UC Irvine
UC Irvine Previously Published Works

Title
Reinforcement learning in signaling game

Permalink
https://escholarship.org/uc/item/37r7t1fz

Authors
Hu, Yilei
Skyrms, Brian
Tarrès, Pierre

Publication Date
2011-03-29

Peer reviewed

Reinforcement learning in signaling game

Yilei Hu[1], Brian Skyrms[2] and Pierre Tarrès[3]

March 31, 2011

Abstract

We consider a signaling game originally introduced by Skyrms, which models how two interacting players learn to signal each other and thus create a common language. The first rigorous analysis was done by Argiento, Pemantle, Skyrms and Volkov (2009) with 2 states, 2 signals and 2 acts. We study the case of M_1 states, M_2 signals and M_1 acts for general M_1, M_2 $\in \mathbb{N}$. We prove that the expected payoff increases in average and thus converges a.s., and that a limit bipartite graph emerges, such that no signal-state correspondence is associated to both a synonym and an informational bottleneck. Finally, we show that any graph correspondence with the above property is a limit configuration with positive probability.

[1]University of Oxford, Mathematical Institute, 24-29 St Giles, Oxford OX1 3LB, United Kingdom. E-mail: `Yilei.Hu@maths.ox.ac.uk`.

[2]School of Social Sciences, University of California at Irvine, CA 92607. E-mail: `bskyrms@uci.edu`

[3]CNRS, Université de Toulouse, Institut de Mathématiques, 118 route de Narbonne, 31062 Toulouse Cedex 9, France. On leave from the Mathematical Institute, University of Oxford. E-mail: `tarres@math.univ-toulouse.fr`

1 Introduction

1.1 Signaling game

Signaling games aim to provide a theoretical framework for the following basic question: how do individuals create a common language? The setting was introduced as follows by Philosopher D. Lewis (1969).

Consider two players, one called Sender, who is regularly given some information that the other does not have and seeks to transmit it, and the other called Receiver. Sender has a fixed set of signals at his disposal throughout the game, which do not have any intrinsic meaning at the very beginning, in the sense that no signal is *a priori* associated to any state; similarly Receiver has a fixed set of possible acts. The game is thereafter repeatedly played according to following steps: (1) Sender observes a certain state of nature, of which Receiver is not aware. (2) Sender chooses a signal and then sends it to Receiver. (3) Receiver observes the signal but not the state, and then chooses an act. (4) Both players receive payoffs at the end of each round, which are functions of state and act.

The process involving the above four steps is called one *communication.* Sender, Receiver, states, signals and acts constitute one basic *communication system.* The game lies in the choices of signals and acts by the agents. Note that we do not fix any strategy at this point; this is the purpose of Section 1.4.

1.2 Mathematical definition

As we will explain at length later on, we adopt a dynamic perspective to investigate this game. The outcome of once-play game is not of much interest to us. What we wonder is whether player can establish a signaling system, within which each signal is uniquely related to a certain state, if they play this game repeatedly.

In this section, we provide solid mathematical definitions for the objects appearing in Section 1.1 and we end with a mathematical definition of repeated signaling game.

Probability space. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, which is a sufficiently rich source of randomness. More specifically, the probability space is at least rich enough for all the random variables appearing in this section to live.

State spaces. Let $\mathcal{S}_1$ be the set of states, $\mathcal{S}_2$ be the set of signals and $\mathcal{A}$ be the set of acts.

Players and strategies. We here introduce Nature as a player in this game who assigns a state of nature to Sender at each round of the game. Three players, Nature, Sender and Receiver respectively generate a random sequence, denoting their strategies throughout this repeated game. More specifically,

(1) Nature generates a sequence $(S_n)_{n \in \mathbb{N}}$ of random variables taking value in the set of states $\mathcal{S}_1$, each denoting which state Nature assigns to Sender at each round of the game.

(2) Sender generates a sequence $(Y_n)_{n \in \mathbb{N}}$ of random variables taking in value in the set of signals $\mathcal{S}_2$, each denoting which signal Sender sends to Receiver at each round of the game.

(3) Receiver generates a sequence $(Z_n)_{n \in \mathbb{N}}$ of random variables taking value in the set of states $\mathcal{A}$, each denoting which state Receiver interpret the signal as at each round of the game.

Payoffs. Let mappings u_n^1 and u_n^2 from $\mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{A}$ to $[0, \infty)$ be payoff functions for Sender and Receiver at the n-th round of the game. In other words, Sender and Receiver gain payoffs $u_n^1(S_n, Y_n, Z_n)$ and $u_n^2(S_n, Y_n, Z_n)$ at the end of n-th round of the game respectively.

Information. Let filtration $(\mathcal{F}_n^1)_{n \in \mathbb{N}}$ (resp. $(\mathcal{F}_n^2)_{n \in \mathbb{N}}$) denote the information available to Sender (resp. Receiver) before he makes his decision at each round of the game: for each $n \in \mathbb{N}$, we let

$$
\begin{aligned}
\mathcal{F}_n^1 &:= \sigma\Big(S_i, 0 \leqslant i \leqslant n, \ Y_i, u_n^1(S_i, Y_i, Z_i), 0 \leqslant i \leqslant n-1\Big), \\
\mathcal{F}_n^2 &:= \sigma\Big(Y_i, 0 \leqslant i \leqslant n, \ Z_i, u_n^2(S_i, Y_i, Z_i), 0 \leqslant i \leqslant n-1\Big).
\end{aligned}
$$

Initial settings. The distributions of i.i.d random variables $(S_n)_{n \in \mathbb{N}}$ and the distributions of Y_0 and Z_0 are given at the beginning of the game, which are called *priori* distribution.

Updating rule for strategies. Let $\mathcal{F}_n^1$-measurable mapping ρ_n^1

$$
dist.(Y_{n+1}) = \rho_n^1\Big(S_i, 1 \leqslant i \leqslant n, \ Y_i, u_i^1(S_i, Y_i, Z_i), 1 \leqslant i \leqslant n-1\Big) \tag{1}
$$

be updating rule of strategies for Sender at n-th round of the game.

Let $\mathcal{F}_n^2$-measurable mapping ρ_n^2

$$dist.(Z_{n+1}) = \rho_n^2 \Big(S_i, 1 \leqslant i \leqslant n,\ Z_i, u_i^2(S_i, Y_i, Z_i), 1 \leqslant i \leqslant n-1 \Big) \qquad (2)$$

be updating rule of strategies for Receiver at n-th round of the game. Different $(\rho_n^1)_{n\in\mathbb{N}}$ and $(\rho_n^2)_{n\in\mathbb{N}}$ represent different learning rules.

Definition 1.1 (Signaling game). *A repeated signaling game G is defined as:*

$$G = \Big((\Omega, \mathcal{F}, \mathbb{P}), \mathcal{S}_1, \mathcal{S}_2, \mathcal{A}, (S_n, Y_n, Z_n, u_n^1, u_n^2, \rho_n^1, \rho_n^2)_{n\in\mathbb{N}} \Big).$$

1.3 Questions

Throughout the paper, we limit our attention to a special, but most common, circumstance under which

(A1) the set of acts matches the set of states of nature by a bijective map; and

(A2) both players only receive fixed payoffs when the act chosen by Receiver matches the state of nature, which is considered as a *successful communication*; otherwise they obtain nothing.

Under these assumptions, the act can be understood as an interpreted state. Note however that the Receiver does not necessarily know the possible states of nature: the mutual goal for the two players is to make the communication succeed, but they are not always aware that they actually coordinate each other, or even that they are involved in a communication game.

The analysis of choices of strategies by the players gives rise to the following

Question 1.1. *What game-theoretical equilibrium is most likely to arise in this repeated game?*

This issue can either be analyzed from a static perspective with classic equilibrium analysis, see Trapa and Nowak (2000), Huttegger (2007) and Pawlowitsch (2008), or in a dynamic perspective, through individual learning models or evolutionary strategies, see for instance Huttegger and Zollman (2010). The latter perspective investigates the evolutionary pathway out of equilibria:

Question 1.2. *Does the communication system asymptotically reach a stable equilibrium state? If so, what are the good candidates, and how does the communication system reach them?*

However, modeling the game through individual learning process is really an issue involving lots of factors, for instance the level of rationality of players. In particular, we are interested in the following question which is first raised by Skyrms,

Question 1.3. *What is the simplest mechanism to ensure the emergence of a signaling system in this repeated game?*

1.4 Reinforcement learning

In this paper we adopt a dynamic perspective, based on the following individual reinforcement learning model.

(A3) Roth-Erev reinforcement learning rule (or Herrnstein's matching law): *the probability of choosing an action is proportional to its accumulated rewards.*

Assumption **(A3)** actually decide the updating rule for distributions of strategies for players. The corresponding behavior is analyzed by Argiento et al. (2009) in the 2-state, 2-signal, 2-act case, who show that an optimal signaling system emerges eventually, in the sense of a one-to-one correspondence between states and signals. We study here the case of M_1 states, M_2 signals and M_1 acts for general M_1, $M_2 \in \mathbb{N}$.

Note that Roth-Erev reinforcement rule is one of many possible strategies of the players, who can have various levels of rationality, each of them leading to a different learning process; for instance the myopic and best response models, see Fudenberg and Levine (1998). Let us briefly motivate the reinforcement condition, corresponding to a low level of rationality. It is natural to believe that individuals with high rationality, devoting themselves to the task of establishing a common language, would rapidly succeed. It is interesting to study whether on the contrary, under the only assumption that these individuals have good memory of their own past experience and aspire to a better payoff, a signaling system would also emerge, and how optimal the limiting system is. This pertains either to individuals with lower cognitive ability, or who do not devote themselves totally to the task of learning the game or take optimal decisions.

1.5   The model

1.5.1   Assumptions

Let $G = \Big((\Omega, \mathcal{F}, \mathbb{P}), \mathcal{S}_1, \mathcal{S}_2, \mathcal{A}, (S_n, Y_n, Z_n, u_n^1, u_n^2, \rho_n^1, \rho_n^2)_{n \in \mathbb{N}} \Big)$ be a signaling game as defined in Section 1.2. Apart from assumptions **(A1)**-**(A3)**, we make further more assumptions for our model. One is about the *priori* distributions and the other is a

more detailed version of assumption **(A2)**.

(A4) States of nature are equiprobable. In other words, for each $n \in \mathbb{N}$, S_n is an independent uniformly distributed random variable. Furthermore, Y_0 and Z_0 are independent uniformly distributed random variable.

(A5) Apart from assumption **(A2)** on payoffs, we here assume that payoffs are constants only dependent on types of states and acts. More precisely, $u_n^1(S_n, Y_n, Z_n) = u_n^2(S_n, Y_n, Z_n) = a_{Z_n,Y_n} \, \mathbb{I}_{\{S_n = Z_n\}}$, where $a_{i,j}$ is a positive constant, for $i \in \mathcal{S}_1$, $j \in \mathcal{S}_2$.

1.5.2 The model

Under assumptions $\mathbf{(A1)} - \mathbf{(A5)}$, we now present the model of the signaling game we are going to study in this paper.

Suppose there are M_1 states, M_2 signals and M_1 acts. Let $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$ and, for all $d \in \mathbb{N}$, let $\mathcal{S}^d := \mathcal{S} \times \ldots \times \mathcal{S}$.

Let $\mathcal{S}_{\text{pair}} := \{(i,j) \, : \, i \in \mathcal{S}_1, \, j \in \mathcal{S}_2\}$ be the set of *strategy pairs*. Note that a strategy pair (i,j) carries different meanings for the Sender and for the Receiver. For the Sender, it means choosing signal j when he observes state i, while for the Receiver it means choosing act i when he receives signal j. This strategy pair (i,j) accumulates the same payoffs for both the Sender and the Receiver, since the corresponding rewards are always received at the same time: let $V(n, i, j)$ denote this accumulated payoff at time n. For each $n \in \mathbb{N}$, let

$$V_n := (V(n, i, j))_{i \in \mathcal{S}_1, \, j \in \mathcal{S}_2}$$

8

be the *payoff vector* at time n.

Let us describe the random process $(V_n)_{n \in \mathbb{N}}$, arising from our model for the game and its strategies in Sections 1.1–1.4:

1. **Initial setting.** For any $(i,j) \in \mathcal{S}_{\text{pair}}$, we assume that $V(0,i,j) > 0$ is fixed.

2. **Reinforcement learning.** At each time step, Sender observes a certain state i from set of states $\mathcal{S}_1$; we assume here that all states arise with equal probability $1/M_1$. Then Sender randomly chooses a signal, his probability of drawing j being

$$\frac{V(n,i,j)}{\sum_{l \in \mathcal{S}_2} V(n,i,l)}.$$

    Receiver observes the signal he receives (let us call it j) and then randomly chooses an act k with probability

$$\frac{V(n,k,j)}{\sum_{l \in \mathcal{S}_1} V(n,l,j)}.$$

3. **Updating rule.** Both Sender and Receiver receive payoffs when the act chosen by Receiver matches the state observed by Sender. For any $i \in \mathcal{S}_1$, $j \in \mathcal{S}_2$,

$$V(n+1,i,j) := \begin{cases} V(n,i,j) + a_{i,j} & \text{if Sender observes state } i \text{ and chooses} \\ & \qquad \text{signal } j, \text{ and Receiver chooses act } i; \\ \\ V(n,i,j) & \qquad \text{if else.} \end{cases}$$

We suppose in this paper that $a_{i,j} := 1$ for all $i \in \mathcal{S}_1$, $j \in \mathcal{S}_2$. We also assume for simplicity that $V(0,i,j) = 1$, for all $i \in \mathcal{S}_1$, $j \in \mathcal{S}_2$, but the proofs carry on to general initial conditions.

