Distributed Fictitious Play for Optimal Behavior of Multi-Agent Systems with Incomplete Information

Ceyhun Eksin and Alejandro Ribeiro

***Abstract*—A multi-agent system operates in an uncertain environment about which agents have different and time varying beliefs that, as time progresses, converge to a common belief. A global utility function that depends on the realized state of the environment and actions of all the agents determines the system's optimal behavior. We define the asymptotically optimal action profile as an equilibrium of the potential game defined by considering the expected utility with respect to the asymptotic belief. At finite time, however, agents have not entirely congruous beliefs about the state of the environment and may select conflicting actions. This paper proposes a variation of the fictitious play algorithm which is proven to converge to equilibrium actions if the state beliefs converge to a common distribution at a rate that is at least linear. In conventional fictitious play, agents build beliefs on others' future behavior by computing histograms of past actions and best respond to their expected payoffs integrated with respect to these histograms. In the variations developed here histograms are built using knowledge of actions taken by nearby nodes and best responses are further integrated with respect to the local beliefs on the state of the environment. We exemplify the use of the algorithm in coordination and target covering games.**

I. Introduction

Our model of a multi-agent autonomous system encompasses an underlying environment, knowledge about the state of the environment that the agents acquire, and a state dependent global objective that agents affect through their individual actions. The optimal action profile maximizes this global objective for the realized environment's state with the optimal action of an agent given by the corresponding action in the profile. The problem addressed in this paper is the determination of suitable actions when the probability distributions that agents have on the state of the environment are possibly different. These not entirely congruous beliefs result in mismatches between the action profiles that different agents deem to be optimal. As a consequence, when a given agent chooses an action to execute, it is important for it to reason about what the beliefs of other agents may be and what are the consequent actions that other agents may take. We propose a solution based on the construction of empirical histograms of past actions and the use of best responses to the utility expectation with respect to these histograms and the state belief. This algorithmic behavior is shown to be asymptotically optimal in the sense that if agents move towards a common belief, the actions they select are optimal with respect to the corresponding expected utility.

Work supported by the NSF award CAREER CCF-0952867, and the ARL grant W911NF-08-2-0004. The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania, 200 South 33rd Street, Philadelphia, PA 19104. Email: {ceksin, aribeiro}@seas.upenn.edu.

While the determination of optimal behavior in multi-agent systems can be considered from different perspectives, the categorization between systems with complete and incomplete information is most germane to this paper [1]. In systems of complete information the environment is either perfectly known or all agents have identical beliefs. In either case, agents can compute the optimal action profile, and determine and execute their corresponding optimal action. This local computation of global solutions is neither scalable nor robust but it can be used as an abstract definition of optimal behavior. This abstraction renders the problem equivalent to the development of distributed methodologies to solve optimization problems [2]–[5], or, more generically, to the determination of Nash equilibria of multiplayer games [6]–[11]. When information is incomplete, the fact that agents have different beliefs implies that they may end up choosing competing actions even if they are intent on cooperating. In a sense, agents are competing against uncertainty, but the manifestation of that competition is in the form of conflicting interests arising from cooperating agents. In this inherent competition Bayesian Nash equilibria are the intrinsic mathematical formulation of optimal behavior [12], [13]. However, determination of these equilibria is computationally intractable except for games with simple beliefs and utilities [14], [15].

If determination of Bayesian equilibria is intractable, the development of approximate methods is necessary. In fact, determining game equilibria is also challenging in games of complete information. This has motivated the development of iterative methods to learn regular – as opposed to Bayesian – equilibrium actions [11], [16]. Of relevance to this paper is the fictitious play algorithm in which agents build beliefs on others' future behavior by computing histograms of past actions and best respond to their expected payoffs integrated with respect to these histograms [17]. When information is complete, fictitious play converges to equilibria of zero sum [16], some other specific games with two players [18], and multiplayer games with aligned interests as determined by a potential function [6]. Recent variations of fictitious play have been developed to expand the class of games for which equilibria can be computed [9], [19] and to guarantee convergence to equilibria with specific properties [8], [20], [21]. Recently, a variant of the fictitious play that operates in a distributed setting where agents observe relevant information from agents that are adjacent in a network is shown to converge in potential games [7].

In this paper, we consider a network of agents with aligned interests. Agents have different and time varying beliefs on

the state of the environment that, as time progresses, converge to a common belief. The asymptotic optimal behavior is therefore formulated as the Nash equilibria of the potential game defined by the expected utility with respect to the asymptotic belief. The goal is to design a distributed mechanism that converges to a Nash equilibrium of this asymptotic game (Section II). The solutions that we propose are variations of the fictitious play algorithm that take into account the distributed nature of the multi-agent system and the fact that the state of the environment is not perfectly known. In a game of incomplete information, expected payoff computation in fictitious play consists of integrating the payoff with respect to both the local belief on the state of the environment and the local beliefs on the behavior of other agents. In a networked setting only local past histories can be available and agents need to reason about the behavior of non-neighboring agents based on past observations of its neighbors only.

In potential games with symmetric payoffs, which are known to admit consensus Nash equilibria, we let agents share their actions with their neighbors, construct empirical histograms of the actions taken by neighbors, and best respond assuming that all agents follow the average population empirical distribution (Section III). This mechanism is shown to converge to a consensus Nash equilibrium when the convergence of individual beliefs on the state of the environment is at least of linear order (Theorem 1). When the potential game is not necessarily symmetric, agents share their empirical beliefs on the behavior of other agents with neighbors in addition to their own actions. Agents keep histograms on the behavior of others by averaging the histograms of neighbors with their own (Section IV). Convergence to a Nash equilibrium follows under the same linear convergence assumption for the beliefs on the state of the environment (Theorem 2).

We numerically analyze the transient and asymptotic equilibrium properties of the algorithms in the beauty contest and the target covering games (Section V). In the beauty contest game, a team of robots tradeoffs between moving toward a target direction and moving in coordination with each other. In the target covering game, a team of robots coordinates to cover a given set of targets and receive payoffs from covering a target that is inversely proportional to the distance to their positions. We observe that Nash equilibrium strategies are successfully determined in both cases.

Notation: For any finite set X, we use $\triangle(X)$ to denote the space of probability distributions over X. For a generic vector $x \in X^n$, x_i denotes the ith element and x_{-i} denotes the vector of elements of x except the ith element, that is, $x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. We use $\|\cdot\|$ to denote the Euclidean norm of a space. $\mathbf{1}_n$ denotes a $n \times 1$ column vector of all ones.

II. Optimal behavior of multi-agent systems with incomplete information

We consider a group of n agents $i \in \mathcal{N} := \{1, \dots, n\}$ that play a stage game over a discrete time index t with simultaneous moves and incomplete information. The important features of this game are the actions a_{it} that agents take at time t, an underlying unknown state of the world θ, local utility functions $u_i(a, \theta)$ that determine the payoffs that different agents receive when the group plays the joint action $a := \{a_1, \dots, a_n\}$ and the state of the world is θ, and time varying local beliefs μ_{it} that assign probabilities to different realizations of the state of the world. We assume that the state of the world is chosen by nature from a space Θ and that actions are chosen from a common, finite, and time invariant set so that we have $a_{it} \in \mathcal{A} := \{1, \dots, m\}$ for all times t and agents i. Local payoffs are then formally defined as functions $u_i : \mathcal{A}^n \times \Theta \to \mathbb{R}$. To emphasize the global dependence of local payoffs we write payoff values as $u_i(a, \theta) = u_i(a_i, a_{-i}, \theta)$ where, we recall, $a_{-i} := \{a_j\}_{j \neq i}$ collects the actions of all agents except i. We assume that the utility values $u_i(a, \theta)$ are finite for all actions a and state realizations θ. The beliefs $\mu_{it} \in \triangle(\Theta)$ are probability distributions on the space Θ.

In general, we allow agent i to maintain a mixed strategy $\sigma_i \in \triangle(\mathcal{A})$ defined as a probability distribution on the action space $\mathcal{A}$ such that $\sigma_i(a_i)$ is the probability that i plays action $a_i \in \mathcal{A}$. The joint mixed strategy profile $\sigma := \{\sigma_1, \dots, \sigma_n\} \in \triangle^n(\mathcal{A})$ is defined as the product distribution of all individual strategies and the mixed strategy of all agents except i is written as $\sigma_{-i} := \{\sigma_j\}_{j \neq i} \in \triangle^{n-1}(\mathcal{A})$. The utility associated with the joint mixed strategy profile σ is then given by

$$u_i(\sigma, \theta) = u_i(\sigma_i, \sigma_{-i}, \theta) = \sum_{a \in \mathcal{A}^n} u_i(a, \theta) \sigma(a). \qquad (1)$$

We further assume that there exists a *global* potential function $u : \mathcal{A}^n \times \theta \mapsto \mathbb{R}$ taking values $u(a, \theta)$ such that for all pairs of action profiles $a = \{a_i, a_{-i}\}$ and $a' = \{a'_i, a_{-i}\}$, state realizations $\theta \in \Theta$, and agents i, the *local* payoffs satisfy

$$u_i(a_i, a_{-i}, \theta) - u_i(a'_i, a_{-i}, \theta) = u(a_i, a_{-i}, \theta) - u(a'_i, a_{-i}, \theta). \qquad (2)$$

The existence of the potential function u is a statement of aligned interests because for a given θ the joint action that maximizes u is a pure Nash equilibrium strategy of the game defined by the u_i utilities [6]. The motivation for considering a potential game is to model a system in which agents play to achieve the game equilibrium in a distributed manner and end up finding the action a that would be selected by a central coordination agent that maximizes the global payoff u. We emphasize, however, that the game may have other equilibria that are not optimal.

The fundamental problem addressed in this paper is that the state of the world θ is unknown to the agents and that different agents have different beliefs μ_{it} on the state of the world. As a consequence, payoffs $u_i(a, \theta)$ cannot be evaluated but estimated and, moreover, estimates of different agents are different. To explain the implications of this latter observation consider the opposite situation in which the agents aggregate their individual beliefs in a common belief

μ. In that case, the payoff estimate

$$u_i(\sigma;\mu) := \int_{\theta\in\Theta} u_i(\sigma,\theta)\, d\mu, \tag{3}$$

can be evaluated by all agents if we assume that the payoff functions u_i are known globally. Assuming global knowledge of payoffs is not always reasonable and it is desirable to devise mechanisms where agents operate without access to the payoff functions of other agents; see, e.g., [7]. Still, an important implication of considering the payoffs in (3) is that optimal behavior is characterized by Nash equilibria. Specifically, for the game defined by the utilities in (3), a Nash equilibrium at time t is a strategy profile σ^* such that no agent has an interest to deviate unilaterally given the common belief μ. I.e., a strategy $\sigma^* = \{\sigma_i^*, \sigma_{-i}^*\}$ such that for all agents i it holds

$$u_i(\sigma_i^*, \sigma_{-i}^*;\mu) \;\geq\; u_i(\sigma_i, \sigma_{-i}^*;\mu), \tag{4}$$

for all other possible strategies σ_i. Given the existence of the potential function u as stated in (2), the Nash equilibrium in (4) with the aggregate belief μ is a proxy for the maximization of the global payoff $u(a;\mu) := \int_{\theta\in\Theta} u(a,\theta)\, d\mu$. In that sense, it represents the best action that the agents could collectively take if they all had access to common information. For future reference we use $\Gamma(\mu)$ to represent the game with players $\mathcal{N}$, action space $\mathcal{A}$ and payoffs $u_i(a;\mu)$,

$$\Gamma(\mu) := \{\mathcal{N}, \mathcal{A}^n, u_i(a;\mu)\}. \tag{5}$$

The game $\Gamma(\mu)$ is said to have complete information.

