

Path to Stochastic Stability: Comparative Analysis of Stochastic Learning Dynamics in Games

Hassan Jaleel and Jeff S. Shamma

Abstract—Stochastic stability is an important solution concept for stochastic learning dynamics in games. However, a limitation of this solution concept is its inability to distinguish between different learning rules that lead to the same steady-state behavior. We identify this limitation and develop a framework for the comparative analysis of the transient behavior of stochastic learning dynamics. We present the framework in the context of two learning dynamics: Log-Linear Learning (LLL) and Metropolis Learning (ML). Although both of these dynamics lead to the same steady-state behavior, they correspond to different behavioral models for decision making. In this work, we propose multiple criteria to analyze and quantify the differences in the short and medium-run behaviors of stochastic learning dynamics. We derive upper bounds on the expected hitting time of the set of Nash equilibria for both LLL and ML. For the medium to long-run behavior, we identify a set of tools from the theory of perturbed Markov chains that result in a hierarchical decomposition of the state space into collections of states called cycles. We compare LLL and ML based on the proposed criteria and develop invaluable insights into the behavior of the two dynamics.

I. INTRODUCTION

Stochastic learning dynamics, like log-linear learning, address the issue of equilibrium selection for a class of games that includes potential games (see, e.g., [1], [2], [3] and [4]). Because of the equilibrium selection property, these learning dynamics have received significant attention, particularly in the context of opinion dynamics in coordination games (see, e.g., [2] and [5]) and game theoretic approaches to the distributed control of multiagent systems [6].

A known issue of stochastic learning dynamics is the slow mixing of their induced Markov chains ([2], [7], and [8]). The mixing time of a Markov chain is the time required by the chain to approach its steady-state behavior. This mixing time is crucial because stochastic stability is defined in terms of the stationary distribution of the Markov chain induced by a learning dynamics. The implication of the slow mixing time is that the behavior of these dynamics in the short and medium run is equally important particularly for engineered systems with a limited lifetime. However, stochastic stability only deals with the steady-state behavior and provides no information about the transient behavior of these dynamics.

The speed of convergence of stochastic learning dynamics has been a focus of research for some time (see e.g., [9], [10],

[11], [12], [13], and [14]). However, there is another aspect related to the slow convergence of these dynamics that has received relatively little research attention. Stochastic stability only explains the steady-state behavior of a learning rule. We show that there are learning dynamics with fairly different update rules that lead to the same steady-state behavior. Since these learning dynamics have the same steady-state behavior, stochastic stability cannot distinguish between them. Nevertheless, the different update rules may result in significantly different behaviors over short and medium run that may be desirable or undesirable but remain entirely unnoticed.

We first establish the implications of having different learning rules with the same steady state. Through an example setup, we demonstrate the differences in short and medium run behaviors of two particular learning dynamics that result in the same steady-state behavior. The example setup is that of sensor coverage game in which we formulate a sensor coverage problem in the framework of a potential game. An important conclusion that we draw from this comparison is that for stochastic learning dynamics, characterization of stochastically stable states is not sufficient. It is also essential to analyze the paths that lead to these stochastically stable states from any given initial condition. Analysis of these paths is critical because they can play a crucial role not only in the short and medium run but also in the long run steady state behavior of the system.

The transient behavior of stochastic dynamics was studied in the context of learning in games in [15] and [16]. Yet the issues related to various learning dynamics leading to the same steady state behavior were not highlighted in these works. Therefore, after motivating the problem, we propose a novel framework for the comparative analysis of a class of stochastic learning dynamics. The proposed framework is based on the theory of Markov chains with rare transitions [17], [18], [19], [20], and [21].

In the proposed framework, we present multiple criteria for comparing the short, medium, and long-run behaviors of a system for different learning dynamics. We refer to the analysis related to the short-run behavior as first order analysis. The first order analysis deals with the expected hitting time of the set of pure Nash equilibria, which is the expected time to reach a Nash equilibrium (NE) for the first time. Because of the known hardness results of computing a NE ([22] and [23]), and the fact that all the Nash equilibria of a potential game are not necessarily potential maximizer, the first order analysis is typically not considered for stochastic learning dynamics. However, for comparative analysis, we show that first-order analysis provides valuable insights into system behavior.

We refer to the analysis related to the medium and long-

H. Jaleel is with the Department of Electrical Engineering, Syed Babar Ali School of Science & Engineering at Lahore Univeristy of Management Sciences (LUMS), Lahore, Pakistan. J.S. Shamma is with the Robotics, Intelligent Systems & Control (RISC) Lab, Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE) at King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. Email: hassan.jaleel@lums.edu.pk, jeff.shamma@kaust.edu.sa. Research supported by funding from KAUST.

run behavior of stochastic learning dynamics as higher-order analysis. For the higher-order analysis, we exploit the fact that the Markov chains induced by stochastic learning dynamics explore the space of joint action profiles hierarchically. This hierarchical exploration of the state space is well explained by an iterative decomposition of the state space into distinct subsets called cycles as was shown in [17], [19], [18], and [24]. Thus, the evolution of Markov chains with rare transitions can be well approximated by transitions among cycles.

The higher order analysis is based on the cycle decomposition of the state space as presented in [25]. We compare the behavior of different learning rules by comparing the exit height (H_e), and the mixing height (H_m) of the cycles generated by the cycle decomposition algorithm applied to these learning rule. The significance of these parameters is that once a Markov chain enters a cycle, the time to exit the cycle is of the order of $e^{\frac{H_e}{T}}$ and the time to visit each state within a cycle before exiting is of the order of $e^{\frac{H_m}{T}}$. Thus, we can efficiently characterize the behavior of each cycle from H_e and H_m .

Our comparative analysis framework applies to the class of learning dynamics in which the induced Markov chains satisfy certain regularity conditions. However, we present the details of the framework in the context of two particular learning dynamics, Log-Linear Learning (LLL) and Metropolis Learning (ML) over potential games. Log-Linear learning is a noisy best response dynamics and is well-known in the game theory literature. The stationary distribution of the induced Markov chain is a Gibbs distribution, which assigns maximum probability to the action profiles that maximize a potential function defined over the state space. It was shown in [26] that LLL is a good behavioral model for decision making when the players have sufficient information and resources to compute their utilities for all the actions in their action set given the actions of other players in the game.

On the other hand, Metropolis learning is a noisy better response dynamics for which the induced Markov chain is a Metropolis chain. It is well established in the statistical mechanics literature that the unique stationary distribution of Metropolis chain is the Gibbs distribution (see, e.g., [18] and [20]). As a behavioral model for decision making, ML is related closely to the pairwise comparison dynamics presented in [27]. Thus, ML is a behavioral model for decision making with low information demand. A player only needs to compare its current payoff with the payoff of a randomly selected action. It does not need to compute the payoffs for all the actions as in LLL.

Hence, we have two learning dynamics, LLL and ML, which correspond to two behavioral models for decision making with different information requirements. However, both of the learning rules lead to the same steady-state behavior. We compare these learning dynamics based on the proposed framework. *The crux of our comparative analysis is that the availability of more information in the case of LLL as compared to ML does not guarantee better performance when the performance criterion is to reach the potential maximizer quickly.*

II. BACKGROUND

A. Preliminaries

We denote the cardinality of a set S by $|S|$. For a vector $x \in \mathbb{R}^n$, x_i is its i^{th} entry, $|x|$ is its Euclidean norm, and x' is its transpose. The Hamming distance between any two vectors x and y in \mathbb{R}^n is

$$d_H(x, y) = |\{i : x_i \neq y_i\}|. \quad (1)$$

$\Delta(n)$ denotes the n -dimensional probability simplex, i.e.,

$$\Delta(n) = \{\mu \in \mathbb{R}^n : \mathbf{1}'\mu = 1, \mu_i \geq 0\},$$

where $\mathbf{1} = (1, 1, \dots, 1)$ is a column vector of appropriate dimension with all the entries equal to 1. For a matrix P , we represent its t^{th} power as $(P)^t$.

B. Markov Chains

Let P be a transition matrix for the Markov chain X and $P(x, y)$ be the transition probability from state x to y . Let $P_x(E)$ be the probability of an event E when the Markov chain is initialized at state x . If a Markov chain is ergodic and reversible, then it has a unique stationary distribution, i.e.,

$$\lim_{t \rightarrow \infty} \mu_0'(P)^t = \pi'$$

for all initial distributions $\mu_0 \in \Delta(n)$. The following definitions are adapted from [2].

Definition 2.1: Let P^0 be a transition matrix for a Markov chain over state space S . Let P^ϵ be a family of perturbed Markov chains on S . We say that P^ϵ is a regular perturbation of P^0 if

- 1) P^ϵ is ergodic for sufficiently small ϵ ,
- 2) $\lim_{\epsilon \rightarrow 0} P^\epsilon(x, y) = P^0(x, y)$, and
- 3) $P^\epsilon(x, y) > 0$ for some $\epsilon > 0$ implies that there exists some function $R(x, y) \geq 0$ such that

$$0 < \lim_{\epsilon \rightarrow 0} \frac{P^\epsilon(x, y)}{\epsilon^{R(x, y)}} < \infty,$$

where $R(x, y) \geq 0$ is the cost of transition from x to y and is normally referred to as resistance.

The Markov process corresponding to P^ϵ is called a regularly perturbed Markov process.

Definition 2.2: Let P^ϵ be any regular perturbation of P^0 with stationary distribution π^ϵ . A state $x \in S$ is a stochastically stable state if and only if

$$\lim_{\epsilon \rightarrow 0} \pi^\epsilon(x) > 0.$$

Thus, any state that is not stochastically stable will have a vanishingly small probability of occurrence in the steady state as $\epsilon \rightarrow 0$.

Given any two states x and y in S , y is *reachable* from x ($x \rightarrow y$) if $(P)^t(x, y) > 0$ for some $t \geq 0$. The neighborhood of x is

$$N(x) = \{y \in S \setminus x : P(x, y) > 0\}.$$

A path $\omega_{x,y}^S$ between any two states x and y in S is a sequence of distinct states $(\omega_0, \omega_1, \dots, \omega_p)$ in S such that $\omega_0 = x$, $\omega_p = y$, $P(\omega_i, \omega_{i+1}) > 0$ and $\omega_i \neq \omega_j$ for all i and j in

$\{0, 1, \dots, p-1\}$. The length of the path is denoted as $|\omega_{x,y}^S|$. The superscript S will be ignored in path notation when the state space is clear from the context. Given a subset A of S and a path $\omega_{x,y}^S$, we say that $\omega_{x,y}^S \in A$ if $\omega_i \in A$ for all $i \in \{0, 1, 2, \dots, p\}$. If there exists an i such that $\omega_i \notin A$, then we say that the path $\omega \notin A$. We define $\Omega^S(x, y)$ as the set of all paths between states x and y in S . States x and y communicate with each other ($x \leftrightarrow y$) if the sets $\Omega^S(x, y)$ and $\Omega^S(y, x)$ are not empty.

Definition 2.3: A set $A \subseteq S$ is connected if $x \leftrightarrow y$ for every $x, y \in A$.

Definition 2.4: The hitting time of $x \in S$ is the first time it is visited, i.e.,

$$\tau_x = \min\{t \geq 0 : X_t = x\}.$$

The hitting time of a subset $A \subset S$ is the first time a state in A is visited, i.e.,

$$\tau_A = \min_{x \in A} \tau_x.$$

Definition 2.5: The exit time of a Markov chain from $A \subset S$ is $\tau_{\partial A}$, where

$$\partial A = \{y \in A^c : P(x, y) > 0 \text{ for some } x \in A\}.$$

Here, $A^c = S \setminus A$, and ∂A will be referred to as the boundary of A . In the above definition it is assumed that $X_0 \in A$.

C. Game Theory

Let $N_p = \{1, 2, \dots, n\}$ be a set of n strategic players. Each player i has a finite set of actions $A_i = \{1, 2, \dots, m_i\}$, and the set $\mathcal{A} = A_1 \times A_2 \times \dots \times A_n$ is the set of joint action profiles. The utility of a player $i \in N_p$ for an action profile $\mathbf{a} \in \mathcal{A}$ is defined by a utility function $U_i : \mathcal{A} \rightarrow \mathbb{R}$. The combination of the action of the i^{th} player and the actions of everyone else is represented by (a_i, \mathbf{a}_{-i}) . The joint action profiles of all the players except i are represented by the set

$$\mathcal{A}_{-i} = A_1 \times A_2 \times \dots \times A_{i-1} \times A_{i+1} \times \dots \times A_n.$$

Player i prefers action profile \mathbf{a} over \mathbf{a}' if and only if $U_i(\mathbf{a}) > U_i(\mathbf{a}')$. If $U_i(\mathbf{a}) = U_i(\mathbf{a}')$, then it is indifferent. An action profile $\mathbf{a}^* \in \mathcal{A}$ is a Nash Equilibrium (NE) if

$$U_i(\mathbf{a}^*) = U_i(a_i^*, \mathbf{a}_{-i}^*) \geq U_i(a_i, \mathbf{a}_{-i}^*)$$

for all $i \in N_p$ and $a_i \in A_i$. The best response set of player i given $\mathbf{a}_{-i} \in \mathcal{A}_{-i}$ is

$$B_i(\mathbf{a}_{-i}) = \{\alpha^* \in A_i : \alpha^* = \operatorname{argmax}_{\bar{\alpha} \in A_i} U_i(\bar{\alpha}, \mathbf{a}_{-i})\}.$$

The neighborhood of an action profile \mathbf{a} is

$$N(\mathbf{a}) = \{\mathbf{a}' \in \mathcal{A} : d_H(\mathbf{a}, \mathbf{a}') = 1\}.$$

The player specific neighborhood set of action profile $\mathbf{a} = (\alpha, \mathbf{a}_{-i})$ is

$$N(\mathbf{a}, i) = \{\mathbf{a}' \in \mathcal{A} : \alpha' \in A_i \setminus \alpha, \mathbf{a}'_{-i} = \mathbf{a}_{-i}\}.$$

Potential Game: A game is a potential game if there exists a real-valued function $\phi : \mathcal{A} \rightarrow \mathbb{R}$, called the potential function, such that

$$U_i(\alpha, \mathbf{a}_{-i}) - U_i(\alpha', \mathbf{a}_{-i}) = \phi(\alpha, \mathbf{a}_{-i}) - \phi(\alpha', \mathbf{a}_{-i})$$

for all $i \in N_p$ and for all $\alpha, \alpha' \in A_i$.