1.5.3 Symmetrization and key processes

Let us symmetrize the notation, which will simplify some proofs (in particular that of Proposition 4.1): for all i $j \in \mathcal{S}$, let

$$
V(n,i,j) := \begin{cases}
0 & i,j \in \mathcal{S}_1 \text{ or } i,j \in \mathcal{S}_2 \\[2ex]
V(n,i,j) & i \in \mathcal{S}_1,\ j \in \mathcal{S}_2 \\[2ex]
V(n,j,i) & j \in \mathcal{S}_1,\ i \in \mathcal{S}_2
\end{cases}
$$

Now, for all $n \in \mathbb{N}$ and $i \in \mathcal{S}$ state or signal, let

$$
T_n^i := \sum_{j \in \mathcal{S}} V(n,i,j)
$$

be its number of successes up to time n.

For all $n \in \mathbb{N}$, let

$$
T_n := \frac{1}{2} \sum_{i \in \mathcal{S}} T_n^i = \sum_{i \in \mathcal{S}_1} T_n^i = \sum_{i \in \mathcal{S}_2} T_n^i.
$$

Then $T_n - T_0$ is the total number of successes of the communication system up to time n.

Let

$$
x_{ij}^n := V(n,i,j)/T_n \quad i,j \in \mathcal{S},\ n \in \mathbb{N};
$$
$$
x_i^n := T_n^i/T_n, \quad i \in \mathcal{S},\ n \in \mathbb{N}.
$$

For all $n \in \mathbb{N}$, let

$$x_n := \left(x_{ij}^n \right)_{i,j \in \mathcal{S}}, \, n \in \mathbb{N}$$

be the *occupation measure* at time n, which takes values in the interior of the simplex

$$\Delta := \left\{ (x_{ij})_{i,j \in \mathcal{S}} : x_{ij} \geqslant 0, \sum_{i \in \mathcal{S}_1, \, j \in \mathcal{S}_2} x_{ij} = 1, x_{ji} = x_{ij} \text{ for all } i, j \in \mathcal{S} \right\}.$$

Let us define

$$\partial \Delta := \left\{ x \in \Delta : \exists \, i \in \mathcal{S}, \, s.t. \, x_i = 0 \right\},$$

which we call *boundary* throughout the paper, although it is not the topological boundary of Δ. One of the technical difficulties in this model is the understanding of the behavior of $(x_n)_{n \in \mathbb{N}}$ near the boundary, as we shall explain later.

Given $x \in \Delta \setminus \partial \Delta$, $i, j \in \mathcal{S}$, let

$$y_{ij} := \frac{x_{ij}}{x_i x_j}$$

be the *efficiency* of the strategy pair (i, j) and let

$$N_i(x) := \sum_{k \in \mathcal{S}} \frac{x_{ik}}{x_i} \cdot y_{ik}$$

be the *efficiency $N_i(x)$ of i.* We will justify this notation in Section 4.1.

Note that processes $(x_n)_{n \in \mathbb{N}}$ and $(T_n)_{n \in \mathbb{N}}$ contain all the important information of the communication system throughout the game. Therefore, to study how our model evolves, we only need to focus on the evolutions of $(x_n)_{n \in \mathbb{N}}$ and $(T_n)_{n \in \mathbb{N}}$.

1.6 Urn setting: Another way to interpret the model

Note that the reinforcement learning and updating rules 2 and 3 in Section 1.5 can be interpreted in an urn setting. Assume Sender has M_1 urns indexed by states, each of them having M_2 colours of balls, one per signal. Similarly assume that Receiver has M_2 urns indexed by signals, each of them having M_1 colours of balls, one per act (or state, since the two sets coincide).

The model corresponds to the following: Sender picks a ball at random in the urn indexed by the state he observes, and sends the signal given by its colour. Then receiver picks a ball at random in the urn indexed by this signal, and chooses the act given by its colour. Both Sender and Receiver put back the balls they picked and, if the act matches the state, add one more ball of the same colour.

2   Main results

Given $x \in \Delta$ distribution of strategy pairs of the communication system, we introduce in Section 2.1 the bipartite graph of state/signal connections associated to x, and its communication potential or efficiency, which measures the corresponding expected payoff up to a multiplicative constant. We present in Section 2.2 the main results of the paper.

2.1   Bipartite graph and communication potential

Definition 2.1. *Given $x \in \Delta$, let $\mathcal{G}_x$ be the weighted bipartite graph with vertices $\mathcal{S} := \mathcal{S}_1 \cup \mathcal{S}_2$, adjacency $\sim$ and weights as follows*

(1) $\forall\, i \in \mathcal{S}_1, j \in \mathcal{S}_2, i \sim j$ *if and only if* $x_{ij} > 0$.

(2) *The weight of edge* $\{i, j\}$ *is its efficiency* $y_{ij} = x_{ij}/(x_i x_j)$.

Note that the new adjacency relation is only justified when x is in the topological boundary of Δ; otherwise, $\mathcal{G}_x$ is then the complete 2-partite graph with partitions $\mathcal{S}_1$ and $\mathcal{S}_2$.

Definition 2.2. *Let* $H : \Delta \longrightarrow \mathbb{R}_+$ *be the function defined by, for all* $x \in \Delta$,

$$H(x) := \sum_{i \in \mathcal{S}_1, j \in \mathcal{S}_2 : x_{ij} > 0} \frac{x_{ij}^2}{x_i x_j} = \frac{1}{2} \sum_{i, j \in \mathcal{S} : x_{ij} > 0} \frac{x_{ij}^2}{x_i x_j}.$$

We call $H(x)$ *communication potential or efficiency of* x.

Note the *communication potential* of x_n at time n can be interpreted - up to a multiplicative constant- as the expected payoff at that time step:

$$\mathbb{P}(T_{n+1} - T_n = 1 \,|\, \mathcal{F}_n) = \frac{1}{M_1} H(x_n), \tag{3}$$

where $\mathbb{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ is the filtration of the past, i.e. $\mathcal{F}_n := \sigma(x_1, \ldots, x_n)$.

Lemma 2.1. *H has minimum 1 and maximum* $\min(M_1, M_2)$ *on* Δ.

Proof. Using Cauchy-Schwartz inequality,

$$H(x) = \sum_{i \in \mathcal{S}_1, j \in \mathcal{S}_2 : x_{ij} > 0} \frac{x_{ij}^2}{x_i x_j} \sum_{i \in \mathcal{S}_1, j \in \mathcal{S}_2} x_i x_j \geqslant \left(\sum_{i \in \mathcal{S}_1, j \in \mathcal{S}_2 : x_{ij} > 0} \frac{x_{ij}}{\sqrt{x_i x_j}} \sqrt{x_i x_j} \right)^2$$

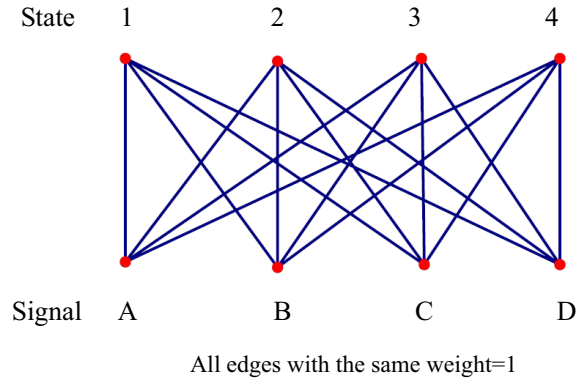$$= \left(\sum_{i \in \mathcal{S}_1, j \in \mathcal{S}_2} x_{ij} \right)^2 = 1$$

provides the first inequality, whereas the second one comes from

$$H(x) = \sum_{i\in\mathcal{S}_1, j\in\mathcal{S}_2 : x_{ij} > 0} \frac{x_{ij}}{x_i}\frac{x_{ij}}{x_j} \leqslant \sup_{i\in\mathcal{S}_1, j\in\mathcal{S}_2 : x_{ij} > 0} \frac{x_{ij}}{x_i} \sum_{i\in\mathcal{S}_1, j\in\mathcal{S}_2 : x_{ij} > 0} \frac{x_{ij}}{x_j}$$

$$\leqslant \sum_{i\in\mathcal{S}_1, j\in\mathcal{S}_2 : x_{ij} > 0} \frac{x_{ij}}{x_j} = M_2;$$

similarly $H(x) \leqslant M_1$.

Now

(a) $H(x)$ reaches the minimum 1 if and only if $\mathcal{G}_x$ is a complete graph on which every edge shares the same weight 1, as displayed on the figure below.

All edges with the same weight=1

(b) If $M_1 \geqslant M_2$ (resp. $M_1 \leqslant M_2$), then $H(x)$ reaches the maximum if and only if every vertex $i \in \mathcal{S}_1$ (resp. $\mathcal{S}_2$) only has one adjacent edge in $\mathcal{G}_x$. In the case $M_1 \leqslant M_2$, this corresponds to a unique meaning for every signal, i.e. perfect efficiency, as displayed on the figure below.
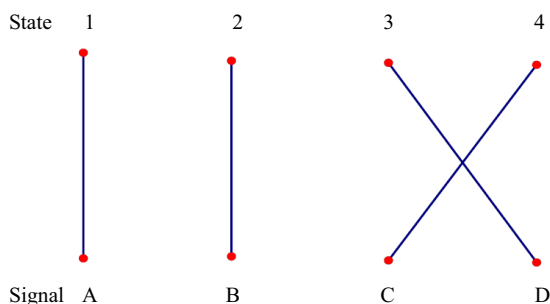
2.2   Main results

Definition 2.3. *Given a graph $\mathcal{G}$ on $\mathcal{S}_1 \cup \mathcal{S}_2$, let $(P)_{\mathcal{G}}$ be the following property:*

- *if we let $\mathcal{C}_1, \ldots \mathcal{C}_d$ be its connected components then, for every $i \in \{1, \ldots, d\}$, $\mathcal{C}_i \cap \mathcal{S}_1$ or $\mathcal{C}_i \cap \mathcal{S}_2$ is a singleton.*

- *each vertex has a corresponding edge.*

We call *synonym* (resp. *informational bottleneck* or *polysemy*) a state (resp. signal) associated to several signals (resp. states or acts), or the corresponding set of adjacent signals (resp. states). Obviously $M_1 \neq M_2$ ensures the existence of at least one synonym or polysemy.

Note that, given $x \in \Delta$, and even if $M_1 = M_2$, property $(P)_{\mathcal{G}_x}$ allows for synonyms or informational bottlenecks, and does not ensure that the system is optimal as a communication system, i.e. that $H(x)$ reaches the maximum of H. Most common languages have such flaws. We show, on the figure below, a graph $\mathcal{G}$ corresponding to

a sub-optimal communication system: it is easy to check that its normalized efficiency is 80%, i.e. that any $x \in \Delta$ such that $\mathcal{G}_x = \mathcal{G}$ is such that $H(x)/\max H = 0.8$.
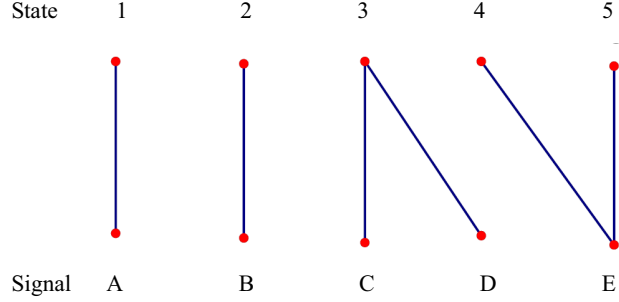
Theorem 2.1. *The communication potential process $(H(x_n))_{n \in \mathbb{N}}$ is a bounded submartingale, and hence converges a.s.*

Theorem 2.2. *$(x_n)_{n \in \mathbb{N}}$ converges to the set of equilibria of mean-field ODE almost surely.*

Remark. We will define the mean-field ODE in Section 3.1. Roughly speaking, the mean-field ODE is derived from the dynamics of expected movement of $(x_n)_{n \in \mathbb{N}}$.

Theorem 2.3. *For all $\mathcal{G}$ on $\mathcal{S}_1 \cup \mathcal{S}_2$ s.t. $(P)_{\mathcal{G}}$ holds, with positive probability*

(a) *$x_n \to x$ s.t. $\mathcal{G}_x = \mathcal{G}$.*

(b) *$\forall i, j \in \mathcal{S}, V(\infty, i, j) = \infty \iff \{i, j\}$ is an edge of $\mathcal{G}$.*

2.3 Contents

The rest of this chapter is devoted to the proof of the main results, as follows. In Section 3 we discuss the stochastic approximation of $(x_n)_{n \in \mathbb{N}}$ by an ordinary differential

equation. In Section 4 we justify some notation from Section 1.5, and show Theorem 2.1 and its deterministic counterpart that H is a Lyapounov function of the associated ODE. In Section 5 we deduce that $(x_n)_{n\in\mathbb{N}}$ almost surely converges to the set of equilibria of this differential equation, and describe the stable equilibria x in terms of graph structure of $\mathcal{G}_x$. In Section 6 we connect our analysis of stability of the reinforcement learning model with the one from the static equilibrium analysis literature. Finally in Section , we show Theorem 2.3 about convergence with positive probability towards subgraphs $\mathcal{G}$ satisfying $(P)_{\mathcal{G}}$.

2.4 Notation

For all u, $v \in \mathbb{R}$, we write $u = \square(v)$ if $|u| \leqslant v$; we let $u \wedge v = \min(u,v)$ (resp. $u \vee v = \max(u,v)$) be the minimum (resp. maximum) of u and v.

We let $\mathsf{Cst}(a_1, a_2, \ldots, a_p)$ denote a positive constant depending only on a_1, a_2, $\ldots$ a_p, and let Cst denote a universal positive constant.

3 Stochastic approximation

3.1 Mean-field ODE

Let $|\cdot|$ be the Euclidean norm on $\mathbb{R}^{M_1 \times M_2}$. Let us calculate the increment of $(x_n)_{n\in\mathbb{N}}$ at time n:

$$x_{n+1} - x_n = \left(\frac{V_{n+1}}{1+T_n} - \frac{V_n}{1+T_n} + \frac{V_n}{1+T_n} - \frac{V_n}{T_n} \right) \ \mathbb{1}_{\{V_{n+1}-V_n \neq 0\}}$$

$$= \frac{\mathbb{1}_{\{V_{n+1}-V_n\neq0\}}}{1+T_n}(V_{n+1}-V_n-x_n). \tag{4}$$

In expectation,

$$\mathbb{E}[\,x_{n+1}-x_n\,|\,\mathcal{F}_n\,] = \frac{1}{(1+T_n)M_1}F(x_n), \tag{5}$$

where F, defined from Δ to $T\Delta$ tangent space of Δ, maps x to

$$F(x) := \left(x_{ij}\left(\frac{x_{ij}}{x_i x_j} - H(x)\right)\right)_{i,j\in\mathcal{S}},$$

with the convention that $F(x)_{ij} = 0$ if $x_{ij} = 0$.