When agents have different beliefs, the equilibrium strategies of (4) cannot be used as a target behavior because agents lack the ability to determine if a strategy σ_i that they may choose satisfies (4) or not. Indeed, while the complete information game serves as an omniscient reference, agents can only evaluate their expected payoffs with respect to their local beliefs μ_{it},

$$u_i(\sigma;\mu_{it}) = u_i(\sigma_i,\sigma_{-i};\mu_{it}) = \int_{\theta\in\Theta} u_i(\sigma_i,\sigma_{-i},\theta)\, d\mu_{it}. \tag{6}$$

Comparing (3) and (6) we see that the fundamental problem of having different beliefs μ_{it} at different agents is that i lacks information needed to evaluate the expected payoff $u_j(\sigma;\mu_{jt})$ of agent j and, for that reason, the game is said to have incomplete information. A way to circumvent this lack of information is for agent i to keep a belief ν_{jt}^i on the strategy profile of player j. If we group these beliefs to define the joint belief $\nu_{-it}^i := \{\nu_{jt}^i\}_{j\neq i}$ that agent i has on the actions of others, it follows that agent i can evaluate the payoff he can expect of different strategies σ_i as

$$u_i(\sigma_i,\nu_{-it}^i;\mu_{it}) = \int_{\theta\in\Theta} u_i(\sigma_i,\nu_{-it}^i,\theta)\, d\mu_{it}. \tag{7}$$

It is then natural to suggest that agent i should choose the strategy σ_i that maximizes the expected payoff in (7). Such strategy can always be chosen to be an individual action that is termed the best response to the beliefs ν_{-it}^i on the actions of others and the belief μ_{it} on the state of the world,

$$a_{it} \in \operatorname*{argmax}_{a_i\in\mathcal{A}} u_i(a_i,\nu_{-it}^i;\mu_{it}). \tag{8}$$

For future reference we also define the corresponding best expected utility that agent i expects to obtain at time t by playing the best response action in (8),

$$v_i(\nu_{-it}^i;\mu_{it}) := \max_{a_i\in\mathcal{A}} u_i(a_i,\nu_{-it}^i;\mu_{it}). \tag{9}$$

We emphasize that $v_i(\nu_{-it}^i;\mu_{it})$ is *not* the utility actually attained by agent i at time t. That utility depends on the actual state of the world and the actions actually taken by others and is explicitly given by $u_i(a_{it}, a_{-it}, \theta)$.

In this paper we consider agents that select best response actions as in (8) and focus on designing decentralized mechanisms to construct the beliefs ν_{jt}^i on the actions of others so that the actions a_{it} attain desirable properties.

In particular, we assume that there is an underlying state learning process so that the local beliefs μ_{it} on the state of the world converge to a common belief μ in terms of total variation. I.e., we suppose that

$$\lim_{t\to\infty} \text{TV}(\mu_{it},\mu) = 0 \quad \text{for all } i\in\mathcal{N}, \tag{10}$$

where the total variation distance $\text{TV}(\mu_{it},\mu) := \sup_{B\in\mathcal{B}(\Theta)} |\mu_{it}(B) - \mu(B)|$ between distributions μ_{it} and μ is defined as the maximum absolute difference between the respective probabilities assigned to elements B of the Borel set $\mathcal{B}(\Theta)$ of the space Θ.

The desirable property that we ask of the process that builds the beliefs ν_{jt}^i on the actions of others is that the actions a_{it} approach one of the Nash equilibrium strategies defined by (4) as the distributions μ_{it} converge to the common distribution μ. The learning process that we propose for the beliefs ν_{jt}^i is based on building empirical histograms of past plays as we explain in sections III and IV. We will show in sections III-A and IV-A that this procedure yields best responses that approach a Nash equilibrium as long as the convergence of μ_{it} to μ is sufficiently fast. We pursue this developments after a pertinent remark.

Remark 1 The game of incomplete information defined by the payoffs in (6) has equilibria that do not necessarily coincide with the equilibria of the complete information game $\Gamma(\mu)$. It is easy to think that the best response a_{it} in (8) yields the best possible utility u_i for agent i. In fact, it is possible for agent i to do better by reasoning that other agents are also playing best response to their beliefs. Strategies that yield an equilibrium point of this strategic reasoning are defined as the Bayesian Nash equilibria of the incomplete information game – see [13], [15] for a formal definition. We utilize the best responses in (8) because determining Bayesian Nash equilibria requires global knowledge of payoffs and information structures.

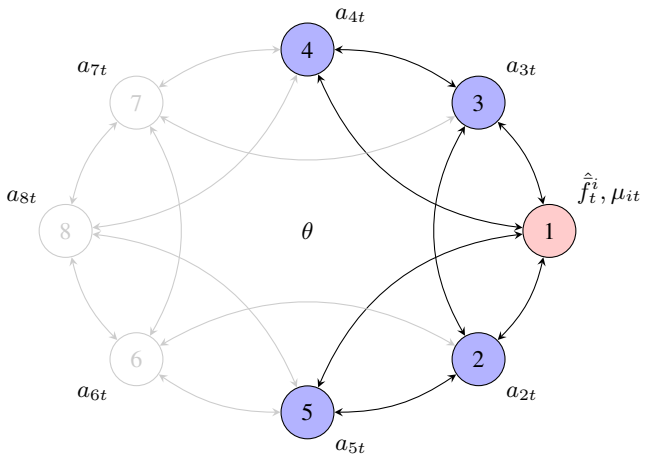


Fig. 1. Distributed fictitious play with observation of neighbors' actions. Agents form beliefs on the strategies of others by keeping a histogram of the average empirical distribution using the observations of actions of their neighbors. E.g., Agent 1 updates an estimate $\hat{\bar{f}}_t^i$ of the average empirical distribution by observing the actions $a_{2t-1}, a_{3t-1}, a_{4t-1}, a_{5t-1}$ played by agents 2, 3, 4, and 5 at time $t-1$ [cf. (16)]. It then selects the best response in (8) which assumes all other agents are playing with respect to the mixed strategy $\nu_{jt}^i = \hat{\bar{f}}_t^i$ given its belief μ_{it} on the state θ.

III. Distributed fictitious play in symmetric potential games

We begin by considering the particular case of symmetric potential games to illustrate concepts, methods, and proof techniques. In a symmetric game, agents' payoffs are permutation invariant in that we have $u_i(a_i, a_j, a_{-i\setminus j}, \theta) = u_j(a_j, a_i, a_{-j\setminus i}, \theta)$ for all pairs of agents i and j. It follows from this assumption that the game admits at least one consensus Nash equilibrium strategy and that, as a consequence, we can utilize a variation of fictitious play [6], [17] in which agents form beliefs on the actions of others by keeping a histogram of actions they have seen taken by other agents in past plays.

Formally, let $f_{it} \in \mathbb{R}^{m\times 1}$ denote the empirical histogram of actions taken by i until time t and define the vector indicator function $\Psi(a_{it}) = [\Psi_1(a_{it}), \dots, \Psi_m(a_{it})] : \mathcal{A} \to \{0,1\}^m$ such that the kth component is $\Psi_k(a_{it}) = 1$ if and only if $a_{it} = k$ and $\Psi_k(a_{it}) = 0$ otherwise. Given the definition of the vector indicator function $\Psi(a_{it})$ it follows that the empirical distribution f_{it} of actions taken by i up until time $t > 1$ is

$$f_{it} := \frac{1}{t-1}\sum_{s=1}^{t-1} \Psi(a_{is}). \tag{11}$$

The expression in (11) is simply a vector arrangement of the average number of times that each of the m possible plays $k \in \{1, \dots, m\}$ has been chosen by i. Since the game, being symmetric, admits at least one symmetric Nash equilibrium, the histogram of the empirical play of the population as a whole is also of interest. Using the definition of the vector indicator $\Psi(a_{it})$ this empirical distribution is written as

$$\bar{f}_t := \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{t-1}\sum_{s=1}^{t-1}\Psi(a_{is})\right]. \tag{12}$$

In conventional fictitious play, agents play best responses to the composite empirical distribution in (12). Here, however, we assume that agents are part of a connected network G with node set $\mathcal{N}$ and edge set $\mathcal{E}$. Agent i can only interact with neighboring agents $j \in \mathcal{N}_i := \{j \in \mathcal{N} : (j,i) \in \mathcal{E}\}$. At time t, the actions of neighboring agents $j \in \mathcal{N}_i$ become known to agent i either through explicit communication or implicit observation. In this setting, agent i cannot keep track of the empirical distribution in (12) because it only observes its neighbors' actions $a_{\mathcal{N}_i t} := \{a_{jt} : j \in \mathcal{N}_i\}$ – see Fig. 1.

What is possible for agent i to compute is an estimate of (12) utilizing the information it has available. This estimate is built by averaging the plays of neighbors so that if we write i's estimate of $\bar{f}_t$ as $\hat{\bar{f}}_t^i$ it follows that for $t \geq 2$ it holds

$$\hat{\bar{f}}_t^i = \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}\left[\frac{1}{t-1}\sum_{s=1}^{t-1}\Psi(a_{js})\right]. \tag{13}$$

In the distributed fictitious play algorithm considered here, agent i has access to the state belief μ_{it} and the estimate on the average empirical distribution $\hat{\bar{f}}_t^i$ in (13). Agent i proceeds to select the best response action a_{it} that maximizes its expected payoff [cf. (8)] assuming that all other agents play with respect to the estimated average empirical distribution. I.e., the action played by agent i is computed as per (8) with $\nu_{jt}^i = \hat{\bar{f}}_t^i$ for all $j \neq i$.

Observe that computation of the histograms in (11) - (13) does not require keeping the history of past plays a_{is} for $s < t$. Indeed, the empirical distribution f_{it} in (11) can be expressed recursively as

$$f_{it+1} = f_{it} \; + \; \frac{1}{t}\Big(\Psi(a_{it}) - f_{it}\Big), \tag{14}$$

Likewise, we can also write the population's empirical distribution $\bar{f}_t$ in (12) recursively as

$$\bar{f}_{t+1} = \bar{f}_t + \frac{1}{t}\left[\frac{1}{n}\sum_{j=1}^{n}\Psi(a_{jt}) - \bar{f}_t\right], \tag{15}$$

and the same rearrangement permits writing the estimate $\hat{\bar{f}}_t^i$ that i keeps of the population's empirical distribution as the recursion

$$\hat{\bar{f}}_{t+1}^i = \hat{\bar{f}}_t^i + \frac{1}{t}\left[\frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}\Psi(a_{jt}) - \hat{\bar{f}}_t^i\right]. \tag{16}$$

In all three cases the recursions are valid for $t \geq 1$ and we have to make $f_{i1} = \bar{f}_1 = \hat{\bar{f}}_1^i = \mathbf{0}$ for the recursions in (14) - (16) to be equivalent to (11) - (13). However, subsequent convergence analyses rely on (14) - (16) and allow for arbitrary initial distributions $\hat{\bar{f}}_1^i$. This is important because, in general, agent i may have side information on the actions a_{j1} that other agents are to chose at time $t = 1$.