III. STOCHASTIC LEARNING DYNAMICS

Stochastic learning dynamics is a class of learning dynamics in games in which the players are modeled to play best/better reply to the actions of other players with high probability. However, with a low probability, the players are modeled to play noisy actions for exploration. The inclusion of noisy actions in players decision model leads to equilibrium selection property for a class of games like potential games. The level of noise depends on a parameter T , which is typically referred to as temperature.

A. Log-Linear Learning (LLL)

Let $\mathbf{a} = (\alpha, \mathbf{a}_{-i})$ be the joint action profile representing the current state of the game. Then, the steps involved in LLL are as follows:

- 1) Activate one of the n players, say player i , uniformly at random.
- 2) All the other players repeat their previous actions.
- 3) Player i selects an action $\alpha' \in A_i$ with probability

$$p_i^{\text{LLL}}(\alpha', \mathbf{a}_{-i}) = \frac{e^{-\frac{1}{T}(U_i(\alpha^*, \mathbf{a}_{-i}) - U_i(\alpha', \mathbf{a}_{-i}))}}{Z_i(\mathbf{a}_{-i})}, \quad (2)$$

$$Z_i(\mathbf{a}_{-i}) = \sum_{\bar{\alpha} \in A_i} e^{-\frac{1}{T}(U_i(\alpha^*, \mathbf{a}_{-i}) - U_i(\bar{\alpha}, \mathbf{a}_{-i}))}.$$

Here $Z_i(\mathbf{a}_{-i})$ is a normalizing constant and $\alpha^* \in B_i(\mathbf{a}_{-i})$, which is the best response set of player i given \mathbf{a}_{-i} . Moreover,

$$\lim_{T \rightarrow 0} Z_i(\mathbf{a}_{-i}) = |B_i(\mathbf{a}_{-i})|.$$

The value of T determines the amount of noise in decision making. For $T = \infty$, the players update their strategies uniformly at random. However, as $T \rightarrow 0$, the probability of the actions yielding higher utilities increases. Thus, LLL induces a Markov chain X^{LLL} over the set of joint action profiles \mathcal{A} with transition matrix P^{LLL} . Note here that the dependence of X^{LLL} and P^{LLL} on T is suppressed for simplifying the notation. The transition probability between any two distinct action profiles \mathbf{a} and \mathbf{a}' is

$$P^{\text{LLL}}(\mathbf{a}, \mathbf{a}') = \frac{1}{n} \begin{cases} p_i^{\text{LLL}}(\alpha', \mathbf{a}_{-i}) & \mathbf{a}' \in N(\mathbf{a}, i), \\ 0 & d_H(\mathbf{a}, \mathbf{a}') > 1. \end{cases}$$

The above expression is valid for all $i \in N_p$. It was shown in [28] that P^{LLL} is a regular perturbation of P^0 with $\epsilon = e^{-1/T}$. Here, P^0 is the transition matrix of the Markov chain induced by asynchronous best response dynamics. Furthermore, it was proved in [4] that in an n -player potential game with a potential function ϕ , if all the agents update their actions using LLL, then the only stochastically stable states are the potential maximizers. The stationary distribution for X^{LLL} is the Gibbs distribution, which is as follows:

$$\pi^{\text{LLL}}(\mathbf{a}) = \frac{1}{Z_G} e^{\frac{1}{T}\phi(\mathbf{a})}, \quad (3)$$

where $Z_G = \sum_{\mathbf{y} \in \mathcal{A}} e^{\frac{1}{T}\phi(\mathbf{y})}$ is a normalizing constant.

B. Metropolis Learning (ML)

We introduce another learning dynamics, which has the same stationary distribution as LLL in (3). We refer to this learning dynamics as Metropolis Learning (ML) because the Markov chain induced by ML is a Metropolis chain, which is well studied in statistical mechanics and in simulated annealing [20]. The steps involved in ML are as follows:

- 1) Activate one of the n players, say player i , uniformly at random.
- 2) All the other players repeat their previous actions.
- 3) Player i selects an action $\alpha' \in A_i$ uniformly at random.
- 4) Player i switches its action from α to α' with probability

$$\min\{1, e^{-\frac{1}{T}(U_i(\alpha, \mathbf{a}_{-i}) - U_i(\alpha', \mathbf{a}_{-i}))}\}.$$

Thus, the probability of transition from $\mathbf{a} = (\alpha, \mathbf{a}_{-i})$ to $\mathbf{a}' = (\alpha', \mathbf{a}_{-i})$ given that player i has been selected is

$$p_i^{\text{ML}}(\mathbf{a}, \mathbf{a}') = \frac{1}{|A_i|} e^{-\frac{1}{T}[U_i(\alpha, \mathbf{a}_{-i}) - U_i(\alpha', \mathbf{a}_{-i})]^+}, \quad (4)$$

where $[c]^+ = c$ if $c > 0$ and is equal to zero otherwise.

In ML, player i switches to a randomly selected action $\alpha' \in A_i$ with probability one as long as $U_i(\alpha', \mathbf{a}_{-i}) \geq U_i(\alpha, \mathbf{a}_{-i})$. Here, α is the action that player i was playing in the previous time slot. Thus, unlike LLL in which a player needs to compute the utilities for all the actions in its action set, the update in ML only requires a player to make a pairwise comparison between a randomly selected action and its previous action. Furthermore, the probability of a noisy action is a function of loss in payoff compared to the previous action.

Metropolis learning generates a Markov Chain X^{ML} over joint action profile \mathcal{A} with transition matrix P^{ML} . The transition probability between any two distinct action profiles \mathbf{a} and \mathbf{a}' is

$$P^{\text{ML}}(\mathbf{a}, \mathbf{a}') = \frac{1}{n} \begin{cases} p_i^{\text{ML}}(\mathbf{a}, \mathbf{a}') & \mathbf{a}' \in N(\mathbf{a}, i), \\ 0 & d_H(\mathbf{a}, \mathbf{a}') > 1. \end{cases}$$

Lemma 3.1: *Transition matrix P^{ML} is a regular perturbation of P^{br} , where P^{br} is the transition matrix for asynchronous better reply dynamics. Moreover, the resistance of any feasible transition, say from $\mathbf{a} = (\alpha, \mathbf{a}_{-i})$ to $\mathbf{a}' = (\alpha', \mathbf{a}_{-i})$, is*

$$R(\mathbf{a}, \mathbf{a}') = [U_i(\alpha, \mathbf{a}_{-i}) - U_i(\alpha', \mathbf{a}_{-i})]^+. \quad (5)$$

for all $\alpha \in A_i$ and for all $i \in N_p$.

Proof: To prove that P^{ML} is a regular perturbation, we first describe the unperturbed process which is asynchronous better reply dynamics. The unperturbed process has the following dynamics.

- 1) A player, say i , is selected at random.
- 2) All the other players repeat their previous actions.
- 3) Player i selects an action α' uniformly at random.
- 4) Player i switches its action from α to α' if

$$U_i(\alpha', \mathbf{a}_{-i}) \geq U_i(\alpha, \mathbf{a}_{-i}).$$

Otherwise, it repeats α . Thus, the transition probability between two distinct action profiles is

$$P^{\text{br}}(\mathbf{a}, \mathbf{a}') = \begin{cases} \frac{1}{n|A_i|} & U_i(\mathbf{a}') \geq U_i(\mathbf{a}), \\ 0 & \text{otherwise.} \end{cases}$$

Similar to LLL, the noise parameter $\epsilon = e^{-\frac{1}{T}}$. The Metropolis chain X^{ML} is ergodic for $\epsilon > 0$ and it satisfies $\lim_{\epsilon \rightarrow 0} P^{\text{ML}}(x, y) = P^{\text{br}}(x, y)$. For the final condition

$$\lim_{\epsilon \rightarrow 0} \frac{P^{\text{ML}}(\mathbf{a}, \mathbf{a}')}{\epsilon^{[U_i(\alpha, \mathbf{a}) - U_i(\alpha', \mathbf{a}_{-i})]^+}} = \frac{1}{n|A_i|} \in (0, \infty),$$

where $0 \leq R(\mathbf{a}, \mathbf{a}') < \infty$ for any given pair \mathbf{a} and \mathbf{a}' . Thus, P^{ML} is a regular perturbation of P^{br} . ■

An important fact regarding ML in the context of this work is that the stationary distribution π^{ML} is also the Gibbs distribution [18], i.e.,

$$\pi^{\text{ML}}(\mathbf{a}) = \pi^{\text{LLL}}(\mathbf{a}) = \frac{1}{Z_G} e^{\frac{1}{T}\phi(\mathbf{a})}. \quad (6)$$

Thus, from stochastic stability perspective, both LLL and ML are the same.

IV. MOTIVATION: SENSOR COVERAGE PROBLEM

To study the difference in behaviors between LLL and ML, which is ignored in stochastic stability analysis, we set up sensor coverage problem as a potential game. Through extensive simulations under various noise conditions, we exhibit the essential differences between the behavior of these learning dynamics in the short and medium runs.

A. Coverage Game Setup

Consider a scenario in which n sensors are deployed randomly to monitor an environment $\mathcal{D} \subset \mathbb{R}^2$ for a long period of time. We approximate \mathcal{D} with a square region defined over intervals $[0, d] \times [0, d]$. To simplify the problem, the area is discretized as a 2-dimensional grid represented by the Cartesian product $\{0, 1, \dots, d\} \times \{0, 1, \dots, d\}$. The location of each sensor $i \in \{1, 2, \dots, n\}$ is $x_i \in \mathbb{R}^2$ where $x_i \sim \text{unif}(\mathcal{D})$, i.e., x_i is a random variable uniformly distributed over the region of interest \mathcal{D} . The footprint of sensor i is a circular disk of radius r , i.e.,

$$F(x_i, r) = \{z \in \mathbb{R}^2 : \|z - x_i\|^2 \leq r\}.$$

We assume that each sensor can choose the radius of its footprint from a finite set, which determines its energy consumption. Let r_c be the communication range of each sensor. We assume that $r_c \geq 2r_{\max}$, where r_{\max} is the maximum sensing radius.

We propose a game-theoretic solution to the sensor coverage problem in which we formulate the problem as a strategic game and implement some local learning rule so that each sensor can learn its schedule based on local information only. The players in this game are the sensors and each player has m_i actions, i.e., $N_p = \{1, 2, \dots, n\}$ and $A_i = \{r_0, r_1, \dots, r_{m_i}\}$. Here, an action of a player is its sensing radius. For each sensor, $r_0 = 0$, which is the off state of a sensor. The joint action profile is the joint state of all the sensors.

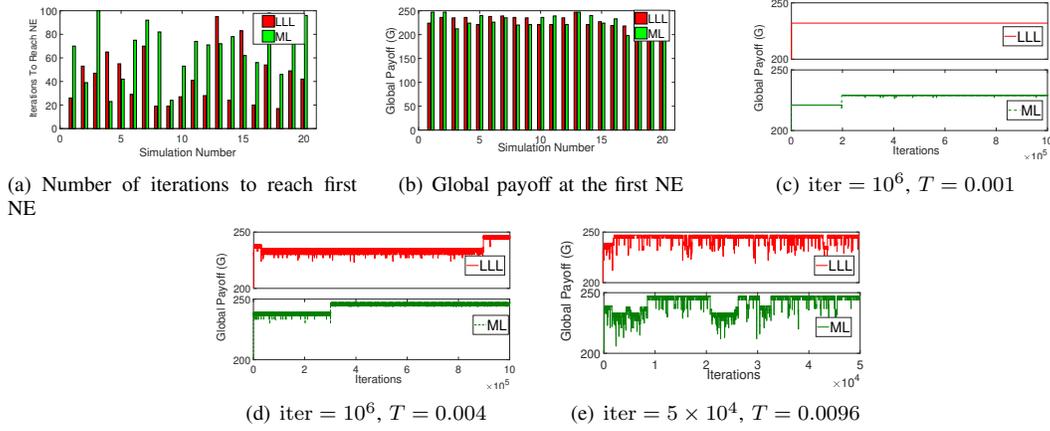


Fig. 1. System performance for LLL and ML with different noise conditions.

Let $p_{kl} = (k, l)$ be a point on the grid where k and l belong to $\{0, 1, \dots, d\}$. Each point in the grid has a binary state and the state is equal to 1 if it is covered by at least one sensor in the on state. Thus, the state of a grid point is whether it is covered or uncovered, i.e.,

$$c(p_{kl}) = \begin{cases} 1 & \exists i \in N_p : x_i \in F(p_{kl}, a_i) \text{ and } a_i \neq r_0, \\ 0 & \text{otherwise.} \end{cases}$$

The objective is to solve the following optimization problem:

$$\begin{aligned} \max_{a \in \mathcal{A}} G(a) &= \max_{a \in \mathcal{A}} (U(a) - C(a)), \text{ where} \\ U(a) &= \sum_{k=0}^d \sum_{l=0}^d c(p_{kl}) \text{ and } C(a) = \sum_{i=1}^n C_i(a_i). \end{aligned}$$

Here $U(a)$ is the total coverage achieved by the network, $C_i(a_i)$ is the cost incurred by sensor i in the on state, and $C(a)$ is the total cost of the system.