Let us consider the following ordinary differential equation, defined on $\Delta \setminus \partial\Delta$:

$$\frac{d\,x}{dt} = F(x). \tag{6}$$

3.2 Results

Lemma 3.1. *There exists an adapted martingale increment process $(\eta_n)_{n\in\mathbb{N}}$ such that, for all $n \in \mathbb{N}$,*

$$x_{n+1}-x_n = \frac{1}{(1+T_n)M_1}F(x_n) + \eta_{n+1}, \tag{7}$$

and $|\eta_{n+1}| \leqslant 2/(1+T_n)$.

Proof. The inequality $|\eta_{n+1}| \leqslant 2/(1+T_n)$ comes from

$$|x_{n+1} - x_n| = \left| \frac{\mathbb{1}_{\{V_{n+1} - V_n \neq 0\}}}{1 + T_n} (V_{n+1} - V_n - x_n) \right| \leqslant \frac{1}{1 + T_n}. \tag{8}$$

$\square$

The following Lemma 3.2, which proves asymptotic linear growth of T_n, will in particular imply that the martingale $(\sum_{k=1}^n \eta_k)_{n \in \mathbb{N}}$ converges a.s., by Doob convergence theorem.

Lemma 3.2. *With Probability 1,*

$$\frac{1}{M_1} \leqslant \liminf_{n \to \infty} \frac{T_n}{n} \leqslant \limsup_{n \to \infty} \frac{T_n}{n} \leqslant \frac{\min(M_1, M_2)}{M_1}.$$

Proof. The result is a direct consequence of (3) and Conditional Borel-Cantelli Lemma, see [4, Theorem I.6] $\square$

Remark. We show later that, as n goes to infinity, $H(x_n)$ converges (Theorem 2.1), and the proof above implies that its limit is also the one of T_n/n.

Formula (7) can be interpreted as a stochastically perturbed Cauchy-Euler approximation scheme for the ODE (6), with step size $1/M_1(1+T_n)$. The step size being $O(1/n)$, $(x_n)_{n \in \mathbb{N}}$ asymptotically shadows solutions of the asymptotic ODE, so that its limit set belongs to a class of possible limit sets of pseudotrajectories of the ODE (see for instance [3]).

Let Γ be the *set of equilibria* of the ODE, i.e.

$$\Gamma := \left\{ x \in \Delta \ : \ F(x) = 0 \right\}.$$

19

4 Lyapunov function

4.1 Deterministic case: Mean-field ODE

Let us start with a heuristical justification of the fact that H is a Lyapounov function, i.e. that it increases along the trajectories of the ODE.

The differential equation (6) can be understood in the language of (non-linear) replicator dynamics, with the following biological perspective. Suppose a population consists of species $(i, j) \in \mathcal{S}_1 \times \mathcal{S}_2$, each corresponding to a strategy pair. Let the *fitness* of (i, j) be its efficiency $y_{ij} = x_{ij}/(x_i x_j)$. The wording is justified by the following interpretation: the probability that population (i, j) increases is

$$\frac{1}{M_1} \cdot \frac{x_{ij}}{x_i} \cdot \frac{x_{ij}}{x_j} = \frac{1}{M_1} \times \text{ proportion of } (i, j) \times \text{ its efficiency.}$$

Then the average fitness of the whole population is the communication potential $H(x)$.

Therefore the mean-field ODE can be understood as follows,

> *Growth rate of x_{ij}*
>
> $= x_{ij} \times ($ *fitness of $(i, j) -$ average fitness of whole population*$)$.

In particular, those species (i.e. strategy pairs) whose fitness (i.e. efficiency) are above the average increase their proportions in the population (i.e. distribution of strategy pairs). Note that the fitness of our species changes over time.

This interpretation makes it reasonable to conjecture that the average fitness of the whole population indeed increases along solutions of ODE (6), as we show in Proposition

4.1 and, in the (discrete) stochastic case, in Theorem 2.1, proved in Section 4.2. The proof of the latter is technically quite long, owing to the non-continuity of H on the boundary $\partial\Delta$, which prevents us from converting the deterministic Proposition 4.1 into a stochastic one via a simple Taylor formula.

Let, for all $x \in \Delta$,

$$p(x) := \frac{1}{2} \sum_{i,j,k\in\mathcal{S}:x_{ij},x_{ik}\neq 0} \frac{x_{ij}x_{ik}}{x_i} \left(y_{ij} - y_{ik}\right)^2.$$

Proposition 4.1. *H is a Lyapunov function on $\Delta \setminus \partial\Delta$ for the mean-field ODE (6); more precisely,*

$$\nabla H \cdot F\left(x\right) \;=\; p(x) \geqslant 0. \tag{9}$$

Remark. We will see later that H is not a strict Lyapunov function; in other words, $\nabla H \cdot F$ does not only vanish on the set of rest points of F.

Proof. We take advantage of the symmetrical notation introduced at the end of Section 1.5: in a mathematical perspective, there is no difference between state and signal in the mean-field ODE. Let us now differentiate H along a path of the ODE: note that, when differentiating with respect to space variables, we are obviously not restricted to Δ, so that x_{ij} and x_{ji} are considered as independent variables (in the calculation below, we use the convention that $x_i = \sum_{j\in\mathcal{S}} x_{ij}$, but any other convention would lead to the same result):

$$\nabla H \cdot F\left(x\right) = \sum_{i,j\in\mathcal{S}} \frac{x_{ij}}{x_i x_j} \left(\frac{(x_{ij})^2}{x_i x_j} - x_{ij} H(x) \right)$$
$$- \frac{(x_{ij})^2}{2(x_i)^2 x_j} \left(\sum_{k\in\mathcal{S}} \frac{(x_{ik})^2}{x_i x_k} - x_{ik} H(x) \right)$$

21

$$-\frac{(x_{ij})^2}{2x_i(x_j)^2}\left(\sum_{k\in\mathcal{S}}\frac{(x_{jk})^2}{x_jx_k} - x_{jk}H(x)\right)$$

$$= \sum_{i,j\in\mathcal{S}}\frac{(x_{ij})^3}{(x_i)^2(x_j)^2} - \sum_{i,j,k\in\mathcal{S}}\frac{(x_{ij})^2(x_{ik})^2}{(x_i)^3x_jx_k} \tag{10}$$

$$-\left(\sum_{i,j\in\mathcal{S}}\frac{(x_{ij})^2}{x_ix_j} - \sum_{i,j,k\in\mathcal{S}}\frac{(x_{ij})^2x_{ik}}{(x_i)^2x_j}\right)H(x) \tag{11}$$

Using that $\sum_{k\in\mathcal{S}}x_{ik}/x_i = 1$,

$$(11) = \sum_{i,j\in\mathcal{S}}\frac{(x_{ij})^2}{x_ix_j}\left(1 - \sum_{k\in\mathcal{S}}\frac{x_{ik}}{x_i}\right)H(x) = 0.$$

Using the symmetry between j and k, we obtain

$$(10) = \sum_{i,j,k\in\mathcal{S}}\frac{(x_{ij})^3x_{ik}}{(x_i)^3(x_j)^2} - \sum_{i,j,k\in\mathcal{S}}\frac{(x_{ij})^2(x_{ik})^2}{(x_i)^3x_jx_k}$$

$$= \frac{1}{2}\sum_{i,j,k\in\mathcal{S}}\frac{x_{ij}x_{ik}}{x_i}\left(\frac{x_{ij}}{x_ix_j} - \frac{x_{ik}}{x_ix_k}\right)^2. \tag{12}$$

$\square$

Lemma 4.1. *For any $x \in \Delta \setminus \partial\Delta$,*

$$\nabla H \cdot F(x) = \sum_{i,j\in\mathcal{S}}x_{ij}(y_{ij} - N_i(x))^2, \tag{13}$$

which can also be written as

$$\nabla H \cdot F(x) = \sum_{i,j\in\mathcal{S}}x_{ij}(y_{ij} - H(x))^2 - \sum_{i\in\mathcal{S}}x_i(N_i(x) - H(x))^2. \tag{14}$$

Remark. In the context of communication systems, the above three formulas (9), (13)–(14) mean that the growth rate of the communication potential is a function

depending on the difference between efficiencies of different strategy pairs.

Proof. Fix i, $j \in \mathcal{S}$, and define $Y : \mathcal{S} \to \mathbb{R}$ by $Y(k) := y_{ik}$, seen as a random variable on $(\mathcal{S}, \mathbb{P}_{z,i})$ (with expectation $\mathbb{E}_{z,i}(.)$), where $\mathbb{P}_{z,i}(\{k\}) := x_{ik}/x_i$. Then

$$\mathbb{E}_{z,i}[(y_{ij} - Y)^2] = \sum_{k \in \mathcal{S}} \frac{x_{ik}}{x_i} (y_{ij} - y_{ik})^2 = (\mathbb{E}_{z,i}(y_{ij} - Y))^2 + \mathbb{E}_{z,i}[(Y - E_{z,i}(Y))^2]$$

$$= (y_{ij} - N_i(x))^2 + \sum_{k \in \mathcal{S}} \frac{x_{ik}}{x_i}(y_{ik} - N_i(x))^2,$$

using that

$$\mathbb{E}_{z,i}(Y) = \sum_{k \in \mathcal{S}} \frac{x_{ik}}{x_i} y_{ik} = N_i(x).$$

Therefore

$$\nabla H \cdot F(x) = \frac{1}{2} \left(\sum_{i,j \in \mathcal{S}} x_{ij}(y_{ij} - N_i(x))^2 + \sum_{i,j,k \in \mathcal{S}} \frac{x_{ij}x_{ik}}{x_i}(y_{ik} - N_i(x))^2 \right)$$

$$= \sum_{i,j \in \mathcal{S}} x_{ij}(y_{ij} - N_i(x))^2,$$

which implies that

$$\nabla H \cdot F(x) := \sum_{i,j \in \mathcal{S}} x_{ij}(y_{ij} - H(x))^2 - \sum_{i \in \mathcal{S}} x_i(N_i(x) - H(x))^2,$$

and completes the proof. $\qquad\square$

Lemma 4.2. *For all $x \in \Delta \setminus \partial\Delta$ and $i \in \mathcal{S}$, $N_i(x) \geqslant 1$.*

Proof. Indeed,

$$\sum_{j \in \mathcal{S}} y_{ij}x_j = \sum_{j \in \mathcal{S}} \frac{x_{ij}}{x_i} = 1$$

23

which subsequently implies, by Cauchy-Schwartz inequality, that

$$1 = \left(\sum_{j \in \mathcal{S}} y_{ij} x_j \right)^2 \leqslant \left(\sum_{j \in \mathcal{S}} y_{ij}^2 x_j \right) \left(\sum_{j \in \mathcal{S}} x_j \right) = N_i(x). \tag{15}$$

$\square$

Let us define

$$\Delta_\epsilon := \{ x \in \Delta \setminus \partial\Delta : p(x) > \epsilon \},$$

and let

$$\Lambda := \{ x \in \Delta : p(x) = 0 \},$$

where p is defined in the statement of Proposition 4.1. The following Lemma 4.3 is straightforward.

Lemma 4.3. $x \in \Lambda$ *if and only if*

$$\frac{x_{ij}}{x_j} = \frac{x_{ik}}{x_k}, \ \text{for all } i, j, k \ \text{s.t. } x_{ij} \neq 0, x_{ik} \neq 0$$

or, equivalently,

$$y_{ij} = y_{ik}, \ \text{for all } i, j, k \ \text{s.t. } x_{ij} \neq 0, x_{ik} \neq 0.$$

Remark. Lemma 4.3 can be phrased as follows: if $x \in \Lambda$ then, in the graph $\mathcal{G}_x$, edges within the same connected component have the same weight. Note that $x \in \Gamma$ is equivalent to all edges of $\mathcal{G}_x$ having the same weight $H(x)$. So the two sets Γ and Λ are different, i.e. H is not a strict Lyapounov function, which justifies the need to prove separately the convergence to the set of equilibria in Section 5.1.

