Distributed fictitious play in potential games of incomplete information

Ceyhun Eksin and Alejandro Ribeiro
Electrical and Systems Engineering Department, University of Pennsylvania
Philadelphia, PA, 19104
E-mail: ceksin@seas.upenn.edu, aribeiro@seas.upenn.edu

Abstract— **Based on the fictitious play algorithm, we introduce the distributed fictitious play algorithm as a decentralized decision-making model in unknown environments with networked interactions. The setup includes a network of agents, each receiving a payoff that depends on own action, actions of others and an unknown state of the world. In this setup, each agent needs to reason about the behavior of others and the unknown state of the world. In the distributed fictitious play algorithm, agents reason about others' behavior by keeping an empirical distribution of the others' actions based on the information received from their neighbors. We consider an information exchange model where agents observe past actions of their neighbors and keep an empirical distribution on the centroid population action. In addition, agents form beliefs on the state of the world through a parallel state learning process. At each stage of the algorithm, agents maximize their expected payoff assuming that others are going to play with respect to their estimated centroid empirical distribution given their belief on the state of the world. We show that the behaviors of agents converge to a consensus Nash equilibrium (NE) strategy of a symmetric potential game – a game with permutation invariant identical payoffs – as long as the state learning process achieves consensus in beliefs fast enough. We exemplify fast enough state learning processes and analyze the convergence behavior of the algorithm in a coordination game.**

I. Introduction

Based on the fictitious play algorithm, we introduce an individual decision-making model for multi-agent systems in uncertain environments which we call the distributed fictitious play algorithm. In fictitious play algorithms, each agent builds a model of future behavior of other agents by forming a histogram on observed actions of the past and best responds to its expected payoff [1], [2]. In setup of this paper, each agent in a network receives a payoff that depends on own action, actions of others and an unknown state of the world. In a networked setting, agents have access to information via their neighbors, that is, all of the past actions are not available. Therefore, in order to act optimally, agents need to reason about the behavior of non-neighboring agents based on past observations of their neighbors only. In addition, agents have uncertainty on the state of the world and update their beliefs on the state using private or local information. Our analysis shows that the agents can do the two processes, namely, reasoning about others' behavior and learning about the state, independently and converge to a Nash equilibrium of a potential game [3].

Belief formation on other agents' behaviors depends on the type of local information exchanged. We consider agents that only observe actions of their neighbors. Agents assume all the other agents follow a 'centroid' empirical distribution which they estimate by keeping account of frequency of observed neighboring actions [4]. Agents take actions that maximize the expected utility at each stage. Expected utility is computed assuming all the other agents independently follow the estimated 'centroid' empirical distribution. We analyze the convergence rate of the estimated empirical distribution in Section II-B and show that agents approach to the true empirical distribution that they estimate at a rate of $O(\log t/t)$ irrespective of the state learning and agent response rules.

The equilibrium convergence result for the model assumes that agents use a local state learning process in which agents agree asymptotically on a distribution on the state of the world at a rate faster than or equal to $O(\log t/t)$. Various decentralized learning models exist in the literature that achieve the desired convergence rate under different assumptions [5]–[8]. We exemplify two such state learning processes, namely averaging and Bayesian learning in Section II-C. The main convergence result states that agents asymptotically reach a consensus Nash equilibrium of any symmetric potential game in which agents have identical beliefs on the state (Theorem 1). At a consensus Nash equilibrium strategy, all agents use the same strategy and play optimal with respect to others' equilibrium strategy. We numerically analyze the transient and asymptotic equilibrium properties of the decentralized fictitious play in the beauty contest game (Section IV). In the beauty contest game, a team of robots tradeoff between moving toward a target direction on which they receive noisy information about and moving in coordination with each other. The communication constraints among robots limit their information sources to their local neighborhood. In addition, robots' beliefs on the target direction are different.