For simplicity, we assume that $C_i(r_0) = 0$, i.e., no cost is incurred by the sensors that are off.

The local utility of each player is computed through marginal contribution utility, as explained in [29], with base action $a_i^b = r_0$, i.e., the base action of each sensor is to be in the off state in which there is no energy consumption. If a_{-i} is the joint state of all the other sensors, then the utility of player i for action a_i is

$$U_i(a_i, a_{-i}) = G(a_i, a_{-i}) - G(a_i^b, a_{-i}).$$

The above equation implies that $U_i(a_i^b, a_{-i}) = 0$. Moreover, for any $k \in \{1, 2, \dots, m_i\}$, we have

$$U_i(r_k, a_{-i}) = [U(r_k, a_{-i}) - U(r_0, a_{-i})] - C_i(a_i).$$

Thus, the marginal contribution utility of sensor i with action a_i is the number of grid points that are covered by the sensor exclusively with footprint of radius a_i minus the cost $C_i(a_i)$.

To make the payoff and the cost terms in U_i compatible, we express the cost of turning a sensor on as a function of the minimum number of grid points that a sensor should cover exclusively. Let $R_{\max}(r)$ be the maximum number of grid

points that a sensor can cover if its footprint has radius r . Then, we define the cost as follows:

$$C_i(a_i) = (\beta R_{\max}(a_i)) \text{ for } \beta \in (0, 1].$$

Thus, the net utility of a sensor is negative if the number of points it covers exclusively is less than $C_i(a_i)$ given a_{-i} .

B. Simulation Results

We simulated the sensor coverage game with $d = 20$, $n = 15$, $\beta = 0.2$, and $A_i = \{0, 15\}$ for all $i \in \{1, 2, \dots, 15\}$. Thus, each sensor could be either in the off state or in the on state with sensing radius of 15. For this setup, the maximum global utility was

$$\max_{a \in \mathcal{A}} G(a) = 247,$$

which was computed numerically based on extensive simulations. To achieve the maximum payoff, we implemented LLL and ML with different values for the noise parameter T and the number of iterations iter . The results of the simulation are presented in Fig. 1.

Initially, all the sensors were in the off state. To compare the short-term behavior of the network with small noise, we set $T = 0.001$ and ran the simulation for twenty times for both LLL and ML with $\text{iter} = 100$. The results of twenty simulations are presented in Fig. 1. In Fig. 1(a), we show the number of iterations to reach a NE for the first time for LLL and ML. Based on the results in Fig. 1(a), the average number of iterations to reach a NE for the first time for LLL and ML were 43.15 and 63.75 respectively. Thus, on average, the system was driven to a NE faster by LLL than ML.

For a system with multiple Nash equilibria, reaching a NE faster is not the only objective. The quality of the NE is also a significant factor. In Fig. 1(b), we present the global payoff at the Nash equilibria reached under LLL and ML in our twenty simulations. The global payoffs at the Nash equilibria under LLL and ML had a mean value of 229.6 and 230.1, and a standard deviation of 8.39 and 12.49 respectively. Although the average global payoffs were almost equal, the higher standard deviation under ML indicates that ML has a higher tendency to explore the state space as compared to LLL.

From the comparisons in Figs. 1(a) and 1(b), LLL seems to be better than ML because it can lead to a NE faster on average. However, ML seems to have a slight edge over LLL if we consider the quality of the Nash equilibria. This observation provides a strong rationale for a comprehensive comparative analysis because we cannot simply declare one learning rule better than the other.

For higher order analysis, the objective was to observe and compare system behavior over an extended period. For $T = 0.001$, optimal configuration could not be achieved under both LLL and ML even in 10^6 iterations. For LLL, the network remained stuck at some suboptimal NE for 10^6 iterations. In ML, there was a single switch in network configuration after approximately 2×10^5 from one NE to another. As we increased the noise to $T = 0.004$, payoff maximizing configurations were reached for both LLL and ML. However, the number of iterations to reach these optimal configurations were huge, particularly in LLL. For $T = 0.0096$, the optimal configurations were reached rapidly.

The ability of ML to stay at an optimal configuration after reaching it is affected more by noise as compared to LLL. In Fig. 1(c) with $T = 0.001$, the network configuration switched from one NE to another under ML, but there was no switch under LLL. In Fig. 1(d) with $T = 0.004$, the network configuration switched to an optimal NE quickly under ML then under LLL. Finally, increasing the noise to $T = 0.0096$ led to an interesting behavior that can be observed in Fig. 1(e). Under ML, the network configuration kept on leaving the payoff-maximizing configurations periodically for significant duration of times. However, under LLL, after reaching an optimal configuration, the network never left the configuration for long durations of time.

From the above observations, we can conclude that the concept of stochastic stability alone is not sufficient to describe the behavior of stochastic learning dynamics. However, these observations are based on the simulation of a particular system under certain conditions, which prohibits us from drawing any general conclusions regarding the behavior of these learning rules. Therefore, we present a general framework to analyze and compare the behavior of different learning rules that have the same stochastically stable states.

V. CYCLE DECOMPOSITION

Consider a family of Markov chains X^T on a finite state space $S = \{1, 2, \dots, n\}$ with transition matrices P^T . We will also refer to any member of this family of Markov chains as an (X^T, P^T) pair. We assume that the transition matrices satisfy the following property.

$$\Gamma(T)e^{-\frac{1}{T}V(x,y)} \leq P^T(x,y) \leq \frac{1}{\Gamma(T)}e^{-\frac{1}{T}V(x,y)}, \quad (7)$$

where $\Gamma(T) > 0$ for $T > 0$ and

$$\lim_{T \rightarrow 0} T \ln \Gamma(T) = 0. \quad (8)$$

Here $V : S \times S \rightarrow \mathbb{R}_+ \cup \infty$ satisfies

$$\begin{cases} 0 \leq V(x,y) < \infty & P^T(x,y) > 0, \\ V(x,y) = \infty & P^T(x,y) = 0. \end{cases} \quad (9)$$

For any (x,y) pair, $V(x,y)$ can be considered as a transition cost from x to y . To better understand the conditions on transition probabilities, we rewrite (7) as

$$\ln \Gamma(T) \leq \ln \left(\frac{P^T(x,y)}{e^{-\frac{1}{T}V(x,y)}} \right) \leq -\ln \Gamma(T).$$

Then, in the limit as T approaches to zero, we use (8) to get

$$\lim_{T \rightarrow 0} T \ln P^T(x,y) = -V(x,y).$$

The above expression implies that in the limit as $T \rightarrow 0$, the natural logarithm of the transition probability between any two states is inversely related to their transition cost. Thus, as $V(x,y)$ increases, the transition probability $P^T(x,y)$ decreases exponentially.

It is assumed the function V is irreducible, which implies that for any state pair (x,y) , there exists a path $\omega_{x,y}^S$ such that

$$V(\omega_{x,y}^S) = \sum_{i=0}^{p-1} V(\omega_i, \omega_{i+1}) < \infty.$$

Definition 5.1: A function $V : S \times S \rightarrow \mathbb{R}_+ \cup \infty$ is induced by a potential function $\phi : S \rightarrow \mathbb{R}$, if the following weak reversibility condition is satisfied for all x and y in S .

$$\phi(x) - V(x,y) = \phi(y) - V(y,x). \quad (10)$$

The following result is from [30] (Prop. 4.1).

Proposition 5.1: Let (X^T, P^T) be a family of Markov chains over state space S such that the transition matrices P^T satisfy (7) and (8). If the function V is induced by a potential ϕ as defined in Def. 5.1, then the stationary distribution π^T is such that

$$\lim_{T \rightarrow 0} T \ln \pi^T(x) = -\max_{y \in S} (\phi(y) - \phi(x)).$$

The result in the proposition implies that in the limit as $T \rightarrow 0$, the natural logarithm of the stationary distribution of x is related inversely to $(\phi(y^*) - \phi(x))$, where y^* is the potential maximizer. Thus, for $x = y^*$, $\pi^T(x) \rightarrow 1$ as $T \rightarrow 0$. For any other state, i.e., $x \neq y^*$, $\pi^T(x)$ decreases exponential at a rate of $(\phi(y^*) - \phi(x))$. Thus, as $T \rightarrow 0$, only the states maximizing the potential function will have a non-zero probability. Based on Prop. 5.1, there is an entire class of Markov chains that lead to potential maximizers.

A. Cycle Decomposition Algorithm

Cycle Decomposition Algorithm (CDA) was presented in [24], based on the ideas originally presented in [17]. The motivation behind this algorithm was to analyze the transient behavior of Markov chains that satisfy (7), (8), and (10) and lead to the stationary distribution defined in Prop. 5.1. In this algorithm, the state space is decomposed into unique cycles in an iterative procedure. The formal definition of cycle as presented in [21] and [30] is as follows:

Definition 5.2: A set $\Pi \subset S$ is a cycle if it is a singleton or it satisfies either of the two conditions.

- 1) For any x, y in Π such that $x \neq y$, we define a set $A = \Pi^c \cup \{y\}$. Then,

$$\lim_{T \rightarrow 0} -T \ln P_x(X_{\tau_A}^T \neq y) > 0,$$

where τ_A is the hitting time of A .

2) For any x, y in Π , $x \neq y$

$$\lim_{T \rightarrow 0} T \ln \mathbb{E}_x(N_\Pi(x, y)) > 0$$

where $N_\Pi(x, y)$ is the number of round trips between x and y performed by the chain X^T before leaving Π .

The first condition says that a subset Π is a cycle if starting from some $x \in \Pi$, the probability of leaving Π before visiting every state $y \in \Pi$ is exponentially small. This condition can also be represented as

$$P_x(X_{\tau(\Pi \cup \{y\})}^T = y) \rightarrow 1 \text{ as } T \rightarrow 0.$$

The second statement states that the expected number of times each $y \in \Pi$ is visited by X^T starting from any $x \in \Pi$ increases exponentially as T approaches to zero.

For the completeness of presentation, we reproduce CDA in Alg. 1. We explain the steps involved in CDA through an example setup presented in Fig. 2. To explain system behavior using CDA, we need the following definitions and results, which are mostly from [24]. The minimum cost of leaving a state x is its exit height $H_e(x)$, where

$$H_e(x) = \min_{\substack{y \in S \\ y \neq x}} V(x, y). \quad (14)$$

A cycle is non-trivial if it has a non-zero exit height. Thus, a singleton is non-trivial cycle if it is a local maxima. For any set of states x and y such that $P^T(x, y) > 0$, we define

$$V_*(x, y) = \begin{cases} V(x, y) - H_e(x) & x \neq y, \\ 0 & x = y. \end{cases}$$

i.e., $V_*(x, y)$ is the excess cost above the minimum transition cost from x . For a path $\omega := (\omega_0, \omega_1, \dots, \omega_k)$

$$V_*(w) = \sum_{i=0}^{k-1} V_*(\omega_i, \omega_{i+1}).$$

To illustrate the working of CDA, we consider a Markov chain over state space $S = \{a, b, c, d, e, f, g, h, i\}$. The possible transitions between the states and the energy/potential landscape over the entire state space are presented in Fig. 2(a). The potentials of the states are as follows:

$$\begin{aligned} \phi(a) &= 2, \phi(b) = 8, \phi(c) = 5, \phi(d) = 12, \phi(e) = 10, \\ \phi(f) &= 11, \phi(g) = 7, \phi(h) = 8, \phi(i) = 16. \end{aligned}$$

Assume that the cost of transition from x to y is

$$V(x, y) = \begin{cases} [\phi(x) - \phi(y)]^+ & P^T(x, y) > 0, \\ \infty & \text{otherwise.} \end{cases} \quad (15)$$

The transition cost is zero if $\phi(y) \geq \phi(x)$ and is equal to the difference in potential between states x and y otherwise. This is the transition cost for standard Metropolis chain in statistical mechanics.

We initialize Alg. 1 with E^0 and $V^0(x, y)$ as defined in (11). To create the first level, we first generate the graph $G(E^0, \mathcal{E}^0)$ as shown in Fig. 2(b) and compute transition costs $V^0(x, y) = [\phi(x) - \phi(y)]^+$ from Fig. 2(a). The edges in $G(E^0, \mathcal{E}^0)$ correspond to the transitions with finite cost. Then,

Algorithm 1 Cycle Decomposition

Require: Define level zero as

$$\begin{aligned} E^0 &:= \{\{x\} : x \in S\} \text{ and} \\ V^0(x, y) &= V(x, y). \end{aligned} \quad (11)$$

Here $V(x, y)$ is defined in (9).

Ensure: The k^{th} level has been constructed with the set E^k and transition costs $V^k(S_i^k, S_j^k)$ for all S_i^k and S_j^k in E^k .

1: **while** $E^k \neq S$ **do**

2: Form a graph $G(E^k, \mathcal{E}^k)$ such that each $S_i^k \in E^k$ is a vertex in G and

$$(S_i^k, S_j^k) \in \mathcal{E}^k \text{ iff } V^k(S_i^k, S_j^k) < \infty.$$

3: For all $S_i^k \in E^k$, compute the minimum exit cost.

$$H_e^k(S_i^k) = \min\{V^k(S_i^k, S_j^k) \forall S_j^k \in E^k, S_j^k \neq S_i^k\}.$$

4: For every S_i^k and $S_j^k \in E^k$, compute

$$V_*^k(S_i^k, S_j^k) = V^k(S_i^k, S_j^k) - H_e^k(S_i^k).$$

5: Construct a graph $G(E^k, \mathcal{E}_*^k)$ such that for each $S_i^k \in E^k$,

$$(S_i^k, S_j^k) \in \mathcal{E}_*^k \text{ iff } V_*^k(S_i^k, S_j^k) = 0.$$

The graph $G(E^k, \mathcal{E}_*^k)$ is a subgraph of $G(E^k, \mathcal{E}^k)$.