We can now summarize the behavior of agent i as follows:

(i) At time t update the state's belief to μ_{it}. (ii) Play the best response action a_{it} in (8) with $\nu^i_{jt} = \hat{\bar{f}}^i_t$ for all $j \neq i$, (iii) Learn the actions a_{it} of neighbors, either through observation or communication, and update $\hat{\bar{f}}^i_{t+1}$ as per (16) – with the empirical histogram $\hat{\bar{f}}^i_1$ initialized to an arbitrary distribution. We show in the following that when all agents follow this behavior, their strategies converge to a consensus Nash equilibrium of the symmetric potential game $\Gamma(\mu)$ [cf. (5)].

A. Convergence

Game equilibria are not unique in general. Consider then the stage game $\Gamma(\mu)$ with a given common belief μ on the state of the world θ. The the set of Nash equilibria of the game $\Gamma(\mu)$ contains all the strategies that satisfy (4),

$$K(\mu) = \{\sigma^* : u_i(\sigma^*; \mu) \geq u_i(\sigma_i, \sigma^*_{-i}; \mu), \text{ for all } i, \sigma_i\}. \tag{17}$$

In the distributed fictitious play process described by (8) and (16) agents assume that all other agents select actions from the same distribution. Therefore, it is reasonable to expect convergence not to an arbitrary equilibrium but to a consensus equilibrium in which all agents play the same strategy. Define then the set of consensus equilibria of the game $\Gamma(\mu)$ as the subset of the Nash equilibria set $K(\mu)$ defined in (17) for which all agents play according to the same strategy,

$$C(\mu) = \{\sigma^* = \{\sigma^*_1, \ldots, \sigma^*_n\} \in K(\mu) : \sigma^*_1 = \ldots = \sigma^*_n\}. \tag{18}$$

We emphasize that not all potential games admit consensus Nash equilibria, but the symmetric potential games considered in this section do have a nonempty set of consensus Nash equilibria; see e.g., [7].

We prove here that the best response actions in (8) are eventually drawn from a consensus equilibrium strategy if the local empirical histograms $\hat{\bar{f}}^i_{t+1}$ are updated according to (16) and the local state beliefs μ_{it} converge to the common belief μ in the sense stated in (10). In the proof of this result we make use of the following assumptions on the network topology and the state learning process.

Assumption 1 *The network $G(\mathcal{N}, \mathcal{E})$ is strongly connected.*

Assumption 2 *For all agents $i \in \mathcal{N}$, the local beliefs μ_{it} converge to a common belief μ at a rate faster than $\log t/t$,*

$$TV(\mu_{it}, \mu) = O\left(\frac{\log t}{t}\right). \tag{19}$$

Assumption 1 is a simple connectivity requirement to ensure that the actions taken by any node eventually become known to all other agents. Assumption 2 requires that agents reach the common belief μ fast enough. This assumption is fundamental to subsequent proofs but is not difficult to satisfy – see Remark 2. We note that the common belief μ is an *arbitrary* belief on the state θ to which all agents converge but is not necessarily the optimal Bayesian aggregate of the information that different agents acquire about the state of the world. Validity of these two assumptions guarantees convergence of the best response actions in (8) as we formally state next.

Theorem 1 *Consider a symmetric potential game $\Gamma(\mu)$ and the distributed fictitious play updates where at each stage agents best respond as in* (8) *with local beliefs $\nu^i_{jt} = \hat{\bar{f}}^i_t$ for all $j \neq i$ formed using* (16) *and belief μ_{it}. If Assumptions 1 and 2 are satisfied and the initial estimated average beliefs are the same for all agents, i.e., if $\hat{\bar{f}}^i_1 = \hat{\bar{f}}^j_1$ for all $i \in \mathcal{N}$ and $j \in \mathcal{N}$, then the average empirical distribution $\bar{f}_t \in \triangle(\mathcal{A})$ converges to a strategy that is an element $\sigma^*_i \in \triangle(\mathcal{A})$ of a strategy profile $\sigma^* \in \triangle^n(\mathcal{A})$ that belongs to the set of consensus Nash equilibria $C(\mu)$ of the symmetric potential game $\Gamma(\mu)$. I.e.,*

$$\lim_{t\to\infty} \|\bar{f}_t - \sigma^*_i\| = 0 \tag{20}$$

where $\sigma^*_i \in \sigma^*$, $\sigma^* \in C(\mu)$, and $\|\cdot\|$ denotes the $\mathcal{L}^2$ norm on the Euclidean space.

Proof: See Appendix B. ■

Since in a consensus Nash equilibrium all agents play according to the same strategy, $\sigma^*_i = \sigma^*_j$ for all i and j, Theorem 1 also means that the n-tuple of the population's average empirical distribution $\bar{f}_t$ is a consensus Nash equilibrium strategy profile $\sigma^* \in C(\mu)$. Notice that this result is not equivalent to showing that each agent's empirical frequency f_{it} is a consensus Nash equilibrium strategy. However, i's model of other agents $\hat{\bar{f}}^i_t$ converges to the average empirical distribution $\bar{f}_t$. In particular, we have $\|\hat{\bar{f}}^i_t - \bar{f}_t\| = O(\log t/t)$ by Lemma 5. Hence, agents do learn to best respond to the consensus equilibrium strategy $\bar{f}_t$. In order for agent i's individual empirical frequency f_{it} to converge to the consensus Nash equilibrium strategy $\bar{f}_t$, the utility function should be such that agents are not indifferent between two actions when they best respond in (8) to the equilibrium strategies of others as we show next.

Corollary 1 *In a distributed fictitious play with action sharing, if the potential function $u(\cdot)$ is such that for $\nu^i_{jt} = \bar{f}_t$ for all $j \neq i$ the maximizing action is unique asymptotically, that is, there exists $a^* \in \mathcal{A}$ such that*

$$\lim_{t\to\infty} u(a^*, \nu^i_{-it}; \mu) - u(a, \nu^i_{-it}; \mu) \geq \epsilon \tag{21}$$

for $\epsilon > 0$ and for all $a \in \mathcal{A} \setminus a^*$ then each agent learns to play according to an empirical frequency that is in equilibrium with others' empirical frequencies for any symmetric potential game $\Gamma(\mu)$, that is,

$$\lim_{t\to\infty} \|f_{it} - \bar{f}_t\| = 0. \tag{22}$$

Proof: See Appendix C. ■

The condition in (21) says that there exists a single distinct action a^* that strictly maximizes the expected utility asymptotically when other agents follow $\bar{f}_t$. We obtain the result

in (22) by leveraging the fact that asymptotically agents' estimates of the average empirical distribution $\hat{\bar{f}}_t^i$ converge to $\bar{f}_t$ and there exists a finite time after which action a^* will be chosen. The result above implies that agents eventually play according to a consensus Nash equilibrium action. Note that the responses of agents during the distributed fictitious play depend on both the state learning process and the process of agents forming their estimates on the average empirical distribution. The results in this section reveal that these two processes can be designed independently as long as both processes converge at a fast enough rate. We make use of this separation in the next section to design a distributed fictitious play process that converges to an equilibrium strategy of any potential game.

Remark 2 The $\log t/t$ convergence rate in Assumption 2 is satisfied by various distributed learning methods including averaging, e.g., [22], [23]; decentralized estimation, e.g., [5], [24]–[28]; social learning models, e.g., [29], [30]; and Bayesian learning, e.g., [31]–[34]. In the way of illustration, averaging models have agent i sharing its previous belief on the state with its neighbors and updating its belief by a weighted averaging of observed distributions that follow the recursion

$$\mu_{it}(\theta) = \sum_{j\in\mathcal{N}_i} w_{ij}\mu_{jt-1}(\theta), \tag{23}$$

for some set of doubly stochastic weights w_{ij}. Convergence to a common distribution follows an exponential rate $O(c^t)$ if all the information available to agents are private observations at time $t = 1$ [23], [35]. In Bayesian learning we can do away with communication altogether and assume that agents keep acquiring private information on θ that they incorporate in the local beliefs μ_{it} using Bayes' law. If the local signals are informative, all agents converge to an atomic belief with all probability in the true state of the world. Linear – meaning $O(1/t)$ – rates of convergence can be achieved with mild assumptions on the rate of novel information [31]. Bayesian updates utilizing neighbors' beliefs are also possible, if computationally cumbersome, and also achieves $O(1/t)$ convergence with mild assumptions [32]–[34].

IV. Distributed Fictitious Play in Generic Potential Games

For generic potential games we consider a distributed fictitious play process in which agents communicate the histograms they keep on the other agents with their neighbors. I.e., Agent i shares its entire belief ν^i_{-it} with its neighbors at each time step in addition to its action a_{it}. When compared to the distributed fictitious play with action observations, the additional information communicated allows agents to keep distinct beliefs on other agents as we explain in the following.

Agent i can keep track of the individual empirical histograms of its neighbors $\{f_{jt}\}_{j\in\mathcal{N}_i}$ by (14) using observations of the actions of its neighbors $\{a_{jt}\}_{j\in\mathcal{N}_i}$, that is, $\nu^i_{jt+1} = f_{jt+1}$ for $j \in \mathcal{N}_i$. Otherwise, agent i can keep an estimate of the empirical histogram of its non-neighbors $j \notin \mathcal{N}_i$ by averaging the estimates of its neighbors on the non-neighboring agent j $\{\nu^k_{jt}\}_{k\in\mathcal{N}_i}$. I.e. the estimate of agent i on $j \notin \mathcal{N}_i$ is given by

$$\nu^i_{jt+1} = \sum_{k\in\mathcal{N}} w^i_{jk}\nu^k_{jt} \tag{24}$$

where $w^i_{jk} > 0$ if and only if $k \in \mathcal{N}_i$ and $\sum_{k\in\mathcal{N}} w^i_{jk} = 1$. Note that in this belief formation, agent i keeps a separate belief on each individual and has the correct estimate of the empirical frequency of its neighbors.

Besides the difference in belief updates the distributed fictitious play is identical to the behavior described in Section III. To summarize, at time t agent i updates its belief on the state μ_{it}, plays with respect to the best response action a_{it} in (8) with beliefs ν^i_{-it}, observe actions and beliefs of neighbors $\{a_{jt}, \nu^j_{-jt}\}_{j\in\mathcal{N}_i}$, and update ν^i_{jt+1} for $j \neq i$ by (14) if $j \in \mathcal{N}_i$ or by (24) if $j \notin \mathcal{N}_i$. In the following, we show the convergence of the empirical frequencies to a Nash equilibrium strategy for any potential game when agents follow the behavior described above.

A. Convergence

Next, we present the main result of the paper that shows that the best responses in the distributed fictitious play with histogram sharing converge to a Nash equilibrium strategy (17) of any potential game $\Gamma(\mu)$ given the same assumptions on network connectivity and on convergence of the state learning process as in Theorem 1.