24

4.2 Proof of Theorem 2.1 and convergence to Λ

4.2.1 Proof of Theorem 2.1

For simplicity, we let $V := V_n$ in the following calculation. Let us compute the expected increment of $(H(x_n))_{n \in \mathbb{N}}$.

$$\mathbb{E}[\, H(x_{n+1}) - H(x_n) \,|\, \mathcal{F}_n \,] \tag{16}$$

$$= \sum_{i \in \mathcal{S}_1, j \in \mathcal{S}_2} \frac{V_{ij}^2}{V_i V_j} \left(\frac{(V_{ij}+1)^2}{(V_i+1)(V_j+1)} - \frac{V_{ij}^2}{V_i V_j} \right.$$

$$+ \sum_{k \in \mathcal{S}, k \neq j} \left(\frac{V_{ik}^2}{(V_i+1)V_k} - \frac{V_{ik}^2}{V_i V_k} \right) + \sum_{k \in \mathcal{S}, k \neq i} \left. \left(\frac{V_{kj}^2}{(V_j+1)V_k} - \frac{V_{kj}^2}{V_j V_k} \right) \right)$$

$$= \frac{1}{2} \sum_{(i,j) \in \mathcal{S}^2} \frac{V_{ij}^2}{V_i V_j} \left(\frac{(V_{ij}+1)^2}{(V_i+1)(V_j+1)} - \frac{V_{ij}^2}{V_i V_j} \right.$$

$$+ \sum_{k \in \mathcal{S}, k \neq j} \left(\frac{V_{ik}^2}{(V_i+1)V_k} - \frac{V_{ik}^2}{V_i V_k} \right) + \sum_{k \in \mathcal{S}, k \neq i} \left. \left(\frac{V_{jk}^2}{(V_j+1)V_k} - \frac{V_{jk}^2}{V_j V_k} \right) \right)$$

$$= \frac{1}{2} \sum_{(i,j) \in \mathcal{S}^2} \frac{V_{ij}^2}{V_i V_j} \left(-\sum_{k \in \mathcal{S}} \frac{V_{ik}^2}{V_i(V_i+1)V_k} - \sum_{k \in \mathcal{S}} \frac{V_{jk}^2}{V_j(V_j+1)V_k} \right.$$

$$\left. + \frac{(V_{ij}+1)^2}{(V_i+1)(V_j+1)} + \frac{V_{ij}^2}{V_i V_j} - \frac{V_{ij}^2}{(V_i+1)V_j} - \frac{V_{ij}^2}{(V_j+1)V_i} \right)$$

$$= - \sum_{(i,j,k) \in \mathcal{S}^3} \frac{V_{ij}^2 V_{ik}^2}{V_i^2(V_i+1)V_j V_k} + \frac{1}{2} \sum_{(i,j) \in \mathcal{S}^2} \frac{V_{ij}^2}{V_i V_j} \frac{2V_i V_j V_{ij} + V_i V_j + V_{ij}^2}{V_i V_j (V_i+1)(V_j+1)}$$

$$= \sum_{(i,j) \in \mathcal{S}^2} \frac{V_{ij}^3}{V_i(V_i+1)V_j^2} - \sum_{(i,j,k) \in \mathcal{S}^3} \frac{V_{ij}^2 V_{ik}^2}{V_i^2(V_i+1)V_j V_k} \tag{17}$$

$$+ \frac{1}{2} \sum_{(i,j) \in \mathcal{S}^2} \frac{V_{ij}^2}{V_i V_j} \frac{2V_i V_j V_{ij} + V_i V_j + V_{ij}^2}{V_i V_j (V_i+1)(V_j+1)} - \sum_{(i,j) \in \mathcal{S}^2} \frac{V_{ij}^3}{V_i(V_i+1)V_j^2} \tag{18}$$

Now let us prove (17) is nonnegative. Indeed,

$$
(17) = \sum_{(i,j,k)\in\mathcal{S}^3} \frac{V_{ij}V_{ik}}{V_i+1} \left(\frac{V_{ij}^2}{V_i^2 V_j^2} - \frac{V_{ij}V_{ik}}{V_i^2 V_j V_k} \right)
$$

$$
= \frac{1}{2} \sum_{(i,j,k)\in\mathcal{S}^3} \frac{V_{ij}V_{ik}}{V_i+1} \left(\frac{V_{ij}}{V_i V_j} - \frac{V_{ik}}{V_i V_k} \right)^2 \geqslant 0.
$$

Next we show that (18) is nonnegative as well:

$$
(18) = -2 \sum_{(i,j)\in\mathcal{S}^2} \frac{V_{ij}^3}{V_i(V_i+1)V_j^2(V_j+1)} + \sum_{(i,j)\in\mathcal{S}^2} \frac{V_{ij}^2}{V_i V_j(V_i+1)(V_j+1)}
$$

$$
+ \sum_{(i,j)\in\mathcal{S}^2} \frac{V_{ij}^4}{V_i^2 V_j^2(V_i+1)(V_j+1)}
$$

$$
= \sum_{(i,j)\in\mathcal{S}^2} \frac{V_{ij}^2}{V_i V_j(V_i+1)(V_j+1)} \left(-\frac{2V_{ij}}{V_j} + 1 + \frac{V_{ij}^2}{V_i V_j} \right)
$$

$$
= \sum_{i\in\mathcal{S}_1, j\in\mathcal{S}_2} \frac{2V_{ij}^2}{V_i V_j(V_i+1)(V_j+1)} \left(1 - \frac{V_{ij}}{V_i} - \frac{V_{ij}}{V_j} + \frac{V_{ij}^2}{V_i V_j} \right)
$$

$$
= \sum_{i\in\mathcal{S}_1, j\in\mathcal{S}_2} \frac{2V_{ij}^2}{V_i V_j(V_i+1)(V_j+1)} \left(1 - \frac{V_{ij}}{V_i} \right) \left(1 - \frac{V_{ij}}{V_j} \right) \geqslant 0.
$$

4.2.2 Convergence to Λ

Let us now prove the following Proposition 4.2.

Proposition 4.2. $(x_n)_{n\in\mathbb{N}}$ *converges to Λ a.s. ; more precisely, $(p(x_n))_{n\in\mathbb{N}}$ converges to 0 a.s.*

Proof. Let us define process $Y_n := H(x_n)$, $n \in \mathbb{N}$. We decompose $(Y_n)_{n\in\mathbb{N}}$ into a martingale $(M_n)_{n\in\mathbb{N}}$ and a predictable process $(A_n)_{n\in\mathbb{N}}$ where $A_{n+1} - A_n = \mathbb{E}[Y_{n+1} - Y_n \,|\, \mathcal{F}_n]$. Since H is bounded, martingale $(M_n)_{n\in\mathbb{N}}$ is upper bounded and hence converges.

Let

$$P_n := \frac{1}{2} \sum_{(i,j,k) \in \mathcal{S}^3} \frac{V_{ij}^n V_{ik}^n}{V_i^n + 1} \left(\frac{V_{ij}^n}{V_i^n V_j^n} - \frac{V_{ik}^n}{V_i^n V_k^n} \right)^2 ;$$

$$Q_n := \sum_{i \in \mathcal{S}_1, j \in \mathcal{S}_2} \frac{(V_{ij}^n)^2}{V_i^n V_j^n (V_i^n + 1)(V_j^n + 1)} \left(1 - \frac{V_{ij}^n}{V_i^n}\right) \left(1 - \frac{V_{ij}^n}{V_j^n}\right).$$

Hence

$$P_n \geqslant \frac{1}{4} \sum_{(i,j,k) \in \mathcal{S}^3} \frac{V_{ij}^n V_{ik}^n}{V_i^n} \left(\frac{V_{ij}^n}{V_i^n V_j^n} - \frac{V_{ik}^n}{V_i^n V_k^n} \right)^2 = \frac{p(x_n)}{2 T_n}.$$

The rest of the argument is similar to the proof of convergence to the set of equilibria in [1]. If $(x_n)_{n \in \mathbb{N}}$ were infinitely often away from Λ, then the drift would cause $(H(x_n))_{n \in \mathbb{N}}$ to go to infinity, hence contradicting the boundedness of H. Indeed, let δ be the distance between Δ_ϵ and the complement of $\Delta_{\epsilon/2}$. Suppose $x_n \in \Delta_\epsilon$, $x_{n+1}, \ldots, x_{n+k-1} \in \Delta_{\frac{\epsilon}{2}} \setminus \Delta_\epsilon$ and $x_{n+k} \in \Delta_{\frac{\epsilon}{2}}^c$,

$$A_{n+k} - A_n \geqslant \sum_{r=n}^{n+k-1} (P_r + Q_r) \geqslant \sum_{r=n}^{n+k-1} P_r = \sum_{r=n}^{n+k-1} \frac{p(x_r)}{2 T_r} \geqslant \frac{\epsilon}{4} \sum_{r=n}^{n+k-1} \frac{1}{T_r + 1}.$$

Therefore,

$$\delta \leqslant \sum_{r=1}^{n+k-1} |x_{r+1} - x_r| \leqslant \sum_{r=1}^{n+k-1} \frac{1}{1 + T_r} \leqslant \frac{4}{\epsilon}(A_{n+k} - A_n).$$

Therefore $x_n \in \Delta_\epsilon$ infinitely often would cause A_n to increase to infinity. By contradiction, $(x_n)_{n \in \mathbb{N}}$ must converge to Λ. $\square$

Remark. The proof of convergence of $p(x_n)$ to 0 would also hold on (deterministic) solutions of the ODE.

Corollary 4.1. *Almost surely,*

$$\frac{T_n}{n} \to \lim_{n\to\infty} \frac{H(x_n)}{M_1} \in \left[\frac{1}{M_1}, \frac{\min(M_1, M_2)}{M_1}\right] \quad \text{as } n \to \infty.$$

Proof. Same as Lemma 3.2. □

5 Equilibria

5.1 Convergence to the set of equilibria : Proof of Theorem 2.2

We already know that the occupation measure $(x_n)_{n\in\mathbb{N}}$ a.s. converges to Λ; the goal of this section is to prove that, more precisely, $(x_n)_{n\in\mathbb{N}}$ a.s. converges to the set of equilibria Γ of the ODE.

Lemma 5.1. *Suppose ϵ small enough and $x \in \Delta_{\epsilon^4}^c$. Then for any $i \in \mathcal{S}_1$, $j \in \mathcal{S}_2$ s.t. $x_{ij} > \epsilon$,*

$$|y_{ij} - N_i(x)| < \frac{\epsilon}{3}, \quad \text{and} \quad |y_{ij} - N_j(x)| < \frac{\epsilon}{3}.$$

Proof. Follows directly from (13). □

Lemma 5.2.

$$y_{ij}^{n+1} - y_{ij}^n = \frac{1}{M_1 T_n} y_{ij}^n \left(y_{ij}^n - N_i(x_n) - N_j(x_n) + H(x_n)\right) + r_{n+1} + \zeta_{ij}^{n+1}, \tag{19}$$

where $(r_n)_{n\geqslant 1}$ *is predictable,* $\mathbb{E}\left[\zeta_{ij}^{n+1}\,|\,\mathcal{F}_n\right]=0$ *and*

$$|\zeta_{ij}^{n+1}| < \frac{12}{T_n x_i^n x_j^n}, \quad |r_{n+1}| < \frac{6(x_i^n + x_j^n)}{(T_n x_i^n x_j^n)^2} \leqslant \frac{12}{(T_n x_i^n x_j^n)^2}. \tag{20}$$

Proof.

$$
\begin{aligned}
y_{ij}^{n+1} - y_{ij}^n &= \frac{T_{n+1} V_{ij}^{n+1}}{V_i^{n+1} V_j^{n+1}} - \frac{T_n V_{ij}^n}{V_i^n V_j^n} \\
&= (T_{n+1} V_{ij}^{n+1} V_i^n V_j^n - T_n V_{ij}^n V_i^{n+1} V_j^{n+1}) \tag{21} \\
&\quad \times \frac{1}{(V_i^n V_j^n)^2}\left(1 + \left(\frac{V_i^n V_j^n}{V_i^{n+1} V_j^{n+1}} - 1\right)\right) \tag{22}
\end{aligned}
$$

and

$$
\begin{aligned}
(21) &= T_n V_i^n V_j^n\,\mathbb{1}_{\Delta V_{ij}^{n+1}>0} + V_{ij}^n V_i^n V_j^n\,\mathbb{1}_{\Delta T_{n+1}>0} + V_i^n V_j^n\,\mathbb{1}_{\Delta V_{ij}^{n+1}>0} \\
&\quad - T_n V_{ij}^n V_i^n \mathbf{1} y_{\Delta V_i^{n+1}>0} - T_n V_{ij}^n V_j^n \mathbf{1} y_{\Delta V_j^{n+1}>0} - T_n V_{ij}^n\,\mathbb{1}_{\Delta V_{ij}^{n+1}>0}.
\end{aligned}
$$

Hence

$$|(21)| \leqslant 6\,T_n V_i^n V_j^n,$$

and

$$\mathbb{E}\left[\frac{(21)}{(V_i^n V_j^n)^2}\,\middle|\,\mathcal{F}_n\right] = \frac{1}{M_1 T_n}\,y_{ij}^n\left(y_{ij}^n - N_i(x_n) - N_j(x_n) + H(x_n)\right).$$

By the following simple estimate

$$\frac{V_i^n V_j^n}{V_i^{n+1} V_j^{n+1}} - 1 \leqslant \frac{1}{V_i^{n+1}} + \frac{1}{V_j^{n+1}},$$

we deduce that

$$|r_{n+1}| = \left| \mathbb{E}\left[\frac{(21)}{(V_i^n V_j^n)^2} \left(\frac{V_i^n V_j^n}{V_i^{n+1} V_j^{n+1}} - 1 \right) \Big| \mathcal{F}_n \right] \right|$$
$$\leqslant \frac{6\, T_n(V_i^n + V_j^n)}{(V_i^n V_j^n)^2} \leqslant \frac{6(x_i^n + x_j^n)}{(T_n x_i^n x_j^n)^2}.$$

and

$$|y_{ij}^{n+1} - y_{ij}^n| = (21) \times (22) \leqslant \frac{6}{T_n x_i^n x_j^n}. \tag{23}$$

Therefore

$$|\zeta_{ij}^{n+1}| = 2\, |y_{ij}^{n+1} - y_{ij}^n - \mathbb{E}[y_{ij}^{n+1} - y_{ij}^n | \mathcal{F}_n]| \leqslant \frac{12}{T_n x_i^n x_j^n}.$$

$\square$

Let

$$U_{ij}(\epsilon) := \left\{ x \in \Delta \,:\, x_{ij} \leqslant \epsilon \text{ or } y_{ij} - H(X) \geqslant -\epsilon \right\}.$$

Lemma 5.3. *Assume $\epsilon > 0$ is small enough and $m_0 \in \mathbb{N}$ is large enough. Let*

$$R_n := \sum_{m=m_0}^{n} \left(y_{ij}^m - y_{ij}^{m-1} - \frac{\epsilon^2}{6m} \right) 1\!\!1_{x_{m-1} \notin U_{ij}(\epsilon) \cup \Delta_{\epsilon^4},\, T_{m-1} \geqslant \frac{m}{2M_1}}, \quad \forall\, n \in \mathbb{N};$$

$$S_n := \sum_{m=m_0}^{n} \left(x_{ij}^m - x_{ij}^{m-1} + \frac{\epsilon}{m} \right) 1\!\!1_{x_{m-1} \notin U_{ij}(\epsilon) \cup \Delta_{\epsilon^4},\, T_{m-1} \geqslant \frac{m}{2M_1}}, \quad \forall\, n \in \mathbb{N}.$$

Then

(a) *$(R_n)_{n \in \mathbb{N}}$ (resp. $(S_n)_{n \in \mathbb{N}}$) is a submartingale (resp. supermartingale).*

30

(b) $\limsup_{n \geqslant m, m \to \infty} (R_n - R_m)^- = \limsup_{n \geqslant m, m \to \infty} (S_n - S_m)^+ = 0.$

Proof. First we note that if $x_n \notin U_{ij}(\epsilon)$,

$$\mathbb{E}[\, x_{ij}^{n+1} - x_{ij}^n \,|\, \mathcal{F}_n \,] = \frac{x_{ij}^n}{1 + T_n}(y_{ij}^n - H(x_n)) \leqslant -\frac{\epsilon}{T_n}.$$

This implies that $(S_n)_{n \in \mathbb{N}}$ is a supermartingale. Now we prove that $(R_n)_{n \in \mathbb{N}}$ is a sub-martingale.