6: Compute the strongly connected components in $G(E^k, \mathcal{E}_*^k)$. A set $F_s^{k+1} \subset E^k$ is a strongly connected component of $G(E^k, \mathcal{E}_*^k)$ if for every S_i^k and S_j^k in F_s^{k+1} , there exists a path $\omega_{S_i^k, S_j^k}^{E^k} \in F_s^{k+1}$ such that $V_*^k(\omega_{S_i^k, S_j^k}^{E^k}) = 0$.

7: Let D^{k+1} be the set of strongly connected components in $G(E^k, \mathcal{E}_*^k)$. Define a minimum set D_m^{k+1} as follows:

$$D_m^{k+1} = \{F_s^{k+1} \in D^{k+1} : V_*^k(S_i^k, S_j^k) > 0 \text{ for all } S_i^k \in F_s^{k+1} \text{ and } S_j^k \in E^k \setminus F_s^{k+1}\}.$$

8: Construct the set E^{k+1}

$$E^{k+1} = D_m^{k+1} \cup \{S_i^k \in E^k : S_i^k \notin F_s^{k+1} \forall F_s^{k+1} \in D_m^{k+1}\}.$$

9: For each $S_i^{k+1} \in E^{k+1}$, define

$$H_m^{k+1}(S_i^{k+1}) = \max\{H_e^k(S_j^k) \forall S_j^k \in E^k, S_j^k \in S_i^{k+1}\}.$$

10: Compute the cost between the sets in E^{k+1} as

$$V^{k+1}(S_i^{k+1}, S_j^{k+1}) = H_m^{k+1}(S_i^{k+1}) + \min_{\substack{S_m^k \in S_i^{k+1} \\ S_n^k \in S_j^{k+1}}} V_*^k(S_m^k, S_n^k). \quad (12)$$

11: $k = k + 1$.

12: **end while**

$$C(S) = \bigcup_{l=0}^k E^l. \quad (13)$$

we compute the exit heights H_e^0 for each state in the graph,

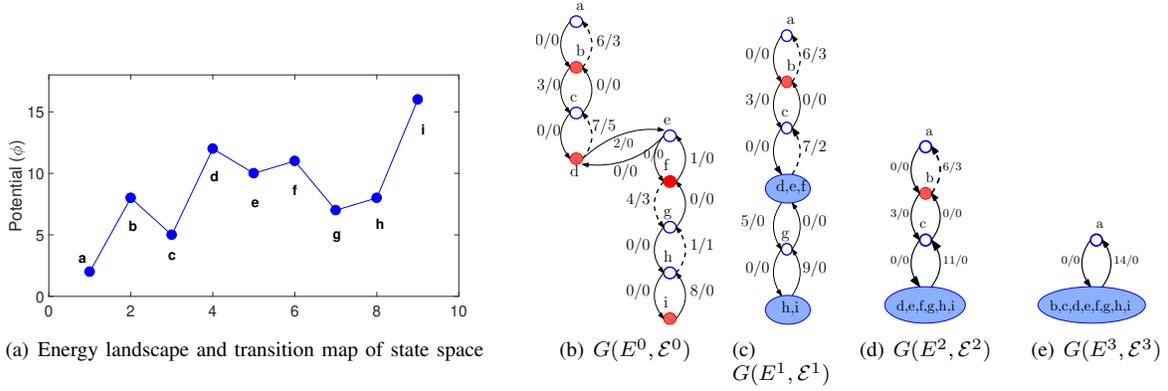


Fig. 2. Fig. 2(a) presents a Markov chain over state space $S = \{a, b, c, d, e, f, g, h, i\}$. Output of CDA for this Markov chain with transition cost in (15) is presented in Figs. 2(b)-2(e). Solid circles correspond to non-trivial cycles and solid edges represent transitions with minimum cost. Moreover, dotted edges represent transitions of higher cost from each state. The edges are labeled with $V^k(S_i^k, S_j^k)/V_*^k(S_i^k, S_j^k)$ for $k \in \{0, 1, 2, 3\}$.

where $H_e^0(x) = H_e(x)$ is defined in (14).

$$H_e^0(a) = 0, H_e^0(b) = 3, H_e^0(c) = 0, H_e^0(d) = 2, H_e^0(e) = 0 \\ H_e^0(f) = 1, H_e^0(g) = 0, H_e^0(h) = 0, H_e^0(i) = 8$$

The next step is to compute the excess communication cost $V_*^0(\cdot, \cdot)$ for each state transition, where $V_*^0(x, y) = V^0(x, y) - H_e^0(x)$. The edges in Fig. 2(b) are labeled with $V^0(x, y)/V_*^0(x, y)$. For instance, for state b , $V^0(b, a) = [\phi(b) - \phi(a)]^+ = 6$ and $V^0(b, c) = [\phi(b) - \phi(c)]^+ = 3$. Since $H_e^0(b) = 3$, the excess communication cost is

$$V_*^0(b, c) = V^0(b, c) - H_e^0(b) = 0, \text{ and} \\ V_*^0(b, a) = V^0(b, a) - H_e^0(b) = 3.$$

In step 5 of the algorithm, a subgraph $G(E^0, \mathcal{E}_*^0)$ of $G(E^0, \mathcal{E}^0)$ is formed in which we only retain the edges that correspond to minimum cost transitions. Thus, in Fig. 2(b), the dotted edges are ignored and the edge set \mathcal{E}_*^0 only contains the solid edges, which have zero excess cost. After forming $G(E^0, \mathcal{E}_*^0)$, we need to identify the strongly connected components in this graph. From Fig. 2(b), the set of strongly connected components is

$$D^1 = \{\{b, c\}, \{d, e, f\}, \{h, i\}\}.$$

In step 7, we form the minimum set D_m^1 that only contains the strongly connected sets from which there are no transitions outside the set with zero excess cost. For instance, for the set $\{b, c\}$, the excess cost of transition from c to d is zero. Therefore, $\{b, c\}$ does not belong to D_m^1 . However, the transitions leaving the set $\{d, e, f\}$ are from d to c and from f to g having excess costs of 5 and 3 respectively. Therefore, $\{d, e, f\}$ belongs to D_m^1 . Similarly, the only transition leaving the set $\{h, i\}$ is from h to g , which has an excess cost of 1. Therefore, $\{h, i\}$ also belongs to D_m^1 .

The next step is to create the state space for the first level, which for the above example is as follows:

$$E^1 = \{\{a\}, \{b\}, \{c\}, \{d, e, f\}, \{g\}, \{h, i\}\},$$

i.e., each strongly connected component in D_m^1 is a single state in E^1 . In step 9, we compute the mixing height H_m^1 of the merged states $\{d, e, f\}$ and $\{h, i\}$. The mixing height

of a merged state is the maximum exit height of the states comprising the merged state. In this example,

$$H_m^1(\{d, e, f\}) = \max\{H_e^0(d), H_e^0(e), H_e^0(f)\}, \\ = \max\{2, 0, 1\} = 2.$$

Similarly, $H_m^1(\{h, i\}) = 8$. For all the other elements in E^1 , $H_m^1(x) = H_e^0(x)$. Finally, we compute the transition cost from the merged states in step 10. Let $S_i^1 = \{d, e, f\}$. Then,

$$V^1(S_i^1, c) = H_m^1(S_i^1) + \min_{x \in S_i^1} V_*^0(x, c), \\ = 2 + \min\{V_*^0(d, c), V_*^0(e, c), V_*^0(f, c)\}, \\ = 2 + \min\{5, \infty, \infty\} = 7.$$

Similarly,

$$V^1(S_i^1, g) = 2 + \min\{V_*^0(d, g), V_*^0(e, g), V_*^0(f, g)\} = 5.$$

For the second merged state $S_j^1 = \{h, i\}$,

$$V^1(S_j^1, g) = H_m^1(S_j^1) + \min\{V_*^0(h, g), V_*^0(i, g)\}, \\ = 8 + 1 = 9.$$

With the computation of the updated transition costs, the first level of CDA is complete. By repeating the same process, we can compute all the levels in CDA as shown in Figs. 2(c)-2(e), till all the states are merged into a single state. The final outcome of CDA is the set $C(S)$ defined in (13). The order of the decomposition of the state space S is

$$n_S = \min\{k \in \mathbb{N} : E^{k+1} = S\}.$$

An increasing family of cycles is defined for each $x \in S$ as follows. Define $x^0 = x$. For each $1 \leq k \leq n_S$,

$$x^{k+1} \in E^{k+1} \text{ such that } x^k \subset x^{k+1}. \quad (16)$$

Given a set $A \subset S$ such that $|A| > 1$, the maximal proper partition $\mathcal{M}(A)$ is

$$\mathcal{M}(A) = \{\Pi \in C_A(S) : \Pi \not\subset \Pi' \text{ for any } \Pi' \in C_A(S)\},$$

where

$$C_A(S) = \{\Pi \in C(S) \mid \Pi \subset A, \Pi \neq A\}.$$

The partition $\mathcal{M}(A)$ is maximal in the sense that each $\Pi \in \mathcal{M}(A)$ is not a subset of any other cycle that is a proper subset of A .

For a cycle $\Pi \in C(S)$,

- order n_Π is

$$n_\Pi = \min\{0 < k < n_S : \Pi \in E^k\}.$$

- exit height $H_e(\Pi)$ is

$$H_e(\Pi) = \begin{cases} \max\{H_e^k(\Pi) : k \leq n_S, \Pi \in E^k\} & \Pi \neq S, \\ \infty & \Pi = S. \end{cases}$$

- mixing height $H_m(\Pi)$ is

$$H_m(\Pi) = \begin{cases} \max\{H_e(\Pi') : \Pi' \in \mathcal{M}(\Pi)\} & |\Pi| > 1, \\ 0 & |\Pi| = 1. \end{cases}$$

- potential $\phi(\Pi)$ is

$$\phi(\Pi) = \max\{\phi(x) : x \in \Pi\}.$$

- communication altitude between any x and y in S is

$$A_c(x, y) = \max_{\omega \in \Omega^S(x, y)} \min_{0 \leq k \leq |\omega|-1} \phi(\omega_k) - V(\omega_k, \omega_{k+1}).$$

- the communication altitude of a cycle $A_c(\Pi)$ is

$$A_c(\Pi) = \min_{x, y \in \Pi} A_c(x, y).$$

The exit and mixing heights of a cycle provide insights into the behavior of a Markov chain inside the cycle. The potential of a cycle is the maximum potential of a state within the cycle. The communication altitude was introduced in [24], and according to Eqs. (11) and (12) in [24], the following relations exist between $A_c(\Pi)$, $H_e(\Pi)$, $H_m(\pi)$ and $\phi(\Pi)$.

$$\begin{aligned} A_c(\Pi) &= \phi(\Pi') - H_e(\Pi'), \text{ and} \\ A_c(\Pi) &= \phi(\Pi) - H_m(\Pi) \end{aligned} \quad (17)$$

for all $\Pi' \in \mathcal{M}(\Pi)$. Since we are maximizing the potential function instead of minimizing as in [24], we have adjusted the above equations accordingly. Let x and y be any two states in S . Then, according to Prop. 2.16 in [24],

$$A_c(x, y) = A_c(y, x) = A_c(\Pi_{xy}), \quad (18)$$

where Π_{xy} is the smallest cycle containing both x and y . We want to note again that the definition of A_c and the corresponding relations are adjusted because we are maximizing a utility instead of minimizing a cost. However, all the results from [24] remain valid, and play an important role in the comparative analysis of LLL and ML.

The main result related to cycles that we will use is from [18] and [30], and is as follows:

Theorem 5.1: *Let $\Pi \in C(S)$. For any $\delta > 0$ and for any x and y in Π , we have*

$$P_x^T \left(e^{\frac{1}{T}(H_e(\Pi)-\delta)} < \tau_{\partial\Pi} < e^{\frac{1}{T}(H_e(\Pi)+\delta)} \right) = 1 - o(1) \quad (19)$$

$$\text{and } P_x^T \left(\tau_y < \tau_{\partial\Pi}, \tau_y < e^{\frac{1}{T}(H_m(\Pi)+\delta)} \right) = 1 - o(1) \quad (20)$$

as $T \rightarrow 0$. Here $\tau_{\partial\Pi}$ is the exit time of X^T from Π and τ_y is the hitting time of y .

Eq. (19) implies that the exit time of a Markov chain X^T from a cycle Π starting from any $x \in \Pi$ is of order $e^{\frac{1}{T}H_e(\Pi)}$, i.e., the exit time is proportional to the exponential of the exit height of Π . Moreover, Eq. (20) suggests that before leaving the cycle Π , X^T will visit all the states in Π within a time of order $e^{\frac{1}{T}H_m(\Pi)}$. Thus, through Eqs. (19) and (20), we can explain the behavior of a Markov chain from its cycles. In the next section, we apply cycle decomposition to ML and LLL and demonstrate the effectiveness of our proposed approach in explaining the behavior of the corresponding evolutionary process.

B. Cycle Decomposition For ML And LLL

Before we can perform comparative analysis using CDA, we need to establish that both ML and LLL satisfy the criteria for CDA. The state space for these learning rules is the set of joint action profiles $\mathcal{A} = A_1 \times A_2 \times \dots \times A_n$, where $A_i = \{1, 2, \dots, m_i\}$. We define

$$|A|_{\max} = \max_{i \in \{1, 2, \dots, n\}} |A_i| \text{ and } |A|_{\min} = \min_{i \in \{1, 2, \dots, n\}} |A_i|.$$

Proposition 5.2: P^{ML} and P^{LLL} satisfy the conditions in (7), (8), and (10) for a potential game setup.

