde{\Delta}(A)$, player i plays a_1^i if x^i is an $\frac{\varepsilon}{2}$-best reply to x^{-i}; otherwise he plays a_2^i. If his payoff is $u^i(1)$, he remembers x^i as the Nash $\frac{\varepsilon}{2}$-equilibrium mixed action. Otherwise he goes to the next action (in lexicographic order).

We define a *state* of a strategy of a player:

State $k.1$: The player knows that there are at least k players, and he did not find an $\frac{\varepsilon}{2}$-Nash equilibrium with k players.

State $k.2$: The player knows that there are at least k players, and he found an $\frac{\varepsilon}{2}$-Nash equilibrium with k players and he remembers his payoff function and the equilibrium.

The player starts his play at state 1.1.

At state $k.1$, the player plays the following chain of steps: "random identification of action profile set of k players" (Theorem 10, condition 2)$\rightarrow$ "identification of payoff function" (Theorem 9)$\rightarrow$ "finding an $\frac{\varepsilon}{2}$-Nash equilibrium". The player remembers the $\frac{\varepsilon}{2}$-Nash equilibrium $(x^1, x^2, ..., x^i, ..., x^n)$, and his payoff function $\widetilde{u}_k^i$ in the game, where he found the equilibrium. The player changes his state to $k.2$.

Let ξ be a number small enough[13] such that for every $\frac{\varepsilon}{2}$-Nash equilibrium $x = (x^i)_{i=1}^n$ the mixed actions profile $y = (y^i)_{i=1}^n$ defined by $y^i = (1-\xi)x^i + \xi(\frac{1}{m^i}, \frac{1}{m^i}, ..., \frac{1}{m^i})$, is an Nash ε-equilibrium.

At state $k.2$, the player plays his mixed action in the $\frac{\varepsilon}{2}$-Nash equilibrium (x^i) with probability $1-\xi$, and he plays all his actions by the uniformly distribution – with probability ξ. If his payoff is not one of the payoffs in $\widetilde{u}_k^i$, he changes his state to $k+1.1$. Otherwise he stays at the state $k.2$.

[13] For example, $\xi = \frac{\varepsilon}{4M}$ (when M is the bound of the payoffs) guarantees the requirement.

The number of players is bounded by P, and the number of actions by T, so every payoff function $\widetilde{u}_k^i$ should be recordd in T^P cells of payoffs. So the strategy is a finite-memory strategy.

If the payoff of player i is not in $\widetilde{u}_k^i$, then the action that was played is not one of the actions in the game $\widetilde{u}_k$. Therefore, the other players will also get some payoff that is not in their payoff function. Therefore, if all the players play by this strategy, then the updates of the states of all the players occur simultaneously.

Let n be the actual number of players. If all the players are at state $k.1$ for every $k \leq n$, then there is a positive probability that they will randomize k different numbers, and then they all will move to state $k.2$. If all the players are at state $k.2$ for every $k < n$, there is a probability ξ^n that all the players will play all their actions with uniform distribution. Hence there is a positive probability that the players will play some action that is not an action in $\widetilde{u}_k$, and then move to the state $k+1.1$. So, finally, the players will get to state $n.2$. In state $n.2$, the players play $(y^i = (1-\xi)x^i + \xi(\frac{1}{m^i}, \frac{1}{m^i}, ..., \frac{1}{m^i}))_{i=1}^n$, which is a Nash ε-equilibrium. The payoff function that every player records in his memory is the actual payoff function of the game, so the players will never get a payoff that is not in their payoff function, and they will stay at state $n.2$ all the time. ∎

Proof ot Theorem 13. Let us assume that such a strategy mapping exists. Consider the matching pennies game where a player with actions set $\{a_1, a_2\}$ plays against a player with actions set $\{a_1^3, a_2^3\}$:

$\Gamma :=$

	a_1^3	a_2^3
a_1	$1,0$	$0,1$
a_2	$0,1$	$1,0$

By Lemma 19 there exists a realizable history $h = ((a(t'), a^3(t'))_{t'=1}^{t}$ such that if h happens, then both players will continue to play actions that are ε close to $(\frac{1}{2}, \frac{1}{2})$ as long as their payoffs remain 0 or 1.