Assume $\epsilon > 0$ small enough and $x_n \notin U_{ij}(\epsilon) \cup \Delta_{\epsilon^4}$. Then Lemma 5.1 implies that

$$y_{ij}^n - N_i(x_n) - N_j(x_n) + H(x_n) \geqslant \frac{\epsilon}{3}.$$

Hence,

$$\begin{aligned}
\mathbb{E}[\, y_{ij}^{n+1} - y_{ij}^n \,|\, \mathcal{F}_n \,] &\geqslant \frac{1}{M_1 T_n x_i^n x_j^n}\left(\frac{x_{ij}^n \epsilon}{3} - \frac{12}{T_n x_i^n x_j^n}\right) \\
&\geqslant \frac{x_{ij}^n \epsilon}{6 M_1 T_n x_i^n x_j^n} \geqslant \frac{\epsilon^2}{6n},
\end{aligned}$$

if $n \geqslant 144 M_1/\epsilon^4$ (which implies $T_n \geqslant 72/\epsilon^4$ and therefore $T_n x_i^n x_j^n \geqslant 72/(x_{ij}^n \epsilon)$).

Let us now prove (b). Let

$$\Pi_n := R_n - \sum_{m=m_0}^{n} \mathbb{E}[\, R_m - R_{m-1} \,|\, \mathcal{F}_{m-1} \,],$$

$$\Xi_n := S_n - \sum_{m=m_0}^{n} \mathbb{E}[\, S_m - S_{m-1} \,|\, \mathcal{F}_{m-1} \,].$$

By (20), we note that for all $n \geqslant m_0$,

$$
\begin{aligned}
\mathbb{E}[\,(\Pi_{n+1} - \Pi_n)^2 \,|\, \mathcal{F}_n\,] &\leqslant \mathbb{E}[\,(y_{ij}^{n+1} - y_{ij}^n)^2 \,|\, \mathcal{F}_n\,] \\
&\leqslant \frac{12}{(T_n x_i^n x_j^n)^2}\; \mathbb{1}_{x_n \notin U_{ij}(\epsilon), T_n \geqslant \frac{n}{2M_1}} \;\leqslant\; \frac{48 M_1^2}{\epsilon^4 n^2}.
\end{aligned}
$$

Therefore $(\Pi_n)_{n \in \mathbb{N}}$ is bounded in $\mathbb{L}^2$ and hence converges. We can obtain similar bounds for $(\Xi_n)_{n \in \mathbb{N}}$, which converges as well. This completes the proof. $\qquad\square$

Lemma 5.4. *Let $\epsilon > 0$, and assume $n \in \mathbb{N}$ is sufficiently large (depending on ϵ). If $x_n \in U_{ij}(\epsilon)$, $|H(x_{n+1}) - H(x_n)| < \epsilon/2$ and $T_n \geqslant n/(2M_1)$, then $x_{n+1} \in U_{ij}(2\epsilon)$.*

Proof. Let $x_n \in U_{ij}(\epsilon)$. Then (8) implies

$$
|x_{ij}^{n+1} - x_{ij}^n| \leqslant \frac{2M_1}{n}.
$$

Assume that n is large enough. If $x_{ij}^n \leqslant \epsilon$, then $x_{ij}^{n+1} \leqslant 2\epsilon$. Otherwise $x_{ij}^n > \epsilon$ and $y_{ij}^n - H(x_n) \geqslant -\epsilon$; assuming $T_n \geqslant n/(2M_1)$ and using (23), we have

$$
|y_{ij}^{n+1} - y_{ij}^n| \leqslant \frac{6}{T_n x_i^n x_j^n} \leqslant \frac{12 M_1}{\epsilon^2 n} < \frac{\epsilon}{2}
$$

and, by assumption $|H(x_{n+1}) - H(x_n)| < \epsilon/2$, so that we conclude that $y_{ij}^{n+1} - H(x_{n+1}) \geqslant -2\epsilon$. $\qquad\square$

Lemma 5.5.

$$
\limsup_{n \to \infty} x_{ij}^n (y_{ij}^n - H(x_n))^- = 0.
$$

Proof. We fix $\epsilon > 0$ and $m_0 \in \mathbb{N}$, and let τ_{m_0} be the stopping time

$$\tau_{m_0} := \inf \left\{ n \geqslant m_0 \, : \, x_n \notin \Delta_{\epsilon^4} \text{ or } T_n < \frac{n}{2M_1} \text{ or } |H(x_n) - H(x_{m_0})| > \frac{\epsilon}{4} \right\}.$$

We only need to show that almost surely, either $\tau_{m_0} < \infty$ or $x_n \in U_{ij}(3\epsilon)$ for all n large enough. This will complete the proof since we know from Theorem 2.1, Lemma 3.2 and Proposition 4.2 that there exists almost surely $m_0 \in \mathbb{N}$ s.t. $\tau_{m_0} = \infty$.

Let σ_{m_0} be the stopping time

$$\sigma_{m_0} := \inf \{ n \geqslant m_0 \, : \, x_n \in U_{ij}(\epsilon) \}.$$

Lemma 5.3(b) implies that there exists a.s. a (random) $m_0 \in \mathbb{N}$ such that, for all $n \geqslant m \geqslant m_0$,

$$(R_n - R_m)^- \leqslant \frac{\epsilon}{2}, \qquad (S_n - S_m)^+ \leqslant \frac{\epsilon}{2}. \tag{24}$$

Therefore $\tau_{m_0} < \infty$ or $\sigma_{m_0} < \infty$, using $\sum_{n \geqslant m_0} 1/n = \infty$ and the observation that x_{ij}^n is bounded.

For all $n \in [\sigma_{m_0}, \tau_{m_0})$, let ρ_n be the largest $k \leqslant n$ such that $x_k \in U_{ij}(\epsilon)$. By (24),

$$y_{ij}^n - y_{ij}^{\rho_n+1} \geqslant -\frac{\epsilon}{2}.$$

Now $x_{\rho_n+1} \in U_{ij}(2\epsilon)$ (by Lemma 5.4): let us assume for instance that $y_{ij}^{\rho_n+1} - H(x_{\rho_n+1}) \geqslant -2\epsilon$. Together with $|H(x_n) - H(x_{\rho_n+1})| \leqslant \epsilon/2$ (by $n < \tau_{m_0}$), we deduce that

$$y_{ij}^n - H(x_n) \geqslant y_{ij}^{\rho_n+1} - H(x_{\rho_n+1}) - \epsilon \geqslant -3\epsilon.$$

With a similar argument, $x_{ij}^{\rho_n+1} \leqslant 2\epsilon$ implies $x_{ij}^n \leqslant 3\epsilon$. So overall, $x_n \in U_{ij}(3\epsilon)$ if $\sigma_{m_0} \leqslant n < \tau_{m_0}$, which enables us to conclude.

$\square$

Proof of Theorem 2.2.

$$2H(x_n) = \sum_{i,j \in \mathcal{S}} y_{ij}^n x_{ij}^n$$

$$= 2H(x_n) + \sum_{i,j \in \mathcal{S}} (y_{ij}^n - H(x_n))^+ x_{ij}^n - \sum_{i,j \in \mathcal{S}} (y_{ij}^n - H(x_n))^- x_{ij}^n$$

Hence, $\lim_{n \to \infty} x_{ij}^n (y_{ij}^n - H(x_n))^- = 0$ implies

$$\lim_{n \to \infty} x_{ij}^n (y_{ij}^n - H(x_n)) = 0.$$

Lemma 5.5 enables us to conclude.

5.2 Bipartite graph structure

Let us recall the bipartite graph defined in Section 3 (see Definition 2.1): any $x \in \Delta$ is associated with a weighted bipartite graph $\mathcal{G}_x$ with vertices $\mathcal{S} := \mathcal{S}_1 \cup \mathcal{S}_2$, adjacency $\sim$ and weights as follows

(1) $\forall i \in \mathcal{S}_1, j \in \mathcal{S}_2, i \sim j$ if and only if $x_{ij} > 0$.

(2) The weight of edge $\{i,j\}$ is $x_{ij}/(x_i x_j)$.

Let $\mathcal{C}_1$, ..., $\mathcal{C}_d$ be the connected components of $\mathcal{G}_x$. Besides the bipartite graph defined above, let us also discuss two other possible ways to assign weights:

Graph $\mathcal{G}_x^1$:   Edge e_{ij} has weight x_{ij}/x_j, $i \in \mathcal{S}_1, j \in \mathcal{S}_2$.

Graph $\mathcal{G}_x^2$:   Edge e_{ij} has weight x_{ij}/x_i, $i \in \mathcal{S}_1, j \in \mathcal{S}_2$.

By Lemma 4.3, we observe some interesting properties of these three graphs when $x \in \Lambda$, in particular $x \in \Gamma$:

On $\mathcal{G}_x$: All edges in a component $\mathcal{C}_k$, $k = 1, \ldots, d$ have the same weight λ_k. Hence,

$$H(x) = \sum_{k=1}^{d} \sum_{i \in \mathcal{S}_1 \cap \mathcal{C}_k} x_i \lambda_k = \sum_{k=1}^{d} \sum_{j \in \mathcal{S}_2 \cap \mathcal{C}_k} x_j \lambda_k.$$

Furthermore, if x is an equilibrium, all the edges in $\mathcal{G}_x$ have the same weight, which equals $H(x)$.

On $\mathcal{G}_x^1$: Every edge linked with the same state i has the same weight, which we denote by k_i. Also for each signal j, the sum of weights of edges linked to j is equal to 1, so that $H(x) = \sum_{i=1}^{M_2} k_i$.

On $\mathcal{G}_x^2$: Every edge linked with the same signal j has the same weight, which we denote by k_j'. Also for each state i, the sum of the weights of edges linked to i is equal to 1. $H(x) = \sum_{j=1}^{M_1} k_j'$.

5.3   Properties of Lyapounov function

We now show that H is constant on each connected component of Γ. Since it is not continuous on the boundary, we first prove in Lemma 5.6 that it takes a constant value on connected subsets of Γ with the same support (defined below) by a differentiability argument, and then conclude in Proposition 5.1 by a continuity argument on the set of equilibria.

Let

$$\Theta := \{\theta \,:\, \theta \subseteq \mathcal{S}_{\text{pair}}\}.$$

For any $x \in \Delta$, we define its support

$$S_x := \{(i,j) \,:\, i \in \mathcal{S}_1,\, j \in \mathcal{S}_2,\, x_{ij} > 0\}.$$

Θ can be used as an index set to divide Δ into several parts in the following sense: for any $\theta \in \Theta$,

$$\Delta_\theta := \{x \in \Delta \,:\, S_x = \theta\},$$

$$\Gamma_\theta := \Delta_\theta \cap \Gamma.$$

Lemma 5.6. *For any $\theta \in \Theta$, H is constant on each component of Γ_θ.*

Proof. Given $q \in \Gamma_\Theta$, let us differentiate H at q with respect to $x_{ij} = x_{ji}$, $(i,j) \in S_q$ without the constraint $x \in \Delta$:

$$
\left(\frac{\partial}{\partial x_{ij}} H(x) \right)_{x=q} = \sum_{(k,l) \in S_q} \frac{\partial}{\partial x_{ij}} \left(\frac{x_{kl}^2}{x_k x_l} \right)
$$

$$
= \frac{\partial}{\partial x_{ij}} \left(\frac{x_{ij}^2}{x_i x_j} \right) + \sum_{k \neq j,\, (i,k) \in S_q} \frac{\partial}{\partial x_{ij}} \left(\frac{x_{ik}^2}{x_i x_k} \right) + \sum_{l \neq i,\, (l,j) \in S_q} \frac{\partial}{\partial x_{ij}} \left(\frac{x_{lj}^2}{x_l x_j} \right)
$$

$$
= \frac{q_{ij}}{q_i q_j} \left(2 - \frac{q_{ij}}{q_i} - \frac{q_{ij}}{q_j} \right) + \sum_{k \neq j,\, (i,k) \in S_q} \frac{q_{ik}}{q_i q_k} \left(-\frac{q_{ik}}{q_i} \right) + \sum_{l \neq i,\, (l,j) \in S_q} \frac{q_{lj}}{q_l q_j} \left(-\frac{q_{lj}}{q_j} \right)
$$

$$
= H(q) \left(2 - \frac{q_{ij}}{q_i} - \frac{q_{ij}}{q_j} \right) + H(q) \left(-\frac{q_i - q_{ij}}{q_i} \right) + H(q) \left(-\frac{q_j - q_{ij}}{q_j} \right)
$$

$$
= 0.
$$

The penultimate equality comes from the fact that $q_{ij}/(q_i q_j) = H(q)$ if $(i,j) \in S_q$, $q \in \Gamma$. $\qquad\square$