Proof: To prove this result, we need to show that there exist cost and $\Gamma(T)$ functions for both ML and LLL that satisfy the three equations in the above statement. We begin with Metropolis learning. Given any pair of distinct action profiles \mathbf{a} and \mathbf{a}' in \mathcal{A} , we define

$$\begin{aligned} V^{\text{ML}}(\mathbf{a}, \mathbf{a}') &= \begin{cases} [U_i(\mathbf{a}) - U_i(\mathbf{a}')]^+ & a_i \neq a'_i, a_{-i} = a'_{-i}, \\ \infty & \text{otherwise,} \end{cases} \\ \text{and } \Gamma^{\text{ML}}(T) &= \frac{1}{n|A|_{\max}}, \end{aligned} \quad (21)$$

where $|A|_{\max}$ is the maximum number of actions of any player in the game. It is straightforward to verify that

$$\lim_{T \rightarrow 0} T \ln(\Gamma^{\text{ML}}(T)) = 0,$$

and V^{ML} and $\Gamma^{\text{ML}}(T)$ satisfy (7).

Next, we show that V^{ML} is induced by a potential function ϕ . Given an action profile pair $(\mathbf{a}, \mathbf{a}')$, two cases need to be considered. The first case is when $P^{\text{ML}}(\mathbf{a}, \mathbf{a}') = 0$, which implies $V^{\text{ML}}(\mathbf{a}, \mathbf{a}') = V^{\text{ML}}(\mathbf{a}', \mathbf{a}) = \infty$. Thus, both the left and right sides of (10) are equal to ∞ . The second case is when $P^{\text{ML}}(\mathbf{a}, \mathbf{a}') > 0$. In this case, the action profiles can be written as $\mathbf{a} = (\alpha, \mathbf{a}_{-i})$, $\mathbf{a}' = (\alpha', \mathbf{a}_{-i})$. Rearranging (10)

$$\phi(\mathbf{a}) - \phi(\mathbf{a}') = V^{\text{ML}}(\mathbf{a}, \mathbf{a}') - V^{\text{ML}}(\mathbf{a}', \mathbf{a}),$$

where

$$\begin{aligned} V^{\text{ML}}(\mathbf{a}, \mathbf{a}') - V^{\text{ML}}(\mathbf{a}', \mathbf{a}) &= [U_i(\alpha, \mathbf{a}_{-i}) - U_i(\alpha', \mathbf{a}_{-i})]^+ - \\ & \quad [U_i(\alpha', \mathbf{a}_{-i}) - U_i(\alpha, \mathbf{a}_{-i})]^+ \\ &= U_i(\alpha, \mathbf{a}_{-i}) - U_i(\alpha', \mathbf{a}_{-i}). \end{aligned}$$

For potential games

$$U_i(\alpha, \mathbf{a}_{-i}) - U_i(\alpha', \mathbf{a}_{-i}) = \phi(\alpha, \mathbf{a}_{-i}) - \phi(\alpha', \mathbf{a}_{-i}).$$

Thus, V^{ML} is induced by a potential function, which concludes the proof for ML.

For Log-Linear Learning, we define

$$V^{\text{LLL}}(\mathbf{a}, \mathbf{a}') = \begin{cases} U_i(\mathbf{a}^*) - U_i(\mathbf{a}') & a_i \neq a'_i, a_{-i} = a'_{-i}, \\ \infty & \text{otherwise,} \end{cases}$$

and $\Gamma^{\text{LLL}}(T) = \frac{1}{nZ_{\max}}$. (22)

Here $\mathbf{a}^* = (\alpha^*, a_{-i})$, where $\alpha^* \in B_i(a_{-i})$. Moreover,

$$Z_{\max} = \max_{i \in \{1, 2, \dots, n\}} Z_{i, \max}, \text{ and} \quad (23)$$

$$Z_{i, \max} = \max_{a_{-i} \in \mathcal{A}_{-i}} Z_i(a_{-i}),$$

where $Z_i(a_{-i})$ is defined in (2).

For any given action profile pair \mathbf{a} and \mathbf{a}' such that $\mathbf{a} \neq \mathbf{a}'$,

$$\lim_{T \rightarrow 0} T \ln(\Gamma^{\text{LLL}}(T)) = 0.$$

Moreover, V^{LLL} and $\Gamma^{\text{LLL}}(T)$ as defined above, satisfy (7). For the weak reversibility condition,

$$\begin{aligned} V^{\text{LLL}}(\mathbf{a}, \mathbf{a}') - V^{\text{LLL}}(\mathbf{a}', \mathbf{a}) &= U_i(\alpha^*, a_{-i}) - U_i(\alpha', a_{-i}) - \\ &\quad U_i(\alpha^*, a_{-i}) + U_i(\alpha, a_{-i}) \\ &= U_i(\alpha, a_{-i}) - U_i(\alpha', a_{-i}). \end{aligned}$$

By following the same series of arguments as for ML, we conclude that V^{LLL} is induced by a potential function ϕ , which concludes the proof. ■

Proposition 5.2 is restricted to LLL and ML. We rewrite (7) as

$$\Gamma(T) \leq \frac{P^T(x, y)}{e^{-\frac{1}{T}V(x, y)}} \leq \frac{1}{\Gamma(T)}.$$

By comparing the above inequalities with Def. 2.1, we can easily verify that for $\epsilon = e^{-1/T}$ and $V(x, y) = R(x, y)$ for every (x, y) pair, any regularly perturbed process P^ϵ satisfies (7) and (8). Moreover, if the game is a potential game, then it satisfies (10). Thus, the framework of cycle decomposition applies to any stochastic learning dynamics on potential games that generate a regularly perturbed Markov process. The condition of $\epsilon = e^{-1/T}$ is not strict and (7) can easily be expressed with a general noise parameter ϵ . Since both LLL and ML have $\epsilon = e^{-1/T}$, we will not go into the details. Thus, from this point onward, we will use the terms cost and resistance for $V(x, y)$ interchangeably.

C. Stochastic Learning Dynamics Explained by CDA

In the previous section, we proved that CDA applies to both ML and LLL. Next, we establish that CDA can effectively explain the medium and long run behaviors of stochastic learning dynamics through the example in Fig. 2. As explained before, Figs. 2(b)-2(e) were the outcome of CDA iterations for the Markov chain in Fig. 2(a) with transition cost in (15). However, the transition cost in (15) is the same as the cost in (21) for potential games. Thus, Figs. 2(b)-2(e) correspond to the outcomes of CDA for ML.

The final iteration results in a single cycle containing the entire state space. The only information conveyed by the last

level is that the chain is irreducible, which is already known. Therefore, we start from $G(E^3, \mathcal{E}^3)$ as presented in Fig. 2(e). The set E^3 comprises a singleton $\{a\}$ with $H_e^3(\{a\}) = 0$ and a non-trivial cycle of order three. The non-trivial cycle contains all the remaining states and has an exit height of 14. Thus, if the initial condition is $\{a\}$, the Markov chain will leave this state quickly and will hit the large cycle containing all the other states. From Thm. 5.1, the time to exit from this cycle will be of the order of $e^{14/T}$. Moreover, the chain will visit all the states within this big cycle in a time of order $e^{11/T}$, where 11 is the mixing height of the cycle. Therefore, for a system with a finite lifetime and initial condition $\{a\}$, we can safely conclude that the system will leave $\{a\}$ quickly and will never revisit it with high probability.

However, what happens when the chain leaves $\{a\}$ or if the initial condition is not in $\{a\}$? These questions cannot be answered adequately from $G(E^3, \mathcal{E}^3)$. To answer these questions, we need to go one level lower to the output of the second iteration $G(E^2, \mathcal{E}^2)$. Graph $G(E^2, \mathcal{E}^2)$ offers more details about the behavior of the Markov chain as compared to $G(E^3, \mathcal{E}^3)$. It reveals that after exiting from $\{a\}$, the Markov chain will hit $\{b\}$ where it will get stuck for a time of the order of $e^{3/T}$. On exiting $\{b\}$, the chain will hit $\{c\}$ with high probability because the transition from $\{b\}$ to $\{c\}$ is the minimum cost transition from $\{b\}$. From $\{c\}$, the chain can either return to $\{b\}$, or move on to $\{d, e, f, g, h, i\}$ with equal probabilities. However, once it hits $\{d, e, f, g, h, i\}$, it will stay within this cycle for most of the time because the exit height of this cycle is 11 which is more than three times the exit height of $\{b\}$. Thus, we can conclude from $G(E^2, \mathcal{E}^2)$ that in the long run, the Markov chain will spend most of its time within the cycle $\{d, e, f, g, h, i\}$ after the short and medium run behavior described above.

Similarly, we can explain the behavior within $\{d, e, f, g, h, i\}$ by examining $G(E^1, \mathcal{E}^1)$ presented in Fig. 2(c). Although switching from $G(E^3, \mathcal{E}^3)$ to $G(E^2, \mathcal{E}^2)$ furnished more information, there is one drawback. Given that the Markov chain is at a state other than $\{a\}$, we cannot easily approximate the hitting time of $\{a\}$ from $G(E^2, \mathcal{E}^2)$. Similarly, switching from $G(E^2, \mathcal{E}^2)$ to $G(E^1, \mathcal{E}^1)$ can provide more information about the behavior of the chain restricted to the set of states $\{d, e, f, g, h, i\}$. However, we can lose high-level information about the transition behavior between $\{d, e, f, g, h, i\}$ and the other states in the state space. Thus, switching from the output of a high-level iteration to a low-level iteration of CDA delivers information of higher resolution. However, this high-resolution information is restricted to small subsets of state space. On the other hand, moving from a lower level to a higher level of CDA yields high-level details on transitions from one set of states to another set of states but abstracts away low-level information.

Regardless of which level of CDA we are analyzing, the key parameter that enables us to describe system behavior through CDA is the exit height of a cycle, which depends on the mixing height of that cycles according to (12). To verify this claim, we apply CDA under LLL on the chain in Fig.2(a), and compare the output with output under ML discussed before. The outputs of the first iteration under ML and the second iteration under

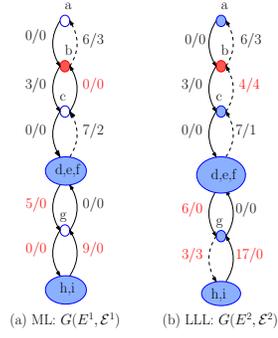


Fig. 3. Comparison of ML and LLL based on CDA

LLL are presented in Fig. 3. The first observation is that both the graphs have the same cycles but different transition costs. The difference in transition costs is highlighted in the figure with a different color.

Assuming that a is the initial condition, both the dynamics have the same behavior till the chain reaches the state c . At c , ML can transition to b or $\{d, e, f\}$ with equal probability. However, LLL will transition to $\{d, e, f\}$ with high probability since the transition to b has a higher cost. Thus, LLL will hit the cycle $\{d, e, f\}$ faster as compared to ML. However, the exit height of this cycle under LLL is six which is one unit higher than the exit height under ML. The difference in exit heights implies that although ML will reach $\{d, e, f\}$ late, it has the ability to leave this cycle quickly as compared to LLL. After exiting from $\{d, e, f\}$, both the chains will hit the state g . From the state g , ML can transition back to $\{d, e, f\}$ or to $\{h, i\}$ with equal probability, where state i is the potential maximizer. However, LLL will return to $\{d, e, f\}$ with high probability because the transition cost to $\{h, i\}$ is high. Based on the comparative analysis of the chain in Fig. 2(a), if a is the initial condition then LLL will initially proceed towards the potential maximizer faster as compared to ML. However, it can get stuck in a cycle longer than ML. Moreover, ML has a zero cost path to potential maximizer from g whereas no such path exists for LLL.

VI. COMPARATIVE ANALYSIS OF ML AND LLL

In this section, we present the main results of this work related to the comparative analysis of stochastic learning dynamics in games. Although the results are presented in the context of LLL and ML, the techniques can be extended to any learning dynamics that satisfy the criteria in (7), (8), and (10).

Note: All the analysis and results presented in this section are for potential game setup.

A. First Order Comparative Analysis

In the first order analysis, we compare the expected hitting times of the set of Nash equilibria for both the learning rules. We assume that for any action profile pair a and a' such that $P^T(a, a') > 0$, we have

$$\phi(a) \neq \phi(a'). \quad (24)$$

This assumption is standard as it simplifies the analysis of stochastic learning dynamics and is not restrictive. Otherwise, if $\phi(a) = \phi(a')$ and $P^T(a, a') > 0$, we can simply merge a and a' into a single state. An important outcome of this assumption is that all the paths that only involve zero cost transitions will not contain any cycles.

Let M be the set of all the Nash equilibria in S , and let τ_M^{LR} be the hitting time of M for a learning rule $LR \in \{\text{LLL}, \text{ML}\}$. Let

$$\Omega^{LR}(a, M) = \{w_{a,a^*}^A : a^* \in M \text{ and } V^{LR}(w_{a,a^*}^A) = 0\},$$

i.e., $\Omega^{LR}(a, M)$ is the set of zero cost paths from state a to set M for LR . The set $\Omega^{LR}(a, M)$ is finite for every state a because the state space is finite. Let

$$\xi^{LR}(a) = \max\{|\omega| : \omega \in \Omega^{LR}(a, M)\}, \text{ and} \\ \sigma^{LR}(a) = \min\{|\omega| : \omega \in \Omega^{LR}(a, M)\}. \quad (25)$$

Thus, $\xi^{LR}(a)$ and $\sigma^{LR}(a)$ are the lengths of the longest and shortest zero cost paths from a to M for LR . Let

$$\Omega_l^{LR}(a, M) = \{\omega \in \Omega^{LR}(a, M) : |\omega| = l\}, \quad (26)$$

i.e., $\Omega_l^{LR}(a, M)$ is the set of all paths in $\Omega^{LR}(a, M)$ having a length of l . The set of zero cost paths from any initial condition is

$$\Omega^{LR} = \bigcup_{a \in \mathcal{A}} \Omega^{LR}(a, M).$$

Before presenting the main results, we need to prove the following propositions.

