Coordination in Learning to Play Nash Equilibrium

Tomasz Sadzik†

Graduate School of Business, Stanford University

November 2007

Abstract

In the paper we analyze learning in the repeated game framework, when agents know only own payoffs (uncoupledness, Hart, MasColell [8]), but observe the common history of play (perfect monitoring). Each is using a simple learning strategy, modeled by a finite automaton. We study learning strategies that can guarantee that the period by period play converges almost surely to an approximate Nash equilibrium. It is shown that the ex-ante coordination on the pairs of learning strategies that actually achieve this is more than exponential in the approximation level, and almost exponential in the complexity of the learning strategies. The same holds if we look at the Bayesian rational learning strategies. We also show that coordination problems persist under the weaker notion of convergence.

---

*I am indebted to Yossi Feinberg for his devoted guidance, support and valuable suggestions. I am grateful to Andy Skrzypacz and Bob Wilson for the discussions and feedback at various stages of the project. I would like to thank Brendan Daley, Jerker Denrell, Matt Jackson and Michael Ostrovsky for their helpful comments.

†Email address: tsadzik@stanford.edu
1 Introduction

Consider two agents playing an infinitely repeated normal form game. At the beginning each player is informed only about own payoffs. As the game progresses agents commonly observe the history of taken actions. Suppose that each player uses a learning strategy, a procedure that determines how to play in a stage game after every history. If agents use an appropriate pair of learning strategies, the play converges to an approximate Nash equilibrium in all stage games - as shown for example by Jordan [10], [12], Kalai and Lehrer [13] sec. 6, Koutsougeras and Yannelis [14], Nyarko [26], Lehrer and Smorodinsky [16], Hart and MasColell [9]. In this paper we investigate what happens if agents have only partial information about opponent’s learning strategies and show that in general, “using an appropriate pair of learning strategies” implies incredibly high degree of ex-ante coordination.

Imagine that agents choose a learning strategy from a class of all learning strategies which converge, i.e. have another learning strategy such that, if paired, guarantee convergence to approximate Nash equilibrium in every stage game. Moreover, restrict the strategies to be represented by bounded-size automata, which we call simple. In other words, we allow the agents to ex-ante restrict attention to learning strategies that actually can converge and that are not too complicated. Our main result is that choosing from this set two learning strategies that would actually succeed in converging to an approximate equilibrium requires coordination (a notion described below) that grows i) at more than exponential rate in the required level of approximation; ii) almost exponentially as we increase the complexity of learning strategies, for a fixed approximation level. The problems persist even if we restrict the set of learning strategies to form Bayesian Nash equilibrium in the repeated game, but do not restrict the prior beliefs. We also show that coordination is necessary when we relax the convergence criterion, and require the players to eventually play the equilibrium only most of the time rather than all the time.

More formally, we model a learning strategy for a set of stage games as a finite automaton.\footnote{For the use of automata in the repeated game framework see e.g. Neyman [24], Rubinstein [27], or Mailath and Samuelson [18]. The automata have also recently been used to model the choice under bounded rationality: see Wilson [30] or Salant [29].} It allows us to consider the procedural complexity/simplicity of a learning strategy, measured by the size of the automaton, and analyze how it constrains the performance of learning strategies. An automaton consists of a finite state space, play and transition functions, as well as a start function. Play and transition functions determine how to play in each state,
and how to change a state after each round of play, given observed action profile. Start function determines for every stage game the starting state for the automaton. We assume that this state can depend only on own payoffs, and so the learning strategies are uncoupled (see Hart and MasColell [8], [9]). Examples of such learning strategies include finitary versions of fictitious play (Brown [1], Example 1), Bayesian learning strategies (see e.g. Jordan [10], [12]) or exhaustive search strategies (Foster and Young [4], Germano and Lugosi [5], Hart and MasColell [9], Example 2).

We say that learning strategies converge to approximate Nash equilibrium if the following holds. For any stage game with payoffs normalized to \([0, 1]\), a pair of mixed strategies is a supported \(\varepsilon\)-Nash equilibrium if every pure action in the support of each mixed strategy is a strict \(\varepsilon\)-best response. If two learning strategies guarantee that in every stage game, with probability one the players’ mixed strategies are supported \(\varepsilon\)-Nash equilibrium in the limit, as the game progresses, then we say that they converge to supported \(\varepsilon\)-Nash equilibrium (or, simply to an approximate equilibrium).

We formalize the notion of coordination in learning to play Nash equilibrium as follows. Each pair of learning strategies converges to a supported \(\varepsilon\)-Nash equilibrium, for some \(\varepsilon\) (possibly greater than 1). Each class of learning strategies defines a meta-coordination game, where the actions are the different learning strategies. The payoffs in this game for a pair of learning strategies is the supremum over \(1 - \varepsilon\), such that the pair converges to supported \(\varepsilon\)-Nash equilibrium. That is, the payoff in the meta-coordination game is a measure of how close to equilibria of the underlying stage games the two learning strategies can get — 1 representing convergence to exact equilibrium and 0 implying lack of convergence for any \(\varepsilon < 1\).

Fix some \(\varepsilon'\) and consider convergence of the learning strategies from the class. The coordination needed to play supported \(\varepsilon'\)-Nash equilibrium in the meta-coordination game, even if players coordinate on actions that can give payoffs at least \(1 - \varepsilon'\) when play with some action of the opponent, is our measure of coordination in the class. It is the size of a minimal set of \(\varepsilon'\)-best responses to the actions that can give payoffs at least \(1 - \varepsilon'\).\(^2\)

We consider two kinds of classes of learning strategies that agents choose from ex-ante. For each class we show that players need a very large amount of coordination on their learning strategies to converge to supported \(\varepsilon\)-Nash equilibrium.

\(^2\)The classes that we look at are, roughly, symmetric, and the choice of the player whose \(\varepsilon'\)-best replies we analyze is irrelevant for the coordination measure.
The first class consists of the learning strategies of minimal complexity that converge to supported $\varepsilon$–Nash equilibrium in all two-by-two games\(^3\) with payoffs normalized to $[0, 1]$. We show that there are more than\(^4\) $\exp(1/\varepsilon)$ many learning strategies in the class such that any learning strategy can converge to a supported $2\varepsilon$–Nash equilibrium with at most one of them. It means that the choice of a right pair from this class, or any class including it, requires more than $\exp(1/\varepsilon)$ amount of coordination.

We do not require those minimal complexity learning strategies to be rational, but rather treat them as descriptions of simple learning behavior, universal for all underlying stage games. They avoid any spurious coordination problems, related to coordinating on some involved pattern of play over time. For example, the simplest coupled learning strategies, i.e. the ones that can depend on opponent’s payoffs, play a fixed equilibrium of a stage game from the first round on. A slightly more complicated strategy, which can adjust in the second round, can converge with all of them\(^5\). This is in a stark contrast to the result above.

Next we consider classes of learning strategies that are restricted to be rational, in the following sense. For any discount factor and a finite $\varepsilon$–dense set of two-by-two games\(^6\) with payoffs normalized to $[0, 1]$ we consider a class of learning strategies, which form a Bayesian equilibrium (when paired with a learning strategy from the class) of the whole repeated incomplete information game, and which converge to Nash equilibrium. A pair of such learning strategies can be modified into a pair that converges to supported $\varepsilon$–Nash equilibrium in all two-by-two games.\(^7\) If we do not constrain the set of players’ priors, the amount of coordination required in each class, or any class including it, is still larger than $\exp(1/\varepsilon)$. Furthermore, increasing the complexity of Bayesian equilibrium strategies, which correspond to more sophisticated priors and detailed sets of games, has similar effects: Even for a fixed approximation level, the coordination problem grows almost at the rate $\exp(1/c)$, where $c$ is the complexity of learning strategies.

One can view these negative results (for coordination needed to guarantee convergence)\(^8\)

\(^3\)i.e. in which each player has only two actions;
\(^4\)More precisely, as $\varepsilon$ converges to zero the number grows at a higher rate than $C^{1/\varepsilon}$, for any $C > 0$. For any positive $\varepsilon$ this number is finite.
\(^5\)It follows from the results by Nachbar [23] that the move to a more complicated strategy is necessary.
\(^6\)i.e. for any two-by-two game there is a game in the set whose Nash equilibria are supported $\varepsilon$–Nash equilibria of the original game, and agents’ payoffs in the games in the set are independent (to prevent implicit coordination on payoffs);
\(^7\)This relaxation of convergence criterion is necessary since Jordan [11] (see also Foster and Young [3]) has proven an impossibility of convergence to exact Nash equilibrium of the stage play over the set of all games under such Bayesian equilibrium dynamics.
as driven by the strong requirement that the agents eventually find the set of approximate equilibria and stay in it forever. In the last part we show that coordination is necessary even if we consider weaker notion of convergence and allow the occasional out of an approximate equilibrium play along the whole time sequence. We consider learning strategies that incorporate a "stochastic search" component together with a "test phase", which allow agents to stochastically search through the space of stage game strategies, and then settle down on the one which is performing well. Despite the stochastic, uncoordinated search, such learning strategies can play supported $\varepsilon$–Nash equilibrium arbitrarily often, with probability one (see Foster and Young [4]). We show that unless the players coordinate on how they test, the frequency of play outside of approximate equilibrium is bounded below by the level of noise in the stochastic part.

One argument against uncoupled learning to play Nash equilibrium focuses on the tension between learning and optimizing along the path of play (see Jordan [11], Foster and Young [3]). The problem lies in the fact that the mixed strategies are by definition the points where the best response correspondence is not lower semicontinuous. Therefore, even as the beliefs converge to the mixed strategy of the actual game, the optimal best responses remain pure. With continuous priors over the set of games this renders convergence to approximate equilibrium under fully rational dynamics impossible.\textsuperscript{8} In our case, when we deal with Bayesian equilibrium strategies, we assume that the agents "coordinate" on the finite supports of their priors. Moreover, we show that the need for coordination persists even without any rationality assumptions.

Nachbar (cf. [22]; see also [21] and [23] for complete information case) has pointed out a different problem of convergence when players are playing Bayesian best responses. It is impossible for each player to have a single strategy that converges to a best response against every opponent’s strategy if they use strategies of comparable complexity. The result establishes that it is not possible to converge to equilibrium when each player at the same time fully hedges against all of the opponent’s strategies (compare Remark 2).

A different argument against uncoupled learning to play equilibrium deals with the class of learning strategies lying on the other side of the rationality spectrum. In recent papers Hart and MasColell (cf. [8], [9]) established that no learning dynamics, which are both stationary and depend only on the recent history of play, can converge to approximate Nash

\textsuperscript{8}In fact, convergence is possible only for a set of games of measure zero, under suitable parametrization.
equilibrium. The class of dynamics includes the adaptive learning-type dynamics based on
the stationarity assumption, such as fictitious play.\footnote{This requires, however, rescaling of the time axis: see Hart and MasColell \cite{8} footnote 3.} It leaves out, however, all of the Bayesian
rational learning dynamics that we study.

On the other hand, it is known that in certain circumstances uncoupled learning in the
repeated play may lead to convergence to the approximate Nash equilibrium (see \cite{10}, \cite{12},
\cite{13}, \cite{14}, \cite{26}, \cite{16}, \cite{9} cited above). Our paper bridges those results in the following sense.
We are not showing an impossibility of convergence under specific assumptions on the learning
strategies. Instead, we focus precisely on the strategies that do converge and identify a new
problem, problem of coordination. Even if we find a converging pair, it is by virtue of an
overwhelming ex-ante coordination on the learning strategies.

Finally, let us point out that we are investigating the convergence to equilibrium of a
stage play, and not the convergence of cumulative frequency of actions to the equilibrium
frequency. It is possible that the learning strategies that do not converge to an approximate
equilibrium nevertheless give rise to frequencies of actions that are approximately the same as
the equilibrium frequencies. The prime example is a fictitious play strategy in the matching
pennies game (see Brown \cite{1}). Convergence of empirical frequencies is relevant from the
observer’s point of view or for such criteria of optimality as "no regret" (see e.g. Hart \cite{6}).
We are interested in the convergence to equilibrium, which implies no incentives to change
the behavior in the limit.

The paper is organized as follows. In Section 2 we develop the framework, define learning
strategies, modes of convergence and the meta-coordination game. In Section 3 we deal with
the minimal complexity and Bayesian equilibrium learning strategies. The weaker notion of
convergence is analyzed in Section 4, and Section 5 provides the proofs of the results.

2 Framework

For every $N \in \mathbb{N}$ we consider the class $\mathcal{G}^N$ of two player normal form stage games with $N$
actions for each player and payoffs normalized to $[0, 1]$. For a fixed $N$ we refer to a game
using its unique pair of payoff functions:

$$\mathcal{G}^N = \{(g_1, g_2) | g_i : \{1, \ldots, N\}^2 \to [0, 1] \}.$$
Let $\Delta^N$ denote the set of mixed strategies in a stage game; for $x \in \Delta^N$ $x(n)$ is the probability assigned to action $n$.

For any $(g_1, g_2) \in \mathcal{G}^N$ let $NE(g_1, g_2)$ be the set of Nash equilibria of $(g_1, g_2)$, i.e. tuples $(x_1, x_2) \in \Delta^N \times \Delta^N$ such that $x_1(n) > 0$ implies $g_1(n, x_2) \geq g_1(y_1, x_2)$ for all $y_1 \in \Delta^N$, and similarly for $x_2$. For any $\varepsilon > 0$ we consider the following approximation: for $(g_1, g_2) \in \mathcal{G}^N$ $(x_1, x_2) \in \Delta^N \times \Delta^N$ is a supported $\varepsilon$-Nash equilibrium if $x_1(n) > 0$ implies $g_1(n, x_2) > g_1(y_1, x_2) - \varepsilon$ for all $y_1 \in \Delta^N$, and similarly for $x_2$. The set of all the supported $\varepsilon$-Nash equilibria of $(g_1, g_2)$ will be denoted $NE^\varepsilon_1(g_1, g_2)$. Action $n$ is dominant for player 1 and payoff function $g$ if $g(n, x_2) \geq g(y_1, x_2)$ for all $y_1, x_2 \in \Delta^N$.

In the dynamic setting a normal form game $(g_1, g_2) \in \mathcal{G}^N$ is played repeatedly at discrete time periods $t = 1, 2, \ldots$. We assume perfect monitoring: after every time period $t$ each player observes everyone’s realized actions. A learning strategy is a general rule or heuristic for a set of normal form games, which can be applied to the repeated play in any game from the set. (In the paper, for simplicity, we will only investigate the learning strategies for the sets of two-by-two games.)

**Definition 1** For any $N \in \mathbb{N}$ and $G \subseteq \mathcal{G}^2$ a learning strategy $L = \{\Omega, s, \phi, \pi\}$ for $G$ consists of a finite state space $\Omega$, a starting state function $s : G \rightarrow \Omega$, a transition function $\phi : \Omega \times \{1, \ldots, N\}^2 \rightarrow \Delta(\Omega)$ and a play function $\pi : \Omega \rightarrow \Delta^N$. Complexity of $L$ is defined as the size of its state space, $|\Omega|$.

In the repeated play of game $(g_1, g_2) \in G$ learning strategy for $G$ starts in the state $s(g_1, g_2)$. If it is in a state $\omega \in \Omega$ at round $t$, it plays $\pi(\omega)$. If at this round actions $(a_1, a_2)$ are realized, the learning strategy enters a new state in round $t+1$ according to the distribution $\phi(\omega, (a_1, a_2))$.

**Example 1** (Finitary Fictitious Play) Since an infinite process of estimation of a binomial parameter cannot be implemented by a finite automaton, fictitious play (Brown [1]) cannot be implemented by a learning strategy. Consider, however, the following finitary version, where, as in fictitious play, agent believes that opponent’s play corresponds to a sequence of i.i.d. random variables, and she best responds to a finite memory estimate of the variable (see Samaniego [28], Leighton, Rivest [17]). For some $N \in \mathbb{N}$, let $L = \{\Omega, s, \phi, \pi\}$ be a learning strategy for player 1 for games in $\mathcal{G}^2$ in which no player has a dominant action. We

---

10 Similarly define dominant actions for player 2 and payoff function $g$. 

6
The novelty in our case is that instead of a single starting state, a learning strategy is equipped with a starting state function. This makes an automaton into a model of a more general, incomplete information strategy.

Example 2 (Search and Hypothesis Testing) (Foster, Young [4]) In this example, the agent does not adjust her estimate of opponent’s i.i.d. play "continuously", as in the case of fictitious play. Now, at each point of time agent holds a point hypothesis about the frequency, and follows the following regime of play. For some \( M \in \mathbb{N} \), agent enters a "test phase" with probability \( 1/M \), and then for \( M \) rounds she does not change her stage play and the hypothesis, gathering the data about opponent’s play. Then she tests her hypothesis. If the estimate passes the test, she keeps the hypothesis and does not change the play. If the estimate fails, she chooses randomly a new hypothesis and a new play. Assume that the agent chooses among approximate best responses, choosing mixed strategies if pure actions have almost the same expected payoff against the estimate.