The setup of this work falls under the literature of learning in games that considers dynamic processes that lead to equilibrium in games [9], [10]. Fictitious play in which it is assumed that past history of the game is public, is one such simple update mechanism that has been shown to converge to a Nash equilibrium strategy in zero sum [9], certain 2 × 2 [11] and identical interest (potential) games [2]. Recently, the convergence results of the fictitious play algorithm has been shown to hold for potential games in a setting where agents only make local observations [4]. Our results leverage on their results and incorporate incomplete and asymmetrical information to the considered environment which is of importance for technological settings. Our motivation stems from the fact that computational burden of Bayesian Nash equilibrium strategies for each agent – optimal decision for each selfish agent given uncertainty about others and state – is not realistic even when the computation is possible [12]. Furthermore, the impossibility of learning 'Bayesian Nash equilibria' strategies in games of incomplete information has been demonstrated in [13]. We circumvent this issue by forcing asymptotic agreement among agents' beliefs on the state of the world via local state learning processes. We then use the fact that an identical interest game with common belief on the state of the world is an identical interest game with complete information with agents' payoffs equal to the expectation over the potential function of the original game with respect to the common belief over the state.

Other variants of the fictitious play algorithm [14], [15] and payoff based learning algorithms, e.g., reinforcement learning, [16] and their combinations [17] are also pertinent to the work here. The focus in these works is to either extend the scope of types of games that admit convergence to its Nash equilibrium through the dynamics proposed [15], or generate dynamics that lead to certain types of Nash equilibrium, e.g., pure (deterministic) Nash equilibrium [17], or optimal equilibrium [18]. Unlike these methods, we do not assume that agents observe their payoffs after each play.

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

II. Learning in Potential Games with Incomplete Information

We consider a simultaneous move incomplete information stage game with N players. Player $i \in \mathcal{N} := \{1, \ldots, N\}$ chooses action a_i from a finite set $\mathcal{A} := \{1, \ldots, m\}$. The payoff relevant state of the world θ is drawn by nature at the beginning of the game from the space Θ. We define $\mathcal{F}$ as the σ-algebra on the set Θ. We let $\mathbf{P}$ denote the set of probability distributions over the space $(\Theta, \mathcal{F})$ and define the total variation distance TV between $P_1 \in \mathbf{P}$ and $P_2 \in \mathbf{P}$ as $TV(P_1, P_2) = \sup_{B \in \mathcal{F}} |P_1(B) - P_2(B)|$.

The payoff to player i $u_i(\cdot)$ depends on the action profile $\mathbf{a} = \{a_1, \ldots, a_N\}$ and the state θ, that is, $u_i(\mathbf{a}, \theta) : \mathcal{A}^N \times \Theta \to \mathbb{R}$. We assume that the utility of each agent is finite for all action profiles and state realization. We consider potential games where there exists a potential function $u : \mathcal{A}^N \times \theta \mapsto \mathbb{R}$ such that for all $i \in \mathcal{N}$ the following relation holds

$$u_i(a_i, a_{-i}, \theta) - u_i(a_i', a_{-i}, \theta) = u(a_i, a_{-i}, \theta) - u(a_i', a_{-i}, \theta) \tag{1}$$

for all $a_i, a_i' \in \mathcal{A}$ and for all $a_{-i} \in \mathcal{A}^{N-1}$ and $\theta \in \Theta$.

The users have common prior belief over the state θ. Given the common belief μ, the expected utility of agent i for the action profile $\mathbf{a} = (a_1, \ldots, a_N)$ is as follows

$$u_i(\mathbf{a}; \mu) := \int_{\theta \in \Theta} u_i(\mathbf{a}, \theta) d\mu(\theta). \tag{2}$$

If there is no additional information provided to the agents, that is, agents do not receive private signals, then the game of incomplete information is equivalent to a complete information game $\Gamma(\mu)$ with players $\mathcal{N}$, action spaces $\mathcal{A}$ and payoffs $u_i(\mathbf{a}; \mu)$, that is, $\Gamma(\mu) = (\mathcal{N}, \mathcal{A}, u_i(\mathbf{a}; \mu))$.