Proposition 6.1: If an action profile $a \notin M$, then $\Omega^{LR}(a, M)$ is non-empty for a finite state space \mathcal{A} for $LR \in \{\text{LLL}, \text{ML}\}$.

Proof: We establish that $\Omega^{LR}(a, M)$ is non-empty by showing that we can always construct a path that belongs to this set.

Let $W_0 = \{a_0\}$, where $a_0 = a$. Assume that there does not exist any $a' \in \mathcal{A}$ such that $V^{LR}(a, a') = 0$. This will imply that $a \in M$, which is a contradiction. Thus, there exists an action profile $a_1 \in \mathcal{A}$ such that $V^{LR}(a, a_1) = 0$. Define $W_1 = W_0 \cup \{a_1\}$. If $a_1 \in M$, we are done and $\omega = (a_0, a_1)$ belong to $\Omega^{LR}(a, M)$. If $a_1 \notin M$, we can argue in the same manner as before that there exists an action profile $a_2 \in \mathcal{A}$ such that $V^{LR}(a_1, a_2) = 0$. Define $W_2 = W_1 \cup \{a_2\}$. By repeating this argument k times, we obtain $W_k = \{a_0, a_1, \dots, a_{k-1}, a_k\}$, where $a_0 = a$. The condition in (24) implies that $\phi(a_l) < \phi(a_{l+1})$ for all $l \in \{0, 1, \dots, k-1\}$. Thus, all the action profiles in the path $\omega = (a_0, a_1, \dots, a_k)$ are unique. Moreover, the state space \mathcal{A} is finite, which ensures that this process will terminate in a finite number of steps, say p . Then, the path (a_0, a_1, \dots, a_p) belongs to $\Omega^{LR}(a, M)$. ■

Proposition 6.2: For LLL and ML,

$$\Omega^{\text{LLL}}(a, M) \subseteq \Omega^{\text{ML}}(a, M).$$

Proof: For any pair of distinct action profiles a and a'

$$V^{\text{LLL}}(a, a') = 0 \implies V^{\text{ML}}(a, a') = 0,$$

but the converse is not true. To prove this statement, let $\mathbf{a} = (\alpha, \mathbf{a}_{-i})$ and $\mathbf{a}' = (\alpha', \mathbf{a}_{-i})$ be any two action profiles in \mathcal{A} . If $V^{\text{LLL}}(\mathbf{a}, \mathbf{a}') = 0$, then $\alpha' \in B_i(\mathbf{a}_{-i})$, i.e., α' is a best response of player i to \mathbf{a}_{-i} . Thus,

$$U_i(\mathbf{a}') - U_i(\mathbf{a}) > 0 \implies V^{\text{ML}}(\mathbf{a}, \mathbf{a}') = 0.$$

To show that the converse is not true, let $V^{\text{ML}}(\mathbf{a}, \mathbf{a}') = 0$, which implies that $U_i(\alpha', \mathbf{a}_{-i}) - U_i(\alpha, \mathbf{a}_{-i}) > 0$. However, if $\alpha' \notin B_i(\mathbf{a}_{-i})$ then $V^{\text{LLL}}(\mathbf{a}, \mathbf{a}') > 0$. Thus, given an action profile \mathbf{a} , all the zero cost paths from \mathbf{a} to M for LLL are also zero cost paths for ML. However, a path having zero cost for ML may not have zero cost for LLL, which concludes the proof. ■

Next, we present our main results for the first order comparative analysis.

Theorem 6.1: *The expected hitting time of the set of Nash equilibria M for a learning dynamics $LR \in \{\text{LLL}, \text{ML}\}$ satisfies*

$$\mathbb{E}(\tau_M^{LR}) \leq k_0 \frac{1}{1 - \rho^{LR}}. \quad (27)$$

where

$$k_0 = \max_{\mathbf{a} \in \mathcal{A}} \sigma^{LR}(\mathbf{a}) \text{ and } \rho^{LR} = 1 - (\Gamma^{LR}(T))^{k_0}. \quad (28)$$

Here $\sigma^{LR}(\mathbf{a})$ is the length of shortest path in $\Omega^{LR}(\mathbf{a}, M)$.

Proof: Let x be a non-negative integer valued countable random variable. Then,

$$\mathbb{E}(x) = \sum_{k=0}^{\infty} P(x > k).$$

Suppose the system is initialized at some action profile \mathbf{a} . Then, from the above definition of expected value,

$$\mathbb{E}_{\mathbf{a}}(\tau_M^{LR}) = \sum_{k=0}^{\infty} P_{\mathbf{a}}(\tau_M^{LR} > k).$$

For $k \leq \sigma^{LR}(\mathbf{a}) - 1$, we have

$$\sum_{k=0}^{\sigma^{LR}(\mathbf{a})-1} P_{\mathbf{a}}(\tau_M^{LR} > k) \leq \sigma^{LR}(\mathbf{a}) \leq k_0.$$

For $k = \sigma^{LR}(\mathbf{a})$, we have

$$\begin{aligned} P_{\mathbf{a}}(\tau_M^{LR} > k) &= 1 - P_{\mathbf{a}}(\tau_M^{LR} \leq k) \leq 1 - P_{\mathbf{a}}(\tau_M^{LR} = k), \\ &\leq 1 - (\Gamma^{LR}(T))^{\sigma^{LR}(\mathbf{a})} \\ &\leq 1 - (\Gamma^{LR}(T))^{k_0} = \rho^{LR}. \end{aligned}$$

Here, $(\Gamma^{LR}(T))^{\sigma^{LR}(\mathbf{a})}$ is a lower bound on the probability of a zero cost path of length $\sigma^{LR}(\mathbf{a})$ and $k_0 = \max_{\mathbf{a}} \sigma^{LR}(\mathbf{a})$. The right side of the above inequality is independent of initial condition \mathbf{a} . Therefore, we will drop the dependence on initial condition from this point onwards.

Given any $k > k_0$, we can express it as $k = qk_0 + c$, where q is some positive integer and $0 \leq c < k_0$. Then, for any $qk_0 \leq k < (q+1)k_0$,

$$\begin{aligned} P(\tau_M^{LR} > k) &= P(\tau_M^{LR} > k \mid \tau_M^{LR} > qk_0)P(\tau_M^{LR} > qk_0), \\ &\leq P(\tau_M^{LR} > qk_0). \end{aligned}$$

Proceeding in a similar fashion, we get

$$\begin{aligned} P(\tau_M^{LR} > qk_0) &= P(\tau_M^{LR} > qk_0 \mid \tau_M^{LR} > (q-1)k_0) \\ &\quad P(\tau_M^{LR} > (q-1)k_0). \end{aligned}$$

For a time-homogeneous Markov chain, we get

$$P(\tau_M^{LR} > qk_0 \mid \tau_M^{LR} > (q-1)k_0) = P(\tau_M^{LR} > k_0).$$

Repeatedly applying this argument q times results in

$$P(\tau_M^{LR} > qk_0) = (P(\tau_M^{LR} > k_0))^q.$$

Thus, for $qk_0 \leq k < (q+1)k_0$

$$P(\tau_M^{LR} > k) \leq (\rho^{LR})^q. \quad (29)$$

By combining the above results, we can compute an upper bound on the expected hitting time of M as follows:

$$\begin{aligned} \mathbb{E}(\tau_M^{LR}) &= \sum_{k=0}^{\infty} P(\tau_M^{LR} > k), \\ &= \sum_{k=0}^{k_0-1} P(\tau_M^{LR} > k) + \sum_{k=k_0}^{2k_0-1} P(\tau_M^{LR} > k) + \dots, \\ &\leq k_0 + k_0\rho^{LR} + k_0(\rho^{LR})^2 + \dots, \\ &= k_0 \frac{1}{1 - \rho^{LR}}, \end{aligned}$$

which is the desired result. ■

Eq. (27) is an upper bound on the expected hitting time of the set of NE for a stochastic learning dynamics. However, based on this result, it is difficult to declare which of the two learning dynamics will lead to a NE faster in general. Next, we derive a condition under which the expected hitting time of M for LLL will be smaller than ML as $T \rightarrow 0$.

Before presenting the next result, we define

$$N_0^{LR}(\mathbf{a}) = \{\mathbf{a}' \in \mathcal{A} \setminus \mathbf{a} : V^{LR}(\mathbf{a}, \mathbf{a}') = 0\}.$$

Here, $N_0^{LR}(\mathbf{a})$ is a subset of the neighborhood of \mathbf{a} and comprises the action profiles having zero cost transitions from \mathbf{a} . Thus, for any $\mathbf{a} \notin M$, $N_0^{LR}(\mathbf{a})$ will always be a non-empty set. For a path $\omega \in \Omega^{LR}(\mathbf{a}, M)$, we define

$$\begin{aligned} n_{\max}^{LR}(\mathbf{a}) &= \max_{\omega \in \Omega^{LR}(\mathbf{a}, M)} \max_{0 \leq k \leq |\omega|-1} |N_0^{LR}(\omega_k)|, \\ n_{\min}^{LR}(\mathbf{a}) &= \min_{\omega \in \Omega^{LR}(\mathbf{a}, M)} \min_{0 \leq k \leq |\omega|-1} |N_0^{LR}(\omega_k)|. \end{aligned} \quad (30)$$

Theorem 6.2: *Let $|A_{\min}|$ be the minimum number of actions of a player in the game. If*

$$|A_{\min}| > \max_{\mathbf{a} \in \mathcal{A}} \frac{\xi^{\text{LLL}}(\mathbf{a})}{\sigma^{\text{ML}}(\mathbf{a})} \frac{n_{\max}^{\text{ML}}(\mathbf{a})}{n_{\min}^{\text{LLL}}(\mathbf{a})} Z_{\max}, \quad (31)$$

then

$$P(\mathbb{E}(\tau_M^{\text{LLL}}) < \mathbb{E}(\tau_M^{\text{ML}})) \rightarrow 1 \text{ as } T \rightarrow 0.$$

Here, $\xi^{\text{LLL}}(\mathbf{a})$ and $\sigma^{\text{ML}}(\mathbf{a})$ are defined in (25), Z_{\max} is defined in (23), and $n_{\max}^{\text{ML}}(\mathbf{a})$ and $n_{\min}^{\text{LLL}}(\mathbf{a})$ are defined in (30).

Proof: Assume that the Markov chain is initialized at some action profile $\mathbf{a} \notin M$ and let $P_{\mathbf{a}\mathbf{a}'}^{LR}$ be the transition probability from \mathbf{a} to \mathbf{a}' for learning dynamics LR . Note here that we have compressed the notation for transition

probability from $P^{LR}(a, a')$ to $P_{aa'}^{LR}$ for convenience. Since we are interested in the limiting case of zero noise, we will restrict our analysis to zero cost paths. Then, we will use the fact that the probability of moving along a path of non-zero cost goes to zero as $T \rightarrow 0$.

To compute the expected hitting time of the set M starting from some $a \in \mathcal{A} \setminus M$, we need to solve the following recursion.

$$\begin{aligned} \mathbb{E}_a(\tau_M^{LR}) &= P_{aa}^{LR} \mathbb{E}_a(\tau_M^{LR}) + \sum_{a^1 \in N_0^{LR}(a)} P_{aa^1}^{LR} \mathbb{E}_{a^1}(\tau_M^{LR}) + 1, \\ &= \frac{1}{\eta_a} + \sum_{a^1 \in N_0^{LR}(a)} \frac{P_{aa^1}^{LR}}{\eta_a} \mathbb{E}_{a^1}(\tau_M^{LR}). \end{aligned}$$

Here

$$\eta_a = (1 - P_{aa}^{LR}) = \sum_{a^1 \in N_0^{LR}(a)} P_{aa^1}^{LR}$$

is the exit probability from state a and $N_0^{LR}(a)$ is the set of action profiles in the neighborhood of state a having zero transition cost. The expected hitting time can be rewritten as follows:

$$\mathbb{E}_a(\tau_M^{LR}) = \sum_{a^1 \in N_0^{LR}(a)} \left(\frac{1/|N_0^{LR}(a)|}{P_{aa^1}^{LR}} + \mathbb{E}_{a^1}(\tau_M^{LR}) \right) \frac{P_{aa^1}^{LR}}{\eta_a}.$$

We can repeat the same steps for $\mathbb{E}_{a^1}(\tau_M^{LR})$ and after some algebraic manipulation, we get

$$\begin{aligned} \mathbb{E}_a(\tau_M^{LR}) &= \sum_{a^1 \in N_0^{LR}(a)} \sum_{a^2 \in N_0^{LR}(a^1)} \left(\frac{1/|N_0^{LR}(a)|}{P_{aa^1}^{LR}} + \right. \\ &\quad \left. \frac{1/|N_0^{LR}(a^1)|}{P_{a^1a^2}^{LR}} + \mathbb{E}_{a^2}(\tau_M^{LR}) \right) \frac{P_{aa^1}^{LR} P_{a^1a^2}^{LR}}{\eta_a \eta_{a^1}}. \end{aligned}$$

These summations include all the zero cost paths of length two that are starting from a . Since $\mathbb{E}_{a'}(\tau_M^{LR}) = 0$ if $a' \in M$, the above recursion terminates after $\xi^{LR}(a)$ iterations when all the paths originating from state a have reached the desired set M . Therefore, we can express $\mathbb{E}_a(\tau_M^{LR})$ in terms of the paths leading to M from a as follows:

$$\mathbb{E}_a(\tau_M^{LR}) = \sum_{\omega \in \Omega^{LR}(a, M)} \kappa_\omega P_\omega^{LR}, \quad (32)$$

where

$$\begin{aligned} \kappa_\omega^{LR} &= \frac{1/|N_0^{LR}(\omega_0)|}{P_{\omega_0\omega_1}^{LR}} + \dots + \frac{1/|N_0^{LR}(\omega_{p-1})|}{P_{\omega_{p-1}\omega_p}^{LR}}, \text{ and} \\ P_\omega^{LR} &= \frac{P_{\omega_0\omega_1}^{LR}}{\eta_{\omega_0}} \frac{P_{\omega_1\omega_2}^{LR}}{\eta_{\omega_1}} \dots \frac{P_{\omega_{p-1}\omega_p}^{LR}}{\eta_{\omega_{p-1}}}. \end{aligned}$$

In the above expression, P_ω^{LR} is the probability of moving along the path ω and κ_ω^{LR} is the weight associated to that path.