Let \( L = \{ \Omega, s, \phi, \pi \} \) be a strategy for player 1 for games in \( G^2 \) in which no player has a dominant action, for which \( \Omega = \bigcup_{k=0}^{2N+1} \{0^k, \ldots, N^k\} \times ((m^k, m^k) | m^k \in \{0, \ldots, M\}, m^k \leq m} \times \{0^k, \ldots, N^k\} \). In the basis, the first coordinate represents the current hypothesis, second coordinate "counts" the \( 1 \) actions in the period of gathering data, and the last coordinate represents how agent plays in a stage game, possibly choosing a mixed strategy. Details are in Section 5.3.

Any pair of learning strategies \( L_1 = \{ \Omega_1, s_1, \phi_1, \pi_1 \} \) and \( L_2 = \{ \Omega_2, s_2, \phi_2, \pi_2 \} \) for some \( G \subseteq G^2 \) together with \( (g_1, g_2) \in G \) defines a random process \( \{(L_1, L_2)^G (g_1, g_2)\}_{t=1}^{\infty} \) with values in \( \Omega_1 \times \Omega_2 \), as well as a random process of stage play \( \{(L_1, L_2)^\Delta (g_1, g_2)\}_{t=1}^{\infty} \) with values in \( \Delta^2 \times \Delta^2 \).

Automata are common tools for representing strategies in the repeated game framework. The novelty in our case is that instead of a single starting state, a learning strategy is equipped with a starting state function. This makes an automaton into a model of a more general, incomplete information strategy.

\*[11] s_{(g_1, g_2)} = 0^{I+k}, \text{ where } I = N \text{ if } g_1(1,1) > g_1(2,1) \text{ and } I = 0 \text{ otherwise, and } k \leq N \text{ is such that } g_1(1,x) = g_1(2,x) \text{ implies } x(1) \in (\max\{0, \frac{1}{N^k}\}, \min\{1, \frac{1}{N^k}\}] \text{. Moreover, } \pi(n^{I+k}) = 1 \text{ iff } I = 0 \text{ and } n < k, \text{ or } I = N \text{ and } n \geq k, \text{ for every } n, k \leq N \text{.}
The main reason why we deal with automata is not their convenient representation, but their explicit complexity measure. On the one hand, learning strategies of finite, and low complexity can be interpreted as heuristics, or rules of behavior, as those are exactly the strategies that can be implemented by a machine. The complexity ordering on the strategies has an immediate interpretation in terms of implementability. On the other hand, the space of learning strategies of finite complexity is amenable for a finitary, complexity based characterization.

Following Hart and MasColell ([8], [9]) we require that the learning strategies be uncoupled. This means that each learning strategy depends only on own payoffs. A pair of uncoupled learning strategies cannot converge to equilibrium with each player "solving an equilibrium" and playing it from the first round on. They have to learn the equilibrium from experience, by observing the common history of play.

**Definition 2** A learning strategy \( L = \{\Omega, s, \phi, \pi\} \) for \( G \) for agent 1 is uncoupled if for any two games \((g_1, g_2), (g_1', g_2') \in G\) we have \( s(g_1, g_2) = s(g_1', g_2') \) (and similarly for agent 2).

Finitary Fictitious Play (Example 1), Search and Hypothesis Testing (Example 2) and the Bayesian rational learning strategies (Section 3.2) are all uncoupled.

**Definition 3** Fix \( 0 < \varepsilon, 0 \leq \delta \). A pair of learning strategies \((L_1, L_2)\) for \( G \) converges to supported \( \varepsilon \)-Nash Equilibrium over \( G' \subseteq G \) if for every \((g_1, g_2) \in G'\), with probability one

\[
\lim_{t \to \infty} (L_1, L_2)\Delta_t(g_1, g_2) \in NE_\varepsilon^\delta(g_1, g_2).
\]

We define analogously convergence to Nash equilibrium.

In order to analyze the coordination needed to converge to supported \( \varepsilon \)-Nash equilibrium in a class of learning strategies we define the following coordination game.

**Definition 4** For any \( G_1 \times G_2 \subseteq G^2 \) and two finite sets \( \mathbb{L}_1 = \{L_1^1, ..., L_1^N\}, \mathbb{L}_2 = \{L_2^1, ..., L_2^N\} \) of the same size \( N \), \( N \in \mathbb{N} \), which consist of learning strategies over \( G_1 \times G_2 \), define the \((\mathbb{L}_1, \mathbb{L}_2)\)-convergence game \((g, g) \in G^N\) such that

\[
g(n, n') = \sup\{1 - \varepsilon | (L_1^n, L_2^{n'}) \text{ converges to supported } \varepsilon - \text{Nash equilibrium over } G_1 \times G_2\}.
\]
For every $\varepsilon > 0$ we define the following measure $Q(\varepsilon, g)$ of coordination in any game $(g, g) \in \mathcal{G}^N$, $N \in \mathbb{N}$. It is the minimal number of $\varepsilon$–best responses against the actions of the opponent, who chooses among actions that can give payoffs above $1 - \varepsilon$ when matched with the right action. Formally, let $Q(\varepsilon, g) = \min\{Q_1(\varepsilon, g), Q_2(\varepsilon, g)\}$, where $Q_1(\varepsilon, g)$ is minimal such that for some $X \subseteq \{1, ..., N\}$, $|X| = Q_1(\varepsilon, g)$,

$$\forall n \leq N \ (\exists n' \leq N \ g(n', n) \geq 1 - \varepsilon) \text{ implies } (\exists n' \in X \ g(n', n) \geq 1 - 2\varepsilon),$$

and $Q_2(\varepsilon, g)$ is defined analogously. Put otherwise, if agents are trying to coordinate in a game $g$ in order to obtain payoffs above $1 - \varepsilon$, and if agent $i$ had no information about which action the opponent is choosing, she can guarantee payoff above $1 - 2\varepsilon$ with probability at most $(Q_1(\varepsilon, g))^{-1}$ (by randomizing uniformly over a minimal set of $\varepsilon$–best responses).

Lemma 1 Fix $0 < \varepsilon$ and $c \in \mathbb{N}$. There is $M(\varepsilon, c) < \infty$ such that for any sets $\mathbb{L}_1$ and $\mathbb{L}_2$ of uncoupled learning strategies of complexity $c$ over some $G_1 \times G_2 \subseteq \mathcal{G}^2$ and the $(\mathbb{L}_1, \mathbb{L}_2)$–convergence game $(g, g)$ we have $Q(\varepsilon, g) \leq M(\varepsilon, c)$.

3 Strong Convergence

3.1 Minimally Complex Strategies

In this section we are considering the classes of strategies defined purely in terms of their complexity. The complexity measure corresponds to, roughly, the number of ways in which the past history of play can influence the future play, which can be interpreted as the long term memory of the learning strategy (see e.g. Mailath, Samuelson [18]). It has direct interpretation in terms of the size of a machine with no outside source for storing information, which can implement the play of the learning strategy. We are investigating a question whether this capacity constraint may serve as a focal point, which allows strategies to coordinate on the convergence to approximate Nash equilibrium.

Theorem 1 Fix $0 < \varepsilon < 1/3$ and suppose that a pair $(L_1^0, L_2^0)$ of uncoupled learning strategies of complexity $c$ converges to supported $\varepsilon$–Nash equilibrium over $\mathcal{G}^2$. Then there are $M \geq \left[\frac{1}{16} - 3\right]^2 \ (> \frac{1}{e^2})$, $D > 0$ independent of $\varepsilon$) pairs of uncoupled learning strategies $(L_1^1, L_2^1), ..., (L_1^M, L_2^M)$, all of complexity $c$, such that

i) each pair $(L_1^m, L_2^m)$ converges to supported $\varepsilon$–Nash equilibrium over $\mathcal{G}^2$, $m \leq M$;
ii) for any uncoupled $L_2$ there is at most one $m \leq M$ such that $(L_1^m, L_2)$ converges to supported $2\varepsilon$–Nash equilibrium over $G^2$, and similarly for any $L_1$.

In particular, for any $m \neq m', m, m' \leq M$, the pair $(L_1^m, L_2^{m'})$ does not converge to supported $2\varepsilon$–Nash equilibrium over $G^2$.

**Corollary 1** Fix $\varepsilon > 0$ and $c$ such that there are uncoupled learning strategies of size $c$ converging to supported $\varepsilon$–Nash equilibrium over $G^2$. There are sets $\mathbb{L}_1$ and $\mathbb{L}_2$ of uncoupled learning strategies of size $c$ such that for any sets $\mathbb{L}_1'$, $\mathbb{L}_2'$ of uncoupled learning strategies, $\mathbb{L}_1' \supseteq \mathbb{L}_1$, $\mathbb{L}_2' \supseteq \mathbb{L}_2$, and the $(\mathbb{L}_1', \mathbb{L}_2')$–convergence game $(g, g)$ we have $Q(\varepsilon, g) \geq \lceil \frac{1}{6c} - 3 \rceil^2 \geq \frac{D}{\varepsilon^2}$ ($> \frac{D}{\varepsilon^2}$, $D > 0$ independent of $\varepsilon$).

**Remark 1** Part ii) is true also if we allow the learning strategy $L_2$ to have infinite state space. Moreover, the statement is true if we only require convergence more than half of the time. The proofs for those cases remain unchanged.

**Remark 2** The following version of the result is true: For $\varepsilon > 0$ and $c$ as in the Theorem, there are $M' \geq \lceil \frac{1}{6c} - 3 \rceil!$ ($> \frac{D'}{\varepsilon^2}$, $D' > 0$ independent of $\varepsilon$) pairs of learning strategies $(L_1^1, L_2^1), \ldots, (L_1^M, L_2^M)$, all of complexity $c$, such that: i) as in the Theorem holds; ii') for any uncoupled $L_2$, if $L_2$ converges to supported $\varepsilon$–Nash equilibrium over $G^2$ with some $L_1^m$, then $L_2$ does not converge to a supported $2\varepsilon$–best response\footnote{$x_2$ is a supported $\varepsilon$–best response against $x_1$ in $(g_1, g_2) \in G^N$ if $x_2(n) > 0$ implies $g_2(x_1, n) > g_1(x_1, y_1) - \varepsilon$ for all $y_1 \in \Delta^N$. Pairs of such $x_1, x_2$ will be denoted $BR^\varepsilon_{2,s}(g_1, g_2)$. $L_2$ converges to supported $\varepsilon$–best response over $G^2$ with $L_1$ if for every $(g_1, g_2) \in G^2$, with probability one $\lim_{t \to \infty} (L_1, L_2)\Delta (g_1, g_2) \in BR^\varepsilon_{2,s}(g_1, g_2)$.} over $G^2$ with any $L_1^{m'}$ for $m' \neq m$, $m' \leq M'$.

The logic of the proof is as follows. Consider any two learning strategies that converge to supported $\varepsilon$–Nash equilibrium over $G^2$. In any game, in which the only supported $\varepsilon$–Nash equilibria consist of mixed strategies, from some time on the converging learning strategies must play only mixed strategies in the stage game. Moreover, each learning strategy must become unresponsive to the history of play, and continue playing own part of the supported $\varepsilon$–Nash equilibrium. On the other hand, the strategy played by player 1 is determined by payoffs of player 2 and vice versa. Given the assumption of uncoupledness, in the first rounds the agents must "communicate" to each other own payoffs: they must "communicate" to each

10
other roughly $\frac{1}{\varepsilon}$ bits of information. Moreover, given that strategies become irresponsible to the history of play, the communication must be accurate, and agents must coordinate perfectly on how they communicate. Finally, there are more than exponentially many (in $\frac{1}{\varepsilon}$) ways to communicate own payoffs, without increasing complexity of the learning strategies.

### 3.2 Bayesian Setting

For a finite set of games $G_1 \times G_2 \subset \mathcal{G}^2$ any pair $(L_1, L_2)$ of learning strategies for $G_1 \times G_2$ and a probability distribution $\nu_G \in \Delta(G_1 \times G_2)$ give rise to the probability distributions $\nu_i(L_1, L_2, \nu_G) \in \Delta(G_i \times \{1,2\}^2)$, $i = 1, 2$. Fix $0 \leq \beta < 1$, a learning strategy $L_{3-i}$ for $G_1 \times G_2$ and $\nu_{G,i} \in \Delta(G_1 \times G_2)$, for some $i \in \{1,2\}$. A Bayesian learning strategy $L_i$ for $L_{3-i}$ and $\nu_{G,i}$ is a learning strategy for $G_1 \times G_2$ such that the following holds: for any other learning strategy $L'_i$ and any $\mathcal{P}_i \in G_i$, $h^T \in \{1,2\}^T$, $T \in \mathbb{N}$, for which $\nu_i(L_1, L_2, \nu_G)\{(\mathcal{P}_i, (a_1^{i+1}, a_2^{i+1}))_{t=1}^\infty | ((a_1^1, a_2^1), (a_1^2, a_2^2))_{t=1}^T = h^T \} > 0$

\[
\int \sum_{(1,2)^2} \mathcal{P}_i(a_1^{i+1}, a_2^{i+1}) \beta^{T-t} d\nu_i(L_i, L_{3-i}, \nu_{G,i})|\mathcal{P}_i, h^T \geq (1)
\]

\[
\int \sum_{(1,2)^2} \mathcal{P}_i(a_1^{i+1}, a_2^{i+1}) \beta^{T-t} d\nu_i(L'_i, L_{3-i}, \nu_{G,i})|\mathcal{P}_i, h^T,
\]

where $\nu_i(L_i, L_{3-i}, \nu_{G,i})|\mathcal{P}_i, h^T$ is the appropriate conditional probability distribution. In other words, a Bayesian learning strategy must be implementing optimal behavior, given prior beliefs over the games, the strategy of the opponent, and posteriors computed via Bayesian updating.

For any $\varepsilon > 0$ we will say that $G_1 \times G_2$ is $\varepsilon$–dense if for every payoff function $g_1 : \{1,2\}^2 \rightarrow [0,1]$ there is $g_1' \in G_1$ such that for every $g_2 : \{1,2\}^2 \rightarrow [0,1]$

\[
NE(g_1', g_2) \subseteq NE_\varepsilon(g_1, g_2),
\]

and similarly for 1 and 2 reversed. Any pair of Bayesian learning strategies $(L_1, L_2)$ for an $\varepsilon$–dense $G_1 \times G_2$ can be modified into a pair of learning strategies $(L_1', L_2')$ over $\mathcal{G}^2$ such that if $(L_1, L_2)$ converges to Nash equilibrium over $G_1 \times G_2$ then $(L_1', L_2')$ converge to supported $\varepsilon$–Nash equilibrium over $\mathcal{G}^2$. For this purpose fix a partition $P_1 = \{p_{1}^g | g_1 \in G_1\}$

\[\text{i.e. } \nu(L_1, L_2, \nu_G)|\mathcal{P}_i, h^T(A) = \frac{\nu(L_1, L_2, \nu_G) \{(g_1, g_2, ((a_1^1, a_2^1), \ldots, a_T^1))_{t=1}^T | g_1 = p_{1}^g \}}{\nu(L_1, L_2, \nu_G) \{(g_1, g_2, ((a_1^1, a_2^1), \ldots, a_T^1))_{t=1}^T | g_1 \}} \cdot
\]
over the set of payoff functions \( \{g_1 : \{1,2\}^2 \rightarrow [0,1]\} \) such that if \( g_1 \in p_1^{G_1} \) then for every \( g_2 : \{1,2\}^2 \rightarrow [0,1] \) (2) is satisfied. For \( L_1 = \{\Omega_1, s_1, \phi_1, \pi_1\} \) define \( L_1' = \{\Omega_1, s'_1, \phi_1, \pi_1\} \) with

\[
s'_1(g_1, g_2) = s_1(g'_1, g_2) \text{ if } g_1 \in p_1^{G_1}.
\]

Define \( P_2 \) and \( L_2' \) in an analogous way.

A pair \((L_1, L_2)\) of Bayesian learning strategies for \( G_1 \times G_2 \) will be called a repeated game (r.g.) Bayesian equilibrium for \( G_1 \times G_2 \) if for some \( \nu_{G,1}, \nu_{G,2} \in \Delta_+(G_1 \times G_2) \) each \( L_i \) is a Bayesian learning strategy for \( L_{3-i} \) and \( \nu_{G,i} \), \( i = 1,2 \).

**Theorem 2** Fix any \( 0 < \varepsilon < 1/3 \) and \( 0 \leq \beta < 1 \). There are repeated game Bayesian equilibria \((L_{1,1}, L_{1,2}), \ldots, (L_{1,M}, L_{2,M})\), \( M \geq \left[ \frac{1}{4\varepsilon} - 2 \right]^2 > \frac{1}{\varepsilon^2} \), \( D'' > 0 \) independent of \( \varepsilon \), for some \( \varepsilon \)-dense \( G_1 \times G_2 \), such that

i) each \((L_{1,m}, L_{2,m})\) converges to Nash equilibrium over \( G_1 \times G_2 \), \( m \leq M \),

ii) for any uncoupled \( L_2 \) there is at most one \( m \leq M \) such that \((L_{1,m}, L_2)\) converges to supported \( 2\varepsilon - \) Nash equilibrium over \( G_1 \times G_2 \), and similarly for any \( L_1 \).

Moreover, for sufficiently low \( \beta \) the complexity of every \( L_{1,m}' \) is lower than \( 5 \times \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} \) if \( \varepsilon < 0.1 \).