The mixed strategy of player i σ_i is a probability distribution on the action space $\mathcal{A}$, that is, $\sigma_i \in \triangle(\mathcal{A})$. Expected utility with respect to the strategy profile $\sigma := (\sigma_1, \ldots, \sigma_N) \in \triangle^N(\mathcal{A}) := \times_{i=1}^N \triangle(\mathcal{A})$ is as follows

$$u_i(\sigma; \mu) = \sum_{\mathbf{a} \in \mathcal{A}^N} u_i(\mathbf{a}; \mu) \sigma(\mathbf{a}). \tag{3}$$

A Nash equilibrium (NE) strategy profile σ^* for the game $\Gamma(\mu)$ is such that for all $i \in \mathcal{N}$ and any $\sigma_i \in \triangle(\mathcal{A})$,

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

A NE strategy is such that assuming all the other agents are playing with respect to their equilibrium strategies it is optimal for each agent to follow its own equilibrium strategy. The left hand side of the NE condition in (4) is equivalently interpreted as the best response of agent i to the equilibrium strategy profile of others σ_{-i}^*. We define the expected utility of agent i when it best responds to a strategy profile of others σ_{-i} given common prior μ on θ as follows

$$v_i(\sigma_{-i}, \mu) := \max_{a_i \in \mathcal{A}} u_i(a_i, \sigma_{-i}; \mu). \tag{5}$$

Then the expected utility of agent i at NE (4) is given by the expected utility when it best responds to the NE strategies of others, $v_i(\sigma_{-i}^*, \mu) = u_i(\sigma_i^*, \sigma_{-i}^*; \mu)$.

We define the set of NE strategies of the stage game $\Gamma(\mu)$ as

$$K(\mu) = \{\sigma^* \in \triangle^N(\mathcal{A}) : u_i(\sigma^*; \mu) \geq u_i(\sigma_i, \sigma_{-i}^*; \mu), \\ \forall \sigma_i \in \triangle(\mathcal{A}), \forall i\}. \tag{6}$$

The set of consensus NE strategies for the game $\Gamma(\mu)$ contain the equilibrium strategies in which all agents use the identical strategy,

$$C(\mu) = \{\sigma \in K(\mu) : \sigma_1 = \sigma_2 = \cdots = \sigma_N\} \tag{7}$$

Observe that for a game $\Gamma(\mu)$ the set of Nash equilibria contains the set of consensus NE by definition, $C(\mu) \subseteq K(\mu)$.

The set of consensus strategies that is ϵ away from the consensus NE set above is the ϵ-Consensus NE strategy set, that is,

$$C_\epsilon(\mu) = \{\sigma \in \triangle^N(\mathcal{A}) : u_i(\sigma^*; \mu) \geq u_i(\sigma_i, \sigma_{-i}^*; \mu) - \epsilon, \\ \forall \sigma_i \in \triangle(\mathcal{A}), \forall i, \sigma_1 = \sigma_2 = \cdots = \sigma_N\} \tag{8}$$

for $\epsilon > 0$. The distance of a strategy $\sigma \in \triangle^N(\mathcal{A})$ from the set of consensus NE $C(\mu)$ is given by $d(\sigma, C(\mu)) = \min_{g \in C(\mu)} ||\sigma - g||$. Using the definition of distance, we define the δ consensus neighborhood of $C(\mu)$ as

$$B_\delta(C(\mu)) = \big\{\sigma \in \triangle^N(\mathcal{A}) : d(\sigma, C(\mu)) < \delta, \\ \sigma_1 = \sigma_2 = \cdots = \sigma_N\big\}. \tag{9}$$

Note that the δ consensus neighborhood is defined as the set of consensus strategies that are close to the set $C(\mu)$.