Let $\bar{\omega}$ and $\underline{\omega}$ be any zero cost paths from $\Omega^{LLL}(a, M)$ and $\Omega^{ML}(a, M)$ respectively. Then, we have

$$\begin{aligned} \kappa_{\bar{\omega}}^{LLL} &\leq \xi^{LLL}(a) \left(\frac{nZ_{\max}}{n_{\min}^{LLL}(a)} \right), \text{ and} \\ \kappa_{\underline{\omega}}^{ML} &\geq \sigma^{ML}(a) \left(\frac{n|A_{\min}|}{n_{\max}^{ML}(a)} \right). \end{aligned}$$

Here $n_{\max}^{ML}(a)$ is the maximum number of zero cost transitions that are possible from any action profile included in some path that belongs to $\Omega^{ML}(a, M)$. Similarly, $n_{\min}^{LLL}(a)$ is the minimum number of zero cost transitions from an action profile on some path in $\Omega^{LLL}(a, M)$. If

$$\sigma^{ML}(a) \left(\frac{n|A_{\min}|}{n_{\max}^{ML}(a)} \right) \geq \xi^{LLL}(a) \left(\frac{nZ_{\max}}{n_{\min}^{LLL}(a)} \right), \quad (33)$$

then the minimum weight of any zero cost path in ML will be greater than the maximum weight of any zero cost path in LLL. Thus, if the above condition is satisfied, then the expected hitting time of M from profile a will be smaller for LLL than for ML. Simple algebraic manipulations on the above inequality and taking maximum over all $a \in \mathcal{A}$ results in the desired condition on $|A_{\min}|$. Since the probability of moving along a path of non-zero cost approaches to zero in the limit as $T \rightarrow 0$,

$$P(\mathbb{E}(\tau_M^{LLL}) < \mathbb{E}(\tau_M^{ML})) \rightarrow 1 \text{ as } T \rightarrow 0$$

holds for all $a \in \mathcal{A}$ if the condition in (31) is satisfied, which concludes the proof. \blacksquare

To understand the result in the above theorem, we rewrite (33) as

$$\frac{n|A_{\min}|}{n_{\max}^{ML}(a)} > \frac{\xi^{LLL}(a)}{\sigma^{ML}(a)} \frac{nZ_{\max}}{n_{\min}^{LLL}(a)}. \quad (34)$$

Consider a zero cost path from action action profile a to M that involves a transition from profile a^1 to a^2 . Assume that player i updates its action in this transition. Then, the expected hitting time of M starting from profile a will involve the following terms:

$$\begin{aligned} \frac{1/|N_0^{ML}(a^1)|}{P_{a^1a^2}^{ML}} &= \frac{n|A_i|}{|N_0^{ML}(a^1)|} \geq \frac{n|A_{\min}|}{|N_0^{ML}(a^1)|}, \\ \frac{1/|N_0^{LLL}(a^1)|}{P_{a^1a^2}^{LLL}} &= \frac{nZ_i(a_{-i}^1)}{|N_0^{LLL}(a^1)|} \leq \frac{nZ_{\max}}{|N_0^{LLL}(a^1)|}. \end{aligned}$$

In ML, any player can be randomly selected and the selected player can select any action from its action. Therefore, $n|A_{\min}|$ is the minimum number of possible transitions from a^1 including the possibility of staying at a^1 . The term $|N_0^{ML}(a^1)|$ is the number of zero cost transitions leaving from a^1 for ML. Therefore, the fraction $n|A_{\min}|/|N_0^{ML}(a^1)|$ can be thought of as the inertia of ML at state a^1 . By inertia of a dynamics at a state, we mean the tendency of the dynamics to remain at that state. Assuming a fixed n , if increasing the number of actions $|A_i|$ does not increase the number of possible zero cost transitions, then the inertia increases. A higher value of inertia will result in a larger number of steps to leave that state.

In LLL, any player can be randomly selected and the selected player say player i selects its best response with probability $1/nZ_i(a_{-i}^1)$, where $Z_i(a_{-i}^1)$ approaches one as noise approaches zero. The best response of a player is always unique because of the assumption in (24). Therefore, the inertia of LLL at state a^1 is less than $nZ_{\max}/|N_0^{LLL}(a^1)|$. Here Z_{\max} incorporates the maximum effect of paths of non-zero cost at each action profile. However, $Z_{\max} < |A_i|$ and approaches to one as $T \rightarrow 0$. Thus, the important difference

between the two dynamics is that the inertia of ML depends on both the number of players and number of actions of each player. However, the inertia of LLL depends primarily on the number of players. It has some dependence on the number of actions through the normalizing constant. However, this dependence on the number of actions vanishes as $T \rightarrow 0$.

In light of the above discussion, the inequality in (34) relates the minimum possible inertia of ML and the maximum possible inertia of LLL in moving along zero cost paths from state \mathbf{a} to M . The condition that we derived states that if the minimum inertia of ML over all the states in the paths that belong to $\Omega^{\text{ML}}(\mathbf{a}, M)$ is $(\xi^{\text{LLL}}(\mathbf{a})/\sigma^{\text{ML}}(\mathbf{a}))$ times greater than the maximum inertia of LLL over all the states in the paths that belong to $\Omega^{\text{LLL}}(\mathbf{a}, M)$, then $\mathbb{E}_{\mathbf{a}}(\tau^{\text{LLL}}(M)) < \mathbb{E}_{\mathbf{a}}(\tau^{\text{ML}}(M))$ as $T \rightarrow 0$. From this condition, we derive the condition on $|A_{\min}|$ in (31).

B. Higher Order Comparative Analysis

In the higher order analysis, we compare the mixing heights and the exit heights of the subsets of state space S that are cycles under both LLL and ML. We show that both the mixing and exit heights of a cycle are smaller for ML than LLL. These results imply that after entering a cycle, ML will visit all the states inside the cycle quickly as compared to LLL and will exit the cycle faster. We start by examining the transition cost between states for both the learning rules.

Proposition 6.3: The cost between any two action profiles \mathbf{a} and \mathbf{a}' satisfies

$$V^{\text{LLL}}(\mathbf{a}, \mathbf{a}') \geq V^{\text{ML}}(\mathbf{a}, \mathbf{a}'). \quad (35)$$

Proof: Let $\mathbf{a} = (\alpha, \mathbf{a}_{-i})$ and $\mathbf{a}' = (\alpha', \mathbf{a}_{-i})$ be any two action profiles in \mathcal{A} . To prove the proposition, we need to analyze three cases based on the definitions in (21) and (22). *Case 1:* $\alpha' \in B_i(\mathbf{a}_{-i})$.

If α' belongs to the best response set of player i for \mathbf{a}_{-i} , then

$$V^{\text{LLL}}(\mathbf{a}, \mathbf{a}') = V^{\text{ML}}(\mathbf{a}, \mathbf{a}') = 0.$$

Case 2: $\alpha' \notin B_i(\mathbf{a}_{-i})$ and $U_i(\alpha', \mathbf{a}_{-i}) \geq U_i(\alpha, \mathbf{a}_{-i})$.

In this case α' is not the best response to \mathbf{a}_{-i} . However, it does not result in a decrease in utility as compared to α . Therefore,

$$V^{\text{ML}}(\mathbf{a}, \mathbf{a}') = 0,$$

Let $\alpha^* \in B_i(\mathbf{a}_{-i})$. Then,

$$\begin{aligned} V^{\text{LLL}}(\mathbf{a}, \mathbf{a}') &= U_i(\alpha^*, \mathbf{a}_{-i}) - U_i(\alpha', \mathbf{a}_{-i}), \\ &> V^{\text{ML}}(\mathbf{a}, \mathbf{a}'). \end{aligned}$$

Case 3: $U_i(\alpha', \mathbf{a}_{-i}) < U_i(\alpha, \mathbf{a}_{-i})$

In this case the target action α' results in a decrease in utility as compared to the current action α .

$$V^{\text{ML}}(\mathbf{a}, \mathbf{a}_{-i}) = U_i(\alpha, \mathbf{a}_{-i}) - U_i(\alpha', \mathbf{a}_{-i}),$$

and

$$\begin{aligned} V^{\text{LLL}}(\mathbf{a}, \mathbf{a}') &= U_i(\alpha^*, \mathbf{a}_{-i}) - U_i(\alpha', \mathbf{a}_{-i}), \quad \alpha^* \in B_i(\mathbf{a}_{-i}), \\ &\geq V^{\text{ML}}(\mathbf{a}, \mathbf{a}_{-i}). \end{aligned}$$

The equality holds if $\alpha \in B_i(\mathbf{a}_{-i})$. ■

Theorem 6.3: *For a cycle Π such that $\Pi \in C^{\text{ML}}$ and $\Pi \in C^{\text{LLL}}$, the following inequality holds*

$$H_m^{\text{LLL}}(\Pi) \geq H_m^{\text{ML}}(\Pi) \quad (36)$$

Proof: From (17), mixing height, potential, and the altitude of communication of a cycle Π are related as follows

$$H_m(\Pi) = \phi(\Pi) - A_c(\Pi).$$

We will show that the altitude of communication satisfies $A_c^{\text{LLL}}(\Pi) \leq A_c^{\text{ML}}(\Pi)$, which will conclude the proof. Let

$$\mathbf{a}_f = \arg \min_{\mathbf{a} \in \Pi} \phi(\mathbf{a}),$$

i.e., \mathbf{a}_f is an action profile in Π with minimum potential. Using the concept of increasing family of cycles with respect to a state in the state space presented in (16), the cycle Π can be represented as

$$\Pi = \mathbf{a}_f^{n_\Pi},$$

where n_Π is the order of Π . Using the same concept, $\mathbf{a}_f^{n_\Pi-1}$ is a cycle of order $n_\Pi - 1$ that belongs to $\mathcal{M}(\Pi)$, the maximal partition of Π , and contains \mathbf{a}_f . Since \mathbf{a}_f was an action profile with minimum potential in Π , it is also a minimum potential action profile in $\mathbf{a}_f^{n_\Pi-1}$. Let $\hat{\mathbf{a}}$ be another action profile in Π such that $\hat{\mathbf{a}} \notin \mathbf{a}_f^{n_\Pi-1}$. Therefore, Π is the minimum cycle containing both \mathbf{a}_f and $\hat{\mathbf{a}}$. The condition in (18) implies that the communication altitude of Π is

$$A_c(\Pi) = A_c(\mathbf{a}_f, \hat{\mathbf{a}}). \quad (37)$$

Given any two action profiles \mathbf{a} and \mathbf{a}' in Π

$$A_c^{\text{LR}}(\mathbf{a}, \mathbf{a}') = \max_{\omega \in \Omega^S(\mathbf{a}, \mathbf{a}')} \min_{0 \leq k \leq |\omega|-1} (\phi(w_k) - V^{\text{LR}}(w_k, w_{k+1})),$$

where ω is a path from \mathbf{a} to \mathbf{a}' , $\omega_k = \mathbf{a}_k$ is the k^{th} action profile in ω , and $(i_1, i_2, \dots, i_{|\omega|})$ be the sequence of players updating their actions. From Prop. 6.3, we know that

$$V^{\text{LLL}}(\mathbf{a}, \mathbf{a}') \geq V^{\text{ML}}(\mathbf{a}, \mathbf{a}'),$$

which implies that $A_c^{\text{LLL}}(\Pi) \leq A_c^{\text{ML}}(\Pi)$. Thus,

$$H_m^{\text{LLL}}(\Pi) \geq H_m^{\text{ML}}(\Pi),$$

which concludes the proof of the theorem. ■

Theorem 6.4: *For a cycle Π such that $\Pi \in C^{\text{ML}}$ and $\Pi \in C^{\text{LLL}}$, the following inequality holds*

$$H_e^{\text{LLL}}(\Pi) \geq H_e^{\text{ML}}(\Pi). \quad (38)$$

Proof: According to Prop. 4.15 in [30], the exit height of a cycle Π can be computed as follows

$$H_e(\Pi) = \min_{\mathbf{a} \in \Pi} \max_{\mathbf{a}' \in \mathcal{A} \setminus \Pi} \phi(\mathbf{a}) - A_c(\mathbf{a}, \mathbf{a}'). \quad (39)$$

For any pair of action profiles \mathbf{a} and \mathbf{a}' , we just proved that

$$A_c^{\text{ML}}(\mathbf{a}, \mathbf{a}') \geq A_c^{\text{LLL}}(\mathbf{a}, \mathbf{a}').$$

Combining this fact with (39) results in

$$\begin{aligned} H_e^{\text{LLL}}(\Pi) &= \min_{a \in \Pi} \max_{a' \in \mathcal{A} \setminus \Pi} \phi(a) - A_c^{\text{LLL}}(a, a'), \\ &\geq \min_{a \in \Pi} \max_{a' \in \mathcal{A} \setminus \Pi} \phi(a) - A_c^{\text{ML}}(a, a'), \\ &= H_e^{\text{ML}}(\Pi), \end{aligned}$$

which concludes the proof of the theorem. In fact, the result in [30] is not restricted to a cycle and is applicable to any subset of the state space. For any $D \subset \mathcal{A}$, the exit height is

$$H_e(D) = \min_{a \in D} \max_{a' \in \mathcal{A} \setminus D} \phi(a) - A_c(a, a').$$

Thus,

$$H_e^{\text{LLL}}(D) \geq H_e^{\text{ML}}(D)$$

for any $D \subset \mathcal{A}$. ■

Theorems 6.3 and 6.4 confirm our observations from the sensor coverage game that ML can exit from any set of action profiles faster as compared to LLL. However, the proofs of these theorems provide more insight related to the comparison of the exit and mixing heights of both the dynamics. In fact, comparing the costs $V^{\text{LLL}}(a, a')$ and $V^{\text{ML}}(a, a')$ reveal that the critical factor contributing to comparatively high exit and mixing heights of LLL is the dependence of transition cost on the best response. In contrast, the transition cost in ML simply depends on pairwise comparison of utilities.