Theorem 2 *Consider a potential game $\Gamma(\mu)$ and the distributed fictitious play updates where at each stage agents best respond as in* (8) *with local beliefs ν^i_{-it} formed using* (14) *if $j \in \mathcal{N}_i$ or using* (24) *if $j \notin \mathcal{N}_i$, and state belief μ_{it}. If Assumptions 1 and 2 are satisfied then the empirical frequencies of agents $f_t := \{f_{jt}\}_{j\in\mathcal{N}} \in \triangle^n(\mathcal{A})$ defined in* (11) *converge to a Nash equilibrium strategy of the potential game with common state of the world beliefs μ, that is,*

$$\min_{\sigma^*\in K(\mu)} \|f_t - \sigma^*\| \to 0 \tag{25}$$

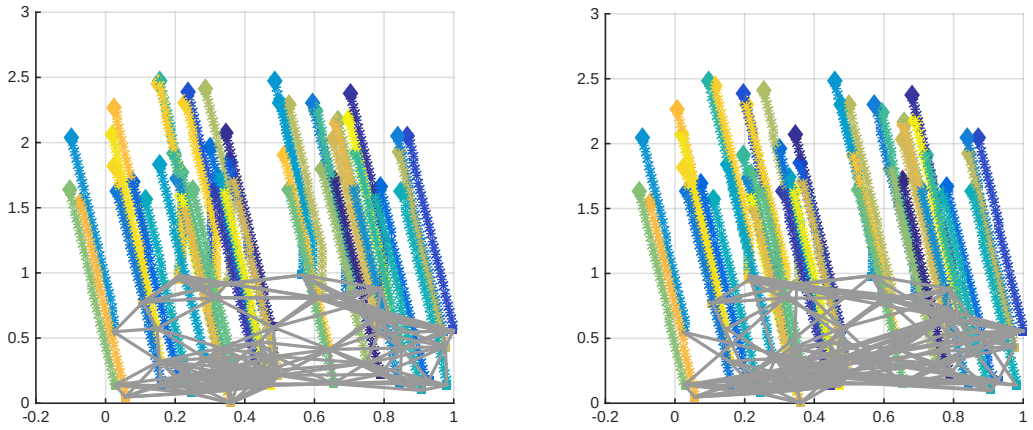
where $K(\mu)$ is the set of Nash equilibria of the game $\Gamma(\mu)$ (17)*.*

Proof: See Appendix D. ∎

The above result implies that when agents share their beliefs on others' histograms and based on this information keep an estimate of the empirical distribution of each agent, their responses converge to a Nash equilibrium of the potential game as long as their beliefs on the state reach consensus at a belief μ fast enough. Theorem 2 generalizes Theorem 1 to the class of potential games given that agents in addition to their actions communicate their beliefs on others with their neighbors.

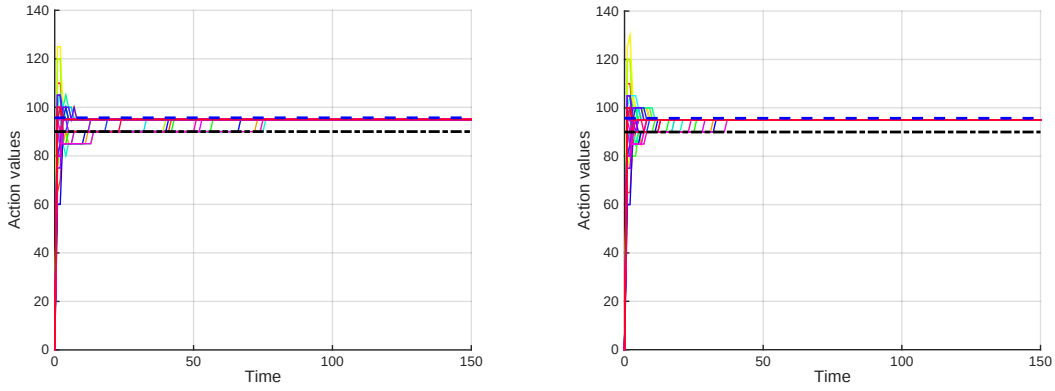
V. Simulations

We explore the effects of the network connectivity, the state learning process and the payoff structure on the per-



(a) (b)

Fig. 2. Position of robots over time for the geometric (a) and small world networks (b). Initial positions and network is illustrated with gray lines. Robots' actions are best responses to their estimates of the state and of the centroid empirical distribution for the payoff in (27). Robots recursively compute their estimates of the state by sharing their estimates of θ with each other and averaging their observations. Their estimates on the centroid empirical distribution is recursively computed using (16). Agents align their movement at the direction 95° while the target direction is $\theta = 90^\circ$.



(a) (b)

Fig. 3. Distributed fictitious play actions of robots over time for the geometric (a) and small world networks (b). Solid lines correspond to each robots' actions over time. The dotted dashed line is equal to value of the state of the world $\theta = 90^\circ$ and the dashed line is the optimal estimate of the state given all of the signals which is equal to 96.1°. Agents reach consensus in the movement direction 95° faster in the small-world network than the geometric network.

formance of the algorithm in two games named the beauty contest game, and the target covering game.

A. *Beauty contest game*

A network of $n = 50$ autonomous robots want to move in coordination and at the same time follow a target direction $\theta = [0^\circ, 180^\circ]$ in a two dimensional topology. Each robot receives an initial noisy signal related to the target direction θ,

$$\pi_i(\theta) = \theta + \epsilon_i \tag{26}$$

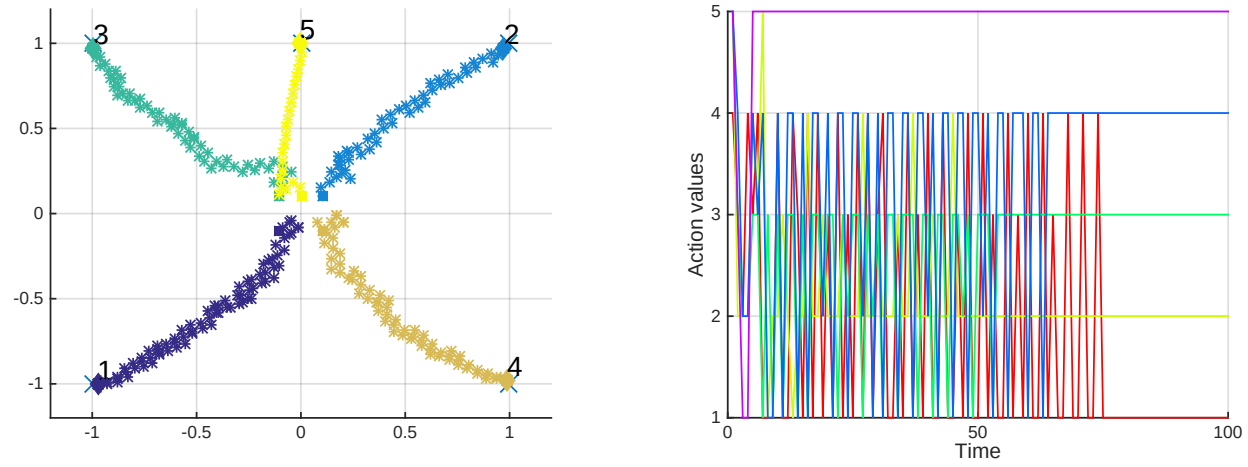
where ϵ_i is drawn from a zero mean normal distribution with standard deviation equal to 20°. Actions of robots determine their direction of movement and belong to the same space as θ but are discretized in increments of 5°, i.e., $\mathcal{A} = (0^\circ, 5^\circ, 10^\circ, \ldots, 180^\circ)$. The estimation and coordination payoff of robot i is given by the following utility function

$$u_i(a, \theta) = -\lambda(a_i - \theta)^2 - (1-\lambda)\left(a_i - \frac{1}{n-1}\sum_{j \neq i} a_j\right)^2 \tag{27}$$

where $\lambda \in (0, 1)$ gauges the relative importance of estimation and coordination. The game is a symmetric potential game and hence admits a consensus equilibrium for any common belief on θ [14].

In the following numerical setup, we choose θ to be equal to 90°. We assume that all robots start with a common prior on the centroid empirical distribution in which they believe that each action is drawn with equal probability. They follow the distributed fictitious play updates with action sharing described in Section III. State learning process is chosen as the averaging model in which robots update their beliefs on the state θ using (23) with initial beliefs formed based on the initial private signal with signal generating function in (26).

In Figs. 2 and 3, we plot robots' positions and their actions,



(a) (b)

Fig. 4. Locations (a) and actions (b) of robots over time for the star network. There are $n = 5$ robots and targets. In (a), the initial positions of the robots are marked with squares. The robots' final positions at the end of 100 steps are marked with a diamond. The positions of the targets are indicated by '$\times$'. Robots follow the histogram sharing distributed fictitious play presented in Section IV. The stars in (a) represent the position of the robots at each step of the algorithm. The solid lines in (b) correspond to the actions of robots over time. All targets are covered by a single robot before 100 steps.

respectively. In Fig. 2, we assume that robot i moves with a displacement of 0.01 meters in the chosen direction a_{it} at stage t. Figs. 2(a) and 3(a) correspond to the behavior in a geometric network when robots are placed on a 1 meter $\times$ 1 meter square randomly and connecting pairs with distance less than 0.3 meter between them. Figs. 2(b) and 3(b) correspond to the behavior in a small-world network when the edges of the geometric network are rewired with random nodes with probability 0.2. The geometric network illustrated in Fig. 2(a) has a diameter of $\Delta_g = 5$ with an average length among users equal to 2.5[1]. The small world network illustrated in Fig. 2(b) has a diameter of $\Delta_r = 4$ with an average length among users equal to 2. We observe that the agents reach consensus at the action 95° in both networks but the convergence is faster in the small-world network (39 steps) than the geometric network (78 steps).

We further investigate the effect of the network structure in convergence time by considering 50 realizations of the geometric network and 50 small-world networks generated from the realized geometric networks with rewire probability of 0.2. The average diameter of the realized geometric networks was 5.1 and the average diameter of the realized small-world networks was 4.1. The mean of the average length of the realized geometric networks was 2.27 while the same value was 1.96 for the realized small-world networks. We considered a maximum of 500 iterations for each network. Among 50 realizations of the geometric network, the distributed fictitious play behavior failed to reach consensus in action within 500 steps in 18 realizations while for small-world networks the number of failures was 5. The average time to convergence among the 50 realizations was 228 steps for the geometric network whereas the convergence took 100 steps for the small-world network on average. In addition, convergence time for the small-world network is observed to be shorter than the corresponding geometric network in all of the runs except one.

[1]Diameter is the longest shortest path among all pairs of nodes in the network. The average length is the average number of steps along the shortest path for all pairs of nodes in the network.

B. Target covering game

n autonomous robots want to cover n targets. The position of a target $k \in \mathcal{T} := \{1, \ldots, n\}$ on the two dimensional space is denoted by $\theta_k \in \mathbb{R}^2$ and are not known by the robots. Robot i starts from an initial location $x_i \in \mathbb{R}^2$ and makes noisy observations s_{ik0} of the location of target k coming from normal distribution with mean θ_k and standard deviation equal to $\sigma\mathbf{I}$ where $\mathbf{I}$ is the 2×2 identity matrix and $\sigma > 0$ for all $k \in \mathcal{T}$. At each stage robots choose one of the targets, that is, $\mathcal{A} = \mathcal{T}$ and receives a payoff from covering that target that is inversely proportional to its distance from the target if no other robot is covering it, that is, the payoff of robot i from covering target $k \in (1, \ldots, n)$ $a_i = k$ is given by

$$u_i(a_i = k, a_{-i}, \theta) = \mathbf{1}\left(\sum_{j \neq i} \mathbf{1}\left(a_j = k\right) = 0\right) h(x_i, \theta_k) \tag{28}$$

where $\mathbf{1}(\cdot)$ is the indicator function and $h(\cdot)$ is a reward function inversely proportional to the distance between the target and the robot's initial position x_i, e.g., $\|x_i - \theta_k\|^{-2}$. The first term in the multiplication above is one if no one else chooses target k otherwise it is zero. The second term in the multiplication decreases with growing distance between robot i's initial position x_i and the target k's position θ_k.