A. Fictitious play

In fictitious play processes, each agent iteratively takes an action $a_{it} \in \mathcal{A}$ and observes actions of other agents over time $t = 1, 2, \dots$. Agents use their observations of actions of others to keep an empirical distribution of others' plays and best respond to this empirical distribution. We use $f_{it} \in \mathbb{R}^{m\times 1}$ to denote the histogram, i.e. the empirical distribution, of agent i's actions until time t. Let $\Psi_{it} : \mathcal{A} \to \{0,1\}^m$ where its kth element is one if $a_{it} = k$ where $k \in \mathcal{A}$, that is, $\Psi_{it}(a_{it})(k) = 1$ if $a_{it} = k$ and $\Psi_{it}(a_{it})(l) = 0$ for $l \neq k$. Given this definition we formally define the empirical distribution of i f_{it} as follows

$$f_{it} = \frac{1}{t}\sum_{s=1}^{t} \Psi_{is}(a_{is}) \tag{10}$$

The empirical distribution can be represented in a recursive manner by reorganizing the above equation

$$f_{it+1} = f_{it} + \frac{1}{t+1}\big(\Psi_{it+1}(a_{it+1}) - f_{it}\big) \tag{11}$$

When actions are publicly observed, agent i computes f_{jt} for all $j \in \mathcal{N}$ and best responds to the empirical distribution $f_{-it} \in \mathbb{R}^{m\times N-1}$ and its belief μ on θ

$$a_{it+1} = \operatorname*{argmax}_{a_i \in \mathcal{A}} u_i(a_i, f_{-it}; \mu) \tag{12}$$

to receive an expected utility of $v_i(f_{-it}; \mu)$ as per (5). We let $f_t \in \mathbb{R}^{m\times N}$ denote the empirical distribution of the population, that is, $f_t := \{f_{1t}, \dots, f_{Nt}\}$.

B. Distributed fictitious play

When actions are *not* public information, agent $i \in \mathcal{N}$ cannot keep track of all agents' empirical distributions. Distributed fictitious play considers the case when interactions are local over a network G with node set $\mathcal{N}$ and edge set $\mathcal{E}$. Agent i's neighborhood defined as $\mathcal{N}_i := \{j : (j,i) \in \mathcal{E}\}$ is its source of information. We make the following assumption on connectivity of agents unless otherwise stated.

Assumption 1 *G is a strongly connected network.*

When agent i only observes actions of his neighbors $a_{\mathcal{N}_i t} := \{a_{jt} : j \in \mathcal{N}_i\}$, one particular quantity it can keep an estimate of is the average empirical play of the population $\bar{f}_t$,

$$\bar{f}_t = \frac{1}{N}\sum_{i=1}^{N} f_{it}. \tag{13}$$

We can equivalently write the above quantity recursively by the recursion for the histogram of i in (11)

$$\bar{f}_{t+1} = \bar{f}_t + \frac{1}{t+1}\big(\bar{\Psi}_{t+1}(\mathbf{a}_{t+1}) - \bar{f}_t\big). \tag{14}$$

where $\bar{\Psi}_t(\mathbf{a}_t) := \frac{1}{N}\sum_{i=1}^{N} \Psi_{it}(a_{it})$ is the centroid best response strategy at time t. We stack $N-1$ of the centroid empirical distributions to define $\bar{f}_{-it} := [\bar{f}_t, \dots, \bar{f}_t] \in \mathbb{R}^{m\times N-1}$ and N centroid distributions to define $\bar{f}_t^N := [\bar{f}_t, \dots, \bar{f}_t] \in \mathbb{R}^{m\times N}$.