It follows that even when agents coordinate ex-ante on playing a r. g. Bayesian equilibrium strategies for some \( G_1 \times G_2 \), a subclass of Bayesian rational strategies, the coordination problem still grows quicker than exponentially for approximation level converging to zero. In the proof we construct both the r. g. Bayesian equilibrium strategies and the prior beliefs that rationalize them. It is instructive to compare the sizes of the learning strategies and the number of miscoordinating pairs: for the approximation level \( 2\varepsilon = 0.05 \) those are around 1000, and above \( 10^9 \), respectively.

The fact that the beliefs about the opponent’s learning strategy are represented by a single learning strategy is not restrictive, as long as the beliefs have finite support. For assume that the beliefs of agent 1 about the games and opponent’s strategies are given by \( \nu_{G',L} \in \Delta(G'_1 \times G'_2 \times \mathbb{L}_2) \), for some finite set of learning strategies \( \mathbb{L}_2 \) over \( G'_1 \times G'_2 \). As above, for any learning strategy \( L_1 \) for agent 1, they induce beliefs \( \nu_1(L_1, \nu_{G',L}) \in \Delta(G'_1 \times (\{1,2\}^2)\infty) \).

Then for any set of payoffs \( G_2 \), with \( G'_2 \subset G_2 \) and \(|G_2| = |G'_2| \times |\mathbb{L}_2| \) there is \( \nu_G \in \Delta(G'_1 \times G_2) \) and a learning strategy \( L_2 \) over \( G'_1 \times G_2 \) with the following property: for any uncoupled
Bayesian equilibria

For every one-to-one functions $f$ to condition their behavior on more details of payoff function aggravates the coordination problem. Even for a fixed approximation level, this introduces much more opportunities for miscoordination.

**Remark 3** The following version of Theorem 2 is true. Fix $0 < \varepsilon < 1/3$, $0 \leq \beta < 1$ and any finite $\varepsilon'$-dense set of games $G_1' \times G_2'$. There is an $\varepsilon$-dense set $G_1 \times G_2$, $G_1' \times G_2' \subseteq G_1 \times G_2$, and repeated game Bayesian equilibria $(L_1^1, L_2^1), \ldots, (L_1^M, L_2^M), M \geq \frac{1}{4\varepsilon^2} - 2)^2$ for $G_1 \times G_2$ with the following properties: i) as in Theorem 2 is true; ii') for any $m \leq M$ there is $X^m \subset \{1, \ldots, M\}$, $m \in X^m$ and $|X^m| \geq \frac{1}{4\varepsilon^2} - 2)^2$, such that for any uncoupled $L_2$ there is at most one $\nu' \in \Omega_m$ such that $(L_1^{m'}, L_2)$ converges to supported $2\varepsilon'$-Nash equilibrium over $G_1' \times G_2'$, and similarly for any $L_1$.

Increasing the size of the Bayesian learning strategies when we simultaneously allow agents to condition their behavior on more details of payoff function aggravates the coordination problem. Even for a fixed approximation level, this introduces much more opportunities for miscoordination.

**Theorem 3** Fix $0 < \varepsilon < 0.1$ and consider any $c > 5 \ast \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$. There are repeated game Bayesian equilibria $(L_1^1, L_2^1), \ldots, (L_1^M, L_2^M), M \geq \frac{1}{\varepsilon} D'' f(c) \geq \sqrt{\varepsilon} < f(c) = h^{-1}(c)$ for $h(x) = x \log x$, $D'' > 0$, for some $\varepsilon$-dense $G_1 \times G_2$, such that i) and ii) as in Theorem 2 hold.

Altogether we have the following

**Corollary 2** For every $0 < \varepsilon$ and $0 \leq \beta < 1$ there are sets $\mathbb{L}_1$ and $\mathbb{L}_2$ of repeated game Bayesian equilibrium strategies for some $\varepsilon$-dense $G_1 \times G_2$ such that for any sets $\mathbb{L}_1'$, $\mathbb{L}_2'$ of uncoupled learning strategies over $G_1 \times G_2$, $\mathbb{L}_1' \supseteq \mathbb{L}_1$, $\mathbb{L}_2' \supseteq \mathbb{L}_2$, and the $(\mathbb{L}_1', \mathbb{L}_2')$-convergence game $(g, g)$ we have $Q(\varepsilon, g) \geq \frac{1}{16} - 2)^2 > \frac{1}{\varepsilon} D''$, $D''$ as in Theorem 2). If additionally $0 < \varepsilon < 0.1$, then for every $c > 5 \ast \frac{1}{\varepsilon} \log \frac{1}{\varepsilon}$ and $\beta$ sufficiently low we can assume that the

---

14 given the assumption of uncoupledness, $L_1$ can be treated both as a learning strategy over $G_1 \times G_2$ or $G_1 \times G_2'$;

15 For this purpose, if $L_2 = \{L_2^1, \ldots, L_2^M\}$ let $G_2 = \bigcup_{m=1}^{M} G_2^m$, for some sets of payoff functions $G_2^m$, with one-to-one functions $f^m : G_2^m \rightarrow G_2^m$, $m \leq M$, such that $f^1$ is an identity. For any $g_1 \in G_1'$ and $f^m(g_2) \in G_2^m$ define $\nu_1(g_1, f^m(g_2)) = \nu_1(g_1, g_2, L_2^m)$. Moreover, if $L_2^{m'} = \{\Omega^{m'} \times \phi^{m'}, \pi^{m'}\}$ with $\Omega^{m} \cap \Omega^{m'} = \emptyset$ for $m \neq m'$, $m \leq M$, then let $L_2 = \{\Omega, s, \phi, \pi\} \Omega = \bigcup_{m=1}^{M} \Omega^{m}$, and for any $m \leq M$, $s(f^m(g_2)) = s^{m}(g_2)$ for any $g_2 \in G_2$, $\phi(\omega, (s, a)) = \phi^{m}(\omega, (a, a'))$ for any $(a, a') \in \{1, 2\}^2$, $\omega \in \Omega^{m}$, and $\pi(\omega) = \pi^{m}(\omega)$ for $\omega \in \Omega^{m}$.
learning strategies are of complexity \(c\) and \(Q(\varepsilon, g) \geq \frac{1}{\varepsilon}D''f(c)\) \((D''\text{ and } f\text{ are as in Theorem 3)}\).

4 Convergence 1-\(\delta\) of the time

In this section we analyze the following, weaker, notion of convergence.

**Definition 5** Fix \(0 < \varepsilon, 0 \leq \delta\). A pair of learning strategies \((L_1, L_2)\) for \(G\) converges \(1 - \delta\) of the time to supported \(\varepsilon\)-Nash Equilibrium over \(G' \subseteq G\) if for every \((g_1, g_2) \in G'\), with probability one

\[
\lim \inf_{t \to \infty} \frac{1}{t} |\{t' \leq t | (L_1, L_2)^A_t(g_1, g_2) \in \text{NE}_\varepsilon(g_1, g_2)\}| \geq 1 - \delta.
\]

The classes that we investigate in this section are motivated by the following two observations. First, as we mentioned in the discussion after Theorem 1, learning strategies that converge with probability one to supported \(\varepsilon\)-Nash equilibrium must stop changing their stage game mixed strategy, even if over time they receive excessive evidence that the opponent is not playing her part of approximate equilibrium. Because of that, the "communication" of payoffs in the first rounds of play must be perfectly reliable. However, it is plausible to introduce some degree of noise in the play, so that each agent is never sure that the initial history establishes with certainty what the opponent’s payoffs are, and that the opponent will play her part of an approximate equilibrium. Second, if agent takes the noise into consideration, she will never settle on the play of a stage game mixed strategy, and will always confront her "hypothesis" about opponent’s stage game mixed strategy with the observed frequency of play. Given that, any level of noise is still compatible with convergence \(1 - \delta\) of the time to supported \(\varepsilon\)-Nash equilibrium, for arbitrarily small \(\delta\) (see Foster, Young [4]).

For some finite sets \(H, P \subset \Delta^2\) and \(K \in \mathbb{N}\) a communication strategy \(\{L, \tau\}\) is a learning strategy \(L = \{\Omega, s, \phi, \pi\}\) together with a mapping

\[
\tau : \{h_z^k | h \in H, z \in P \cup \{u, d\}, k \leq K\} \to \Omega
\]

such that \(\tau(h^k_p) = \tau(h^{k'}_{p'})\) only if \((k, p) = (k', p')\), \(k \leq K, p \in P\). In other words, it is a learning strategy with distinguished \(K\) copies of a "start" state \(h_p\), for every \((h, p) \in H \times P\).
as well as "rejection states" $h_u$ and $h_d$ for every $h \in H$ (see below). Consider the following assumption. For any $x \in \Delta^2_+$ let $W_\varepsilon(x) = \{x' \in \Delta^2_+ | |x(1) - x'(1)| > \varepsilon\}.$

$\gamma \in \text{TREMBLE.}$ For any $\omega \in \Omega$ and $k \leq K$,
i) For every $(p, h) \in P \times H$, $p \in \Delta^2_+$

$$\phi(\omega, (a_1, a_2))(\{h_p^k | p' \in W_\varepsilon(p^k) \cap P\}) > \frac{\gamma}{1 - \gamma} \ast \phi(\omega, (a_1, a_2))(h_p^k);$$

(3)

ii) For every $x \in \Delta^2_+$

$$\phi(\omega, (a_1, a_2))(\{\omega' | \pi(\omega') \in W_\varepsilon(x)\}) > \frac{\gamma}{1 - \gamma} \ast \phi(\omega, (a_1, a_2))(\{\omega' | \pi(\omega') = x\});$$

(4)

For any finite $H \subset \Delta^2$ consider a set $\{(S_h, \phi_h)\}_{h \in H}$ of $H$-tests, where each $S_h$ is a set including three distinguished states $\{h, h_u, h_d\}$, and $\phi_h : S_h \setminus \{h_u, h_d\} \times \{1, 2\} \to \Delta(S_h)$, satisfying\(^{16}\)

$$E_h(\min\{t | \phi_h^t(h) \in \{h_u, h_d\}\}) \geq E_h(\min\{t | \phi_h^{t'}(h') \in \{h_u, h_d\}\}), \ h, h' \in H$$

(5)

$$P_h(\min\{t | \phi_h^t(\omega) \in \{h_u, h_d\}\} < \infty) > 0, \ h \in H, \ \omega \in S_h$$

For any communication strategy $L = \{\Omega, s, \phi, \pi\}$ and $H$-tests $\{(S_h, \phi_h)\}_{h \in H}$, we define a $L$-learning strategy $\{(\Omega', s', \phi', \pi')\}$ as follows. Each agent starts playing as in the communication strategy $L$. Upon entering a state $h_p^k$, $(h, p) \in H_i \times P_i, k \leq K$, she assumes a tentative hypothesis that the other agent plays $h$, enters a test phase, and while holding this hypothesis she plays $p$. After certain histories of play the hypothesis may be rejected, and the agent resumes the communication at one of the two possible rejection states $h_u^k$ or $h_d^k$.

The formal definition of the $L$-learning strategy is in Section 5.3.

**Example 3 (Noisy signals)** For any $\varepsilon > 0$ let $G_1 \times G_2$ be a finite set of games in which no player has a dominant action and such that $g_1(1, x) = g_1(2, x)$ implies $g'_1(1, x) \neq g'_1(2, x)$ for $g_1 \neq g'_1, g_1, g'_1 \in G_1$, and similarly for player 2. Consider the following pair of learning

\(^{16}\)where for any $h, h' \in H$ and $\omega, \omega' \in S_h$, we have

$$P_h(\phi_{h'}(\omega)) = \omega = 1,$$

$$P_h(\phi_{h'}(\omega) = \omega') = \sum_{\omega'' \in S_h, a \in \{1, 2\}} P_h(\phi_{h'}^{-1}(\omega) = \omega'' \ast \phi_{h'}(\omega'', a)(\omega') \ast h(a), t > 1,$$
strategies \( L_i = \{ \Omega_i, s_i, \phi_i, \pi_i \} \) over \( G_1 \times G_2, \ i \in \{1, 2\} \). For some \( M \in \mathbb{N} \), in the first \( M \) rounds of play each strategy "encodes" own payoff function, playing deterministically in each round. After \( M \) rounds of play each learning strategy \( L_i, \ i \in \{1, 2\} \), is in one of the \( \{G_1 \times G_2\} \) absorbing states \( \omega_i^{(g_1, g_2)}, g_1 \in G_1, g_2 \in G_2 \), in which it plays a strategy \( x^{g_3 - i} \in \Delta_+^2 \) such that \((x^{g_2}, x^{g_1}) \in NE(g_1, g_2).\) The pair \((L_1, L_2)\) converges to Nash equilibrium over \( G_1 \times G_2.\)

Let \( L_i^{M'} = \{ \Omega_i^{M'}, s_i^{M'}, \phi_i^{M'}, \pi_i^{M'} \} \) be the following modifications of \( L_i, \ i \in \{1, 2\} \). Now, after \( M \) rounds of playing a game \((g_1, g_2)\) a learning strategy enters a "correct" state \( \omega_i^{(g_1, g_2), g_3 - i} \) with probability \( 1 - \gamma \), and one of the "incorrect" states \( \omega_i^{(g_1, g_2), g_{3} - i}, g_{3} - i \neq g_3 - i \), with probability \( \frac{\gamma}{|G_i| - 1} \). For \( M' \) rounds learning strategy continues playing \( x^{g_3 - i} \) if it entered a state \( \omega_i^{(g_1, g_2), g_{3} - i} \), and computes the empirical frequency of opponent’s play. Then, if frequency is consistent with the hypothesis that the opponent plays \( x^{g_3} \), it continues the same play for \( M + M' \) rounds, estimating frequency after first \( M \) rounds. Otherwise, learning strategy moves to the state \( s_i^{M'}(g_1, g_2).\)

For any \( \delta > 0 \) if we choose \( M' \) large enough and accordingly powerful tests of opponents play, the pair \((L_1^{M'}, L_2^{M'})\) converges \( 1 - \delta \) of the time to Nash equilibrium.

In fact, in order to achieve convergence \( 1 - \delta \) of the time, we do not need to assume any explicit communication phase. Consider Example 2. The learning strategies chose the hypothesis about opponent’s play and own play fully randomly. For any \( \varepsilon, \delta > 0 \), for appropriate parameter values, if the tests are sufficiently powerful, the learning strategies from the Example converge \( 1 - \delta \) time to supported \( \varepsilon - \)Nash equilibrium (Foster, Young [4], see also Germano and Lugosi [5]).

**Example 4** A different example of \( H- \) tests \( \{ \{ S_h, \phi_h \}\} \) for \( H \) that satisfy (5) is based on the estimation procedure that we used for the finitary model of Fictitious Play (Example 1, Samaniego [28]). For \( \overline{N}, N \in \mathbb{N} \) such that for every \( h \in H, 0 < h(1) - \overline{N} < h(1) + \frac{N}{N} < 1 \) let \( S_h = \{ h_d, h(1) - \frac{N - 1}{N}, ..., h(1) - \frac{1}{N}, h(1) + \frac{1}{N}, ..., h(1) + \frac{N - 1}{N}, h_u \} \), and \( \phi_h(\omega, 2)(\omega - \frac{1}{N}) = \omega, \phi(\omega, 2)(\omega) = 1 - \omega, \phi(\omega, 1)(\omega) = \omega, \phi(\omega, 1)(\omega + \frac{1}{N}) = 1 - \omega.\) For sufficiently large \( \overline{N} \) and \( N \) such tests satisfy (5).

**Theorem 4** For each \( i = 1, 2 \) consider a communication strategy \( L_i \) satisfying \( \gamma \in TREMBLE \) and a sequence of \( L_i \) - learning strategies \( \{ L_n^i \}_{n \in \mathbb{N}} \). Suppose that every pair \((L_1^0, L_2^0)\), \( n \in \mathbb{N} \), converges to supported \( \varepsilon - \)Nash equilibrium over \( G^2 \) \( 1 - \delta_n \) of the time, and \( \delta_n \to 0.\)

\( ^{17} \) a state \( \omega \in \Omega_i \) is absorbing if \( \phi_1(\omega, (a, a'))(\omega) = 1 \), for any \( a, a' \in \{1, 2\}; \)

\( ^{18} \) and let \( \phi(h, \cdot) = \phi(h(1), \cdot); \)
Then for any \( N \in \mathbb{N} \) and any sufficiently large \( n \), there is a nonempty set of games \( G^{(N,n)} \subset G^2 \) such that for any \((g_1, g_2) \in G^{(N,n)}\)

\[
\text{Prob}(\forall i \in \{1, 2\} \lim \inf_{\ell \to \infty} \frac{1}{\ell} |\{t' \leq t | (L_i^N, L_{3-i}^n)_{t'}(g_1, g_2) \in \text{NE}_e(g_1, g_2)\}| > 1 - \gamma) = 0.
\]

If, due to noise, a pair of learning strategies does not coordinate on the procedure in which it enters a pair of states in which approximated Nash equilibrium is played, it must coordinate on how it exits the tentative states, to guarantee weak convergence. Heuristically, the result can be seen as an approximation result. Under the strong convergence to approximated equilibrium, the learning strategies must at some point of time become fully irresponsible to the data, and so no noise from the initial period of "communication" can be eradicated. If we look at a convergence \( 1 - \delta \) of the time, but for decreasing \( \delta \), without additional coordination essentially the same phenomenon occurs.