When all of the robots start from the same location, that is, $x_i = x$ for all $i \in \mathcal{N}$, the game with payoffs above can be shown to be a potential game by using the definition of potential games in (2). Furthermore, the game is symmetric. In this setup, we would like each robot to assign itself to

a single target different from the rest of the robots, that is, we are interested in convergence to a pure strategy Nash equilibrium in which each robot picks a single action similar to the target assignment games considered in [20]. Observe that the target covering game can not have a pure consensus equilibrium strategy. To see this, assume that all robots cover the same target then they all receive a payoff of zero. Any robot that deviates to another target receives a positive payoff. Therefore, there cannot be a pure consensus strategy equilibrium. As a result, instead of the action sharing scheme, we consider the histogram sharing distributed fictitious play by which it is possible but not guaranteed that the robots converge to a pure strategy Nash equilibrium.

In the numerical setup, we consider $n = 5$ robots with the payoffs in (28) and n targets. The locations of targets are respectively given as follows $\theta_1 = (-1, -1)$, $\theta_2 = (1, 1)$, $\theta_3 = (-1, 1)$, $\theta_4 = (1, -1)$, $\theta_5 = (0, 1)$. We consider the case that the initial positions of robots are different from each other with the reward function $h(x_i, \theta_k) = \|x_i - \theta_k\|^{-2}$. Specifically, the initial positions of the robots equal to $x_1 = (-0.1, -0.1)$, $x_2 = (0.1, 0.1)$, $x_3 = (-0.1, 0.1)$, $x_4 = (0.1, -0.1)$, and $x_5 = (0, 0.1)$. Robots make noisy observations s_{ikt} for all $k \in \mathcal{T}$ after each step. The observations have the same distribution as s_{ik0} with $\sigma = 0.2$ meters. We assume that the robots update their beliefs on the positions of targets using the Bayes' rule based on the observations.

Figs. 4(a)-(b) shows the movement of robots and actions of robots over time, respectively, for the star network. We assume that robots move by a distance of 0.02 meters along the estimated direction of the target they choose at each step of the distributed fictitious play. The estimated direction is a straight line from the current position to the estimated position of the chosen target. I.e., the robots make observations and decisions in every 0.02 meters. Finally, we assume that the robot covers the target if it is 0.05 meters away from a target and no other robot covers it. In figs. 4(a)-(b), we observe that each robot comes to 0.05 meters neighborhood of a target within 100 steps. Furthermore, the robots cover all of the targets, that is, they converge to a pure Nash equilibrium.

Next, we compare the distributed fictitious play algorithm to the centralized (optimal) algorithm. In the centralized algorithm, at the beginning of each step agents aggregate their signals and then take the action to maximize the expected global objective defined as the sum of the utilities of all (28). Since there exists multiple equilibria in the complete information target coverage game, it is not guaranteed that the distributed fictitious play algorithm converges to the optimal equilibrium at each run. For this purpose, we considered 50 runs of the algorithm where in each run signals are generated from different seeds. We assume that the algorithm has converged when each target is covered by a robot within 0.05 meters from the target. In Fig. 5, we plot the evolution of the global utility with respect to time for the distributed fictitious play algorithm runs with the best and the worst final payoff, and for the centralized algorithm. The best final configuration overlaps with the final centralized solution which is given by

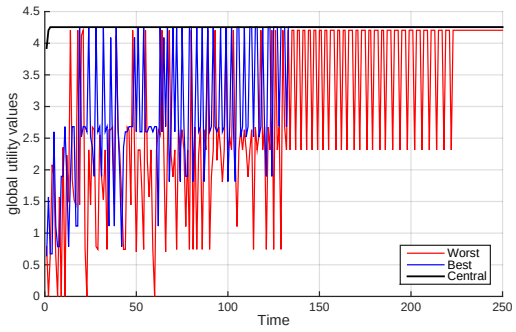


Fig. 5. Comparison of the distributed fictitious play algorithm with the centralized optimal solution. Best and worst correspond to the runs with the highest and lowest global utility in the distributed fictitious play algorithm. The algorithm converges to the global optimal point in 40 runs out of a total of 50 runs.

$a = [1, 2, 3, 4, 5]$ resulting in a global objective value of 4.25. The worst final configuration is given by $a = [1, 5, 3, 4, 2]$ resulting in a global objective value of 4.20.

VI. Conclusion

This paper considered the optimal behavior of multi-agent systems with uncertainty on the state of the world. The fundamental problem of interest was to have a model of optimal agent behavior given a global objective that depends on the state and actions of all the agents when agents have different beliefs on the state. We posed the setup as a potential game with the global objective as the common utility of the multi-agent system and set the optimal behavior as a Nash equilibrium strategy of the game when agents have common beliefs on the state of the environment. We presented a class of distributed algorithms based on the fictitious play algorithm in which agents reach an agreement on their state beliefs asymptotically through an exogenous process, build beliefs on the behavior of others based on information from neighbors and best respond to their expected utility given their beliefs on the state and others.

We considered two information exchange scenarios for the algorithm where in the first scenario agents communicated their actions. For this scenario we showed that when the agents keep track of the population's average empirical frequency of actions as a belief on the behavior of every other individual, their behavior converges to a consensus Nash equilibrium of any symmetric potential game with common beliefs on the state. In the second scenario we considered agents exchanging their entire beliefs on others in addition to their actions. For this scenario we proposed averaging of the observed histograms as a model for keeping beliefs on the behavior of others and showed that their empirical frequency converges to a Nash equilibrium of any potential game. We exemplified the algorithm in a coordination game – a symmetric potential game – and a target covering game – a potential game. In these examples, we observed that the diameter of the network is influential in convergence rate where the shorter the diameter is, the faster the convergence is.

APPENDIX A
DEFINITIONS AND INTERMEDIATE CONVERGENCE RESULTS

We define notions that relate closeness of a strategy to the set of consensus Nash equilibria of the game $\Gamma(\mu)$.

The distance of a strategy $\sigma \in \triangle^n(\mathcal{A})$ from the set of consensus Nash equilibria $C(\mu)$ is given by

$$d(\sigma, C(\mu)) = \min_{g \in C(\mu)} \|\sigma - g\|. \tag{29}$$

The set of consensus strategies that are ϵ away from the consensus Nash equilibrium set (18) is the ϵ-consensus Nash equilibrium strategy set, that is,

$$C_\epsilon(\mu) = \{\sigma^* \in \triangle^n(\mathcal{A}) : u_i(\sigma^*; \mu) \geq u_i(\sigma_i, \sigma^*_{-i}; \mu) - \epsilon, \text{ for all } \sigma_i \in \triangle(\mathcal{A}), \text{for all } i, \sigma_1 = \sigma_2 = \cdots = \sigma_n\} \tag{30}$$

for $\epsilon > 0$.

We define the δ-consensus neighborhood of $C(\mu)$ as

$$B_\delta(C(\mu)) = \big\{\sigma \in \triangle^n(\mathcal{A}) : d(\sigma, C(\mu)) < \delta, \sigma_1 = \sigma_2 = \cdots = \sigma_n\big\}. \tag{31}$$

Note that the δ consensus neighborhood is defined as the set of consensus strategies that are close to the set $C(\mu)$. We can similarly define the ϵ Nash equilibrium set $K_\epsilon(\mu)$ and δ neighborhood of $K(\mu)$ in (17) as $B_\delta(K(\mu))$ by removing the agreement constraint on the equilibrium strategies [7].

The following intermediate results can be found in Appendix B in [7]. They are stated here for completeness.

Lemma 1 *If the processes $g_t \in \triangle^n(\mathcal{A})$ and $h_t \in \triangle^n(\mathcal{A})$ are such that $\|g_{-it} - h_{-it}\| = O(\log t/t)$ for all $i \in \mathcal{N}$ and the state learning process for all $i \in \mathcal{N}$ generates estimate beliefs $\{\{\mu_{it}\}_{t=0}^\infty\}_{i\in\mathcal{N}}$ that satisfy Assumption 2, then for a potential payoff u in* (2) *the following is true for the expected utility of best response behavior $v(\cdot)$ in* (9)*,*

$$\|v(g_{-it}; \mu_{it}) - v(h_{-it}; \mu)\| = O(\frac{\log t}{t}). \tag{32}$$

Proof: The proof is detailed in Lemma 4 in [7]. The proof follows by first making the observation that the expected utility defined in (1) for the potential function is Lipschitz continuous, and second using the definition of the Lipschitz continuity to bound the difference between the best response expected utilities in (9) for g_{-it}, μ_{it} and h_{-it}, μ by the distance between g_{-it}, μ_{it} and h_{-it}, μ multiplied by the Lipschitz constant. ■

Lemma 2 *If $\sum_{t=1}^T \frac{\alpha_t}{t} < \infty$ for all $T > 0$, $\|\alpha_t - \beta_t\| = O(\frac{\log t}{t})$ and $\beta_{t+1} \geq 0$ then $\sum_{t=1}^T \frac{\beta_t}{t} < \infty$ as $T \to \infty$.*

Proof: Refer to the proof of Lemma 5 in [7]. ■

Lemma 3 *Denote the n-tuple of the average empirical distribution with $\bar{f}_t^n := \{\bar{f}_t, \ldots, \bar{f}_t\}$. If for any $\epsilon > 0$ the following holds*

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : \bar{f}_t^n \notin C_\epsilon(\mu)\}}{T} = 0 \tag{33}$$

then $\lim_{t\to\infty} d(\bar{f}_t^n, C(\mu)) = 0$ where $d(\cdot,\cdot)$ is the distance defined in (29)*.*

Proof: By Lemma 7 in [7], (33) implies that for a given $\delta > 0$ there exists an $\epsilon > 0$ such that

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : \bar{f}_t^n \notin B_\delta(C(\mu))\}}{T} = 0 \tag{34}$$

Using above equation, the result follows by Lemma 1 in [36]. ■

Lemma 4 *For the potential game with function $u(\cdot)$ in* (2) *and expected best response utility $v(\cdot)$* (9)*, consider a sequence of distributions $f_t \in \triangle^n(\mathcal{A})$ for $t = 1, 2, \ldots$ and a common belief on the state $\mu \in \triangle(\Theta)$. Define the process $\beta_t := \sum_{i=1}^n v(f_{-it}; \mu) - u(f_{it}, f_{-it}; \mu)$ for $t = 1, 2, \ldots$. If*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^T \frac{\beta_t}{t} = 0 \tag{35}$$

then $\lim_{t\to\infty} d(f_t, K(\mu)) = 0$ where $f_t = \{f_{1t}, \ldots, f_{nt}\}$.

Proof: By Lemma 6 in [7], the condition (35) implies that for all $\epsilon > 0$

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : f_t \notin K_\epsilon(\mu)\}}{T} = 0. \tag{36}$$

By Lemma 7 in [7], (36) implies that for all $\delta > 0$ the following is true

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : f_t \notin B_\delta(K(\mu))\}}{T} = 0 \tag{37}$$

The above convergence result yields desired convergence result by Lemma 1 in [36]. ■

APPENDIX B
PROOF OF THEOREM 1

Before we prove the theorem, we present an intermediate result that shows the convergence rate of the belief of agent i on the population's average empirical distribution $\hat{\bar{f}}_t^i$ to the true average empirical distribution of the population $\bar{f}_t^i$.