Agent i keeps an estimate of the average empirical play of the population by averaging the observations of its neighbors, that is, i's estimate of $\bar{f}_t$ is written as follows

$$\hat{\bar{f}}_t^i = \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i} \frac{1}{t}\sum_{s=1}^{t} \Psi_{js}(a_{js}) \tag{15}$$

We can equivalently write i's estimate of average empirical distribution as follows

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

Since agent i cannot keep an estimate of individual empirical distributions in the local observation setting, it, incorrectly, assumes that others are playing with respect to $\hat{\bar{f}}_t^i$. In consequence, agent i plays a best response to $\hat{\bar{f}}_{-it}^i := [\hat{\bar{f}}_t^i, \dots, \hat{\bar{f}}_t^i] \in \mathbb{R}^{m\times N-1}$ in distributed fictitious play.

Next, 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 1 *Consider the distributed fictitious play in which the centroid empirical distribution of the population $\bar{f}_t$ evolves according to* (14) *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., $\hat{\bar{f}}_0^i = \bar{f}_0$ 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 [4] 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 in [4] leverages on the fact that the change in the centroid empirical distribution is at most $1/t$ by the recursion in (14). 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.

C. State Relevant Information

The belief of agent i on the state θ at time t is denoted by $\hat{\mu}_t^i \in \mathbf{P}$ and is formed by a *state learning* process SL_i. Denoting the information of agent i at time t by I_{it} the state learning process is a mapping from I_{it} to a belief on $\theta \in \Theta$, $SL_i : I_{it} \mapsto \mathbf{P}$. Throughout the paper, we make the following assumption on the state learning process.

Assumption 2 *For any agent $i \in \mathcal{N}$, the state learning process SL_i and information set I_{it} are such that the belief of i converges to a belief $\hat{\mu}^* \in \mathbf{P}$, that is,*

$$\lim_{t\to\infty} TV\left(SL_i(I_{it}), \hat{\mu}^*\right) = O\left(\frac{\log t}{t}\right) \quad \forall i \in \mathcal{N}. \tag{17}$$

The assumption above states that the total variation distance between the belief of agent i on the state θ at time t formed by the state learning process SL_i and a distribution on θ $\hat{\mu}^* \in \mathbf{P}$ shrinks in the order of $\log t/t$. This means that agents aggregate information fast enough and agree on their belief on the state θ using the local state learning process. We remark that $\hat{\mu}^*$ is not necessarily the optimal belief on the state, it is a belief on the state to which all agents' beliefs converge.

Note that the assumption does not restrict the information received by agents and information exchange among agents. As a result, we can use various social learning [5], [6], decentralized estimation [19]–[24] and averaging models [25], [26] existing in the literature depending on the information exchange model, as long as the convergence rate in the above assumption is satisfied. Here we present two examples of state learning processes that satisfies the above assumption.

Averaging. The state belongs to a finite space Θ and agent i starts with initial beliefs $\mu_{i0} \in \mathbf{P}$. At each step t agent i shares its previous belief on the state with its neighbors and updates its belief by weighted averaging the observed distributions,

$$\hat{\mu}^i_t(\theta) = \sum_{j \in \mathcal{N}} w_{ij} \hat{\mu}^j_{t-1}(\theta) \tag{18}$$

for all $\theta \in \Theta$ where $w_{ij} \geq 0$ if $j \in \mathcal{N}_i$ and $\sum_{j \in \mathcal{N}} w_{ij} = 1$. In this information of agent i at time t is given by $I_{it} = \{\{\hat{\mu}^j_l\}_{j \in \mathcal{N}_i, l=0,1,\ldots,t-1}, \mu_{i0}\}$. The convergence rate of averaging models have been analyzed in various generalized scenarios such as quantization or time varying connectivity [26], [27].

Bayesian Learning. Agent i starts with prior on θ $\hat{\mu}^i_0$ and at each step t update their belief on the state $\hat{\mu}^i_t$ using the Bayes' law upon observing noisy signals $s_{it} \in S$ generated according to a signal generating distribution $\pi_i : \Theta \mapsto S$. The information of agent i at time t is given by $I_{it} = \{\hat{\mu}^i_0, \{s_{il}\}_{l=1,\ldots,t}\}$. If the signals are informative and Gaussian then the uncertainty over θ decreases with $O(1/t^r)$ for $r > 0$ [28]. Furthermore, agents can also exchange beliefs on θ among each other and use the additional information to update their beliefs according to Bayes' law [8], [29], [30].