5 Appendix

For any \( g : \{1, 2\}^2 \to [0, 1] \) with no dominant action define \( d(g) \in \{-1, 1\} \), \( d(g) = 1 \) if and only if \( g(a, a) > g(a, a') \) for \( a \neq a' \), and \( I_i(g) \in (0, 1), i = 1, 2 \), such that

\[
g(1, x_{3-i}) = g(2, x_{3-i}) \text{ for } x_{3-i}(1) = I_i(g).
\]

For any \( \varepsilon > 0 \) action \( n \) is \( \varepsilon \)-dominant for player 1 and payoff function \( g \) if \( g(n, x_2) > g(y_1, x_2) - \varepsilon \), and \( \varepsilon \)-dominant if \( g(n, x_2) > g(y_1, x_2) + \varepsilon \), for all \( y_1, x_2 \in \Delta^N \). Let \( G^1_\varepsilon \) and \( G^{-1}_\varepsilon \) be the sets of payoff functions such that for every \( g^1 \in G^1_\varepsilon \) and \( g^{-1} \in G^{-1}_\varepsilon \)

\[
\max\{g^1((1, 1)), g^1((2, 2))\} = \max\{g^{-1}((1, 2)), g^{-1}((2, 1))\} = 1,
\]

\[
\min\{g^1((1, 1)), g^1((2, 2))\}, \min\{g^{-1}((1, 2)), g^{-1}((2, 1))\} \geq \varepsilon,
\]

\[
g^1((2, 1)) = g^1((1, 2)) = g^{-1}((1, 1)) = g^{-1}((2, 2)) = 0.
\]

In words, for any payoffs \( g^u \in G^u_\varepsilon \) agent has no \( \varepsilon \)-dominant action, and she wants to "match" in case \( u = 1 \) and "mismatch" in case \( u = -1 \), with minimal payoff if she fails to do so.

---

19 Throughout this section we follow the following notational convention: the tuples of actions \((a_i, a_{3-i})\), \((a_i, x)\) for \( x \in \Delta^2 \) and payoffs \((g_i, g_{3-i})\), \((g_i, g)\), will denote \((a_{3-i}, a_i)\), \((x, a_i)\) and \((g_{3-i}, g_i)\), \((g, g_i)\) for the case when \( i = 2 \).

20 We analogously define \( \varepsilon \)-dominant and \( \varepsilon \)-dominant actions for player 2 and payoff function \( g \).
Proof. (Lemma 1) Fix $0 < \varepsilon$, $G_1 \times G_2 \subseteq G^2$ and $c \in \mathbb{N}$ such that there are learning strategies of complexity $c$ that converge to supported $\varepsilon$–Nash equilibrium over $G_1 \times G_2$ (otherwise $M(\varepsilon, c) = 0$). We can assume that all the learning strategies of complexity $c$ have the same state space $\{1, \ldots, c\}$. Associate with every such learning strategy $L = \{(1, \ldots, c), s, \phi, \pi\}$ for player 1 value $\chi(L) = (\chi_1(L), \chi_2(L), \chi_3(L)) \in [0, 1]^c \times (\text{Pow}([0, 1]^{1,2})^c \times (\text{Pow}(c))^{c\times\{1,2\}}$. The first coordinate encodes the play function, the second one encodes the start function, and the last one encodes the (nondeterministic) transition function. For any $\chi_i(L) = \pi(n)(1)$ and let $\chi_{2,n}$ be the closure of $\{(x_{(a,a')}(a,a')\in\{1,2\}^2)|g,g'(a,a') = x_{(a,a')}, (a,a') \in \{1,2\}^2, s(g,g') = n\}$. Finally, for any $n, n' \leq c$ and $(a,a') \in \{1,2\}^2$ let $n' \in \chi_3(n,a,a')$ if and only if $\phi(n, (a,a'))(n') > 0$.

Since $[0, 1]^c \times (\text{Pow}([0, 1]^{1,2})^c \times (\text{Pow}(c))^{c\times\{1,2\}}$ is a compact metric space (with the maximum metric over Euclidean, Hausdorff, and discrete metrics), there are open sets $a_1, \ldots, a_m$ such that $\cup_{m=1}^M o_m = [0, 1]^c \times (\text{Pow}([0, 1]^{1,2})^c \times (\text{Pow}(c))^{c\times\{1,2\}}$ and for any $m \leq M$ $(\chi_1, \chi_2, \chi_3), (\chi'_1, \chi'_2, \chi'_3) \in o_m$ implies that: i) for any $n \leq c |\chi_{1,n} - \chi'_{1,n}| < \varepsilon/4$; ii) for any $n \leq c$ if $(x_{(a,a')}(a,a')\in\{1,2\}^2 \in \chi_{2,n}$ and $(x'_{(a,a')}(a,a')\in\{1,2\}^2 \in \chi'_{2,n}$ then $|x_{(a,a')}-x'_{(a,a')}| < \varepsilon/4$, $(a,a') \in \{1,2\}^2$; iii) for any $n, n' \leq c$ and $(a,a') \in \{1,2\}^2$ $n' \in \chi_3(n,a,a')$ iff $n' \in \chi_3'(n,a,a')$. It follows that for any two learning strategies $L_1$ and $L_2$ for player 1 such that $\chi(L_1), \chi(L_2) \in o_m$, $m \leq M$, and any $L_2$ for player 2 if $(L^k, L_2)$ converges to supported $\varepsilon$–Nash equilibrium over $G_1 \times G_2$ then $(L^{3-k}, L_2)$ converges to supported $2\varepsilon$–Nash equilibrium over $G_1 \times G$, $k = 1, 2$. Analogous analysis for roles of players 1 and 2 reversed establishes the proof, with $M(\varepsilon, c)$ bounded above by $M$. ■

5.1 Theorem 1

Lemma 2 For any $\varepsilon > 0$ there exist $x^n_{\varepsilon} \in \Delta^2_+$, $n \leq N_\varepsilon$, such that the following holds:

i) for any $\varepsilon > \varepsilon$ and $g$ for which no action is $\varepsilon$–dominant for player 1, for some $n \leq N_\varepsilon$ we have\(^ {21}\)

$$|g(1, x^n_{\varepsilon}) - g(2, x^n_{\varepsilon})| < \varepsilon,$$

(8)

ii) there are functions $g^n_{i,\varepsilon} \in G^n_\varepsilon$, $u = -1, 1$, $i \in \{1, 2\}$, $n \leq N_\varepsilon$, such that, with $x^n_{\varepsilon}(1) = 1,$

$$|g^n_{i,\varepsilon}(1, x_{3-i}) - g^n_{i,\varepsilon}(2, x_{3-i})| < \varepsilon$$ implies $x_{3-i}(1) \in (x^n_{\varepsilon}(1), x^n_{\varepsilon}(1)),$

iii) for any $1 < k \leq N_\varepsilon$ there are functions $g^n_{i,\varepsilon}$, $u = -1, 1$, $i \in \{1, 2\}$, $n \leq \lfloor N_\varepsilon/k \rfloor - 1,$

\(^ {21}\)by symmetry, same holds for player 2.
such that for any $\varepsilon \geq \varepsilon$ with $(\frac{1-\varepsilon}{1+\varepsilon})^{k} \leq \frac{1-\varepsilon}{1+\varepsilon}$ we have $g^{n,u,k}_{i,\varepsilon} \in G^{u}_{i,\varepsilon}$ and

$$|g^{n,u,k}_{i,\varepsilon}(1, x_{3-i}) - g^{n,u,k}_{i,\varepsilon}(2, x_{3-i})| < \varepsilon \text{ implies } x_{3-i}(1) \in (x_{\varepsilon}^{nk}(1), x_{\varepsilon}^{(n-1)k}(1)).$$

Moreover, $\left|\frac{1}{\varepsilon} - 2\right| \leq N_{\varepsilon} \leq \frac{1}{\varepsilon} + 2$.

**Proof.** (Lemma 2) For any $\varepsilon > 0$ we define $x^{n}_{\varepsilon}$, $n \leq N_{\varepsilon}$, as follows. Let $N_{\varepsilon}' \in \mathbb{N}$ be such that

$$\left(1 - \frac{\varepsilon}{1+\varepsilon}\right)^{N_{\varepsilon}'} - \frac{1}{1+\varepsilon} > \left(1 - \frac{\varepsilon}{1+\varepsilon}\right)^{N_{\varepsilon}'} - \frac{1}{1+\varepsilon} + 1.$$ 

For $n = 1, ..., N_{\varepsilon}'$ let

$$x^{n}_{\varepsilon} = \left(1 - \frac{\varepsilon}{1+\varepsilon}\right)^{n}. \quad (9)$$

Let $N_{\varepsilon} \in \mathbb{N}$ be such that

$$\left(1 - x^{N_{\varepsilon}_n}\right) \cdot \left(1 + \varepsilon\right)^{N_{\varepsilon}'} - \frac{1}{1+\varepsilon} < \left(1 - x^{N_{\varepsilon}_n}\right) \cdot \left(1 + \varepsilon\right)^{N_{\varepsilon}' - 1}.$$ 

For $n = N_{\varepsilon}' + 1, ..., N_{\varepsilon}$ let $x^{n}_{\varepsilon}$ be such that

$$\left(1 - x^{n}_{\varepsilon}\right) = \left(1 - x^{N_{\varepsilon}_n}\right) \cdot \left(1 + \varepsilon\right)^{N_{\varepsilon}' - 1}. \quad (10)$$

To prove i), given $g$ with no $\varepsilon$-dominant action for 1 note that $I_{1}(g) \in \left[\frac{\varepsilon}{1+\varepsilon}, \frac{1}{1+\varepsilon}\right]$. (Otherwise, when e.g. $g(2, 2) > g(1, 2)$ and $I_{1}(g) \in \left(\frac{1}{1+\varepsilon}, 1\right]$ we have $\frac{1 - I_{1}(g)}{I_{1}(g)} = \frac{g(2, 1) - g(1, 1)}{g(2, 2) - g(1, 2)}$, implying $g(2, 1) > g(1, 1) - \varepsilon$, i.e. 2 is $\varepsilon$-dominant for 1 for $g$.) The values of $x^{n}_{\varepsilon}$, $1 \leq n \leq N_{\varepsilon}$, are constructed in such a way that

$$\left[\frac{\varepsilon}{1+\varepsilon}, \frac{1}{1+\varepsilon}\right] \subset \bigcup_{n=1,\ldots,N_{\varepsilon}} (\overline{x^{n}_{\varepsilon}}, \overline{x^{n}_{\varepsilon}}, \overline{x^{n}_{\varepsilon}}), \quad (11)$$

where

$$\overline{x^{n}_{\varepsilon}} = \begin{cases} 
\frac{x^{n}_{\varepsilon}}{1+\varepsilon} & \text{if } x^{n}_{\varepsilon} \geq \frac{1}{2}(1+\varepsilon), \\
\frac{x^{n}_{\varepsilon} - \varepsilon}{1 - \varepsilon} & \text{if } x^{n}_{\varepsilon} < \frac{1}{2}(1+\varepsilon), 
\end{cases} \quad \overline{x^{n}_{\varepsilon}} = \begin{cases} 
\frac{x^{n}_{\varepsilon}}{1-\varepsilon} & \text{if } x^{n}_{\varepsilon} \geq \frac{1}{2}, \\
\frac{x^{n}_{\varepsilon} + \varepsilon}{1+\varepsilon} & \text{if } x^{n}_{\varepsilon} < \frac{1}{2}. 
\end{cases}$$
Suppose that $I_1(g) \in (x^0_\varepsilon, x^n)$ and $x^n_\varepsilon \geq \frac{1}{2}(1 + \varepsilon)$ (the other three cases are proven analogously). We have
\[
x^n_\varepsilon g(1, 1) + (1 - x^n_\varepsilon) g(1, 2) - x^n_\varepsilon g(2, 1) - (1 - x^n_\varepsilon) g(2, 2)
\]
\[
= (x^n_\varepsilon - I_1(g))(g(1, 1) - g(1, 2) + g(2, 2) - g(2, 1))
\]
\[
= (x^n_\varepsilon - I_1(g))(g(2, 2) - g(2, 1))(\frac{1 - I_1(g)}{I_1(g)} + 1) \leq \frac{x^n_\varepsilon}{I_1(g)} - 1 \in (0, \varepsilon),
\]
where both equalities follow from (6). This finishes the proof of i).

We will prove ii) and iii) simultaneously, and then let $g_i(\varepsilon, u) = g_i(\varepsilon, 1), n \leq N_\varepsilon, u = -1, 1$. Fix $\varepsilon' > 0$ such that $(\frac{1 - \varepsilon}{1 + \varepsilon})^k = \frac{1 - \varepsilon'}{1 + \varepsilon'}$. Define the functions $g_i(\varepsilon, u), g_i(\varepsilon, 1), i \in \{1, 2\}, n \leq \lceil N_\varepsilon/k \rceil$, in such a way that
\[
g_{i, \varepsilon}^{n, 1, k}(1, 1) = g_{i, \varepsilon}^{n, 1, k}(2, 1), g_{i, \varepsilon}^{n, 1, k}(2, 2) = g_{i, \varepsilon}^{n, 1, k}(1, 2),
\]
\[
g_{i, \varepsilon}^{n, 1, k}(a, a') = g_{i, \varepsilon}^{n, 1, k}(a, a'), g_{i, \varepsilon}^{n, 1, k}(a, a') = g_{i, \varepsilon}^{n, 1, k}(a', a). a, a' \in \{1, 2\}
\]
The payoffs $g_{i, \varepsilon}^{n, 1, k}, n \leq \lceil N_\varepsilon/k \rceil$ are
\[
x_{\varepsilon}^{nk}(1) \geq \frac{1}{2} \Rightarrow g_{i, \varepsilon}^{n, 1, k}(2, 2) = 1, g_{i, \varepsilon}^{n, 1, k}(1, 1) = \frac{1 - \varepsilon - x_{\varepsilon}^{nk}(1)}{x_{\varepsilon}^{nk}(1)},
\]
\[
\frac{1}{2} < x_{\varepsilon}^{nk}(1) \geq \frac{1}{2} - \frac{\varepsilon'}{2} \Rightarrow g_{i, \varepsilon}^{n, 1, k}(2, 2) = 1, g_{i, \varepsilon}^{n, 1, k}(1, 1) = \frac{1 - x_{\varepsilon}^{nk}(1)}{x_{\varepsilon}^{nk}(1)} + \varepsilon',
\]
\[
\frac{1}{2} - \frac{\varepsilon'}{2} > x_{\varepsilon}^{nk}(1) \Rightarrow g_{i, \varepsilon}^{n, 1, k}(1, 1) = 1, g_{i, \varepsilon}^{n, 1, k}(2, 2) = \frac{x_{\varepsilon}^{nk}(1) + \varepsilon'}{1 - x_{\varepsilon}^{nk}(1)}.
\]
We verify conditions for payoffs for player 1 (with other case analogous). The payoffs $g_{i, \varepsilon}^{n, 1, k}$ are well defined, i.e. $g_{i, \varepsilon}^{n, 1, k}(a, a2) \in [0, 1], a_i \in \{1, 2\}$.

From $x^n_i \in [\frac{2}{1 + \varepsilon}, 1 + \varepsilon]$ follows that $g_{i, \varepsilon}^{n, 1, k}(1, 1), g_{i, \varepsilon}^{n, 1, k}(2, 2) \geq \varepsilon$, i.e. $g_{i, \varepsilon}^{n, 1, k} \in G^1_\varepsilon, n \leq N_\varepsilon$.

Moreover, for $k > 1$ we have $x^n_{\varepsilon} \in (\frac{2}{1 + \varepsilon}, 1 + \varepsilon)$ for $n \leq \lceil N_\varepsilon/k \rceil - 1$, and so $g_{i, \varepsilon}^{n, 1, k} \in G^1_\varepsilon$. Now, $g_{i, \varepsilon}^{n, 1, k}, g_{i, \varepsilon}^{n, 1, k}, x^n_{\varepsilon}$ and $x^n_{\varepsilon}(n-1)^k, n \leq \lceil N_\varepsilon/k \rceil$, (with $x^n_{\varepsilon}(1) = 1$) are defined in such a way that $g_{i, \varepsilon}^{n, 1, k}(2, x^n_{\varepsilon}) - g_{i, \varepsilon}^{n, 1, k}(1, x^n_{\varepsilon}) = g_{i, \varepsilon}^{n, 1, k}(1, x^n_{\varepsilon}) - g_{i, \varepsilon}^{n, 1, k}(2, x^n_{\varepsilon}) = \varepsilon$ and $g_{i, \varepsilon}^{n, 1, k}(1, x^n_{\varepsilon}(n-1)^k) - g_{i, \varepsilon}^{n, 1, k}(2, x^n_{\varepsilon}(n-1)^k) = g_{i, \varepsilon}^{n, 1, k}(2, x^n_{\varepsilon}(n-1)^k) - g_{i, \varepsilon}^{n, 1, k}(1, x^n_{\varepsilon}(n-1)^k) \geq \varepsilon'$. This establishes ii) and iii).