Now consider the following three-player game where players 1 and 2 have the same action set $\{a_1, a_2\}$ and player 3 has actions set $\{a_1^3, a_2^3\}$:

$\Gamma' :=$

	a_1^2	a_2^2
a_1^1	$1,1,0$	$1,0,1^*$
a_2^1	$0,1,1^*$	$0,0,1$

a_1^3

	a_1^2	a_2^2
a_1^1	$0,0,1$	$1,0,1^*$
a_2^1	$0,1,1^*$	$1,1,0$

a_2^3

where on the diagonal $a^1 = a^2$ players 1 and 2 play matching pennies against player 3. Out of the diagonal player $i = 1, 2$ gets 1 if he plays a_1^i, and 0 if he plays a_2^i. Player 3 gets 1 out of the diagonal in any case. Neither $((\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{2}))$ nor an ε-perturbation of it is a Nash ε-eqilibrium for $\varepsilon \leq \frac{1}{8}$ (because a deviation of player 1 to the pure action a_1^1 increases his payoff in more than $\frac{1}{8}$).

$\widetilde{h} := ((a(t'), a(t'), a^3(t'))$ is a realizable history in Γ', because there is a positive probability that all the players will play the history h. If $\widetilde{h}$ occurs then all the players will continue to play actions that are ε close to $(\frac{1}{2}, \frac{1}{2})$ forever, because all the payoffs in Γ' are 0 or 1. $\blacksquare$

Lemma 19 *Let $f = (f^1, f^2)$ be a strategy profile, where f^1, f^2 are finite-memory strategies that guarantee almost sure convergence of the play to Nash ε-equilibria in the matching pennies game. Then there exists a finite history $h = (a^1(t'), a^2(t))_{t'=1}^{t}$, realizable by f, such that starting from time $t+1$ the players play the Nash ε-equilibrium as long as their payoffs remain stay* 0 *or* 1.

This lemma claims more than the existence of a history that leads to a play of $((\frac{1}{2},\frac{1}{2}),(\frac{1}{2},\frac{1}{2}))\pm\varepsilon$ in the matching pennies game. It claims that once the players are convinced that they should play $((\frac{1}{2},\frac{1}{2}),(\frac{1}{2},\frac{1}{2}))$, they will play it forever, no matter what their payoffs in the future will be (even in the extreme situations where their payoffs will be 0 forever).

Proof. Let Λ^i be the set of all the possible memory states of player i, i.e., all the states of the strategy automaton. Let $\Lambda=\Lambda^1\times\Lambda^2$.

The strategies f^1,f^2 induce a Markov process on the finite Markov chain Λ. Let $\Omega\subset\Lambda$ be the minimal reachable invariant set, such that once it is reached, the Markov process stays there. For every state $\omega\in\Omega$, the played action is an ε-preturbation of $((\frac{1}{2},\frac{1}{2}),(\frac{1}{2},\frac{1}{2}))$; otherwise, there is a positive frequency of steps where the Nash ε-equilibrium is not played. Let α be some realizable path that lead to Ω, and h the realized actions on the path.

Assume, by contradiction, that there exists some finite sequence of actions and payoffs of 0 and 1 that lead player 1 not to play an ε-preturbation of $(\frac{1}{2},\frac{1}{2})$. This sequence occurs with a positive probability in the matching pennies because for every action of player 1 there exists action of player 2 that leads to the right payoff (0 or 1). So there is a positive probability of leaving Ω once it is reached, which contradicts the fact that Ω is a minimal invariant set. ■

Proof of Proposition 14. For every player i that has fewer actions $m^i<T$ we can define actions $a^i_{m^i+1},a^i_{m^i+2},...,a^i_T$ to be identical to his first action a^i_1. Denote this game by $\widetilde{\Gamma}$. Given strategies that lead to a Nash ε-equilibrium in $\widetilde{\Gamma}$, we can derive strategies that lead to a Nash ε-equilibrium in the original game, by playing a^i_1 each step when $a^i_{m^i+1},a^i_{m^i+2},...,a^i_T$ are played. In addition, note that if $\Gamma\in\mathcal{I}_\delta$, then $\widetilde{\Gamma}\in\mathcal{I}_\delta$. So we will construct strategies that lead to a Nash ε-equilibrium in every game where all the

players have T actions.