III. Convergence in Symmetric Potential Games with Incomplete Information

In this section, we restrict our attention to games in which agents interests are symmetric, that is, we assume $u_i(a_i, a_j, a_{-i\setminus j}, \theta) = u_j(a_j, a_i, a_{j\setminus i}, \theta)$ for all i and j. These games can be shown to admit NE with symmetric strategies, that is, for any belief on the state $\mu \in \mathbf{P}$, the set of consensus NE strategies $C(\mu)$ is not empty [4]. Note that in the distributed fictitious play, agents observe local actions, keep track of the centroid empirical distribution $\bar{f}_t$ and assume that this is the mixed strategy that all agents play with respect to. Therefore, the process can only converge to an empirical distribution over the action profile space $\triangle^N(\mathcal{A})$ such that each agent is playing with respect to the same distribution, i.e., it can only converge to a consensus strategy. That is, if the game does not admit a consensus NE then the distributed fictitious play will not converge to a NE of the game.

Below, we present our main result for the symmetric games that shows that distributed fictitious play with local action observations converges to a consensus NE of the potential game $\Gamma(\hat{\mu}^*)$. The proof presented follows the same outline of the proof of Theorem 1 in [4] which follows a similar outline to the proof in [31].

Theorem 1 *Consider the distributed fictitious play updates where agents at each stage best respond to their local beliefs on the population's empirical distribution in* (15)*. Then the centroid empirical distribution $\bar{f}^N_t$ converges to a consensus NE of the identical interest game with common state of the world belief $\hat{\mu}^*$ if assumptions of Lemma 1 and Assumption 2 are satisfied.*

Proof: Given the recursion for the centroid empirical distribution in (14), we can write the expected utility when all agents follow the centroid empirical distribution $\bar{f}_t$ and have identical beliefs $\hat{\mu}^*$ as follows

$$u(\bar{f}^N_{t+1};\, \hat{\mu}^*) = u\left(\bar{f}^N_t + \frac{1}{t+1}(\bar{\Psi}^N_{t+1}(\mathbf{a}_{t+1}) - \bar{f}^N_t);\, \hat{\mu}^*\right) \tag{19}$$

By the multi-linearity of the expected utility, we expand the above expected utility as follows [31]

$$\begin{aligned} u(\bar{f}^N_{t+1}; \hat{\mu}^*) = u(\bar{f}^N_t; \hat{\mu}^*) + \\ \frac{1}{1+t}\sum_{i=1}^N u(\bar{\Psi}_{t+1}(\mathbf{a}_{t+1}), \bar{f}_{-it}; \hat{\mu}^*) - u(\bar{f}_{it}, \bar{f}_{-it}; \hat{\mu}^*) \\ + \frac{\delta}{(1+t)^2} \end{aligned} \tag{20}$$

where the first order terms of the expansion are explicitly written and the remaining higher order terms are collected to the term $\delta/(1+t)^2$.

Consider the total utility term in (20) where agent i is playing with respect to the centroid best response strategy at time $t+1$ $\bar{\Psi}_{t+1}(\mathbf{a}_{t+1})$ and other agents use the centroid empirical distribution, $\sum_{i=1}^N u(\bar{\Psi}_{t+1}(\mathbf{a}_{t+1}), \bar{f}_{-it}; \hat{\mu}^*)$. By the definition of the centroid best response strategy given in Section II-B, we write the term in consideration as

$$\begin{aligned} &\sum_{i=1}^N u(\bar{\Psi}_{t+1}(\mathbf{a}_{t+1}), \bar{f}_{-it}; \hat{\mu}^*) \\ &\qquad = \sum_{i=1}^N u(\frac{1}{N}\sum_{i=1}^N \Psi_{it}(a_{it+1}), \bar{f}_{-it}; \hat{\mu}^*). \end{aligned} \tag{21}$$