As for the bounds on $N_\varepsilon$, we have $N_\varepsilon \geq 2 * \left[ \log(1 + \varepsilon) \frac{1}{2} \right] - 1 \geq 2 * \left[ \frac{1}{2} \right] - 1 \geq \left[ \frac{1}{2} - 2 \right]$, where the second inequality follows from the fact that $(\frac{1 - \varepsilon}{1 + \varepsilon})^n \geq 1 - 2n\varepsilon$. We also have $N_\varepsilon \leq 2 * \log(1 + \varepsilon) \frac{1}{2} + 1 \leq 2 * (\frac{1 + \varepsilon}{2}) + 1 = \frac{1}{2} + 2,$ where the first inequality follows from
\((1-\varepsilon)^n \leq 1 - n\varepsilon(1 + \varepsilon)\), for \(n \leq \log\left(\frac{1}{1-\varepsilon}\right)\).

**Proof.** (Theorem 1) Consider any uncoupled learning strategies \(L_1^0 = \{\Omega_1, s_1, \phi_1, \pi_1\}\) and \(L_2^0 = \{\Omega_2, s_2, \phi_2, \pi_2\}\) of complexity \(c \in \mathbb{N}\). Let \(\varepsilon < \varepsilon\) be such that \(L_1\) and \(L_2\) converge to supported \(\varepsilon\)-Nash equilibrium over \(G^2\), and \((1-\varepsilon)^3 < \frac{1-2\varepsilon}{1-\varepsilon}\), which is possible as long as \(\varepsilon < 1/3\). For any game \((g_1, g_2)\) \(\in G^2\) define the sets

\[
E_i(g_1, g_2) := \{\omega \in \Omega_i | \Pr \exists T \exists t (L_1, L_2)^T(g_1, g_2) \mid \omega \times \Omega_3-i) > 0\}, \quad i = 1, 2.
\]

**Claim 1.** For the payoff functions \(g_{i,\varepsilon}^{n,u}\) as in Lemma 2, \(u = -1, 1, n \leq N\), and the sets \(E_i^n = \bigcup_{n' \leq N} E_i(g_{i,\varepsilon}^{n'-1}, g_{2,\varepsilon}^{n,1}) \cup E_i(g_{i,\varepsilon}^{n,1}, g_{2,\varepsilon}^{n-1}), \quad i \in \{1, 2\}, \quad n \leq N\), we have

\[
E_i^n \cap E_i^{n'} = \emptyset \text{ for all } n \neq n', \; n, n' \leq N, \; i \in \{1, 2\};
\]

\[
\pi_i(E_i^n) \subset \Delta_{N,K}^2, \; n \leq N.
\]

(Proof Claim 1) From the definition of sets \(E_i^n, \; n \leq N\), and the fact \(L_1\) and \(L_2\) converge to supported \(\varepsilon\)-Nash equilibrium over \(G^2\) follows that \(x_1 \in \pi_1(E_i^n)\) implies \(x_1 \in (\bigcup N \varepsilon E_{\pi_i}(g_{i,\varepsilon}^{n'-1}, g_{2,\varepsilon}^{n,1}) \cup N \varepsilon E_{\pi_i}(g_{i,\varepsilon}^{n,1}, g_{2,\varepsilon}^{n-1}))_1\). The proof of the Claim follows then from ii) in Lemma 2.

Fix two partitions \(P_i = \{p_i^{1d}, p_i^{2d}, p_i^{1s}, \ldots, p_i^{N_s}\}\) over the set of payoff functions such that if 1 is \(\varepsilon\)-dominant for 1 in \(g\) then \(g \in p_i^{1d}\), if only 2 is \(\varepsilon\)-dominant for 1 in \(g\) then \(g \in p_i^{2d}\) and if \(g \in p_i^{1s}\) then \(|g(1, x_i^n) - g(2, x_i^n)| < \varepsilon\). Define \(P_2\) analogously. The existence of such partitions follows from i) in Lemma 2.

For any pair of permutations \(f_1, f_2 : \{1, \ldots, N\} \rightarrow \{1, \ldots, N\}\) define a pair of uncoupled learning strategies \(L_i^{f_1,f_2} = \{\Omega_i, s_i^{f_1,f_2}, \phi_i, \pi_i^{f_1,f_2}\}, \; i = 1, 2\), of complexity \(c\), where:

\[
\pi_i^{f_1,f_2}|_{\Omega_i \setminus (E_i^n \cup \ldots \cup E_i^{N_s})}(1) = 1, \text{ if } \pi_i(1) > 0 \text{ and}
\]

\[
\pi_i^{f_1,f_2}|_{\Omega_i \setminus (E_i^n \cup \ldots \cup E_i^{N_s})}(2) = 1 \text{ otherwise,}
\]

\[
\pi_i^{f_1,f_2}(\omega) = x_i^{f_1-1}(n) \text{ for } \omega \in E_i^n, \; n \leq N_i,
\]

\[
s_i^{f_1,f_2}|_{p_i^{1d}} = s_i(g_{i,\varepsilon}^{1d} \cdot) \text{ for some } g_{i,\varepsilon}^{1d} \text{ for which } 1 \text{ is } \varepsilon \text{ + dominant for } i,
\]

\[
s_i^{f_1,f_2}|_{p_i^{2d}} = s_i(g_{2,\varepsilon}^{2d} \cdot) \text{ for some } g_{2,\varepsilon}^{2d} \text{ for which } 2 \text{ is } \varepsilon \text{ + dominant for } i,
\]

\[
s_i^{f_1,f_2}(g_i, \cdot) = s_i(g_i^{f_1(n), u} \cdot) \text{ if } g_i \in p_i^{1s} \text{ and } d(g) = u, \; u = -1, 1, \; n \leq N_i.
\]
Claim 2. Learning strategies $L_1^{f_1,f_2}$ and $L_2^{f_1,f_2}$ converge to $\varepsilon-$supported Nash equilibrium over $\mathcal{G}^2$.

(Proof Claim 2) First, observe that due to (14) the sets $E^n_i$ are absorbing, i.e. $\phi_i(\omega, (a, b)) (E^n_{i}) = 1$ for $\omega \in E^n_{i}$, $(a, b) \in \{1, 2\}^2$, $1 \leq n \leq N_{\varepsilon}$. Pick any game $(g_1, g_2) \in \mathcal{G}^2$. We consider three cases.

Case 1) Both agents have an $\varepsilon-$dominant action, say action 1 for either. In this case $(L_1^{f_1,f_2}, L_2^{f_1,f_2})_1^{e} (g_1, g_2) \equiv (L_1, L_2)_1^{e} (g_{1\varepsilon}^{ld}, g_{2\varepsilon}^{ld})$, and since with probability one

$$\exists T \forall t > T (L_1, L_2)_1^{e} (g_{1\varepsilon}^{ld}, g_{2\varepsilon}^{ld}) \in S_1 \times S_2,$$

where $\pi_1(S_1^1) = \pi_1^{f_1,f_2}(S_1^1) \in \{1\}$ and $\pi_2(S_2^1) = \pi_2^{f_1,f_2}(S_2^1) \in \{1\}$, the same holds for the process $\{L_1^{f_1,f_2}, L_2^{f_1,f_2}\}$ $E_i^{\phi}(g_1, g_2)_{t > 0}$.

Case 2) Only one agent has an $\varepsilon-$dominant action. In case e.g. 1 is an $\varepsilon-$dominant action for 1 and $d(g_2) = 1$, we have $(L_1^{f_1,f_2}, L_2^{f_1,f_2})_1^{e} (g_1, g_2) \equiv (L_1, L_2)_1^{e} (g_{1\varepsilon}^{ld}, g_{2\varepsilon}^{nld})$ for some $n \leq N_{\varepsilon}$ and the proof proceeds as in Case 1.

Case 3) No agent has an $\varepsilon-$dominant action. Suppose that $(g_1, g_2) \in (p_1^{0}, p_2^{0})$, $n, n' \leq N_{\varepsilon}$, i.e. $(x_{1\varepsilon}^{n'}, x_{1\varepsilon}^{n}) \in N E_{\varepsilon}^n (g_1, g_2)$. Then $(L_1^{f_1,f_2}, L_2^{f_1,f_2})_1^{e} (g_1, g_2) \equiv (L_1, L_2)_1^{e} (g_{1\varepsilon}^{n}, g_{1\varepsilon}^{n'})$ for some $u, u' \in \{-1, 1\}$. Since with probability one

$$\exists T \forall t > T (L_1, L_2)_1^{e} (g_{1\varepsilon}^{n}, g_{2\varepsilon}^{n'}) \in E_1^{f_2(n')} \times E_2^{f_1(n)},$$

the same holds for the process $\{L_1^{f_1,f_2}, L_2^{f_1,f_2}\}$ $E_i^{\phi}(g_1, g_2)_{t > 0}$. The results follows, since for any $(\omega_1, \omega_2) \in E_1^{f_2(n')} \times E_2^{f_1(n)}$ we have $\pi_1^{f_1,f_2}(\omega_1) = x_{1\varepsilon}^{n'}(f_2(n')) = x_{1\varepsilon}^{n'}$ and $\pi_2^{f_1,f_2}(\omega_2) = x_{1\varepsilon}^{n}(f_1(n)) = x_{1\varepsilon}^{n}$. This finishes the proof of Claim 2.

Consider the payoff functions $g_{1,\varepsilon}^{n,u,3}$, $u = -1, 1$, $n \leq \lfloor N_{\varepsilon}/3 \rfloor - 1$ defined in Lemma 3. From the definition it follows that

$$g_{1,\varepsilon}^{n,u,3} \in \left( \bigcup_{l=1,2,3} p_{i}^{3(n-1)+l} \right) \cap \bigcap_{l \leq 3(n-1), l > 3n} (p_{i}^{l})^{c},$$

\[22\]
$u = -1, 1 \leq \left| N_{\Delta}/3 \right| - 1, i = 1, 2$. Define for $i = 1, 2$ the sets of functions

$$F' = \left\{ f : \{1, \ldots, N_{\Delta}\} \rightarrow \text{sur} \{1, \ldots, N_{\Delta}\} \right\}$$

\forall n \leq \left| N_{\Delta}/3 \right| - 1 \exists n' \leq \left| N_{\Delta}/3 \right| - 1 f(3n - l) = 3n' - l,$

for $l \in \{0, 1, 2\}, f(n) = n$ for $n > 3 \left| N_{\Delta}/3 \right| - 3$.

Since $|F' \times F'| \geq \left| \frac{N_{\Delta}}{3} - 2 \right|^2 \geq \left| \frac{1}{6\epsilon} - 3 \right|^2 > (\left(\frac{1}{6\epsilon} - 3\right)/3)^2 > \frac{1}{\epsilon^2}$ for an appropriate $D > 0$, the following Claim establishes the proof of Theorem.

**Claim 3.** No uncoupled learning strategy $L_2$ converges to supported $\epsilon$–Nash equilibrium over $G^2$ with $L_1^{f_1, f_2}$ as well as $L_1^{f_1', f_2'}$ for $f_i, f_i' \in F'$, $i = 1, 2$, such that $(f_1, f_2) \neq (f_1', f_2')$.

(Proof Claim 3) Suppose that a learning strategy $L_2 = \{\Omega, s, \phi, \pi\}$ converges to supported $\epsilon$–Nash equilibrium over $G^2$ with $L_1^{f_1, f_2}$ and consider any $L_1^{f_1', f_2'}$ with $(f_1, f_2) \neq (f_1', f_2')$, $f_i, f_i' \in F'$, $i = 1, 2$. Let $n, n' \leq \left| N_{\Delta}/3 \right| - 3$ be such that $(f_1(3n), f_2(3n')) \neq (f_1'(3n), f_2'(3n'))$. With probability one\(^{22}\)

\begin{align*}
\exists \forall t &> T \left( L_1^{f_1, f_2}; L_2 \right)_l^\Delta \left( g_{1, \epsilon}^{n, 1, 3}, g_{2, \epsilon}^{n', 1, 3} \right) \in \bigcup_{l=0,1,2} \{ x_{\Delta}^{3n' - l} \} \times \pi_2 \left( E_m^2 \right), \quad (16) \\
\exists \forall t &> T \left( L_1^{f_1, f_2}; L_2 \right)_l^\Delta \left( g_{1, \epsilon}^{n, 1, 3}, g_{2, \epsilon}^{n', 1, 3} \right) \in \bigcup_{l=0,1,2} E_2^{f_2(3n' - l)} \times E_m^2, \quad (17)
\end{align*}

(where $E_m^2 \subset \Omega_2$ is such that $x \in \pi_2 \left( E_m^2 \right)$ implies $x(1) \in (x_{\Delta}^{3n}(1), x_{\Delta}^{3(n-1)}(1))$). From the definition of $L_1^{f_1', f_2'}$ it follows that with probability one

\begin{align*}
\exists \forall t &> T \left( L_1^{f_1, f_2}; L_2 \right)_l^\Delta \left( g_{1, \epsilon}^{f_1^{-1}(f_1(3n))/3, 1, 3}, g_{2, \epsilon}^{n', 1, 3} \right) \in \bigcup_{l=0,1,2} E_2^{f_2(3n' - l)} \times E_m^2, \quad (17) \\
\exists \forall t &> T \left( L_1^{f_1, f_2}; L_2 \right)_l^\Delta \left( g_{1, \epsilon}^{f_1^{-1}(f_1(3n))/3, 1, 3}, g_{2, \epsilon}^{n', 1, 3} \right) \in \bigcup_{l=0,1,2} \{ x_{\Delta}^{f_2^{-1}(f_2(3n' - l))} \} \times \pi_2 \left( E_m^2 \right).
\end{align*}

In the case when $f_2(3n') \neq f_2'(3n')$, i.e. $\bigcup_{l=0,1,2} \{ x_{\Delta}^{3n' - l} \} \neq \bigcup_{l=0,1,2} \{ x_{\Delta}^{f_2^{-1}(f_2(3n' - l))} \}$, the claim follows from the last line and the fact that $\bigcup_{l=0,1,2} \{ x_{\Delta}^{f_2^{-1}(f_2(3n' - l))} \} \cap N E_{\Delta}^2 \left( g_1^{f_1^{-1}(f_1(3n))/3, 1, 3}, g_{2, \epsilon}^{n', 1, 3} \right) = \emptyset$. In case when $f_2(3n') = f_2'(3n')$ (and $f_1(3n) \neq f_1'(3n)$, $g_1^{f_1^{-1}(f_1(3n))/3, 1, 3} \neq g_{1, \epsilon}^{n', 1, 3}$) the process $(L_1^{f_1, f_2}; L_2)(g_{1, \epsilon}^{f_1^{-1}(f_1(3n))/3, 1, 3}, g_{2, \epsilon}^{n', 1, 3})$ has exactly the same distribution as $(L_1^{f_1, f_2}; L_2)(g_{1, \epsilon}^{n, 1, 3}, g_{2, \epsilon}^{n', 1, 3})$. The result follows from the fact that $NE_{\Delta}^2 \left( g_1^{n, 1, 3}, g_{2, \epsilon}^{n', 1, 3} \right) \cap \bigcup_{l=0,1,2} \{ x_{\Delta}^{f_2^{-1}(f_2(3n' - l))} \} \neq \emptyset$.

\(^{22}\)the contradiction follows also if we require only that the following holds more than half of the time - see Remark 1.

23
\[ NE_\varepsilon'(g_{1,\varepsilon}(f_1(3n))/3,1,3, g_{2,\varepsilon}(f_2(3n))/3,1,3) = \emptyset. \]

**Proof.** (Remark 2) For the \( F' \) as in the proof of the Theorem 1, suppose that a learning strategy \( L_2 = \{\Omega, s, \phi, \pi\} \) converges to supported \( \varepsilon \)–Nash equilibrium over \( G^2 \) with \( L_1^{f_1,f_2} \) and consider any \( L_1^{f_1',f_2'} \) with \( f_2 \neq f_2', f_2', f_2' \in F' \). Since for any \( x_k^k, k \leq |N_\varepsilon| - 3, \]
\[ |g_{2,\varepsilon}^{-1,3}(x_k^k,1) - g_{2,\varepsilon}^{-1,3}(x_k^k,2)| > 2\varepsilon \]
unless \( k \in \{3n' - 2, 3n' - 1, 3n'\} \), it follows from (16) and (17), as well as \( \bigcup_{i=0,1,2} \{x_{3i}^{3i-1}\} \neq \bigcup_{i=0,1,2} \{x_{2i}^{2i-1}(f_2(3n'-i))\} \) that \( L_2 \) is not converging to a supported \( 2\varepsilon \)–best response over \( G^2 \) with \( L_1^{f_1,f_2'} \).

### 5.2 Theorems 2 and 3

For any \( t \in \mathbb{N} \) define the set of histories for time \( t \) as \( H^{N,0} = \{\emptyset\} \) and \( H^{2,t} = (\{1,2\}^2)^t, t > 0 \).