Proposition 5.1. *H is constant on each connected component of Γ.*

Proof. Let us show that H is continuous on Γ, which will enable us to conclude. Indeed, suppose that $q \in \Gamma$, and that $x \in \Gamma$ is in the neighbourhood of $q \in \Gamma$ within Δ, then $S_x \supseteq S_q$ and, using $x \in \Gamma$,

$$H(x) = \sum_{(i,j)\in S_q} \frac{x_{ij}^2}{x_i x_j} + \sum_{(i,j)\in S_x \setminus S_q} x_{ij} H(x),$$

so that

$$H(x) = \frac{1}{1 - \sum_{(i,j)\in S_x \setminus S_q} x_{ij}} \sum_{(i,j)\in S_q} \frac{x_{ij}^2}{x_i x_j},$$

and the conclusion follows. $\qquad\square$

5.4 Classification of equilibria and stability

5.4.1 Jacobian matrix

At any equilibrium $x \in (\Delta \setminus \partial\Delta) \cap \Gamma$ (F is not differentiable on $\partial\Delta$), we calculate the Jacobian matrix

$$J_x = \left(\frac{\partial F_{lk}}{\partial x_{ij}} \right)_{(i,j),(l,k)\in\mathcal{S}_{\text{pair}}}$$

where, by a slight abuse of notation,

$$F(x) = (F_{ij}(x))_{(i,j)\in\mathcal{S}_{\text{pair}}}.$$

For all (i,j), $(l,k) \in \mathcal{S}_{\text{pair}}$, a simple extension of the calculation in the proof of Lemma 5.6 yields $\frac{\partial H}{\partial x_{ij}}(x) = -\ \mathbb{1}_{\{x_{ij}=0\}} 2H(x)$, so that

$$
\begin{aligned}
\frac{\partial F_{lk}}{\partial x_{ij}} &=\ \mathbb{1}_{\{(i,j)=(l,k),x_{ij}=0\}}(y_{lk} - H(x)) + x_{lk}\frac{\partial\, y_{lk}}{\partial x_{ij}} - x_{lk}\frac{\partial\, H}{\partial x_{ij}}(x) \\
&=-\ \mathbb{1}_{\{(i,j)=(l,k),x_{ij}=0\}}H(x) + x_{lk}\frac{\partial\, y_{lk}}{\partial x_{ij}} + 2x_{lk}\ \mathbb{1}_{\{x_{ij}=0\}}H(x) \\
&= H(x)\ \mathbb{1}_{\{(i,j)=(l,k)\}}\big(\ \mathbb{1}_{\{x_{ij}\neq 0\}} -\ \mathbb{1}_{\{x_{ij}=0\}}\big) \\
&\quad + H(x)\ \mathbb{1}_{\{x_{lk}\neq 0\}}\left(-\frac{x_{ik}}{x_i}\ \mathbb{1}_{\{i=l\}} - \frac{x_{lj}}{x_j}\ \mathbb{1}_{\{j=k\}}\right) + 2x_{lk}\ \mathbb{1}_{\{x_{ij}=0\}}H(x)
\end{aligned}
$$

Therefore, for any $(i,j) \in \mathcal{S}_{\text{pair}}$ s.t. $x_{ij} \neq 0$,

$$
\frac{\partial F_{ij}}{\partial x_{ij}} = H(x)\left(1 - \frac{x_{ij}}{x_i} - \frac{x_{ij}}{x_j}\right); \tag{25}
$$

$$
\frac{\partial F_{ik}}{\partial x_{ij}} = H(x)\left(-\frac{x_{ik}}{x_i}\right), \qquad k \in \mathcal{S}_2 \setminus \{j\}; \tag{26}
$$

$$
\frac{\partial F_{lj}}{\partial x_{ij}} = H(x)\left(-\frac{x_{lj}}{x_j}\right), \qquad l \in \mathcal{S}_1 \setminus \{i\}; \tag{27}
$$

$$
\frac{\partial F_{lk}}{\partial x_{ij}} = 0, \qquad l \in \mathcal{S}_1 \setminus \{i\},\ k \in \mathcal{S}_2 \setminus \{j\}; \tag{28}
$$

for any $(i,j) \in \mathcal{S}_{\text{pair}}$ s.t. $x_{ij} = 0$,

$$
\frac{\partial F_{ij}}{\partial x_{ij}} = -H(x); \tag{29}
$$

$$
\frac{\partial F_{lk}}{\partial x_{ij}} = 0, \qquad l \in \mathcal{S}_1,\ k \in \mathcal{S}_2,\ (l,k) \neq (i,j),\ x_{lk} = 0. \tag{30}
$$

Let $\mathcal{C}_1, \ldots, \mathcal{C}_d$ be the connected components of the edges of $\mathcal{G}_x$. Let

$$
J_x^m := \left(\frac{\partial\, F_{lk}}{\partial x_{ij}}\right)_{(i,j),(l,k)\in\mathcal{C}_m}.
$$

38

Therefore, using (25)-(30), J_x can be written as follows, by putting first (i,j) and (l,k) coordinates such that $x_{ij} \neq 0$ and $x_{lk} \neq 0$ (in the same order, with increasing connected components $\mathcal{C}_1, \ldots, \mathcal{C}_d$)

$$
\begin{pmatrix}
J_x^1 & & & & & (0) \\
& \ddots & & & & \\
& & J_x^d & & & \\
& & & -H(x) & & \\
& & & & \ddots & \\
(*) & & & & & -H(x)
\end{pmatrix}
$$

5.4.2 Classification of equilibria based on stability

Let us introduce a few definitions of stability for ordinary differential equations.

Definition 5.1. x is Lyapounov stable *if for any neighborhood U_1 of x, there exists a neighborhood $U_2 \subseteq U_1$ of x such that any solution $x(t)$ starting in U_2 is such that $x(t)$ remains in U_2 for all $t \geqslant 0$.*

Definition 5.2. x *is* asymptotically stable *if it is* Lyapounov stable *and there exists a neighbourhood U_1 such that any solution $x(t)$ starting in U_1 is such that $x(t)$ converges to x.*

An equilibrium that is Lyapunov stable but not asymptotically stable is called *neutrally stable* sometimes.

Definition 5.3. x *is* linearly stable *if all eigenvalues of the Jacobian matrix at x have nonpositive real part; otherwise, x is called* linearly unstable.

Remark that, with these definitions, linear stability allows for eigenvalues to have zero real part, and therefore does not necessarily imply Lyapounov stability. However the dynamics considered here makes these stable equilibria indeed Lyapounov stable when they do not lie outside the boundary of Δ are studying here, as it can be shown by the help of an entropy function; Section 7 on convergence with positive probability to stable configurations can be understood as a consequence of this propoerty in the nondeterministic case.

Definition 5.4. *Let*

$$\Gamma_0 := \Gamma \cap \Delta \setminus \partial\Delta,$$

$$\Gamma_b := \Gamma \cap \partial\Delta,$$

and let Γ_s (resp. Γ_u) be the set of linearly stable (resp. unstable) equilibria in Γ_0 for the mean-field ODE.

For any $x \in \Gamma_u$, let

$$\mathcal{E}_x := \{\theta \in \mathbb{R}^{M_1 \times M_2} : |\theta| = 1 \text{ and } \exists\, (i,j) \in S_x \text{ s.t. } \theta \cdot \mathbf{e}_{ij} > 0\}.$$

Proposition 5.2. *We have*

(a) $\Gamma_s = \{x \in \Gamma_0 : (P)_{\mathcal{G}_x} \text{ holds}\}.$

(b) *If $x \in \Gamma_u$, then there exists an eigenvector in $\mathcal{E}_x$ whose eigenvalue has positive real*

part.

(c) *For all $x \in \Gamma_u$ and $\theta \in \mathcal{E}_x \setminus \{0\}$, there exists a neighbourhood $\mathcal{N}(x)$ of x such that, if $x_n \in \mathcal{N}(x)$, then*

$$\mathbb{E}[\, (\eta_{n+1} \cdot \theta)^2 \,|\, \mathcal{F}_n \,] > \frac{\mathsf{Cst}(x)}{n^2},$$

where η_{n+1} is the martingale increment defined in (7).

Note that **(b)-(c)** will be used to that $(x_n)_{n \in \mathbb{N}}$ stochastically "perturbs enough" the ODE (6), in order to prove nonconvergence to unstable equilibria.

To prove Proposition 5.2, we need following Lemma 5.7 on the structure of $\mathcal{G}_x$ when $x \in \Gamma_0$, and the elementary Lemma 5.8.

Lemma 5.7. *For any $x \in \Gamma_0$ such that $\mathcal{P}_{\mathcal{G}_x}$ does not hold, $\mathcal{G}_x$ has at least one connected component on which every vertex has at least two edges.*

Proof. First, $x \in \Gamma_u$ implies that there exists a connected component $\mathcal{C}$ with at least two states and two signals. Assume, by contradiction, that signal j is in $\mathcal{C}$ and j is only linked to one state, for instance state i. Then $x_{ij}/x_j = 1$, and for all $k \in \mathcal{S}_2$, s.t. $x_{ik} > 0$,

$$\frac{x_{ik}}{x_k} = \frac{x_{ij}}{x_j} = 1,$$

i.e. k is only linked with one edge. It implies $\mathcal{C} \cap \mathcal{S}_1 = \{i\}$, which contradicts our assumption. $\qquad\square$

Lemma 5.8. *If a random variable ω satisfies that $\mathbb{E}[\omega] < \infty$, $\mathbb{P}(\omega = a) \geqslant p$ and $\mathbb{P}(\omega = b) \geqslant p$, then*

$$\mathrm{Var}(\omega) \geqslant \frac{(b-a)^2 p}{4}.$$

41

Proof.

$$\text{Var}(\omega) \geqslant p\big((a - \mathbb{E}[\omega])^2 \vee (b - \mathbb{E}[\omega])^2\big) \geqslant \frac{(b-a)^2 p}{4} \,.$$

□

Proof of Proposition 5.2. (a)-(b) Suppose that $x \in \Gamma_0$ and that $\mathcal{P}_{\mathcal{G}_x}$ does not hold, and let us prove that $x \in \Gamma_u$. Lemma 5.7 implies that $\mathcal{G}_x$ has a connected component on which every vertex has at least two edges, which we assume to be $\mathcal{C}_1$ w.l.o.g. Let $V(\mathcal{C}_1)$ (resp. $E(\mathcal{C}_1)$) be its set of vertices (resp. edges). Let us show that J_x^1 has at least one eigenvalue with positive real part.

Indeed, let us compute the trace of J_x^1:

$$Tr(J_x^1) = H(x) \sum_{i,j \in \mathcal{C}_1 : i \sim j} (1 - \frac{x_{ij}}{x_i} - \frac{x_{ij}}{x_j})$$
$$= H(x) \left(|E(\mathcal{C}_1)| - |V(\mathcal{C}_1)| \right) \geqslant 0.$$

The last inequality comes from the fact that the number of edges is greater than or equal to the number of vertices in $\mathcal{C}_1$ because every vertex has at least two edges in $\mathcal{C}_1$.

Now it is easy to check that $(J_x^1)\mathbf{1} = -H(x)\,\mathbb{1}$ where $\mathbf{1} = (1, \ldots, 1)^T$, and therefore that $-H(x)$ is an eigenvalue of J_x^1, which enables us to conclude (b) and the first part of (a).

Now suppose that $x \in \Gamma_0$, and that $\mathcal{P}_{\mathcal{G}_x}$ holds. Then each component of $\mathcal{G}_x$ has only one state or only one signal. Let us assume for instance that $\mathcal{C}_1$ consists of states

$1, \ldots, k$ and signal A. Then J_x^1 equals

$$
-\frac{H(x)}{x_A}
\begin{pmatrix}
x_1 & x_2 & \ldots & x_k \\
x_1 & x_2 & \ldots & x_k \\
\vdots & \vdots & \ddots & \vdots \\
x_1 & x_2 & \ldots & x_k
\end{pmatrix}.
$$

The rank of J_x^1 is 1, and its eigenvalues are 0 and $-H(x)$, which completes the proof.

(c) Let $x \in \Gamma_u$ and $\theta \in \mathcal{E}_x \setminus \{0\}$. Let $\{\mathbf{e}_{ij}\}$ be a orthogonal basis set in $\mathbb{R}^{M_1 \times M_2}$.
Let

$$
W_{n+1} := \mathbb{1}_{|V_{n+1}-V_n|=1}\big(V_{n+1} - V_n - x_n\big) = (1 + T_n)(x_{n+1} - x_n).
$$

We note that

$$
W_{n+1} \cdot \theta = 0 , \quad \text{with probability } 1 - \frac{H(x_n)}{M_1} ,
$$

$$
\forall i \in \mathcal{S}_1, j \in \mathcal{S}_2, \ W_{n+1} \cdot \theta = (1 - x_{ij}^n)\mathbf{e}_{ij} \cdot \theta , \quad \text{with probability } \frac{x_{ij}^n y_{ij}^n}{M_1} .
$$

Note that $x \in \Gamma_u$ implies that $H(x) \neq M_1$ and that, for all $(i,j) \in S_x$, $x_{ij} \neq 1$.

Therefore, assuming that x_n is in the neighbourhood of x (for which $y_{ij} = H(x)$),
Lemma 5.8 implies

$$
\mathrm{Var}\big(W_{n+1} \cdot \theta \mid \mathcal{F}_n\big) \geqslant \max_{i,j \in \mathcal{C}_1 : i \sim j} \left\{ \min\left(1 - \frac{H(x_n)}{M_1}, \frac{x_{ij}^n y_{ij}^n}{M_1}\right) \frac{((1 - x_{ij}^n)\mathbf{e}_{ij} \cdot \theta)^2}{4} \right\} \quad (31)
$$

$$
\geqslant \mathsf{Cst}(x) > 0, \quad (32)
$$

where we use $\theta \in \mathcal{E}_x$ in the penultimate inequality. This completes the proof. $\square$

If $M_1 = M_2$, let us define the set of equilibria Γ_{sig} as follows: $x \in \Gamma_{\text{sig}}$ if and only if there exists a bijective map ϕ from $\mathcal{S}_1$ to $\mathcal{S}_2$ such that

$$
x_{ij} := \begin{cases} \frac{1}{M_1} & \text{if } j = \phi(i), \\[2mm] 0 & \text{otherwise.} \end{cases}
$$

Equilibria in Γ_{sig} correspond to a perfectly efficient signaling system, in the sense that they bear no synonyms or informational bottlenecks. It is easy to check from the calculation in the proof of (b) of Proposition 5.2 that the Jacobian matrix at such an x has the only eigenvalue $-H(x)$, which implies asymptotic stability.