VII. CDA FOR SENSOR COVERAGE GAME

To compare the performance of sensor coverage game, we developed a setup with three sensors in a grid of size 10×10 . The sensors were located at $\{(9.03, 3.98), (8.4, 1.4), (1.96, 6.35)\}$. Each sensor had a fixed sensing range, which enabled us to represent the action set as on or off. Thus, the size of the state space was eight from s_0 to s_7 . The state s_i corresponded to the joint action profile that is the binary equivalent of $i \in \{0, 1, \dots, 7\}$. There were two equilibrium configurations s_3 and s_5 , where $s_3 = \{0, 1, 1\}$ and $s_5 = \{1, 0, 1\}$. Moreover, s_5 is the potential maximizer. The results are presented in Fig. 4.

For three sensors, there were eight possible configurations. The utility of each configuration along with a state transition diagram is presented in 4(a), which is same for both the dynamics. The outputs of all the iterations of CDA for ML and LLL are presented in the figures ranging from 4(b)-4(f) and 4(g)-4(l) respectively. A simple comparison of the two sets of figures verified the analysis presented in the previous section. The cycles in the iterations 2–5 for ML and 3–6 for LLL were the same, as shown in figures 4(c)-4(f) and 4(i)-4(l) respectively. For each of these cycles, $H_e^{\text{LLL}} \geq H_e^{\text{ML}}$.

For the initial condition $s_0 = (0, 0, 0)$, the difference in the paths to stochastically stable states can also be observed from Figs. 4(b) and 4(g)-4(h). In the case of ML, the sensor configuration can reach either of the two equilibrium configurations through s_1 , resulting in the formation of the cycle $\{s_1, s_3, s_5\}$. However, under LLL, the sensor configuration will first hit the potential maximizer s_5 through s_1 as shown by the cycle $\{s_1, s_5\}$. In the next iteration, s_4 is added to the cycle instead of s_3 , which signifies that the network configuration will cycle

between the states s_5 , s_1 , and s_4 exponentially many times before hitting s_3 for the first time.

VIII. CONCLUSIONS

We highlighted a critical issue with stochastic stability as a solution concept, which is its inability to distinguish between learning rules that lead to the same steady-state behavior. To address this problem, we presented a comprehensive framework for analyzing and comparing the transient performance of such learning dynamics. In the proposed framework, the main contribution was to identify cycle decomposition of Markov chains as a set of tools that enabled the comparative analysis of the stochastic learning dynamics. Moreover, we selected the expected hitting time to the set of Nash equilibria and the exit time from a subset of state space as important parameters to compare the performance of stochastic learning rules. We selected LLL and ML as representative members of the class of stochastic learning dynamics and showed that both of these dynamics have the same stochastically stable states, but significantly different short and medium run behavior. Based on the proposed comparative analysis framework, we identified critical factors, which effect the expected hitting time to the set of Nash equilibria for LLL and ML. We also proved that the exit time from a subset of state space will always be higher for LLL as compared to ML.

REFERENCES

- [1] D. Foster and P. Young, "Stochastic evolutionary game dynamics," *Theoretical population biology*, vol. 38, no. 2, pp. 219–232, 1990.
- [2] H. P. Young, "The evolution of conventions," *Econometrica*, vol. 61, no. 1, pp. 57–84, January 1993.
- [3] M. Kandori, G. J. Mailath, and R. Rob, "Learning, mutation, and long run equilibria in games," *Econometrica: Journal of the Econometric Society*, pp. 29–56, 1993.
- [4] L. E. Blume, "The statistical mechanics of strategic interaction," *Games and economic behavior*, vol. 5, no. 3, pp. 387–424, 1993.
- [5] A. Montanari and A. Saberi, "Convergence to equilibrium in local interaction games," in *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*. IEEE, 2009, pp. 303–312.
- [6] J. R. Marden and J. S. Shamma, "Game theory and distributed control," *Handbook of game theory*, vol. 4, pp. 861–900, 2012.
- [7] V. Auletta, D. Ferraioli, F. Pasquale, and G. Persiano, "Metastability of logit dynamics for coordination games," in *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*. SIAM, 2012, pp. 1006–1024.
- [8] V. Auletta, D. Ferraioli, F. Pasquale, P. Penna, and G. Persiano, "Convergence to equilibrium of logit dynamics for strategic games," *Algorithmica*, vol. 76, no. 1, pp. 110–142, 2016.
- [9] G. E. Kreindler and H. P. Young, "Fast convergence in evolutionary equilibrium selection," *Games and Economic Behavior*, vol. 80, pp. 39–67, 2013.
- [10] —, "Rapid innovation diffusion in social networks," *Proceedings of the National Academy of Sciences*, vol. 111, no. Supplement 3, pp. 10 881–10 888, 2014.
- [11] G. Ellison, D. Fudenberg, and L. A. Imhof, "Fast convergence in evolutionary models: A lyapunov approach," *Journal of Economic Theory*, vol. 161, pp. 1–36, 2016.
- [12] D. Shah and J. Shin, "Dynamics in congestion games," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 38, no. 1. ACM, 2010, pp. 107–118.
- [13] I. Arieli and H. P. Young, *Fast convergence in population games*. Department of Economics, University of Oxford, 2011.
- [14] —, "Stochastic learning dynamics and speed of convergence in population games," *Econometrica*, vol. 84, no. 2, pp. 627–676, 2016.
- [15] G. Ellison, "Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution," *The Review of Economic Studies*, vol. 67, no. 1, pp. 17–45, 2000.

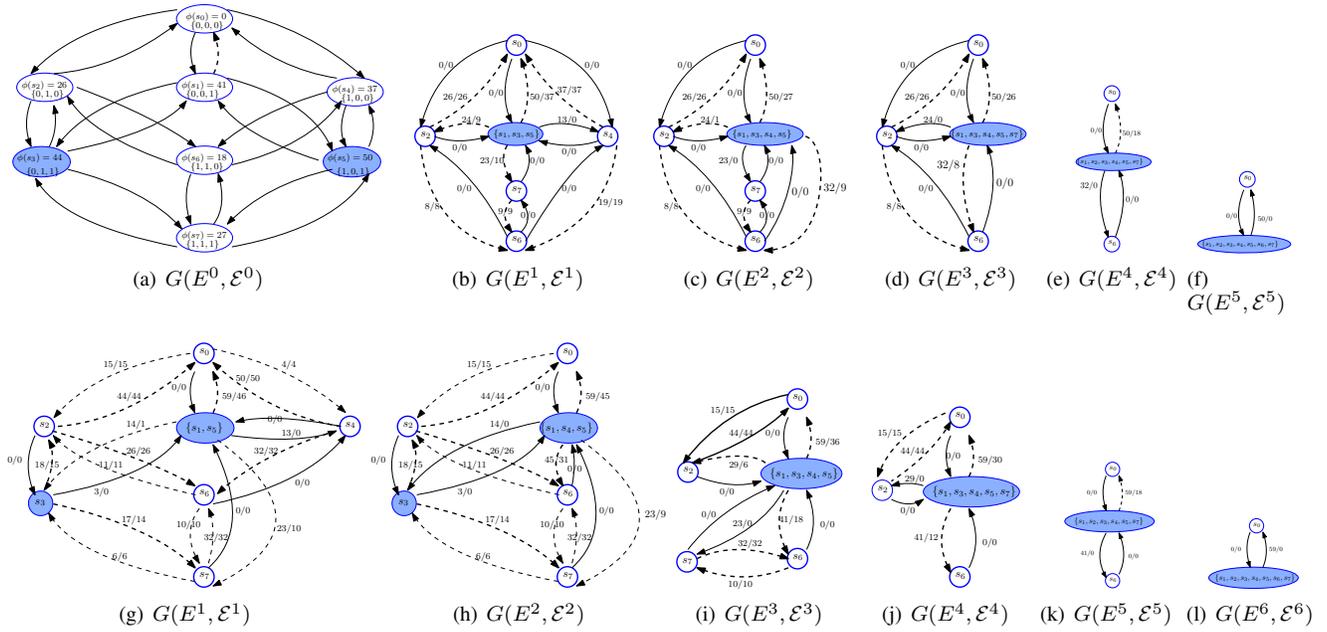


Fig. 4. Output of CDA for the sensor coverage game with $N = 3$ and $A_i = \{0, 6\}$, under ML and LLL. Fig. 4(a) shows the zeroth level, which is same for both the dynamics. Figs. 4(b)-Fig. 4(e) present the outputs under ML and Fig. 4(g)-4(l) present the outputs for LLL.

- [16] D. K. Levine and S. Modica, "Dynamics in stochastic evolutionary models," *Theoretical Economics*, vol. 11, no. 1, pp. 89–131, 2016.
- [17] M. I. Freidlin and A. D. Wentzell, "Random perturbations," in *Random Perturbations of Dynamical Systems*. Springer, 1984, pp. 15–43.
- [18] E. Olivieri and E. Scoppola, "Markov chains with exponentially small transition probabilities: First exit problem from a general domain. I. The reversible case," *Journal of statistical physics*, vol. 79, no. 3–4, pp. 613–647, 1995.
- [19] —, "Markov chains with exponentially small transition probabilities: first exit problem from a general domain. II. The general case," *Journal of statistical physics*, vol. 84, no. 5, pp. 987–1041, 1996.
- [20] O. Catoni, "Metropolis, simulated annealing, and iterated energy transformation algorithms: Theory and experiments," *Journal of Complexity*, vol. 12, no. 4, pp. 595–623, 1996.
- [21] O. Catoni and R. Cerf, "The exit path of a markov chain with rare transitions," *ESAIM: Probability and Statistics*, vol. 1, pp. 95–144, 1997.
- [22] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, "The complexity of computing a nash equilibrium," *SIAM Journal on Computing*, vol. 39, no. 1, pp. 195–259, 2009.
- [23] X. Chen, X. Deng, and S.-H. Teng, "Settling the complexity of computing two-player nash equilibria," *Journal of the ACM (JACM)*, vol. 56, no. 3, p. 14, 2009.
- [24] A. Trounev, "Cycle decompositions and simulated annealing," *SIAM Journal on Control and Optimization*, vol. 34, no. 3, pp. 966–986, 1996.
- [25] H. Jaleel and J. S. Shamma, "Transient response analysis of metropolis learning in games," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 9661–9667, 2017.
- [26] M. Mäs and H. H. Nax, "A behavioral study of "noise" in coordination games," *Journal of Economic Theory*, vol. 162, pp. 195–208, 2016.
- [27] W. H. Sandholm, "Pairwise comparison dynamics and evolutionary foundations for nash equilibrium," *Games*, vol. 1, no. 1, pp. 3–17, 2009.
- [28] J. R. Marden and J. S. Shamma, "Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation," *Games and Economic Behavior*, vol. 75, no. 2, pp. 788–808, 2012.
- [29] J. R. Marden and A. Wierman, "Distributed welfare games," *Operations Research*, vol. 61, no. 1, pp. 155–168, 2013.
- [30] O. Catoni, "Simulated annealing algorithms and markov chains with rare transitions," in *Séminaire de probabilités XXXIII*. Springer, 1999, pp. 69–119.



Hassan Jaleel Hassan Jaleel is an Assistant Professor at the Lahore University of Management Sciences (LUMS) in Lahore, Pakistan. He received his M.S. and Ph.D. degrees in Electrical and Computer Engineering (ECE) with specialization in Systems and Control from the Georgia Institute of Technology, Atlanta, GA, USA. His research interests are in the areas of complex networks, game theory, stochastic geometry, and online distributed optimization. Jaleel was a Fulbright scholar from 2009–2013 and is a member of IEEE.



Jeff Shamma Jeff S. Shamma is a Professor of Electrical Engineering at the King Abdullah University of Science and Technology (KAUST), where he is also the Director of the Center of Excellence for NEOM Research at KAUST and principal investigator of the Robotics, Intelligent Systems & Control laboratory (RISC). Shamma is the former Julian T. Hightower Chair in Systems & Control in the School of Electrical and Computer Engineering at Georgia Tech, and he has held faculty positions at the University of Minnesota, The University of Texas at Austin, and the University of California, Los Angeles. Shamma received a Ph.D. in systems science and engineering from MIT in 1988. He is the recipient of an NSF Young Investigator Award, the American Automatic Control Council Donald P. Eckman Award, and the Mohammed Dahleh Award, and he is a Fellow of the IEEE and of the IFAC (International Federation of Automatic Control). Shamma is the Editor-in-Chief for the IEEE Transactions on Control of Network Systems and an associate editor of the IEEE Transactions on Robotics.