For any \( h^t = ((a_1^1, a_2^1), ..., (a_1^n, a_2^n)) \in \mathcal{H}^{2,t} \) let \( h^t\mid_{t-1} = ((a_1^1, a_2^1), ..., (a_1^{t-1}, a_2^{t-1})) \) for \( t > 1 \) and \( h^t; (a_1, a_2) = ((a_1^1, a_2^1), ..., (a_1^n, a_2^n), (a_1, a_2)) \in \mathcal{H}^{2,t+1} \), for any \( (a_1, a_2) \in \{1,2\}^2 \). For any pair \( (L_1, L_2) \) of learning strategies over \( G \subseteq G^2 \) and \( (g_1, g_2) \in G \) let \( \{G_1, G_2\}_{t=0}^{\infty} = (g_1, g_2) \}

For finite \( G_1 \times G_2 \subset G^2 \) and any partial function \( c : \bigcup_{t=0}^{\infty} \mathcal{H}^{2,t} \rightarrow \mathcal{P}(G_1) \times \mathcal{P}(G_2) \) let \( \text{Dom}^c := \{h^t \in \bigcup_{t=0}^{\infty} \mathcal{H}^{2,t} | c \text{ is defined on } h^t \} \) and \( \text{Dom}^{c'} := \{h^t \in \bigcup_{t=0}^{\infty} \mathcal{H}^{2,t} | h^t \in \text{Dom}^c \}

For any finite \( G_1 \times G_2 \subset G^2 \) a protocol for \( G_1 \times G_2 \) a partial function \( c : \bigcup_{t=0}^{\infty} \mathcal{H}^{2,t} \rightarrow \mathcal{P}(G_1) \times \mathcal{P}(G_2) \) such that

(D1) \( \text{Dom}^c \) is prefix closed: \( h^t \in \text{Dom}^c \) implies \( h^t\mid_{t-1} \in \text{Dom}^c, t > 1 \),

(D2) if \( h^t \in \text{Dom}^c \) then for some \( (x_1, x_2) \in \Delta^2 \times \Delta^2 \) \( (x_1, x_2) \in NE(g_1, g_2) \) for all \( (a_1, a_2) \in c(h^t), \)

(A1) \( c(\emptyset) = G_1 \times G_2, \)

(A2) if \( h^t \notin \text{Dom}^c \) then \( c(h^t; (a_1, a_2)) = c(h^t; (a_1, 1)) \cup \{x_1^t, x_2^t\} \times (c_2(h^t; (1, a_2)) \cup c_2(h^t; (2, a_2))), \)

(A3) if \( h^t \notin \text{Dom}^c \) then \( \{c(h^{t+1})\}_{t=0}^{\infty} \mid h^t \) is a partition\(^{23}\) of \( c(h^t) \) and \( c(h^t; (a_1, a_2)) \neq \emptyset \) for every \( (a_1, a_2) \in \{1,2\}^2 \),

(dom) If \( a_i \) is dominant for \( i \) for \( g_i \) then \( g_i \in c((a_1^1, a_2^1), ..., (a_1^n, a_2^n)) \) implies \( a_i^t = a_i, t \leq t. \)

A pair of learning strategies \( (L_1, L_2) \) over \( G_1 \times G_2 \) implements a protocol \( c \) if for every

\(^{23}\text{i.e. } c(h^t) = \bigcup_{(a_1, a_2) \in \{1,2\}^2} c(h^t; (a_1, a_2)) \text{ and } c(h^t; (a_1, a_2)) \cap c(h^t; (a'_1, a'_2)) = \emptyset \text{ for } (a_1, a_2) \neq (a'_1, a'_2).\)
\[ h^t = ((a_1^t, a_2^t), \ldots, (a_1^t, a_2^t)) \in \text{Dom}^c \]

\[ (g_1, g_2) \in c(h^t) \text{ iff } \text{Prob}((L_1, L_2)^{\{1,2\}}_t(g_1, g_2) = (a_1^t, a_2^t) \text{ for } l \leq t) > 0, \quad (18) \]

and if \( h^t \in \text{Dom}^c \) then \((g_1, g_2) \in c(h^t)\) implies \( \text{Prob}(\forall T > t (L_1, L_2)^{\Lambda} T(g_1, g_2) \in \text{NE}(g_1, g_2)) \)

\[(L_1, L_2)^{\{1,2\}}_t(g_1, g_2) = (a_1^t, a_2^t) \text{ for } l \leq t) = 1.\]

**Lemma 3** For any protocol \( c \) over \( G_1 \times G_2 \) and sufficiently low \( \beta \) there is a r.g. Bayesian Equilibrium \((L'_1, L'_2)\) for \( G_1 \times G_2 \) that implements \( c \).

**Proof.** For each learning strategy \( L'_i = \{\Omega_i, s_i, \phi_i, \pi_i\} \) the state space \( \Omega_i \), starting state function \( s_i \) and the transition function \( \phi_i \) will be defined as

\[
\begin{align*}
\Omega_i &= \{(g_i, h^t) | g_i \in G_i, h^t \in \text{Dom}^c, g_i \in c_i(h^t)\}, \\
s_i(g_i, g_{3-i}) &= (g_i, \emptyset), \\
\phi_i((g_i, h^t), (a_i, a_{3-i})) &= (g_i, (h^t; (a_i, a_{3-i}))) \text{ if } (h^t; (a_i, a_{3-i})) \in \text{Dom}^c, \\
&= (g_i, h^t) \text{ otherwise.}
\end{align*}
\]

In the following for a given \( c \) we define priors \( \nu^c_i, \pi \) and play functions \( \pi_i, i = 1, 2, \) and show that \((L'_1, L'_2)\) is a r.g. Bayesian equilibrium for \( G_1 \times G_2 \) for sufficiently low \( \beta \) with associated priors \( \nu^c_{G,1}, \nu^c_{G,2} \), which implements \( c \).

Let \( \gamma = \min\{I_i(g_i), 1 - I_i(g_i) | g_i \in G_i, i \text{ has no dominant action for } g_i, i = 1, 2\} > 0.\)

**Step 1.** For each \( i \in \{1,2\} \) and \( g_i \in G_i \) for which \( i \) has no dominant action we define inductively a function \( p_{g_i} : \{h^t \in \text{Dom}^c | g_i \in c_i(h^t)\} \rightarrow [0,1] \) in such a way that if \( h^t \notin \text{Dom}^c \) then

\[ p_{g_i}(h^t) = \sum_{(a_i, a_{3-i}) \in \{1,2\}^2} p_{g_i}(h^t; (a_i, a_{3-i})). \quad (20) \]

Let \( p_{g_i}(\emptyset) = 1 \). If \( p_{g_i}(h^t) \) is defined and \( h^t \notin \text{Dom}^c \) let \( \pi_i \) be such that \( g_i \in c_i(h^t; (\pi_i, a_{3-i})) \).

If \( \pi_i = 1 \) and \( d(g_i) = -1 \), or \( \pi_i = 2 \) and \( d(g_i) = 1 \) then put \( p_{g_i}(h^t; (\pi_i, 1)) = \gamma^2 \cdot p_{g_i}(h^t) \) (and \( p_{g_i}(h^t; (\pi_i, 2)) = (1 - \gamma^2) \cdot p_{g_i}(h^t) \)).

If \( \pi_i = 2 \) and \( d(g_i) = -1 \), or \( \pi_i = 1 \) and \( d(g_i) = 1 \) then put \( p_{g_i}(h^t; (\pi_i, 1)) = (1 - \gamma^2) \cdot p_{g_i}(h^t) \) (and \( p_{g_i}(h^t; (\pi_i, 2)) = \gamma^2 \cdot p_{g_i}(h^t) \)).

**Step 2.** Given the function \( p_{g_i} \) from Step 1 we define the priors \( \nu^c_{G,i} \) and the play functions \( \pi_i, i = 1, 2, \) and verify that \((L'_1, L'_2)\) implements \( c \). For any \( g_i \in G_i, g_{3-i} \in G_{3-i} \)
define \( \nu^c_{G,i}(g_i, g_{3-i}) = \frac{1}{|c_{G,i}||c_3-i|} \) if \( i \) has a dominant action for \( g_i \) and otherwise

\[
\nu^c_{G,i}(g_i, g_{3-i}) = \frac{1}{|G_i|} \frac{1}{|c_{3-i}|(h^t)} * p_g(h^t),
\]

where \( h^t \) is the unique \( h^t \in Dom^c \) such that \((g_i, g_{3-i}) \in c(h^t)\). For any \( h^t \in Dom^c \) fix some \((x_1, x_2) \in NE(g_1, g_2)\) for every \((g_1, g_2)\) and let

\[
\pi_i(g_i, h^t) = x_i.
\]

For \( h^t \notin Dom^c \) let

\[
\pi_i(g_i, h^t)(\pi_i) = 1,
\]

for \( \pi_i \) such that \( g_i \in c_i(h^t; (\pi_i, a_{3-i})) \) for some \( a_{3-i} \in \{1, 2\} \).

Due to \( p_{g_i}(\emptyset) = 1 \) and (20) we have that \( \sum_{g_i \in G_i} \sum_{g_{3-i} \in G_{3-i}} \nu^c_{G,i}(g_i, g_{3-i}) = \sum_{g_i \in G_i} \frac{1}{|G_i|} = 1 \). From Step 1. we know that for any \( g_i \) if \( g_i \in c_i(h^t) \) then \( p_{g_i}(h^t) > 0 \), which implies that \( \nu^c_{G,i}(g_i, g_{3-i}) > 0 \), \((g_i, g_{3-i}) \in G_i \times G_{-i} \), and so \( \nu^c_i, i = 1, 2 \), are well defined. From (23) follows also that for any \( g_i \in G_i, \pi_i \in \{1, 2\} \) and \( h^t \notin Dom^c \) we have \( g_i \in c_i(h^t; (\pi_i, a_{3-i})) \) for some \( a_{3-i} \in \{1, 2\} \) if and only if \( \pi_i(g_i, h^t)(\pi_i) = 1 \). Together with the definition of \( \phi_i \) this implies that \((L^c_1, L^c_2)\) implements \( c \).

**Step 3.** \((L^c_1, L^c_2)\) is a r.g. Bayesian equilibrium for \( G_1 \times G_2 \) with associated priors \( \nu^c_{G,1}, \nu^c_{G,2} \). Since \((L^c_1, L^c_2)\) implements \( c \) it is enough to show that for any \( g_i \in G_i, h^T \in Dom^c \) such that \( g_i \in c_i(h^T) \) and \( L_i^t \) we have

\[
\int_{\{(1,2)^\infty\}} \sum_{t=T}^\infty \sum_{g_i(a_i^{t+1}, a_{3-i}^{t+1})} \beta^{t-T} d\nu_i(L_i, L_{3-i}, \nu_{G,i})|g_i, h^T \geq
\]

\[
\int_{\{(1,2)^\infty\}} \sum_{t=T}^\infty \sum_{g_i(a_i^{t+1}, a_{3-i}^{t+1})} \beta^{t-T} d\nu_i(L_i^t, L_{3-i}, \nu_{G,i})|g_i, h^T.
\]

If \( i \) has a dominant action for \( g_i \) then (24) holds trivially due to (dom), and otherwise we proceed inductively as follows. If \( h^T \in Dom^c \) then (24) is implied by the definition of \( \pi_i(g_i, h^T) \) in (22) and \( \phi_i \) in (19). For \( h^T \notin Dom^c \), by induction, we must only consider the alternative strategies \( L_i^t \) that deviate in period \( T + 1 \) to pure actions, i.e. such that \( \nu_i(L_i, L_{3-i}, \nu_{G,i})|g_i, h^T(\{a_i^{T+1} = 1\}) \neq \nu_i(L_i^t, L_{3-i}, \nu_{G,i})|g_i, h^T(\{a_i^{T+1} = 1\}) \in \{0, 1\} \).
From the definitions (21) and (23), for any \( a_{3-i} \in \{1, 2 \} \)
\[
\nu_i(L_i', L_{3-i}, \nu_{G_i})|g_i, h^T(\{a_{3-i}^T = a_{3-i}\}) = \frac{p_{g_i}(h^T; (\overline{a_i}, a_{3-i}))}{p_{g_i}(h^T)},
\]
for \( \overline{a_i} \) such that \( g_i \in c_i(h^T; (\overline{a_i}, a_{3-i})) \) for some \( a_{3-i} \in \{1, 2\} \). (25) is equal to \( 1 - \gamma^2 \) when \( g_i(\overline{a_i}, a_{3-i}) \geq g_i(\overline{a_i}', a_{3-i}) \), due to Step 1. From the definition of \( \gamma \) it also follows that (24) holds with strict inequality for \( \beta = 0 \), and so holds when \( \beta \) is sufficiently small, i.e. when the upper bound on the discounted continuation payoff \( \frac{\beta}{1-\beta} \) is small.

**Proof.** (Theorem 2) Fix \( N \in \mathbb{N} \) such that \( 2^{N-2} \leq N_\epsilon < 2^{N-1} \) and \( \underline{\epsilon} < \epsilon \) such that \( (\frac{1-\epsilon^3}{1+\epsilon^2}) \leq \frac{1-\epsilon^2}{1+\epsilon^2} < (\frac{1-\epsilon^2}{1+\epsilon^2})^2 \) and \( 2 * N_\epsilon \leq 2^N \). For each \( i = 1, 2 \) consider the sets \( G_i \) consisting of two payoff functions \( g_i^{1d} \), \( g_i^{2d} \) with dominant actions 1 and 2, respectively, and \( 2^{N+1} \) functions \( g_i^{m,u} \in G_i^u \), \( n \leq 2^N, \ u = -1, 1, \) such that
i) \( g_i^{m,u} = g_{i,2e}^{k,u} \) for \( n = 4k, \ k \leq N_{2e} \);

ii) for every \( n' \leq N_\epsilon \) there is \( n \leq 2^N \) such that \( I_i(g_i^{m,u}) = x_i^{n'}(1) \);

iii) for every \( n \in \{4k+1, \ldots, 4k+4\}, \ k = 0, \ldots, N_{2e} - 1 \), we have \( I_i(g_i^{m,u}) \in [x_{2e}^{k+1}(1), x_{2e}^{k}(1)] \);

iv) \( I_i(g_i^{m,u}) > x_{2e}^{N_{2e}}(1) \) for \( 4N_{2e} < n \leq 2^N \), where \( x_i^{n'} \), \( x_i^{k} \) and \( g_i^{k,u} \) are defined in Lemma 2. The existence of such functions follows from \( (\frac{1-\epsilon^3}{1+\epsilon^2}) \leq \frac{1-\epsilon^2}{1+\epsilon^2} < (\frac{1-\epsilon^2}{1+\epsilon^2})^2 \) and the definitions in Lemma 2. From ii) above and i) in Lemma 2 follows that \( G_1 \times G_2 \) is \( \epsilon \)-dense.

Fix \( \beta > 0 \) for which Lemma 3 holds for \( G_1 \times G_2 \). For the sake of transparency we consider separately two cases.

**1. Case \( \beta \leq \beta \). Step 1.1.** Let \( F = \{ f : 2^N \rightarrow_{sur} \{1, 2\}^N \} \) and \( C = \{ e^{f_1f_2} | (f_1, f_2) \in F^2 \} \) consist of the protocols for \( G_1 \times G_2 \) such that

\[
c_i^{f_1f_2}((1, a_{3-i}); (1, a_{3-i}')) = \{g_i^{1d}\},
\]
\[
c_i^{f_1f_2}((2, a_{3-i}); (2, a_{3-i}')) = \{g_i^{2d}\},
\]
\[
c_i^{f_1f_2}((1, a_{3-i}); (2, a_{3-i}')) = \{g_i^{m,1}, n \leq 2^N\},
\]
\[
c_i^{f_1f_2}((2, a_{3-i}); (1, a_{3-i}')) = \{g_i^{m,1}, n \leq 2^N\}.
\]

For \( t > 2 \) \( e^{f_1f_2}(h^t) \) is defined only if \( h^t|_2 \in \{(2, 1); (1, 2); (1, 2); (2, 1)\} \), and then for any \( (a_1^{t}, a_2^{t}) \), \ldots, \( (a_1^{t}, a_2^{t}) \in \{1, 2\}^2, \ t \leq N, \)
\[
c_i^{f_1f_2}((a_1, a_2); (a_1', a_2'); (a_1^{t}, a_2^{t}); \ldots; (a_1^{t}, a_2^{t})) = \{g_i^{m,u}|(f_i(n))_{t'} = a_i^{t'}, \ t' \leq t\}.
\]
where \( u = 1 \) if \((a_i, a'_i) = (1, 2)\) and \( u = -1 \) if \((a_i, a'_i) = (2, 1)\), \( i = 1, 2 \). Note that \( c^{f_1f_2}(h^t) \) is not defined for \( t > N + 2 \).

**Step 1.2.** From Lemma 3 we know that for each \( c^{f_1f_2} \in C \) there is a r.g. Bayesian Equilibrium \((L_1^{c^{f_1f_2}}, L_2^{c^{f_1f_2}})\) for \( G_1 \times G_2 \) implementing \( c^{f_1f_2} \). We now verify that each \( L_i^{c^{f_1f_2}} \) can be chosen of complexity \((5 + N) \times 2^N\). For a given \( L_i^{c^{f_1f_2}} = \{ \Omega_i, s_i, \phi_i, \pi_i \} \) as in (19) we construct \( L_i^{c^{f_1f_2}} = \{ \Omega'_i, s'_i, \phi'_i, \pi'_i \} \) with \( \Omega'_i = \bigcup_{t=0,1,2} \{ (g_i, h^t) | g_i \in G_{i\xi}, h^t \in Dom_{g_i}^{c_i} \} \cup \{ \omega_{t,k} | t = 3, ..., N + 2, k \in \{ 1, 2 \} \} \), and a projection \( \iota : \Omega_i \rightarrow \Omega'_i \) such that \( s'_i(\omega) = s_i(\iota^{-1}(\omega)), \phi'_i(\omega, (a_1, a_2))(\omega') = \phi_i(\iota^{-1}(\omega), (a_1, a_2))(\iota^{-1}(\omega')) \) and \( \pi'_i(\omega) = \pi_i(\iota^{-1}(\omega)) \) are well defined. The projection \( \iota \) is an identity on \( \bigcup_{t=0,1,2} \{ (g_i, h^t) | g_i \in G_{i\xi}, h^t \in Dom^{c_i}, g_i \in c_i(h^t) \} \).