The following equality can be shown by using the multi-linearity of expectation and permutation invariance of the utility [4],

$$\sum_{i=1}^N u(\bar{\Psi}_{t+1}(\mathbf{a}_{t+1}), \bar{f}_{-it}; \hat{\mu}^*) = \sum_{i=1}^N u(\Psi_{it+1}, \bar{f}_{-it}; \hat{\mu}^*). \tag{22}$$

The above equality means that the total expected utility when agents play with the centroid best response at time $t+1$ against the centroid empirical distribution at time t is equal to the total expected utility when agents best respond to the centroid empirical distribution at time t.

We substitute in the above equality (22) for the corresponding term in (20) to get the following

$$\begin{aligned} u(\bar{f}^N_{t+1};\hat{\mu}^*) &= u(\bar{f}^N_t;\hat{\mu}^*) + \\ &\frac{1}{1+t}\sum_{i=1}^N u(\Psi_{it+1},\bar{f}_{-it};\hat{\mu}^*) - u(\bar{f}_{it},\bar{f}_{-it};\hat{\mu}^*) \\ &+ \frac{\delta}{(1+t)^2}. \end{aligned} \tag{23}$$

We can upper bound the right hand side by adding $|\delta|/(1+t)^2$ to the left hand side.

$$\begin{aligned} u(\bar{f}^N_{t+1};\hat{\mu}^*) - u(\bar{f}^N_t;\hat{\mu}^*) + \frac{|\delta|}{(1+t)^2} \geq \\ \frac{1}{1+t}\sum_{i=1}^N u(\Psi_{it+1},\bar{f}_{-it};\hat{\mu}^*) - u(\bar{f}_{it},\bar{f}_{-it};\hat{\mu}^*) \end{aligned} \tag{24}$$

Define $L_{it+1} := v_i(\hat{\bar{f}}^i_{-it};\hat{\mu}^i_{t+1}) - u(\Psi_{i,t+1},\bar{f}_{-it};\hat{\mu}^*)$. Note that since agents have identical interests, 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 $\sum_{i=1}^N L_{it+1}/t+1$ to both sides of the above equation to get the following inequality,

$$\begin{aligned} u(\bar{f}^N_{t+1};\hat{\mu}^*) - u(\bar{f}^N_t;\hat{\mu}^*) + \frac{|\delta|}{(1+t)^2} \\ + \frac{1}{1+t}\sum_{i=1}^N v(\hat{\bar{f}}^i_{-it};\hat{\mu}^i_{t+1}) - u(\Psi_{it+1},\bar{f}_{-it};\hat{\mu}^*) \\ \geq \frac{1}{1+t}\sum_{i=1}^N v(\hat{\bar{f}}^i_{-it};\hat{\mu}^i_{t+1}) - u(\bar{f}_{it},\bar{f}_{-it};\hat{\mu}^*). \end{aligned} \tag{25}$$

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

$$\begin{aligned} u(\bar{f}^N_{T+1};\hat{\mu}^*) - u(\bar{f}^N_1;\hat{\mu}^*) \\ + \sum_{t=1}^{T+1}\frac{|\delta|}{(1+t)^2} + \sum_{t=1}^{T+1}\sum_{i=1}^N \frac{L_{it+1}}{1+t} \\ \geq \sum_{t=1}^{T+1}\frac{1}{1+t}\sum_{i=1}^N v(\hat{\bar{f}}^i_{-it};\hat{\mu}^i_{t+1}) - u(\bar{f}_{it},\bar{f}_{-it};\hat{\mu}^*). \end{aligned} \tag{26}$$

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

$$\alpha_{t+1} := \sum_{i=1}^N v(\hat{\bar{f}}^i_{-it};\hat{\mu}^i_{t+1}) - u(\bar{f}_{it},\bar{f}_{-it};\hat{\mu}^*). \tag{27}$$