The covergnce process of the strategies from the proof of Theorem 12 can be described as follows:

Each player randomizes for himself an index. $\rightarrow$ Given this specific configuration of indexes, the players find their own payoff function. $\rightarrow$ For every discretized mixed action each player sends a bit (by playing a_1^i or a_2^i) to the other players. This bit reveals whether the player is ε-best replying or not. $\rightarrow$ Eventually they find a Nash ε-equilibrium and play it as long as they conclude that this is not the real game. The conclusion made when a player gets a payoff that does not exist in his payoff function. In this case, they repeat the chain of steps described above.

Let us explain how each of the steps in the process can be done in the case of game $\widetilde{\Gamma} \in \mathcal{I}_\delta$.

Given a configuration of indexes, the players find the payoff function as follows:

The players assume that there are P players. So they know the action set A, and they play all the actions in lexicographic order. Finally, every player removes from his list of existing players all the players that do not influence him.

When the configuration of indexes is a permutation of N, all the players will remove the players $(N+1, N+2, ..., T)$.

Player i can send a bit to player j in the following way:

All the players except player i will play all the actions (the actions from A, not from A^{-i}) in lexicographic order, while player i will play constantly a_1^i if he sends the bit 0, and he will play all the actions in lexicographic order if he sends the bit 1. Player $j \neq i$ will distinguish between the behaviors of player i (when he sends 0 or 1) because player i influences j; i.e., player j

receives the bit. So every player i can sends a bit to every player j in finite number of steps[14].

Finally, how can the players conclude that the current game is not the real one?

Assume that the procedure for finding an equilibrium described above takes q steps. Then the players can play the equilibrium $\frac{q}{\varepsilon}$ times, and afterward search for an equilibrium by new randomizations. If in the new search they find a game with more players, they start to play the new equilibrium; otherwise, they stay with the old one.

This procedure eventually detects the right number of players, and hence leads to playing a Nash ε-equilibrium with a limit frequency $1-\varepsilon$ of the time. ■

Sketch of the proof of Theorem 15. Given a game, we can define a directed graph whose vertices are the players, and with an edge from i to j if player i influences player j; this graph will be called the *influence graph*. The influence graph induces a directed tree (or forest) of the strong components.

The strategy is as follows. First, every player randomizes an index for himself. Then the players try to find out what the influence graph is by going through all the sequences of different players in lexicographic order. For each sequence, each player sends bits 00 or 01 (by the method described in proposition 14). For example, for the sequence 3,2,5, player 3 send bits 01 to player 2. If player 2 distinguishes between the two bits, he sends to player 5 bits 01, otherwise he sends bits 00. If player 5 receives two different bits, he knows that player 3 influences player 2 and player 2 influences him. At

[14] Clearly all the players $1,2,...,n$, even players $k \neq i,j$, must play as was described above, and "waite" untill the bit will be sent.

the end of this process each player can deduce the structure of the influence graph in all the vertices that influence him.

Next, the players find a Nash ε-equilibrium. First the players in the strong component on the head of the tree find a Nash ε-equilibrium. It can be done by sending a bit to the other players (directly or through other players in the component), and then by sending the information of what their actions are in the equilibrium to all the other players[15]. The players in the lower part of the tree (the sons) find a Nash ε-equilibrium, given the actions of the players in the head of the tree, and so on.

Denote by q the number of steps that this procedure takes. Now each player plays $\frac{q}{\varepsilon}$ times the Nash ε-equilibrium that he found in the randomizations where the number of players that influence him was maximal.

Finally, all the players will randomize a permutation, where the number of players that influence each player is maximal. Therefore, the players will play a Nash ε-equilibrium with frequency $1-\varepsilon$. ■

8 Discussion

1. The only case not discussed above is that of a Nash ε-equilibrium with additional information.

In Theorem 10 two examples of additional information were given: the index and total number of players.

For the case of the total number of players, we can easily generalize the impossibility result of Theorem 13. In the proof, instead of considering the matching pennies game, we consider matching pennies with $n-2$ fictitious players (i.e., players with a constant payoff of 0, and not influence the payoffs

[15] The ε-equilibrium is an integer multiplication of ν and so it is finite information.

of others), and instead of Γ' we consider Γ' with $n-3$ fictitious players.

Hence, additional information about the total number of players does not improve the optimal results.