For \( t = 3, ..., N + 2 \) and \( h^t = ((a_1^t, a_2^t), ..., (a_1^t, a_2^t)) \) let:

\[ \iota(g_{i,k}^{m,u}, h^t) = \omega_{t,k}, \text{ with } k(t' - 2) = a_{3-t}^{t'} \text{ for } 3 \leq t' \leq t, \text{ } k(t' - 2) = (f_i(n))_{t'} \text{ for } t < t' \leq N + 2, \]

Since \( 2^{N-2} \leq N_\varepsilon \), we also have \( |\Omega'_i| \leq (5 + N) \times 2^N \leq (5 + \log 4N_\varepsilon + 4N_\varepsilon) \leq (5 + \log 4(\frac{1}{\varepsilon} + 2) \times 4(\frac{1}{\varepsilon} + 2)) \leq 5 \times \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} \) as long as \( \varepsilon < 0.1 \).

**Step 1.3.** From the fact that \( c(h^t) \) is not defined for \( t > N + 2 \), the definition of protocols and implementability it follows that the r.g. Bayesian Equilibria \((L_1^{c}, L_2^{c})\) converge to Nash equilibrium over \( G_1 \times G_2 \), proving i) in the Theorem.

Now, we proceed similarly as in Theorem 1 to prove ii). Fix some \( \overline{f} \in F \) and define the sets of functions

\[
F' = \{ f : 2^N \rightarrow_{sur} \{1, 2\}^N \} \quad \forall n \leq N_{2c} \ \exists n' \leq N_{2c} \ f(4n - l) = \overline{f}(4n' - l)
\]

\[
\text{for } l \in \{0, 1, 2, 3\}, \ f(n) = \overline{f}(n) \text{ for } n > 4N_{2c} \}.
\]

We check exactly as in Claim 3 of Theorem 1 that for any \( f_i, f'_i \in F' \), \( i = 1, 2 \), no learning strategy \( L \) can converge to supported \( 2\varepsilon - \text{Nash equilibrium} \) over \( G_1 \times G_2 \) against \( L_1^{c^{f_1f_2}} \) and \( L_2^{c^{f_1f_2}} \) unless \( (f_1, f_2) = (f'_1, f'_2) \), where \( (L_1^{c^{f_1f_2}}, L_2^{c^{f_1f_2}}) \) and \( (L_1^{c^{f'_1f'_2}}, L_2^{c^{f'_1f'_2}}) \) are any r.g. Bayesian equilibria over \( G_1 \times G_2 \) implementing \( c^{f_1f_2} \) and \( c^{f'_1f'_2} \), respectively. The number of such protocols is equal to \( |F' \times F'| \geq (N_{2c})^2 > (\lfloor \frac{1}{2\varepsilon} - 2 \rfloor)^2 > \left( \frac{1}{2\varepsilon} - 2 \right)^2 > \frac{1}{\varepsilon^2}D^m \), for an appropriate \( D^m > 0 \).

2. **Case \( \beta > \beta_1 \)**. **Step 2.1.** For parameters \( N' \in \mathbb{N}, \ N' > 0 \) and \( \theta > 0 \) define sets \( G_1^{N'} = \{ g_{i,k}^{m,u} | g_{i,k}^{m,u} \in G_i, \ k \in \{1, 2\}^{N'} \} \cup \{ g_{i,a}^{d} | a \in \{1, 2\}, \ k \in \{1, 2\}^{N'+N} \} \) such that
1) \( a \) is dominant for \( g_i^{ad}(a,...,a) \), \( a \in \{1,2\} \);

2) \( G_i^N \setminus \{g_i^{ld}(1,...,1), g_i^{2d}(2,...,2)\} \subset G_i^{-1} \cup G_i^1 ;

3) for any \( g_i^{m,u} \in G_i \) we have \( d(g_i^{m,u}) = d(g_i^{m,u}) \) for all \( k \), \( I_i(g_i^{m,u}) = I_i(g_i^{m,u}) \) for some \( k \) and

\[ |\min_k I_i(g_i^{m,u}) - \max_k I_i(g_i^{m,u})| < \theta; \]  \hspace{1cm} (26)

4) for \( g_i^{ad} \in G_i, a \in \{1,2\} \), we have

\[ |\min_{k \neq (a,...,a)} I_i(g_i^{ad}) - \max_{k \neq (a,...,a)} I_i(g_i^{ad})| < \theta^2, \]  \hspace{1cm} (27)

as well as \( d(g_i^{ad}) = 1 \) for \( k \neq (a,...,a) \), \( a \in \{1,2\} \), and \( 1 - \frac{\theta}{2} > I_i(g_i^{ad}) \geq 1 - \theta \) for \( a = 1 \) and \( \frac{\theta}{2} < I_i(g_i^{ad}) \leq \theta \) for \( a = 2 \). The exact values of the parameters \( N' \) and \( \theta \) will be specified in Step 2.4.

**Step 2.2.** Let \( C^{N'} = \{c_i^{f_1f_2} \mid (f_1, f_2) \in F^2 \} \) consist of the protocols for \( G_1^{N'} \times G_2^{N'} \) such that

\[ c_i^{f_1f_2}((1, a_{3-i}), (1, a'_{3-i})) = \{g_i^{ld} \mid k \in \{1,2\}^{N'+N} \}, \]

\[ c_i^{f_1f_2}((2, a_{3-i}), (2, a'_{3-i})) = \{g_i^{2d} \mid k \in \{1,2\}^{N'+N} \}, \]

\[ c_i^{f_1f_2}((1, a_{3-i}), (2, a'_{3-i})) = \{g_i^{m,1} \mid t \leq \min\{t, N\}, n \leq N, k \in \{1,2\}^{N'} \}, \]

\[ c_i^{f_1f_2}((2, a_{3-i}), (1, a'_{3-i})) = \{g_i^{m,-1} \mid t \leq \min\{t, N\}, n \leq N, k \in \{1,2\}^{N'} \}. \]

For any \( (a_1, a_2), ..., (a_1', a_2') \in \{1,2\}^2, t \leq N + N' \) let

\[ c_i^{f_1f_2}((a_i, a_{3-i}); (a_i', a'_{3-i}); (a_1, a_2); ...; (a_1', a_2')) = \{g_{i,k}^{m,u} \mid (f_i(n)) = a_i' \text{ for } t' \leq \min\{t, N\}, \]

\[ k(t'-N) = a_i', N < t' \leq t \}, \]

where \( u = 1 \) if \( (a_i, a_i') = (1,2) \) and \( u = -1 \) if \( (a_i, a_i') = (2,1) \), \( i = 1,2. \) For any \( (a_1, a_2), ..., (a_1', a_2') \in \{1,2\}^2, t \leq N' + N \) let

\[ c_i^{f_1f_2}((a_i, a_{3-i}); (a_i, a'_{3-i}); (a_1, a_2); ...; (a_1', a_2')) = \{g_{i,k}^{a,d} \mid k(t') = a_i' \text{ for } t' \leq t \}, \]

for \( a_i \in \{1,2\} \).

**Step 2.3.** For every \( c_i^{f_1f_2} \) we consider the following pair of learning strategies \( (L_i^{c_i^{f_1f_2}}, L_i^{c_i^{f_1f_2}}) \) for \( G_1^{N'} \times G_2^{N'} \). Let \( L_i^{c_i^{f_1f_2}} = \{\Omega_i, s_i, \phi_i, \pi_i \} \) be the learning strategies constructed in
Lemma 3, and let

\[ \Omega_i^a = \{(g_i, h^t) \in \Omega_i|h_t \in \text{Dom}^c, (g_i, g_{i-d_i}^{ad}(a, \ldots, a)) \in c_{1,2}^f(h^t)\}, a \in \{1, 2\}. \]

Define \( L_{i}^f = \{\Omega_i', s_i', \phi_i', \pi_i'\} \) such that \( \Omega_i' = \Omega_i \cup \{1, 2\}, s_i' = s_i, \pi_i'|_{\Omega_i} = \pi_i, \pi_i'|_{\Omega_i'} = a, a \in \{1, 2\}, \) and

\[
\begin{align*}
\phi_i'|_{\Omega_i \cup \Omega_i'} &= \phi_i, \\
\phi_i'((g_i, h^t), (a_i, a_{3-i}))(g_i, h^t) &= 1 \text { if } a_{3-i} = a, (g_i, h^t) \in \Omega_i^a, a \in \{1, 2\} \\
\phi_i'((g_i, h^t), (a_i, a_{3-i}))\langle\pi\rangle &= 1 \text { if } a_{3-i} \neq a, (g_i, h^t) \in \Omega_i^a, a \in \{1, 2\} \\
\phi_i'\langle\pi, (a_i, a_{3-i})\rangle\langle\pi\rangle &= 1.
\end{align*}
\]

**Step 2.4.** We now show that for every protocol \( c_{1,2}^f \in C^N \) the pair of learning strategies \( (L_{i}^{f_1}, L_{i}^{f_2}) \) for \( G_1^N \star G_2^N \) constitute a r.g. Bayesian equilibrium, for the appropriate choice of \( N' \) and \( \theta \).

We first show that player \( i \) with payoffs \( g_{i,k}^{n,u} \) is best off playing \( k(m') \) in rounds \( 2+N+m' \), \( m' \leq N' \). For \( m' = N' \) this is obvious as players repeat playing stage game Nash equilibrium, and for \( m' < N' \) by deviating to the other pure action \( i \) increases her expected discounted payoff by at most

\[
[1 \ast \gamma^2 - (\varepsilon - \theta) \ast (1 - \gamma^2)] + \theta \frac{\beta}{1 - \beta} \leq [\theta^2 - (\varepsilon - \theta) \ast (1 - \theta^2)] + \theta \frac{\beta}{1 - \beta},
\]

where the first part is her instantaneous loss (see Lemma 3, Step 1 and Step 3) and the second part is her discounted surplus from playing outside of Nash equilibrium from time \( N' + N + 2 \) on (see (26)). The second inequality follows from \( \gamma = \text{min}\{I_i(g_i), 1 - I_i(g_i)\} g_i \in G_i^N \), \( i \) has no dominant action for \( g_i, i = 1, 2 \) < \( \theta \).

Now we show that player \( i \) with payoffs \( g_{i,k}^{d} \) is best off playing \( k(m') \) in rounds \( 2+m', m' \leq N' + N \). Without loss of generality assume that \( a = 1 \). Note that \( \theta \frac{\beta}{2-\theta} < g_{i,k}^{d}(1,1) \leq \theta \frac{\beta}{1-\theta} \) (and \( g_{i,k}^{d}(1,2) = g_{i,k}^{d}(2,1) = 0, g_{i,k}^{d}(2,2) = 1 \)). If \( i \) mimics \( g_{i,k'}^{d}, k' \neq (1, \ldots, 1) \) then she increases her expected discounted payoff by at most

\[
[\theta^2 - \theta \frac{\theta}{2-\theta} \ast (1 - \theta^2)] + \theta^2 \frac{\beta}{1 - \beta},
\]

(see (27)). If \( i \) deviates to \( g_{i,k'}^{d}, k' = (a, \ldots, a) \), and, for example \( k = (1, \ldots, 1, 2) \), then the
respective bound is
\[\begin{align*}
[
\theta^2 - (1 - \theta^2) + \beta(1 - \theta) + \frac{\beta^2}{1 - \beta} \left( \frac{\theta}{1 - \theta} - \theta \right) &= 2\theta^2 - 1 + \frac{\beta}{1 - \beta} (1 - \beta + \beta \frac{\theta}{1 - \theta} - \theta),
\end{align*}\]

along the optimal deviation, when \(i\) plays 1 in round \(1 + N' + N, 2\) in round \(2 + N' + N\) and 1 forever after. In all three cases, when \(\theta\) is sufficiently low, the bound is strictly negative.

We now consider deviations in rounds \(1 + n, n = 0, \ldots, N + 1\) for payoffs \(g_{i,k}^{m,u}\) and rounds 1, 2 for payoffs \(g_{i,k}^{\beta d}\). For any such deviation, agent incurs a strictly positive instantaneous loss (see Lemma 3 Step 3), and the discounted benefit from the future is equal at most \(\frac{\beta N' + 1}{1 - \beta}\). For sufficiently large \(N'\) the deviations are unprofitable.

**Step 2.5.** We proceed exactly as in Step 1.3 to prove i) and ii) in the Theorem. ■

**Proof.** (Remark 3 - sketch) The first difference with respect to the proof of Theorem 2 is in the definition of the set \(G_1 \times G_2\). Fix a set \(G_1' \times G_2'\) and let \(s\) be such that for any \(n \leq N_{2s'}\) we have \(|\{g_i \in G_i' | d(g_i) = u, I_i(g_i) \in [x_{2s'}, x_{2s'}^u]\} | \leq s\), for \(u = -1, 1, i = 1, 2\).
Let also \(l\) be such that \((1 - \varepsilon)^l \leq \frac{1 - 2s'}{1 + 2s'} < (1 - \varepsilon)^{-l - 1}\). For each \(i = 1, 2\) define \(G_i = \{g_i^{\beta d}, g_i^{\beta d}\} \cup \{g_i^{s_0, 1}, \ldots, g_i^{s_{s_0}, 1}\} \cup \{g_i^{s_0, 1}, \ldots, g_i^{s_{s_0}, 1}\}\) for large enough \(N\), such that the following holds. \(g_i^{\beta d}\) and \(g_i^{\beta d}\) are payoff functions for which actions 1 and 2 are dominant for agent \(i\), respectively.
Moreover, \(G_i' \subset G_i\), and \(G_i\) satisfies the following, \(u = -1, 1\),

i) \(g_{i}^{m,u} = g_{i,2s'}^{k,u}\) for \(n = (l + s + 1)k, k \leq N_{2s'}\);

ii) for every \(n' \leq N_{2s'}\) there is \(g_{i}^{m,u} \in G_i\) such that \(I_i(g_{i}^{m,u}) = x_{2s'}^n(1)\);

iii) for every \(n \in \{(l + s + 1)k + 1, \ldots, (l + s + 1)k + l + s + 1\}, k = 0, \ldots, N_{2s'} - 1\), we have \(I_i(g_{i}^{m,u}) \in [x_{2s'}^{k+1}(-1), x_{2s'}^k(-1)]\);

iv) \(I_i(g_{i}^{m,u}) > x_{2s'}^{-(l + s + 1)N_{2s'}}\) for \((l + s + 1)N_{2s'} < n < 2^N\).

The r. g. Bayesian equilibria \((L_1, L_2), \ldots, (L_1, L_2)\) are constructed exactly as in Theorem 2. For any \(m \leq M\) such that \(L_1 = L_1^{F_1}, L_2 = L_2^{F_2}\), where \(f_1, f_2 \in \{f : 2^N \rightarrow_{sur} \{1, 2\}^N\}\), we define the set \(X^m\) as follows. For

\[F_i' = \{f : 2^N \rightarrow_{sur} \{1, 2\}^N\} \forall n \leq N_{2s'} \exists n' \leq N_{2s'} f((l + s + 1)n - l) = f_i((l + s + 1)n' - l) \text{ for } l \in \{0, 1, \ldots, l + s\}, f(n) = f_i(n) \text{ for } n > (l + s + 1)N_{2s'}, i = 1, 2\]

\[\text{i.e. } X^m = \{m' \leq M | L_1^{m'} = L_1^{F_1} f_1 \text{ such that } f_1 \in F_1' \text{ and } f_2' \in F_2'\}. \text{ The rest of the proof follows as in Theorems 1 and 2. ■}\]
Proof. (Theorem 3) Consider the lowest $\varepsilon < \varepsilon$ such that $2 \ast \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} < c$ and $2N = 2N, N \in N$. We construct the sets $X^1, \ldots, X^{2N}$ each with $N_1(\varepsilon, \underline{a})$ elements, $\frac{1}{9} \leq \frac{N_1(\varepsilon, \underline{a})}{N_2(\varepsilon, \underline{a})}$, $\frac{1}{N_2(\varepsilon, \underline{a})} \leq 1$, such that for every $g \in G_{2N}^1 \cup G_{2N}^{-1}$ and any $n \in X^n, x' \in X^{n'}, n' \neq \varepsilon$, $n, n' \leq N_1(\varepsilon, \varepsilon)$,

$$\text{either } |g(a_i, x) - g(a_i^c, x)| > 2\varepsilon \text{ or } |g(a_i, x') - g(a_i^c, x')| > 2\varepsilon.$$ (28)

For this purpose, choose the lowest $k$ such that $\frac{1 - \varepsilon}{1 + 2\varepsilon} < \varepsilon < \frac{1 - 2\varepsilon}{2\varepsilon}$, and for $N_1(\varepsilon, \varepsilon) = \frac{N_2(\varepsilon, \varepsilon)}{N_2(\varepsilon, \varepsilon)} - 1$ define $X^n = \{x_1, \ldots, \ell x_{k+1}, \ldots, x_{2k+1}, \ldots, x_{2n+1} \}$, $n \leq \frac{N_2(\varepsilon, \varepsilon)}{N_2(\varepsilon, \varepsilon)} - 1$. Since $x_{k+1}^2(2n+1+1) \leq x_{k+1}^2(2n+1+1), n \leq \frac{N_2(\varepsilon, \varepsilon)}{N_2(\varepsilon, \varepsilon)} - 1$, this guarantees that $1/8 \leq \frac{N_1(\varepsilon, \varepsilon)+1}{N_2(\varepsilon, \varepsilon)} \leq 1$ and $1/3 \leq \frac{N_2(\varepsilon, \varepsilon)+1}{N_2(\varepsilon, \varepsilon)} \leq 1/2$. The property (28) follows from iii) in Lemma 2.