Lemma 5 *Consider the distributed fictitious play in which the centroid empirical distribution of the population $\bar{f}_t$ evolves according to* (15) *and agents update their estimates on the empirical play of the population $\hat{\bar{f}}_t^i$ according to* (16)*. If the network satisfies Assumption 1 and the initial beliefs are the same for all agents, i.e., $\hat{\bar{f}}_1^i = \bar{f}_1$ for all $i \in \mathcal{N}$, then $\hat{\bar{f}}_t^i$ converges in norm to $\bar{f}_t$ at the rate $O(\log t/t)$, that is, $\|\hat{\bar{f}}_t^i - \bar{f}_t\| = O(\frac{\log t}{t})$*

Proof: See Appendix A in [7] for a proof. ■

Observe that the above result is true irrespective of the game that the agents are playing and uncertainty in the state.

The proof leverages on the fact that the change in the centroid empirical distribution is at most $1/t$ by the recursion in (15). Then by averaging observed actions of neighbors in a strongly connected network the beliefs of agent i on the centroid empirical distribution evolves faster than the change in the centroid empirical distribution.

Proof Theorem 1: Given the recursion for the average empirical distribution in (15), we can write the expected utility for the potential function $u(\cdot)$ when all agents follow the centroid empirical distribution $\bar{f}_{t+1}$ and have identical beliefs μ as follows

$$u(\bar{f}^n_{t+1};\,\mu) = u\left(\bar{f}^n_t + \frac{1}{t}\left(\frac{1}{n}\sum_{i=1}^n \Psi(a_{it}) - \bar{f}^n_t\right);\,\mu\right) \quad (38)$$

where $\bar{f}^n_t := \{\bar{f}_t,\ldots,\bar{f}_t\} \in \triangle^n(\mathcal{A})$ is the n-tuple of the average population empirical distribution. Define the average best response strategy at time t $\bar{\Psi}(a_t) := \frac{1}{n}\sum_{i=1}^n \Psi(a_{it})$. By the multi-linearity of the expected utility, we expand the above expected utility as follows [36]

$$\begin{aligned} u(\bar{f}^n_{t+1};\mu) = u(\bar{f}^n_t;\mu)+ \\ \frac{1}{t}\sum_{i=1}^n u(\bar{\Psi}(a_t),\bar{f}^{n-1}_t;\mu) - u(\bar{f}_t,\bar{f}^{n-1}_t;\mu) + \frac{\delta}{t^2} \end{aligned} \quad (39)$$

where the first order terms of the expansion are explicitly written and the remaining higher order terms are collected to the term δ/t^2.

Consider the total utility term in (39) where agent i is playing with respect to the average best response strategy at time $t+1$ $\bar{\Psi}(a_t)$ and remaining agents use the average empirical distribution $\bar{f}^{n-1}_t$, $\sum_{i=1}^n u(\bar{\Psi}(a_t),\bar{f}^{n-1}_t;\mu)$. By the definition of the average best response strategy, we write the term in consideration as

$$\sum_{i=1}^n u\left(\bar{\Psi}(a_t),\bar{f}^{n-1}_t;\mu\right) = \sum_{i=1}^n u\left(\frac{1}{n}\sum_{i=1}^n \Psi(a_{it}),\bar{f}^{n-1}_t;\mu\right). \quad (40)$$

The following equality can be shown by using the multi-linearity of expectation and permutation invariance of the utility [7],

$$\sum_{i=1}^n u(\bar{\Psi}(a_t),\bar{f}^{n-1}_t;\mu) = \sum_{i=1}^n u(\Psi(a_{it}),\bar{f}^{n-1}_t;\mu). \quad (41)$$

The above equality means that the total expected utility when agents play with the average best response strategy at time t $\bar{\Psi}(a_t)$ against the average empirical distribution $\bar{f}^{n-1}_t$ at time t is equal to the total expected utility when agents best respond to the average population empirical distribution at time t.

We substitute in the above equality (41) for the corresponding term in (39) to get the following

$$\begin{aligned} u(\bar{f}^n_{t+1};\mu) = u(\bar{f}^n_t;\mu)+ \\ \frac{1}{t}\sum_{i=1}^n u(\Psi(a_{it}),\bar{f}^{n-1}_t;\mu) - u(\bar{f}_t,\bar{f}^{n-1}_t;\mu) + \frac{\delta}{t^2}. \end{aligned} \quad (42)$$

We can upper bound the right hand side by adding $|\delta|/t^2$ to the left hand side.

$$\begin{aligned} u(\bar{f}^n_{t+1};\mu) - u(\bar{f}^n_t;\mu) + \frac{|\delta|}{t^2} \geq \\ \frac{1}{t}\sum_{i=1}^n u(\Psi(a_{it}),\bar{f}^{n-1}_t;\mu) - u(\bar{f}_t,\bar{f}^{n-1}_t;\mu) \end{aligned} \quad (43)$$

Define $L_{it} := v_i(\nu^i_{-it};\mu_{it}) - u(\Psi(a_{it}),\bar{f}^{n-1}_t;\mu)$ where $\nu^i_{jt} = \hat{\bar{f}}^i_t$ for $j \neq i$. Note that since agents have identical payoffs, we can drop the subindex of the expected utility of agent i when it best responds to the strategy profile of others $v_i(\cdot)$ defined in Section II to write it as $v(\cdot)$. Now we add and subtract $\sum_{i=1}^n L_{it}/t$ to both sides of the above equation to get the following inequality,

$$\begin{aligned} u(\bar{f}^n_{t+1};\mu) - u(\bar{f}^n_t;\mu) + \frac{|\delta|}{t^2} \\ + \frac{1}{t}\sum_{i=1}^n v(\nu^i_{-it};\mu_{it}) - u(\Psi(a_{it}),\bar{f}^{n-1}_t;\mu) \\ \geq \frac{1}{t}\sum_{i=1}^n v(\nu^i_{-it};\mu_{it}) - u(\bar{f}_t,\bar{f}^{n-1}_t;\mu). \end{aligned} \quad (44)$$

Summing the inequalities above from time $t=1$ to time $t=T$, we get

$$\begin{aligned} u(\bar{f}^n_{T+1};\mu) - u(\bar{f}^n_1;\mu) + \sum_{t=1}^{T+1}\frac{|\delta|}{t^2} + \sum_{t=1}^{T+1}\sum_{i=1}^n \frac{L_{it}}{t} \\ \geq \sum_{t=1}^{T+1}\frac{1}{t}\sum_{i=1}^n v(\nu^i_{-it};\mu_{it}) - u(\bar{f}_t,\bar{f}^{n-1}_t;\mu). \end{aligned} \quad (45)$$

Next we define the following term that corresponds to the inside summation on the right hand side of the above inequality,

$$\alpha_t := \sum_{i=1}^n v(\nu^i_{-it};\mu_{it}) - u(\bar{f}_t,\bar{f}^{n-1}_t;\mu). \quad (46)$$

The term α_t captures the total difference between expected utility when agents best respond to their beliefs on the average population empirical distribution $\nu^i_{jt} = \hat{\bar{f}}^i_t$ and their beliefs on θ μ_{it}, and when they follow the current centroid empirical distribution $\bar{f}_t$ with common beliefs on the state μ. Note that by Lemma 5 and Assumption 2 the conditions of Lemma 1 are satisfied, that is, $\|v(\nu^i_{-it};\mu_{it}) - u(\bar{f}_t,\bar{f}^{n-1}_t;\mu)\| = O(\log t/t)$. By the assumption that utility value is finite and Lemma 1, the left hand side of (45) is

finite. That is, there exists a $\bar{B} > 0$ such that

$$\bar{B} \geq \sum_{t=1}^{T+1} \frac{\alpha_t}{t}. \tag{47}$$

for all $T > 0$. Next, we define the following term

$$\beta_t := \sum_{i=1}^{n} v(\bar{f}_t^{n-1}; \mu) - u(\bar{f}_t, \bar{f}_t^{n-1}; \mu) \tag{48}$$

that captures the difference in expected payoffs when agents best respond to the centroid empirical distribution $\bar{f}_t^{n-1}$ for others given the common asymptotic belief μ, and when everyone follows the current centroid empirical distribution $\bar{f}_t$ with common beliefs on the state μ. When we consider the difference between α_t and β_t, the following equality is true by Lemma 1,

$$\|\alpha_t - \beta_t\| = \|\sum_{i=1}^{n} v(\nu_{-it}^i; \mu_{it}) - v(\bar{f}_t^{n-1}; \mu)\| = O(\frac{\log t}{t}). \tag{49}$$

Further $\beta_t \geq 0$. Hence, the conditions of Lemma 2 are satisfied which implies that the following holds

$$\sum_{t=1}^{T} \frac{\beta_t}{t} < \infty. \tag{50}$$

From the above equation it follows by the Kronecker's Lemma that [37, Thm. 2.5.5]

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \beta_t = 0. \tag{51}$$

The above convergence result implies that by Lemma 6 in [7], for any $\epsilon > 0$, the number of centroid empirical frequencies away from the ϵ consensus NE is finite for any time T, that is,

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : \bar{f}_t^n \notin C_\epsilon(\mu)\}}{T} = 0. \tag{52}$$

The relation above implies that the distance between the empirical frequencies and the set of symmetric NE diminishes by Lemma 3, that is,

$$\lim_{t\to\infty} d(\bar{f}_t^n, C(\mu)) = 0. \tag{53}$$

where $d(\cdot,\cdot)$ is the distance defined in (29). The result in (20) follows from above. ∎

Appendix C
Proof of Corollary 1

Denote the $n-1$ tuple of the average empirical distribution $\bar{f}_t$ by $\bar{f}_t^{n-1}$. By the Lipschitz continuity of the multilinear utility expectation we have that $\|u(a_i, \bar{f}_t^{n-1}; \mu) - u(a_i, \nu_{-it}^i; \mu_{it})\| \leq K\|(\bar{f}_t^{n-1}, \mu) - (\nu_{-it}^i, \mu_{it})\|$ for all a_i where $\nu_{jt}^i = \hat{\bar{f}}_t^i$ for all $j \neq i$ and $K \geq 0$ is the Lipschitz constant. By Lemma 5 and Assumption 2, we have $\|(\bar{f}_t^{n-1}, \mu) - (\nu_{-it}^i, \mu_{it})\| = O(\log t/t)$. Then using (21), we have for all $a \in \mathcal{A} \setminus a^*$

$$u(a^*, \nu_{-it}^i; \mu_{it}) - u(a, \nu_{-it}^i; \mu_{it}) \geq \epsilon - \delta_t \tag{54}$$

for $\nu_{jt}^i = \hat{\bar{f}}_t^i$ for all $j \neq i$, $\delta_t \geq 0$ and $\delta_t = O(\log t/t)$. Therefore, there exists a finite time $T > 0$ such that $\epsilon - \delta_t > 0$ for all $t > T$. This means that a^* is the best response action of i after time T. Then the empirical frequency of $i \in \mathcal{N}$ f_{it} converges to $\Psi(a^*)$ which implies (22).

Appendix D
Proof of Theorem 2

Before we prove the theorem, we first analyze the convergence rate of the histogram sharing presented in Section IV where we defined $\nu_{jt}^i = f_{jt}$ if $j \in \mathcal{N}_i$ or ν_{jt}^i is given by (24) if $j \notin \mathcal{N}_i$.

Denote the lth element of ν_{jt}^i by $\nu_{jt}^i(l)$. Define the matrix that captures population's estimate on j's empirical distribution, $\hat{F}_{jt} := [\nu_{jt}^1, \ldots, \nu_{jt}^n]^T \in \mathbb{R}^{n\times m}$. The lth column of $\hat{F}_{jt}$ represents the population's estimate on j's lth local action denoted by $\hat{F}_{jt}(l) := [\nu_{jt}^1(l), \ldots, \nu_{jt}^n(l)]^T \in \mathbb{R}^{n\times 1}$.