In the case of the index, we can improve the result of Theorem 15, to convergence with frequency 1. The idea is that when the players know their index, they do not make randomizations. They play according to the procedure described in the proof of Theorem 15; thereafter it all the players will know the Nash ε-equilibrium and play it.

So, by using strategies with additional information of index, we can guarantee the optimal result, namley, convergence with limit frequency 1, that works in a general game.

2. We also did not consider the case of a mixed Nash equilibrium (excepting the impossibility result of Theorem 2). For finite-memory strategies, a discussion of the mixed Nash equilibrium is problematic, because even for games with integer payoffs the equilibrium actions can include irrational probabilities. The cardinality of possible plays with finite-memory strategies is countable, whereas the cardinality of possible mixed equilibrium in generic or general games is a continuum. This proves the impossibility of convergence to mixed equilibrium.

3. On the face of it, the pure Nash equilibrium seems to be simpler solution than the Nash ε-equilibrium. Although in generic games the convergence with frequency 1 can be guaranteed for a Nash ε-equilibrium, this is not the case for a pure Nash equilibrium. The reason is that even if a player thinks that he is playing an equilibrium, there still has to be some probability of deviation. Otherwise the players may "stuck" at the point where they think that it is an equilibrium but really it is not (this idea proved formally in Theorems 1 and 13). The player can accept this deviation in the case of a Nash ε-equilibrium: the player plays a Nash $\frac{\varepsilon}{2}$-equilibrium and deviates with probability $\frac{\varepsilon}{2}$ which generates a Nash ε-equilibrium. A pure Nash equilibrium is an exact solution, so this idea is not implamentable.

4. All the dynamics presented in the proofs of the theorems are very "ugly" dynamics in many senses. For example:

- The players do not know who the participants of the game are, although the players are fully correlated. For example, they go through some set in lexicographic order, that requires full correlation between the players.

- Playing this dynamics is not reasonable: the strategies do not reflect a reasonable behavior; i.e., one that is adaptive or rationalizable, and so on...

- The dynamic in Theorem 15, requires a huge number of steps to find the equilibrium even for small games.

A direction for future research is the question of finding a reasonable dynamics that achieves the optimal results for a Nash ε-equilibrium in both optimal cases:

- Dynamics that lead to a Nash ε-equilibrium in generic games with frequency 1.

- Dynamics that lead to a Nash ε-equilibrium in general games with frequency $1 - \varepsilon$. For this latter case the regret-testing of Germano&Lugosi may satisfy the requirements. It is an open question whether it converges for general games.

References

[1] Arslan, G., Marden, J.R., Shamma J.S., Young, H.P., 2007. Payoff-based dynamics for multi-player weakly acyclic games. Decision and Control, 2007 46th IEEE Conference. 3422-3427.

[2] Babichenko, Y., 2007. Uncoupled automata and pure Nash equilibria. Discussion paper #459. Center for the Study of Rationality, The Hebrew University.

[3] Foster, D.P., Kakade, S.M. 2008. Deterministic calibration and Nash equilibrium. Journal of Computer System Sciences 74. 115-130.

[4] Foster, D.P., Young, H.P., 2001. On the impossibility of predicting the behavior of rational agents. Proceedings of the National Academy of Sciences of the U.S.A. 98. 12848-12853.

[5] Foster, D.P., Young, H.P., 2003. Learning, hypothesis testing, and Nash equilibrium. Games and Economic Behavior 45, 73-96.

[6] Foster, D.P., Young, H.P., 2006. Regret-testing: learning to play Nash equilibrium without knowing you have an opponent. Theoretical Economics 1, 341-367.

[7] Germano, F., Lugosi, G., 2007. Global Nash convergence of Foster and Young's regret-testing. Games and Economic Behavior 60. 135–154.

[8] Hart, S., Mas-Colell, A., 2003. Uncoupled dynamics do not lead to Nash equilibrium. American Economic Review 93, 1830-1836.

[9] Hart, S., Mas-Colell, A., 2006. Stochastic uncoupled dynamics and Nash equilibrium. Games and Economic Behavior 57, 286-303.

[10] Young, H.P., 2007. The possible and the impossible in multi-agent learning. Artificial Intelligence 171, 429-433.

[11] Young, H.P., 2009. Learning by trial-and-error. Games and Economic Behavior 65, 626-643.