Corollary 5.1. *Suppose the game has the same number of states and signals, i.e. $M_1 = M_2$ and that $x \in \Gamma_{\text{sig}}$. Then x is asymptotically stable for the mean-field ODE.*

6   Links with static equilibrium analysis

We present in this section some results on the characterization of *evolutionarily stable strategies* (ESS) and *neutrally stable strategies* (NSS) for signaling games. Note that Taylor and Jonker (1978), and Zeeman (1979) propose, in a general setting, conditions under which an evolutionarily stable strategy is indeed a (dynamically) stable equilibrium. However the condition is quite strong, and only applies to special cases of the signaling game.

We show here an equivalence between this static context and the underlying rein-

forcement learning dynamics, i.e. that

(1) the set of ESS matches the set of asymptotically stable equilibria of the mean-field ODE;

(2) the set of neutrally stable strategies matches the set of linearly stable equilibria of the mean-field ODE;

(3) the set of Nash equilibria matches the set Λ where $\nabla H \cdot F$ vanishes.

In Sections 6.1, 6.2 and 6.3 we successively present usual notions in the static setting of signaling games, results on equilibrium selection and the connection between stability of the dynamics and the static setting of the game mentioned above.

6.1 Static setting

Let

$$\mathcal{P} = \{P \in \mathbb{R}_+^{M_1 \times M_2} : \forall\, i \in \mathcal{S}_1, \sum_{j \in \mathcal{S}_2} p_{ij} = 1\},$$

$$\mathcal{Q} = \{Q \in \mathbb{R}_+^{M_2 \times M_1} : \forall\, j \in \mathcal{S}_2, \sum_{i \in \mathcal{S}_1} q_{ji} = 1\}.$$

A Sender's strategy can be represented by a $M_1 \times M_2$ matrix $P \in \mathcal{P}$ in the sense that if he sees state i, he chooses signal j with probability p_{ij}. Also we can represent Receiver's strategy by a $M_2 \times M_1$ matrix $Q \in \mathcal{Q}$, i.e. if he sees signal j he chooses act i with probability q_{ji}. The payoff function is

$$\text{Payoff}(P, Q) = \sum_{i \in \mathcal{S}_1, j \in \mathcal{S}_2} p_{ij} q_{ji} = tr(PQ).$$

As in the real world, where somebody can be sender at one time and receiver at another, the authors assume that each individual can be Sender or Receiver with equal probability at each time, which symmetrizes the game. Thus the (mixed) strategy of any of the two players of the symmetrized game is a pair of Sender and Receiver matrices (P, Q), and the payoff function Υ can be written as

$$\Upsilon[(P, Q), (P', Q')] = \Upsilon[(P', Q'), (P, Q)] = \frac{1}{2} tr(PQ') + \frac{1}{2} tr(P'Q).$$

6.2 Equilibrium selection

Definition 6.1. *A strategy* $(P, Q) \in \mathcal{P} \times \mathcal{Q}$ *is called a Nash strategy if*

$$\Upsilon[(P, Q), (P, Q)] \geqslant \Upsilon[(P', Q'), (P, Q)], \; \forall \, (P', Q') \in \mathcal{P} \times \mathcal{Q}.$$

There are uncountably many Nash strategies in signaling games. Let us recall the following notions of evolutionary and neutrally stable equilibria, which enable one to distinguish the relevant limiting strategies of the game; note that the notions are purely static here.

Definition 6.2. *A strategy* $(P, Q) \in \mathcal{P} \times \mathcal{Q}$ *is* evolutionarily stable *if*

(i) *it is a Nash strategy, and*

(ii) $\Upsilon[(P, Q), (P', Q')] > \Upsilon[(P', Q'), (P', Q')]$ *for all* $(P', Q') \neq (P, Q)$.

Definition 6.3. *A strategy* $(P, Q) \in \mathcal{P} \times \mathcal{Q}$ *is* neutrally stable *if*

(i) *it is a Nash strategy, and*

(ii) *if* $\Upsilon[(P,Q),(P,Q)] = \Upsilon[(P',Q'),(P,Q)]$ *for some* $(P',Q') \in \mathcal{P} \times \mathcal{Q}$, *then*

$$\Upsilon[(P,Q),(P',Q')] \geqslant \Upsilon[(P',Q'),(P',Q')].$$

The following Propositions 6.1, 6.2 and 6.3 characterize ESSs and NSSs in the signaling game.

Proposition 6.1 (Trapa and Nowak (2000)). *Let* $(P,Q) \in \mathcal{P} \times \mathcal{Q}$ *be such that neither* P *nor* Q *contains a column that consists entirely of zeros. Then* (P,Q) *is a Nash strategy if and only if there exist positive numbers* $p_1, \ldots, p_n$ *and* $q_1, \ldots, q_m$ *such that*

(1) *for each* j, *the* j-*th column of* P *has its entries drawn from* $\{0, p_j\}$,

(2) *for each* i, *the* i-*th column of* Q *has its entries drawn from* $\{0, q_i\}$

(3) *for all* i, j, $p_{ij} \neq 0$ *if and only if* $q_{ji} \neq 0$.

Remark. The assumption that neither P nor Q contains a column consisting entirely of zeros corresponds to the requirement that no signal or act falls out of use.

Proposition 6.2 (Pawlowitsch (2007)). *Let* $(P,Q) \in \mathcal{P} \times \mathcal{Q}$ *be a Nash strategy. Then* (P,Q) *is a neutrally stable strategy if and only if*

(1) *at least one of the two matrices,* P *or* Q, *has no zero column, and*

(2) *if* P *or* Q *contains a column with more than one positive element, then*

all the elements in this column take values in $\{0, 1\}$.

Proposition 6.3 (Trapa and Nowak (2000)). $(P,Q) \in \mathcal{P} \times \mathcal{Q}$ *is an evolutionarily stable strategy if and only if* $M_1 = M_2$, P *is a permutation matrix and* $Q = P^T$.

6.3 Connection

Let us now define a map between the static and dynamic learning models, in order to emphasize the correspondence between stability in either settings.

Let s_1 be a bijective map from $\mathcal{S}_1$ to $\{1, \ldots, M_1\}$ and s_2 be a bijective map from $\mathcal{S}_2$ to $\{1, \ldots, M_2\}$. Let us define map Ψ from $\Delta \setminus \partial\Delta$ to $\mathcal{P} \times \mathcal{Q}$: $\Psi(x) = (P, Q)$ where

$$p_{s_1(i)s_2(j)} = \frac{x_{ij}}{x_i}, \ q_{s_2(j)s_1(i)} = \frac{x_{ij}}{x_j}, \qquad \forall\, i \in \mathcal{S}_1,\, j \in \mathcal{S}_2.$$

Proposition 6.4. *Let*

$$\Gamma := \Big\{ (P, Q) \in \mathcal{P} \times \mathcal{Q} : \textit{neither } P \textit{ nor } Q \textit{ contains}$$
$$\textit{any column that consists entirely of zeros} \Big\},$$

and let

$$\mathcal{L}_{\mathcal{P} \times \mathcal{Q}} := \{ (P, Q) \in \mathcal{P} \times \mathcal{Q} \textit{ is a Nash strategy } \} \cap \Gamma,$$
$$\bar{\mathcal{L}}_{\mathcal{P} \times \mathcal{Q}} := \{ (P, Q) \in \mathcal{P} \times \mathcal{Q} \textit{ is a neutrally stable strategy } \} \cap \Gamma,$$
$$\tilde{\mathcal{L}}_{\mathcal{P} \times \mathcal{Q}} := \{ (P, Q) \in \mathcal{P} \times \mathcal{Q} \textit{ is an evolutionarily stable strategy } \} \cap \Gamma.$$

Then

(a) $\Psi((\Delta \setminus \partial\Delta) \cap \Lambda) = \mathcal{L}_{\mathcal{P} \times \mathcal{Q}}$ *and* $\Psi((\Delta \setminus \partial\Delta) \setminus \Lambda)) \cap \mathcal{L}_{\mathcal{P} \times \mathcal{Q}} = \emptyset$.

(b) $\Psi(\Gamma_s) = \bar{\mathcal{L}}_{\mathcal{P} \times \mathcal{Q}}$.

(c) $\Psi(\Gamma_{\text{sig}}) = \tilde{\mathcal{L}}_{\mathcal{P} \times \mathcal{Q}}$.

Proof. (a) is a direct consequence of Lemma 4.3 and Proposition 6.1.

Conversely, given $(P, Q) \in \mathcal{L}_{\mathcal{P} \times \mathcal{Q}}$, let $p_1, \ldots p_n$ and $q_1, \ldots q_m$ be defined as in Proposition 6.1. Note that

$$p_{s_1(i)s_2(j)} q_{s_2(j)s_1(i)} = p_{s_2(j)} q_{s_1(i)} \, \mathbb{1}_{p_{s_1(i)s_2(j)} \neq 0} = p_{s_1(i)s_2(j)} q_{s_1(i)} = p_{s_2(j)} q_{s_2(j)s_1(i)}.$$

Define

$$Z := \sum_{(i,j) \in \mathcal{S}_{pair}} p_{s_1(i)s_2(j)} q_{s_2(j)s_1(i)} = \sum_{i \in \mathcal{S}_1} q_{s_1(i)} = \sum_{j \in \mathcal{S}_2} p_{s_2(j)},$$

and let

$$x_{ij} := \frac{p_{s_1(i)s_2(j)} q_{s_2(j)s_1(i)}}{Z}, \quad i \in \mathcal{S}_1, j \in \mathcal{S}_2;$$

$$x_i := \frac{q_i}{\sum_{k \in \mathcal{S}_1} q_k} = \sum_{j \in \mathcal{S}_2} x_{ij}, \quad i \in \mathcal{S}_1;$$

$$x_j := \frac{p_j}{\sum_{k \in \mathcal{S}_2} p_k} = \sum_{i \in \mathcal{S}_1} x_{ij}, \quad j \in \mathcal{S}_2.$$

Then $(P, Q) = \Psi(x)$, and $x \in \Delta \setminus \partial \Delta) \cap \Lambda$ is again a direct consequence of Lemma 4.3 and Proposition 6.1.

Let us now prove (b), and assume $(P, Q) = \Psi(x)$: if one column of P (resp. Q) has more than one element, this corresponds to an informational bottleneck (resp. synonym). Hence Condition (2) in Proposition 6.2 means that a state(act)-signal correspondence cannot be associated both to a synonym and an informational bottleneck, which translates into $(P)_{\mathcal{G}_x}$. (c) follows from Corollary 5.1 and Proposition 6.3. $\qquad \square$

7 Convergence with positive probability to stable configurations

Let us prove the more general

Theorem 7.1. *Let $q \in \Gamma_s$, and let $\mathcal{N}(q)$ neighbourhood of q in Δ. Then, with positive probability,*

(a) $x_n \to x \in \mathcal{N}(q)$ *s.t.* $\mathcal{G}_x = \mathcal{G}_q$.

(b) $V(\infty, i, j) = \infty \iff \{i, j\}$ *is an edge of $\mathcal{G}$.*

Theorem 7.1 is an obvious consequence of the following Proposition 7.1. Given $\mathcal{G} = (\mathcal{S}, \sim)$, assume that $(P)_{\mathcal{G}}$ holds: then each connected component of $\mathcal{G}$ contains either only a single state or a single signal. Let $\pi := \pi_{\mathcal{G}} : \mathcal{S} \longrightarrow \mathcal{R}_q^1$ be the function mapping $i \in \mathcal{S}$ to the single state/signal in the same connected component as i, with the convention that we choose the state if the component consists only of one state and one signal.

For all $i \in \mathcal{S}$, $n \in \mathbb{N}$ and $\epsilon > 0$, let us define

$$\alpha_i^n := x_i^n / x_{\pi(i)}^n$$

$$H_n^1 := \bigcap_{i \in \mathcal{S}, i = \pi(i)} \{ V_i^n \geqslant 2\epsilon n \},$$

$$H_n^2 := \bigcap_{i \in \mathcal{S}} \{ \alpha_i^n \geqslant \epsilon \},$$

$$H_n^3 := \bigcap_{i,j \in \mathcal{S}, \pi(i) \neq \pi(j)} \{ V_{ij}^n \leqslant \sqrt{n} \}.$$

Proposition 7.1. *Let $\mathcal{G}$ be such that $(P)_{\mathcal{G}}$ holds, and let $\pi := \pi_{\mathcal{G}}$. For all $\epsilon \in (0, 1/M_1)$,*

if H_n^1, H_n^2 and H_n^3 hold, and $n \geqslant \mathsf{Cst}(\epsilon, M_1, M_2)$, then, with lower bounded probability (only depending on ϵ, M_1 and M_2), for all $i, j \in \mathcal{S}$, $k \geqslant n$,

$$V_{ij}^\infty = V_{ij}^n, \quad when \ \pi(i) \neq \pi(j); \tag{33}$$

$$\alpha_i^k / \alpha_i^n \in (1 - \epsilon, 1 + \epsilon); \tag{34}$$

$$V_i^k \geqslant \epsilon k, \quad when \ \pi(i) = i. \tag{35}$$

In the remainder of this section, we fix the graph $(\mathcal{G}, \sim)$ (and thus $\pi = \pi_\mathcal{G}$) and $\epsilon > 0$. The proof consists of the following Lemmas 7.1–7.3.