Fix sets $G_i$ of payoff functions consisting of two payoff functions $g_i^{1d}, g_i^{2d}$ with dominant actions 1 and 2, respectively, and $2 \ast 2N$ functions $g_i^{m,u} \in G_{2N}$ such that $d(g_i^{m,u}) = u$ and $I_i(g_i^{m,u}) = x_{k+1}^1(1)$, for $n \leq N_2(\varepsilon, \varepsilon), i \in \{1, 2\}, u = -1, 1$. Let $\{(L_1^{f_1f_2}, L_2^{f_1f_2})|f_i : 2N \rightarrow_{\text{sur}} \{1, 2\}^N, \text{ } i \in \{1, 2\}\}$ be the set of the repeated game Bayesian equilibria over $G_1 \times G_2$, for $\beta$ sufficiently low so that each $L_i^{f_1f_2}$ has complexity $5 \ast \frac{1}{\varepsilon} \log \frac{1}{\varepsilon} < c$, constructed as in Steps 1.1 and 1.2 in Theorem 2. Each pair $(L_1^{f_1f_2}, L_2^{f_1f_2})$ converges to Nash equilibrium over $G_1 \times G_2$.

Fix some $\mathcal{F} : 2N \rightarrow_{\text{sur}} \{1, 2\}^N$ and let

$$F' = \{f : 2N \rightarrow_{\text{sur}} \{1, 2\}^N | f|_{X^n} \leq X^{n'} \} \text{ iff } \mathcal{F}(n') > \mathcal{F}(n'), n' \leq N_1(\varepsilon, \varepsilon).$$

We have

$$|F'| = \frac{(N_1(\varepsilon, \varepsilon) + N_2(\varepsilon, \varepsilon))!}{N_2(\varepsilon, \varepsilon)! N_1(\varepsilon, \varepsilon)!} \geq d_1 \frac{(N_1(\varepsilon, \varepsilon)N_2(\varepsilon, \varepsilon))N_2(\varepsilon, \varepsilon)}{N_2(\varepsilon, \varepsilon)N_1(\varepsilon, \varepsilon)N_2(\varepsilon, \varepsilon)} = d_1 N_1(\varepsilon, \varepsilon)N_2(\varepsilon, \varepsilon) \geq d_2 \frac{d_3}{\varepsilon} \geq d_2 \frac{d_3}{\varepsilon} f(c) > \frac{1}{\varepsilon} d_6 f(c),$$

where $d_1, \ldots, d_6$ are constants independent of $\varepsilon$ and $c$, and $f(c) = h^{-1}(c)$ for $h(x) = x \log x$ (and so $\sqrt{c} < f(c)$). Due to (28) we show as in the proof of Theorem 2 that if any $L_1'$ converges to supported $2\varepsilon - \text{Nash equilibrium over } G_1 \times G_2$ with $L_1^{f_1f_2}$ and $L_1^{f_1f_2}$ for $f_1, f'_1 \in F'$, then $f_1 = f'_1$. For the roles of players 1 and 2 reversed the proof is analogous. $\blacksquare$

$24$ for any linear strict ordering $"<" \text{ over } \{1, 2\}^N$;
5.3 Definitions for Section 4

(L-learning strategy) For any communication strategy $L = \{\Omega, s, \phi, \pi\}$ for agent $i$ and $H$—tests $\{\{S_x, \phi_x\}\}_{x \in H}$, we define an L-learning strategy $\{\Omega', s', \phi', \pi'\}$ as follows.

\[
\Omega' = \Omega \cup \bigcup_{h \in H, p \in P, k \leq K} h_h \omega \in S_h \setminus \{h_u, h_d\}, k \leq K ,
\]

\[
s' = s,
\]

\[
\pi'(\omega) = \pi(\omega), \omega \in \Omega h_h \{h_p | h \in H, k \leq K ,
\]

\[
\pi'(\omega_p) = p, \omega \in S_h \{h_u, h_d\}, h \in H, k \leq K ,
\]

\[
\phi'(\omega^k_p, (a_{i}, a_{3-1})) = \phi_h(\omega, (a_{3-1}))(\omega'), \omega, \omega' \in S_h \{h_u, h_d\}, h \in H, k \leq K ,
\]

\[
\phi'(\omega^k_p) = \phi_x(\omega, (a_{3-1}))(\omega'), \omega \in S_h \{h_u, h_d\}, \omega^k \in \{h_u, h_d\}, h \in H, k \leq K ,
\]

\[
\phi'(\omega, (a_1, a_2))(\omega') = \phi(\omega, (a_1, a_2))(\omega'), \omega \in \Omega \{h^k_p | (h, p) \in H \times P, k \leq K\}, \omega' \in \Omega.
\]

(Example 2) Fix a parameter $\varepsilon, \tau > 0$, which will be a measure of strength of agent’s test. Define $\phi$ in such a way that for every $k$

- (test start) $\phi((n^k, (0^k), n^k), (\cdot, \cdot))((n^k, (0^k), n^k)) = \frac{M-1}{M}$,
- $\phi((n^k, (0^k), n^k), (\cdot, 1))((n^k, (1^k), n^k)) = \frac{1}{M}$,
- $\phi((n^k, (0^k), n^k), (\cdot, 2))((n^k, (1^k), n^k)) = \frac{1}{M}$,

- (test) $\phi((n^k, (m^k), n^k), (\cdot, 1))((n^k, ((m+1)^k), (m+1)^k)) = 1$,
- $\phi((n^k, (m^k), n^k), (\cdot, 2))((n^k, ((m+1)^k), m^k)) = 1$ for any $m^k \leq m^k < M$, $n^k, n^k \in \{0^k, \ldots, N^k\}$;

- (test passed) $\phi((n^k, (M^k), n^k), (\cdot, \cdot))((n^k, (0^k), n^k))$, when $\frac{m^k}{M} \in (\frac{n}{N} - \tau, \frac{n}{N} + \tau)$,
- (test failed) $\phi((n^k, (M^k), n^k), (\cdot, \cdot))((n^k, (0^k), n^k)) = \frac{1}{M(\tau + 2N-1)}$, where $n^k \in \{0^k, \ldots, N^k\}$ and $n^k = 0$ for $n < k - 1 \leq N$ or $n > N + k$, $n^k = N^k$ for $k \leq n \leq N$ or $N \leq n < N + k$, and $n^k \in \{0^k, \ldots, N^k\}$ otherwise.

Moreover, let $\frac{n}{N} = \frac{1}{2N-1}$ and $s(g_1, g_2) = ((n^I+N^I), (M^I+N^I), n^I+k))$ for arbitrary $n^k$, where $I = N$ if $g_1(1, 1) > g_1(2, 1)$ and $I = 0$ otherwise, and $k \leq N$ is such that $g_1(1, x) = g_1(2, x)$ implies $x(1) \in \{\max\{0, \frac{k-1}{N}\}, \min\{1, \frac{k}{N}\}\}$.

5.4 Theorem 4

For a pair of learning strategies $L_1 = \{\Omega_1, s_1, \phi_1, \pi_1\}$ and $L_2 = \{\Omega_2, s_2, \phi_2, \pi_2\}$ for $G^2$ together with $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$ define random processes $\{(L_1, L_2)_t|^\infty_{t=1}(\omega_1, \omega_2)\}$ with values in $\Omega_1 \times \Omega_2$,
as well as \( \{(L_1, L_2)\}^\infty_{i=1} \) with values in \( \Delta^N \times \Delta^N \), which start with probability one in states \( (\omega_1, \omega_2) \) and \( (\pi_1(1), \pi_2(\omega_2)) \), respectively.

**Proof.** (Theorem 4) Fix \( \{L^n_i\}_{n \in \mathbb{N}}, i = 1, 2 \), as in the Theorem and let \( \overline{g} \) be the "normalized" matching pennies game, i.e. \( \overline{g}_1(1, 1) = \overline{g}_1(2, 2) = \overline{g}_2(1, 2) = \overline{g}_2(2, 1) = 1 \), and otherwise the payoffs are 0.

Consider any \( L_i \)-learning strategy \( L_i^N = \{\Omega_i^N, s_i^N, \phi_i^N, \pi_i^N\}, i \in \{1, 2\}, N \in \mathbb{N} \).

**Claim 1.** For any \( h_k^z, h \in H_{3-i}, z = u, d, k \leq K_{3-i}, \) and \( \omega \in \Omega_i^N \)

\[
P(\lim_{T \to \infty} \frac{|\{t \leq T | (L_i^N, L_{3-i})^{h_k^z}(\omega, h_k^z) \in NE_\varepsilon^z(\overline{g})\}|}{T} < 1 - \gamma) = 1
\]

The process \( \{(L_i^N, L_{3-i})^{h_k^z}(\omega, h_k^z)\}^\infty_{i=1} \) is a finite state Markov process over the state space \( \Omega_i^N \times \Omega_{3-i} \). Therefore it follows from the ergodic Theorem that for any \( (\omega_i, \omega_{3-i}) \in \Omega_i^N \times \Omega_{3-i} \)

\[
P(\lim_{T \to \infty} \frac{|\{t \leq T | (L_i^N, L_{3-i})^{h_k^z}(\omega, h_k^z) = (\omega_i, \omega_{3-i})\}|}{T} = \psi(\omega_i, \omega_{3-i})) = 1,
\]

where \( \psi \) is an ergodic distribution over \( \Omega_i^N \times \Omega_{3-i} \). On the other hand, from (4) and the definition of ergodic distributions follows that for any \( x \in \Delta^N_\varepsilon \), \( \psi\{\{\omega_1, \omega_2\}|(\pi_{3-i}(\omega_{3-i}) \in W^\varepsilon(x)\} > \frac{1}{1 - \gamma} * \psi\{\{\omega_1, \omega_2\}|(\pi_{3-i}(\omega_{3-i}) = x\} \). Since \( |\overline{g}_i(1, x_{3-i}) - \overline{g}_i(2, x_{3-i})| < \varepsilon \) implies \( |\overline{g}_i(1, x'_{3-i}) - \overline{g}_i(2, x'_{3-i})| > \varepsilon \), it follows that \( \psi\{\{\omega_1, \omega_2\}|(\pi_i^N(\omega_i), \pi_{3-i}(\omega_{3-i}) \in NE_\varepsilon(\overline{g})\}) < 1 - \gamma \). This establishes Claim 1.

**Claim 2.** For any \( (h, p) \in H_{3-i} \times P_{3-i}, k \leq K_{3-i}, \omega \in \Omega_i^N \) and \( n \in \mathbb{N} \)

\[
P(\lim_{T \to \infty} \frac{|\{t \leq T | (L_i^N, L_{3-i})^{h_k^z}(\omega, h_k^z) \in NE_\varepsilon^z(\overline{g})\}|}{T} < 1 - \gamma | \forall t \geq 1(L_i^N, L_{3-i})^{h_k^z}(\omega, h_k^z) \notin \{h^z_u, h^z_d\}) = 1.
\]

Conditional on the event \( \forall t \geq 1(L_i^N, L_{3-i})^{h_k^z}(\omega, h_k^z) \notin \{h^z_u, h^z_d\} \) the process \( \{(L_i^N, L_{3-i})^{h_k^z}(\omega, h_k^z)\}^\infty_{i=1} \) is a finite state Markov process over the state space \( \Omega_i^N \), with transition probabilities

\[
p(\omega, \omega') = \sum_{a_1, a_2 \in \{1, 2\}} p(a_2) * \pi_i^N(\omega)(a_1) * \phi_i^N((\omega, (a_1, a_2))(\omega')),
\]

and so for any \( \omega_i \in \Omega_i^N \)

34
where \( \psi \) is an ergodic distribution over \( \Omega_i^N \). On the other hand, from the definition of \( \phi_i^N \) for any \( h \in H_i, \ p, p' \in P_i, \ k' \leq K_i \) and \( \omega, \omega' \in S_i^N \{ h_u, h_d \} \) we have

\[
p((\omega^{k'}, p^{k'}), (\omega^{k''}, p^{k''})) = p((\omega^{k'}, p^{k'}), (\omega^{k''}, p^{k''}))
\]

From this as well as (3), (4) and the second line in (5) it follows that for any \( \psi \) that theorem that dent draws from the distribution \( \pi \) and all \( \omega, \omega' \in S_i^N \{ h_u, h_d \} \) we have

\[
P(\lim_{T \to \infty} \frac{|\{t \leq T|(L_i^T, L_{3-i}^T)_{i,j}(\omega, h_p^{k}) = \omega_i\}|}{T} = \psi(\omega_i))
\]

\[
\forall t \geq 1(L_i^T, L_{3-i}^T)_{i,j}(\omega, h_p^{k}) \notin \{h_u^{k}, h_d^{k}\} = 1,
\]

where \( \chi > 0 \), for sufficiently large \( n \) and some \( i \in \{1, 2\} \) for all sequences \( (\overrightarrow{m}, ..., \overrightarrow{m}) \in \{1, 2\}^T \), and all \( h \in H_i \) we have

\[
\text{Prob} \left( \min \{t|\phi_{i,h,n}(x^n) \notin \{h_u, h_d\} \} \geq T(\overrightarrow{m}, ..., \overrightarrow{m}) \right) \geq 1 - \chi,
\]

where \( L_i^T \) is a \( L_i \)-learning strategy for \( L_i \) and tests \( \{\{S_{i,h,n}, \phi_{i,h,n}\}\} \in H_i \).

First, it is enough to prove it for a single \( h \in H_i \), due to (5). Fix any finite sequence \( (\overrightarrow{m}, ..., \overrightarrow{m}) \) and \( n \in N \). If

\[
\text{Prob} \left( \min \{t|\phi_{i,h,n}(h) \notin \{h_u, h_d\} \} \geq T(\overrightarrow{m}, ..., \overrightarrow{m}) \right) < 1 - \chi,
\]

for each \( i \in \{1, 2\} \) and all \( x \in H_i \), then for any \( (\omega_1, \omega_2) \in \Omega_1^N \times \Omega_2^N \)

\[
\text{Prob} \left( \{t \leq T + 1|(L_i^T, L_{3-i}^T)_{i,j}(\omega_1, \omega_2) \notin NE^\varepsilon_0(\overrightarrow{m})\} \right) > \frac{T}{T + 1} < (1 - \chi) * e(T) * \gamma,
\]

where \( e(T) \) is a lower bound on probability of sampling a sequence \( (\overrightarrow{m}, ..., \overrightarrow{m}) \) in \( T \) independent draws from the distribution \( x \in \Delta^2 \) with \( x(1) \in (\frac{1-\delta}{2}, \frac{1+\delta}{2}) \). It follows from the ergodic theorem that \( (L_i^T, L_{3-i}^T) \) does not converge to supported \( \varepsilon \)-Nash equilibrium \( 1 - \delta \) of the time for \( \delta \) sufficiently close to 0. Choosing \( n \) sufficiently high we establish the desired contradiction and prove Claim 3.

Fix \( N \in N \). Consider \( T_1 \) such that for any \( h_z, h \in H_{3-i}, \ z = u, d, \) and \( \omega \in \Omega_i^N \), for every
\[ T > T_1 \text{ (see Claim 1)} \]

\[
E(\frac{|\{t \leq T | (L_i^N, L_{3-i})_t^\Delta(\omega, h_p^k) \in NE^\varepsilon(\gamma)\}|}{T}) < 1 - \gamma. \tag{29}
\]

Let \( T_2 \in \mathbb{N} \) and \( \chi > 0 \) be such that such that for any \( i \in \{1, 2\}, n \in \mathbb{N}, \omega \in \Omega_i^N, (h, p) \in H_{3-i} \times P_{3-i}, k \leq K_{3-i} \) and \( T > T_2 \) (see Claim 2)

\[
E(\frac{|\{t \leq T | (L_i^N, L_{3-i})_t^\Delta(\omega, h_p^k) \in NE^\varepsilon(\gamma)\}|}{T_1 + T}) < 1 - \gamma. \tag{30}
\]

Finally, for \( T = T_2 \) and \( \chi \) as above consider any \( n \) sufficiently large and \( i \in \{1, 2\} \) as in Claim 3. The proof follows from the application of the ergodic theorem, when we use Claim 1 for the limiting ergodic distributions \( \psi \) on \( \Omega_i^N \times \Omega_{3-i}^N \) with \( \psi(\{h_p^k | (h, p) \in H_{3-i} \times P_{3-i}, k \leq K_{3-i}\}) \times \Omega_i^N = 0 \), Claim 2 for the case \( \psi(\{h_p^k | h \in H_{3-i}, k \leq K_{3-i}\} \cup \{h_d^k | h \in H_{3-1}, k \leq K_{3-i}\}) \times \Omega_i^N = 0 \) and (29) and (30) otherwise.

\section*{References}