Observe that j's estimate of the frequency of its own action l is correct, that is, $\nu_{jt}^j(l) = f_{jt}(l)$. Define the vector $\mathbf{x}_{jlt} \in \mathbb{R}^{n\times 1}$ where its jth element is equal to the empirical frequency of agent j taking action $l \in \mathcal{A}$, that is, $\mathbf{x}_{jlt}(j) = f_{jt}(l)$, and its other elements are zero. Further define the weighted adjacency matrix for belief update on the frequency of agent j's lth action $W_{jl} \in \mathbb{R}^{n\times n}$ with $W_{jl}(i,k) = w_{jk}^i$ for all i and k. We remind that w_{jk}^i is the weight that i uses to mix agent $j \in \mathcal{N}_i$'s belief on agent $k \notin \mathcal{N}_i$'s empirical distribution in (24). Also note that there are m weight matrices W_{jl} each corresponding to one action $l \in \mathcal{A}$.

The matrix W_{jl} is row stochastic, that is, the sum of row elements of W_{jl} is equal to one for each row by $\sum_{k\in\mathcal{N}_i} w_{jk}^i = 1$ and we have that $W_{jl}(i,j) = 1$ for $j \in \mathcal{N}_i \bigcup i$. The latter condition on W_{jl} is by the fact that if $j \in \mathcal{N}_i$, j sends its action to its neighbor i and hence $\nu_{jt}^i = f_{jt}$. Given these definitions we can write a linear recursion for population's estimate of j's empirical frequency of its lth action

$$\hat{F}_{jt+1}(l) = W_{jl}(\hat{F}_{jt}(l) + \mathbf{x}_{jlt+1} - \mathbf{x}_{jlt}). \tag{55}$$

Note that if the above linear system converges to the true empirical frequency of $f_{jt}(l)$ in all of its elements then it implies that all agents learned its true value.

Next, we analyze the linear update in (55) and show the convergence of the belief of agent i on the population's empirical distribution ν_{-it}^i to the true average empirical distribution of the rest of the population f_{-it} at rate $O(\log t/t)$.

Lemma 6 *Consider the distributed fictitious play in which the empirical distribution of agent j f_{jt} evolves according to* (14) *and agent i updates its estimate on the empirical play of the population ν_{-it}^i according* (14) *if $j \in \mathcal{N}_i$ or using* (24) *if $j \notin \mathcal{N}_i$. If the network satisfies Assumption 1 and the initial beliefs are the same for all agents, i.e., $\nu_{j1}^i = f_{j1}$ for all $i \in \mathcal{N}$, then ν_{jt}^i converges in norm to f_{jt} at the rate $O(\log t/t)$, that is, $\|\nu_{jt}^i - f_{jt}\| = O(\frac{\log t}{t})$ for all $j \in \mathcal{N}$.*

Proof: We consider the difference between the population's estimate of the empirical frequency of j taking action $l \in \mathcal{A}$ and j's true empirical distribution $f_{jt}(l)\mathbf{1}$ by subtracting $f_{jt+1}(l)\mathbf{1}$ from both sides of (55) to get

$$\hat{F}_{jt+1}(l) - f_{jt+1}(l)\mathbf{1} = W_{jl}(\hat{F}_{jt}(l) + \mathbf{x}_{jlt+1} - \mathbf{x}_{jlt}) - f_{jt+1}(l)\mathbf{1}. \tag{56}$$

Since W_{jl} is row stochastic, we can move the last term on the right hand side inside the matrix multiplication,

$$\hat{F}_{jt+1}(l) - f_{jt+1}(l)\mathbf{1} = W_{jl}(\hat{F}_{jt}(l) + \mathbf{x}_{jlt+1} - \mathbf{x}_{jlt} - f_{jt+1}(l)\mathbf{1}). \tag{57}$$

We can equivalently express $f_{jt+1}(l) = f_{jt}(l) + \mathbf{x}_{jlt+1}(j) - \mathbf{x}_{jlt}(j)$ by the definition of the vector $\mathbf{x}_{jlt}$. Substituting this expression for the $f_{jt+1}(l)$ on the right hand side of the above equation we have

$$\hat{F}_{jt+1}(l) - f_{jt+1}(l)\mathbf{1} = W_{jl}\Big(\hat{F}_{jt}(l) + \mathbf{x}_{jlt+1} - \mathbf{x}_{jlt} - (f_{jt}(l) + \mathbf{x}_{jlt+1}(j) - \mathbf{x}_{jlt}(j))\mathbf{1}\Big). \tag{58}$$

Let $\mathbf{y}_t := \hat{F}_{jt}(l) - f_{jt}(l)\mathbf{1}$, then

$$\mathbf{y}_{t+1} = W_{jl}(\mathbf{y}_t + \mathbf{x}_{jlt+1} - \mathbf{x}_{jlt} - (\mathbf{x}_{jlt+1}(j) - \mathbf{x}_{jlt}(j))\mathbf{1}). \tag{59}$$

Let $\boldsymbol{\delta}_t := \mathbf{x}_{jlt+1} - \mathbf{x}_{jlt} - (\mathbf{x}_{jlt+1}(j) - \mathbf{x}_{jlt}(j))\mathbf{1}$. Next, we provide an upper bound for $\|\boldsymbol{\delta}_t\|$ by using the triangle inequality and observing the fact that recursion for fictitious play in (14) can change only the jth element of the vector $\mathbf{x}_{jlt}$ by $1/t$, that is, $\mathbf{x}_{jlt+1}(j) - \mathbf{x}_{jlt}(j) = \frac{1}{t}(\Psi(a_{jt})(l) - f_{jt}(l))$, as follows

$$\|\boldsymbol{\delta}_t\| = \|\mathbf{x}_{jlt+1} - \mathbf{x}_{jlt} - (\mathbf{x}_{jlt+1}(j) - \mathbf{x}_{jlt}(j))\mathbf{1}\| \tag{60}$$

$$\leq \|\mathbf{x}_{jlt+1} - \mathbf{x}_{jlt}\| + \|\mathbf{x}_{jlt+1}(j)\mathbf{1} - \mathbf{x}_{jlt}(j)\mathbf{1}\| \tag{61}$$

$$\leq \frac{1}{t} + \frac{n}{t} = \frac{n+1}{t} = O(\frac{1}{t}). \tag{62}$$

Now consider the row stochastic matrix W_{jl}. Its largest eigenvalue is $\lambda_1 = 1$ and its right eigenvector is equal to column vector of ones $\mathbf{1}$ by the Perron-Frobenius theorem [38, Ch. 2.2]. The left eigenvector associated with the eigenvalue λ_1 is given by $\mathbf{e}_j^T$. This is easy to see when we interpret W_{jl} as representing a Markov chain where state j is an absorbing state and there is a positive transition probability from any other state to state j. Note that once a state i that is a neighbor of j is reached, the transition to state j is with probability 1 due to the update rule. Because the graph $\mathcal{G}$ is strongly connected, for any $i \notin \mathcal{N}_j$ there exists a path to a node $k \in \mathcal{N}_j \bigcup j$. As a result the absorbing state j is reached with positive probability which implies the stationary distribution of the Markov chain is given by $\mathbf{e}_j$, that is, with probability 1 the state is j. Moreover, $\lim_{t\to\infty} W_{jl}^t \to \mathbf{1}\mathbf{e}_j^T$.

Now define the matrix $\overline{W}_{jl} = W_{jl} - \mathbf{1}\mathbf{e}_j^T$. By the fact that the limiting power sequence of the matrix is $\mathbf{1}\mathbf{e}_j^T$, $\lim_{t\to\infty} \overline{W}_{jl}^t \to \mathbf{0}$. By its construction the sum of the row elements of $\overline{W}_{jl}$ is zero for any row, that is, $\overline{W}_{jl}\mathbf{1} = \mathbf{0}_{n\times 1}$. Further note that the jth row of $\overline{W}_{jl}$ is all zeros as well as all the rows k such that $j \in \mathcal{N}_k$.

By using the definition of $\boldsymbol{\delta}_t$, we can equivalently write (59) as

$$\mathbf{y}_{t+1} = W_{jl}(\mathbf{y}_t + \boldsymbol{\delta}_t) = \sum_{s=0}^{t} W_{jl}^{s+1}\boldsymbol{\delta}_{t-s} + W_{jl}^t\mathbf{y}_1 \tag{63}$$

for $t = 1, 2, \ldots$. The second equality follows by writing the first equality for $\{\mathbf{y}_s\}_{s=1,\ldots,t}$ and iteratively substituting each term. Note that by the assumption $\nu_{j1}^i = f_{j1}$, $\mathbf{y}_1 = \mathbf{0}$. So when we consider the norm of $\mathbf{y}_{t+1}$, $\|\mathbf{y}_{t+1}\|$, we are left with

$$\|\mathbf{y}_{t+1}\| = \|\sum_{s=0}^{t} W_{jl}^{s+1}\boldsymbol{\delta}_{t-s}\| \leq \sum_{s=0}^{t} \|W_{jl}^{s+1}\boldsymbol{\delta}_{t-s}\| \tag{64}$$

Now use the decomposition $W_{jl} = \overline{W}_{jl} + \mathbf{1}\mathbf{e}_j^T$ in the above line to get

$$\|\mathbf{y}_{t+1}\| \leq \sum_{s=0}^{t} \|(\overline{W}_{jl} + \mathbf{1}\mathbf{e}_j^T)^{s+1}\boldsymbol{\delta}_{t-s}\| \tag{65}$$

Since $\overline{W}_{jl}\mathbf{1} = \mathbf{0}$, $\mathbf{e}_j^T\overline{W}_{jl} = \mathbf{0}$ and $\mathbf{1}\mathbf{e}_j^T = (\mathbf{1}\mathbf{e}_j^T)^s$ for any $s = 1, 2, \ldots$, we have $W_{jl}^s = \overline{W}_{jl}^s + \mathbf{1}\mathbf{e}_j^T$. Then we can upper bound $\|\mathbf{y}_{t+1}\|$ by using the triangle inequality as follows

$$\|\mathbf{y}_{t+1}\| \leq \sum_{s=0}^{t} \|\overline{W}_{jl}^{s+1}\boldsymbol{\delta}_{t-s}\| + \|(\mathbf{1}\mathbf{e}_j^T)^{s+1}\boldsymbol{\delta}_{t-s}\| \tag{66}$$

Further note $\boldsymbol{\delta}_t(j) = 0$ for any $t = 1, 2, \ldots$ by the definition of $\mathbf{x}_{jlt+1}$ and $\mathbf{x}_{jlt}$, and therefore $\mathbf{e}_j^T\boldsymbol{\delta}_t = 0$, which means the second term on the right hand side of the inequality is zero, that is,

$$\|\mathbf{y}_{t+1}\| \leq \sum_{s=0}^{t} \|\overline{W}_{jl}^{s+1}\boldsymbol{\delta}_{t-s}\|. \tag{67}$$

Furthermore, the spectral radius of $\overline{W}_{jl}$ is strictly less than 1, that is, $\bar{\lambda}_1 := \rho(\overline{W}_{jl}) < 1$ because $\lim_{t\to\infty} \overline{W}_{jl}^t = \mathbf{0}$ [39, Thm. 1.10]. As a result, we have

$$\|\mathbf{y}_{t+1}\| \leq \sum_{s=0}^{t} \|\overline{W}_{jl}^{s+1}\boldsymbol{\delta}_{t-s}\| \leq \sum_{s=0}^{t} \rho(\overline{W}_{jl})^{s+1}\|\boldsymbol{\delta}_{t-s}\| \tag{68}$$

Note that by (62), we have $\|\boldsymbol{\delta}_{t-s}\| = n + 1/t - s$. Define $\delta_{avg}(t) := \frac{1}{t}\sum_{s=1}^{t} \frac{n+1}{s}$. By Chebychev's sum inequality [40] (p. 43-44), we have

$$\|\mathbf{y}_{t+1}\| \leq \delta_{avg}(t) \sum_{s=0}^{t} \bar{\lambda}_1^{s+1} \tag{69}$$

$$= \delta_{avg}(t)(\bar{\lambda}_1 \frac{1 - \bar{\lambda}_1^{t+1}}{1 - \bar{\lambda}_1}) \leq \frac{\delta_{avg}(t)}{1 - \bar{\lambda}_1} \tag{70}$$

Noting that $\delta_{avg}(t) := \frac{1}{t}\sum_{s=1}^{t} \frac{n+1}{s} = O(\frac{\log t}{t})$, we have $\|\mathbf{y}_{t+1}\| = \|\hat{F}_{jt}(l) - f_{jt}(l)\mathbf{1}\| = O(\frac{\log t}{t})$ for any $l \in \mathcal{A}$. Consequently, $\|\nu_{jt}^i - f_{jt}\| = O(\frac{\log t}{t})$. ■

Similar to Lemma 5 the above result is true irrespective of the game that the agents are playing. The result leverages on the fact that the change in the empirical distribution of agent j is at most $1/t$ by the recursion in (14) and the belief updates of i on j's empirical frequency evolves faster than the change in agent j's empirical distribution. We continue with the proof of the Theorem.