Let, for all $i, j \in \mathcal{S}$, $n \in \mathbb{N}$,

$$\tau_n^{1,i,j} := \inf\{k \geqslant n \ : \ V_{ij}^k \neq V_{ij}^n\};$$

$$\tau_n^{2,i} := \inf\{k \geqslant n \ : \ \alpha_i^k / \alpha_i^n \notin (1 - \epsilon, 1 + \epsilon)\};$$

$$\tau_n^{3,i} := \inf\{k \geqslant n \ : \ V_i^k < \epsilon k\},$$

and let

$$\tau_n^1 := \inf_{i,j \in \mathcal{S}, \pi(i) \neq \pi(j)} \tau_n^{1,i,j}, \quad \tau_n^2 := \inf_{i \in \mathcal{S}} \tau_n^{2,i}, \quad \tau_n^3 := \inf_{i \in \mathcal{S}, \pi(i) = i} \tau_n^{3,i}, \quad \tau_n := \tau_n^1 \wedge \tau_n^2 \wedge \tau_n^3.$$

Lemma 7.1. If $n \geqslant \mathsf{Cst}(\epsilon, M_1, M_2)$, then

$$\mathbb{P}\left(\tau_n^1 > \tau_n^2 \wedge \tau_n^3 \,|\, \mathcal{F}_n, H_n^1, H_n^2, H_n^3 \right) \geqslant \exp\left(-2\epsilon^{-4} M_1 M_2\right).$$

Proof. Assuming $n \geqslant \mathsf{Cst}(\epsilon, M_1, M_2)$,

$$\mathbb{P}\left(\tau_n^1 > \tau_n^2 \wedge \tau_n^3 \mid \mathcal{F}_n, H_n^1, H_n^2, H_n^3 \right) \geqslant \prod_{k \geqslant n} \left(1 - \sum_{\pi(i) \neq \pi(j)} \frac{(V_{ij}^n)^2}{V_i^k V_j^k} \right)$$

$$\geqslant \exp\left(-\frac{3}{2} \sum_{\pi(i) \neq \pi(j), k \geqslant n} \frac{n}{\epsilon^4 k^2} \right)$$

$$\geqslant \exp\left(-2\epsilon^{-4} M_1 M_2 \right).$$

$\square$

Lemma 7.2. *If $n \geqslant \mathsf{Cst}(\epsilon, M_1, M_2)$ then, for all $i \in \mathcal{S}$,*

$$\mathbb{P}\left(\tau_n^{2,i} > \tau_n^1 \wedge \tau_n^3 \mid \mathcal{F}_n, H_n^1, H_n^2, H_n^3 \right) \geqslant 1 - 2\exp(-\mathsf{Cst}(\epsilon)n).$$

Proof. Fix $i \in \mathcal{S}$, $n \in \mathbb{N}$, and assume w.l.o.g. that $\pi(i) \neq i$.

Let, for all $j \in \mathcal{S}$ and $k \geqslant n$,

$$\hat{V}_j^k := \hat{V}_j^n + \sum_{l \sim j} (V_{jl}^k - V_{jl}^n),$$

which is equal to V_j^k as long as $k < \tau_n^1$.

Let, for all $k \geqslant n$,

$$W_k := \log \frac{\hat{V}_i^k}{\hat{V}_{\pi(i)}^k},$$

and let us consider the Doob decomposition of $(W_k)_{k \geqslant n}$:

$$W_k = W_n + \Delta_k + \Psi_k$$

$$\Delta_k := \sum_{j \geqslant n+1} \mathbb{E}(W_j - W_{j-1} | \mathcal{F}_{j-1}).$$

Assume that H_n^1, H_n^2 and H_n^3 hold, and that $k < \tau_n$: then

$$
\begin{aligned}
|\Delta_{k+1} - \Delta_k| &= |\,\mathbb{E}\left[W_{k+1} - W_k \,|\, \mathcal{F}_k\right]| \\
&= \left| \frac{1}{M_1} \left(\frac{V_{i\pi(i)}^k}{V_i^k} \right)^2 \frac{1}{V_{\pi(i)}^k} (1 + \Box((V_i^k)^{-1}) - \frac{1}{M_1} \sum_{j \sim \pi(i)} \frac{V_{\pi(i)j}^k}{V_j^k} \frac{V_{\pi(i)j}^k}{(V_{\pi(i)}^k)^2} (1 + \Box((V_{\pi(i)}^k)^{-1}) \right| \\
&= \frac{1}{M_1 V_{\pi(i)}^k} \left| 1 + k^{-1/2}\Box(\mathsf{Cst}(\epsilon, M_1, M_2) - \sum_{j \sim \pi(i)} \left(1 + k^{-1/2}\Box(\mathsf{Cst}(\epsilon, M_1, M_2)) \right) \frac{V_{\pi(i)j}^k}{V_{\pi(i)}^k} \right| \\
&= k^{-3/2}\Box\left(\mathsf{Cst}(\epsilon, M_1, M_2)\right),
\end{aligned}
$$

where we use that, for all $j \sim \pi(i)$,

$$
\left| \frac{V_{\pi(i)j}^k}{V_j^k} - 1 \right|, \quad \left| \sum_{l \sim \pi(i)} \frac{V_{\pi(i)l}^k}{V_{\pi(i)}^k} - 1 \right| \leqslant k^{-1/2}\mathsf{Cst}(\epsilon, M_1, M_2). \tag{36}
$$

Therefore, for all $k \geqslant n$,

$$
|\Delta_k| \leqslant n^{-1/2}\mathsf{Cst}(\epsilon, M_1, M_2).
$$

Let us now estimate the martingale increment: $|\Psi_{k+1} - \Psi_k| \leqslant \mathsf{Cst}(\epsilon)k^{-1}$ (since $|W_{k+1} - W_k| \leqslant \mathsf{Cst}(\epsilon)k^{-1}$), so that Lemma 7.4 implies

$$
\mathbb{P}\left(\sup_{k \geqslant n} |\Psi_k - \Psi_n| \, \mathbb{1}_{\{k \leqslant \tau_n\}} \geqslant \epsilon/2 \right) \geqslant 1 - 2\exp(-\mathsf{Cst}(\epsilon)n),
$$

which completes the proof. $\qquad\square$

Lemma 7.3. *If $\epsilon \in (0, 1/M_1)$ and $n \geqslant \mathsf{Cst}(\epsilon, M_1, M_2)$ then, for all $i \in \mathcal{S}$ such that*

$\pi(i) = i$,

$$\mathbb{P}\left(\tau_n^{3,i} > \tau_n^1 \wedge \tau_n^2 \,\middle|\, \mathcal{F}_n, H_n^1, H_n^2\right) > 1 - 2\exp(-\mathsf{Cst}(\epsilon)n).$$

Proof. Let $n \in \mathbb{N}$, assume that H_n^1, H_n^2 and H_n^3 hold, and fix $i \in \mathcal{S}$ such that $\pi(i) = i$. Let us consider the Doob decomposition of $(V_i^k)_{k \geqslant n}$:

$$V_i^k := V_i^n + \Phi_k + \Xi_k$$

$$\Phi_k := \sum_{j \geqslant n+1} \mathbb{E}\left(V_i^j - V_i^{j-1} \,\middle|\, \mathcal{F}_{j-1}\right).$$

Now, for all $\eta > 0$, if $n \geqslant \mathsf{Cst}(\eta, \epsilon)$ and $k < \tau_n$, (36) implies

$$\Phi_{k+1} - \Phi_k = \mathbb{E}\left(V_i^{k+1} - V_i^k \,\middle|\, \mathcal{F}_k\right) \geqslant \frac{1}{M_1} \sum_{j \sim i} \frac{(V_{ij}^k)^2}{V_i^k V_j^k} \geqslant \frac{1}{M_1} - \eta.$$

Let us now estimate the martingale increment: let, for all $p \geqslant n$,

$$\chi_p := \sum_{k=n}^{p-1} \frac{\Xi_{k+1} - \Xi_k}{k}.$$

Then, for all $p \geqslant n$,

$$\Xi_p = \sum_{n \leqslant k \leqslant p-1} (\chi_{k+1} - \chi_k)k = -\sum_{n \leqslant k \leqslant p-1} \chi_k + (p-1)\chi_p.$$

This implies, using Lemma 7.4 (and $|\Xi_{k+1} - \Xi_k| \leqslant 1$ for all $k \geqslant n$) that, for all $\epsilon > 0$,

$$\mathbb{P}\left(\forall k \geqslant n, \ V_i^k \geqslant (2\epsilon - \eta)n + (k-n)(1/M_1 - \eta) \,\middle|\, \mathcal{F}_n\right)$$

54

$$\geqslant \mathbb{P}\left(\sup_{p\geqslant n}\left|\frac{\Xi_p}{p}\right|\leqslant \eta \,|\,\mathcal{F}_n\right) \geqslant \mathbb{P}\left(\sup_{k\geqslant n}|\chi_k|\leqslant \frac{\eta}{2}\,|\,\mathcal{F}_n\right) \geqslant 1 - 2\exp(-\mathsf{Cst}(\eta)n);$$

we choose $\eta = \min(\epsilon/2, 1/M_1 - \epsilon)$, which completes the proof.

$\square$

Lemma 7.4. *Let $(\gamma_k)_{k\in\mathbb{N}}$ be a deterministic sequence of positive reals, let $\mathbb{G} := (\mathcal{G}_n)_{n\in\mathbb{N}}$ be a filtration, and let $(M_n)_{n\in\mathbb{N}}$ be a $\mathbb{G}$-adapted martingale such that $|M_{n+1} - M_n| \leqslant \gamma_n$ for all $n \in \mathbb{N}$. Then, for all $n \in \mathbb{N}$,*

$$\mathbb{P}\left(\sup_{k\geqslant n}(M_k - M_n)\geqslant \lambda \,|\,\mathcal{G}_n\right) \leqslant \exp\left(-\frac{\lambda^2}{2\sum_{k\geqslant n}\gamma_k^2}\right).$$

Proof. Let, for all $\theta \in \mathbb{R}$,

$$Z_n(\theta) := \exp\left(\theta M_n - \frac{\theta^2}{2}\sum_{k=1}^{n}\gamma_k^2\right).$$

Then $(Z_n(\theta))_{n\in\mathbb{N}}$ is supermartingale, so that

$$\mathbb{P}\left(\sup_{k\geqslant n}(M_k - M_n)\geqslant \lambda \,|\,\mathcal{G}_n\right)$$
$$\leqslant \mathbb{P}\left(\sup_{k\geqslant n} Z_k(\theta)\geqslant Z_n(\theta)\exp\left(\lambda\theta - \frac{\theta^2}{2}\sum_{k=1}^{n}\gamma_k^2\right)\,|\,\mathcal{G}_n\right)$$
$$\leqslant \exp\left(\frac{\theta^2}{2}\sum_{k=1}^{n}\gamma_k^2 - \lambda\theta\right) = \exp\left(-\frac{\lambda^2}{2\sum_{k\geqslant n}\gamma_k^2}\right)$$

if we choose $\theta := \lambda/\sum_{k\geqslant n}\gamma_k^2$.

$\square$

References

[1] R. Argiento, R. Pemantle, B. Skyrms, and S. Volkov. Learning to signal : Analysis of a micro-level reinforcement model. *Stochastic processes and their applications*, 119(2):373–390, 2009.

[2] A. W. Beggs. On the convergence of reinforcement learning. *Journal of Economic Theory*, 122:1–36, 2005.

[3] M. Benaïm. *Dynamics of stochastic approximation algorithms*. In Seminaires de Probabilités XXXIII, volume 1709 of Lecture notes in mathematics. Springer-Verlag, 1999.

[4] R. Durrett. *Probability: Theory and Examples*. Duxbury Press, Belmont, CA, Third Edition, 2004.

[5] I. Erev and A. E. Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review*, 88:848–881, 1998.

[6] D. Fudenberg and D. K. Levine. *The Theory of Learning in Games*. Cambridge: MIT Press, 1998.

[7] R. J. Herrnstein. On the law of effect. *Journal of the Experimental Analysis of Behavior*, 15:245–266, 1970.

[8] J. Hofbauer and S. Huttegger. Feasibility of communication in binary signaling games. *Journal of Theoretical Biology*, 254:842–849, 2008.

[9] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press, 1998.

[10] E. Hopkins and M. Posch. Attainability of boundary points under reinforcement learning. *Games and Economic Behavior*, 53:110–125, 2005.

[11] S. Huttegger. Evolution and the explanation of meaning. *Philosophy of Science*, 74:1–27, 2007.

[12] S. Huttegger and K. Zollman. Signaling games: dynamics of evolution and learning. *To appear in Language, Games, and Evolution*, 2010.

[13] D. Lewis. *Convention: A Philosophical Study*. Harvard: Harvard University Press, 1969.

[14] V. Losert and E. Akin. Dynamics of games and genes : Discrete versus continuous time. *J. Math. Biology*, 17:241–251, 1983.

[15] J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press, 1982.

[16] C. Pawlowitsch. Why evolution does not always lead to an optimal signaling system. *Games and Economic Behavior*, 63:203–226, 2008.

[17] R. Pemantle. *Random processes with reinforcement*. Massachussets Institute of Technology doctoral dissertation, 1988.

[18] R. Pemantle. Nonconvergence to unstable points in urn models and stochastic approximations. *Annals of Probability*, 18:698–712, 1990.

[19] A. E. Roth and I. Erev. Learning in extensive-form games: Experimental data and simple dynamics models in the intermediate term. *Games and Economic Behavior*, 8:164–212, 1995.

[20] B. Skyrms. *Signals: Evolution, Learning, and Information*. Oxford: Oxford University Press, 2010.

[21] P. Tarrès. Pièges répulsifs. *C.R.Acad.Sci.Paris Sér. I Math*, 330:125–130, 2000.

[22] P. Tarrès. Vertex-reinforced random walk on $\mathbb{Z}$ eventually gets stuck on five points. *Ann. Probab.*, 32(3B):2650–2701, 2004.

[23] P. D. Taylor and L. Jonker. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156, 1978.

[24] P. Trapa and M. Nowak. Nash equilibria for an evolutionary language game. *Journal of Mathematical Biology*, 41(2):172–188, 2000.

[25] E.C. Zeeman. Population dynamics from game theory. *Proc. Int. Conf. Global Theory of Dynamical System, Northwestern*, 1979.