Proof of Theorem 2: Proof follows the same proof outline in Theorem 1. Start by exploiting the multi-linearity of the expected utility when all individuals play with respect to their empirical distributions [36], that is,

$$u(f_{t+1};\mu) = u(f_t;\mu) + \frac{1}{t}\sum_{i=1}^{n} u(\Psi(a_{it}), f_{-it};\mu) - u(f_{it}, f_{-it};\mu) + \frac{\delta}{t^2}. \tag{71}$$

for some $\delta > 0$ which we collect higher order terms. We move the first term of the RHS to the left and add $|\delta|/t^2$ to the left hand side and get rid of the last term on the right hand side,

$$u(f_{t+1};\mu) - u(f_t;\mu) + \frac{|\delta|}{t^2} \geq \frac{1}{t}\sum_{i=1}^{n} u(\Psi(a_{it}), f_{-it};\mu) - u(f_{it}, f_{-it};\mu). \tag{72}$$

Now define $L_{it} := v(\nu^i_{-it};\mu_{it}) - u(\Psi(a_{it}), f_{-it};\mu)$. Add $\sum_{i=1}^n L_{it}/t$ to both sides of the above equation to get

$$u(f_{t+1};\mu) - u(f_t;\mu) + \frac{|\delta|}{t^2} + \frac{1}{t}\sum_{i=1}^{n} L_{it} \geq \frac{1}{t}\sum_{i=1}^{n} v(\nu^i_{-it};\mu_{it}) - u(f_{it}, f_{-it};\mu). \tag{73}$$

Now we sum up the terms above from time $t = 1$ to T,

$$u(f_{T+1};\mu) - u(f_1;\mu) + \sum_{t=1}^{T+1}\frac{|\delta|}{t^2} + \sum_{t=1}^{T+1}\frac{1}{t}\sum_{i=1}^{n} L_{it} \geq \sum_{t=1}^{T+1}\frac{1}{t}\sum_{i=1}^{n} v(\nu^i_{-it};\mu_{it}) - u(f_{it}, f_{-it};\mu). \tag{74}$$

Consider the left hand side of the above equation. The utility and therefore the expected utility is bounded. The third term is summable. By Lemma 6 and Assumption 2, the conditions of Lemma 1 are satisfied. Lemma 1 yields that the last term on the left hand side of (74) is summable. Hence, the left hand side of (74) is bounded.

Define $\alpha_t := \sum_{i=1}^n v(\nu^i_{-it};\mu_{it}) - u(f_{it}, f_{-it};\mu)$. Using the definition of α_t and the boundedness of the left hand side of the above equation, it follows from (74) that there exists some bounded parameter $0 < \bar{B} < \infty$ such that

$$\bar{B} > \sum_{t=1}^{\infty}\frac{\alpha_t}{t}. \tag{75}$$

Define $\beta_t := \sum_{i=1}^n v(f_{-it};\mu) - u(f_{it}, f_{-it};\mu)$ and consider the difference between α_{t+1} and β_{t+1}

$$\|\alpha_t - \beta_t\| = \|\sum_{i=1}^{n} v(\nu^i_{-it};\mu_{it}) - v(f_{-it};\mu)\| \tag{76}$$

Lemma 1 implies that the above equality is equal to $\|\alpha_t - \beta_t\| = O(\log t/t)$. By noting that $\beta_t \geq 0$, the conditions of Lemma 2 are satisfied which implies that

$$\sum_{t=1}^{T}\frac{\beta_t}{t} < \infty \tag{77}$$

for any $T > 0$. As a result the time average of the above sum converges to zero by Kronecker's Lemma [37, Thm. 2.5.5], that is,

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\frac{\beta_t}{t} = 0. \tag{78}$$

We remark that β_t captures the difference in expected payoffs when agent i best responds to others' empirical distribution f_{-it} given the common asymptotic belief μ, and when agent i follows its own empirical distribution f_{it} with common beliefs on the state μ. The convergence in (25) follows from the above equation by Lemma 4. ■

REFERENCES

[1] Y. Shoham and K. Leyton-Brown, *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press, 2008, vol. 1.

[2] C. Eksin and A. Ribeiro, "Distributed network optimization with heuristic rational agents," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5396–5411, October 2012.

[3] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multiagent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, 2009.

[4] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, 1986.

[5] J. Chen and A. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *Signal Processing, IEEE Transactions on*, vol. 60, no. 8, pp. 4289–4305, 2012.

[6] D. Monderer and L. Shapley, "Fictitious play property for games with identical interests," *Journal of economic theory*, vol. 68, no. 1, pp. 258–265, 1996.

[7] B. Swenson, S. Kar, and J. Xavier, "Empirical centroid fictitious play: An approach for distributed learning in multi-agent games," *IEEE Trans. Signal Process.*, vol. PP, no. 99, p. 1, 2015.

[8] J. Marden, G. Arslan, and J. Shamma, "Joint strategy fictitious play with inertia for potential games," *IEEE Trans. Automatic Control*, vol. 54, no. 2, pp. 208–220, 2009.

[9] J. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to nash equilibria," *IEEE Trans. Automatic Control*, vol. 50, no. 3, pp. 312–327, 2005.

[10] S. Hart, "Adaptive heuristics," *Econometrica*, vol. 73, no. 5, pp. 1401–1430, 2005.

[11] H. Young, *Strategic learning and its limits*. Oxford University Press, 2004.

[12] J. Harsanyi, "Games with incomplete information played by bayesian players - part ii. bayesian equilibrium points." *Management Science*, vol. 14, no. 5, pp. 320–334, 1968.

[13] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie, "Learning in networks with incomplete information:asymptotic analysis and tractable implementation of rational behavior," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 30–42, May 2013.

[14] ——, "Bayesian quadratic network game filters," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2250 – 2264, May 2014.

[15] E. Dekel, D. Fudenberg, and D. Levine, "Learning to play bayesian games," *Games and Economic Behavior*, vol. 46, no. 2, pp. 282–303, 2004.

[16] D. Fudenberg and D. Levine, *The Theory of Learning in Games*, 1st ed. Cambridge, MA: MIT Press, 1998.

[17] G. W. Brown, "Iterative solution of games by fictitious play," *Activity analysis of production and allocation*, vol. 13, no. 1, pp. 374–376, 1951.

[18] D. Fudenberg and D. Kreps, "Learning mixed equilibria," *Games and Economic Behavior*, vol. 5, no. 3, pp. 320–367, 1993.

[19] D. Fudenberg and S. Takahashi, "Heterogeneous beliefs and local information in stochastic fictitious play," *Games and Economic Behavior*, vol. 71, no. 1, pp. 100–120, 2011.

[20] G. Arslan, J. Marden, and J. Shamma, "Autonomous vehicle-target assignment: A game-theoretical formulation," *Journal of Dynamic Systems, Measurement, and Control*, vol. 129, no. 5, pp. 584–596, 2007.

[21] J. Marden and J. Shamma, "Revisiting log-linear learning: Asynchrony, completeness and a payoff-based implementation," *Games and Economic Behavior*, vol. 75, no. 2, pp. 788–808, 2012.

[22] A. Jadbabaie, J. Lin, and A. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[23] A. Kashyap, T. Basar, and R. Srikant, "Quantized consensus," *Automatica*, vol. 43, no. 7, pp. 1192–1203, 2007.

[24] I. Schizas, A. Ribeiro, and G. Giannakis, "Consensus in ad hoc wsns with noisy links - part i: distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 1650–1666, January 2008.

[25] S. Stankovic, M. Stankovic, and D. Stipanovic, "Decentralized parameter estimation by consensus based stochastic approximation," in *Proc. of the 46th IEEE Conference on Decision and Control (CDC)*, New Orleans, LA, USA, Dec. 2007, pp. 1535–1540.

[26] R. Olfati-Saber, "Distributed kalman filtering for sensor networks," in *46th IEEE Conference on Decision and Control, 2007*. IEEE, 2007, pp. 5492–5498.

[27] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma, "Belief consensus and distributed hypothesis testing in sensor networks," in *Networked Embedded Sensing and Control*. Springer Berlin Heidelberg, 2006, pp. 169–182.

[28] S. Kar, J. M. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Tran. Information Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.

[29] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, "Distributed detection: Finite-time analysis and impact of network topology," 2014.

[30] A. Jadbabaie, P. Molavi, and A. Tahbaz-Salehi, "Information heterogeneity and the speed of learning in social networks," *Columbia Business School Research Paper*, pp. 13–28, 2013.

[31] X. Vives, "Learning from others: a welfare analysis." *Games Econ. Behav.*, vol. 20, no. 2, pp. 177–200, 1997.

[32] D. Gale and S. Kariv, "Bayesian learning in social networks," *Games Econ. Behav.*, vol. 45, no. 2, pp. 329–346, 2003.

[33] P. Djuric and Y. Wang, "Distributed bayesian learning in multiagent systems," *IEEE Signal Process. Mag.*, vol. 29, pp. 65–76, March, 2012.

[34] M. Mueller-Frank, "A general framework for rational learning in social networks," *The Theoretical Economics*, vol. 8, pp. 1–40, 2013.

[35] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Trans. Autom. Control*, vol. 54, no. 11, 2009.

[36] D. Monderer and L. Shapley, "Fictitious play property for games with identical interests," *Journal of economic theory*, vol. 68, no. 1, pp. 258–265, 1996.

[37] R. Durrett, *Probability: Theory and Examples*, 3rd ed. Cambridge Series in Statistical and Probabilistic Mathematics, 2005.

[38] A. E. Brouwer and W. H. Haemers, *Spectra of graphs*. Springer, 2011.

[39] R. S. Varga, *Matrix iterative analysis*. Springer Science & Business, 2009, vol. 27.

[40] G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities, Cambridge Mathematical Library*. Cambridge University Press, 1988.