The term α_t captures the total difference between expected utility when agents best respond to their beliefs on the centroid empirical distribution and their beliefs on θ, and when they follow the current centroid empirical distribution with common beliefs on the state $\hat{\mu}^*$. Note that by Lemma 1 and Assumption 2 the conditions of Lemma 2 are satisfied. By the assumption that utility value is finite and Lemma 2, the left hand side of (26) is finite. That is, there exists a $\bar{B} > 0$ such that

$$\bar{B} \geq \sum_{t=1}^{T+1}\frac{\alpha_{t+1}}{1+t}. \tag{28}$$

Next, we define the following term

$$\beta_{t+1} := \sum_{i=1}^N v(\bar{f}_{-it};\hat{\mu}^*) - u(\bar{f}_{it},\bar{f}_{-it};\hat{\mu}^*) \tag{29}$$

that captures the difference in expected payoffs when agents best respond to the centroid empirical distribution and the common asymptotic belief $\hat{\mu}^*$, and when they follow the current centroid empirical distribution with common beliefs on the state $\hat{\mu}^*$. When we consider the difference between α_{t+1} and β_{t+1}, the following equality is true by Lemma 2,

$$\begin{aligned} ||\alpha_{t+1} - \beta_{t+1}|| &= ||\sum_{i=1}^N v(\hat{\bar{f}}^i_{-it};\hat{\mu}^i_{t+1}) - v(\bar{f}_{-it};\hat{\mu}^*)|| \\ &= O(\frac{\log t}{t}). \end{aligned} \tag{30}$$

Further $\beta_{t+1} \geq 0$. Hence, the conditions of Lemma 3 are satisfied which implies that the following holds

$$\sum_{t=1}^T \frac{\beta_{t+1}}{t+1} < \infty. \tag{31}$$

From the above equation it follows by the Kronecker's Lemma that [32, Thm. 2.5.5]

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T \beta_t = 0. \tag{32}$$

The above convergence result implies that by Lemma 6 in [4], for any $\epsilon > 0$, the number of centroid empirical frequencies away from the ϵ consensus NE is finite for any time T, that is,

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

The relation above implies that the distance between the empirical frequencies and the set of symmetric NE diminishes by Lemma 4, that is,

$$\lim_{t\to\infty} d(\bar{f}^N_t, C(\hat{\mu}^*)) = 0. \tag{34}$$

■

The above result implies that when agents share their actions and based on this information keep an estimate of the empirical distribution of the population, their responses converge to a consensus NE of the symmetric potential game as long as their beliefs on the state reach consensus fast enough. The result also indicates that the state learning process and acquiring of information regarding population's play can be designed separately. 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 empirical centroid distribution. The analysis above reveals that these two processes can be designed independently as long as they converge at a fast enough rate.

IV. Simulations

We numerically analyze the performance of the algorithm in the beauty contest game and explore the effects of the network connectivity structure.

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 $\theta = [0°, 180°]$ in a two dimensional topology. Each robot receives an initial noisy signal related to the target direction θ,

$$\pi_i(\theta) = \theta + \epsilon_i \tag{35}$$

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., $\mathcal{A} = (0°, 5°, 10°, \dots, 180°)$. The estimation and coordination payoff of robot i is given by the following utility function

$$u_i(a, \theta) = -\lambda(a_i - \theta)^2 - (1-\lambda)(a_i - \frac{1}{N-1}\sum_{j \neq i} a_j)^2 \tag{36}$$

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 [12].

In the following numerical setup, we choose θ to be equal to 90°. We assume that all robots start with a common prior on each others' empirical frequency of actions such that they all believe others are going to play each action with equal probability. Then they update their beliefs according to the recursion in (16) upon observing actions of their neighbors. Robot i moves with a displacement of 0.01 meters in the chosen direction a_{it}.

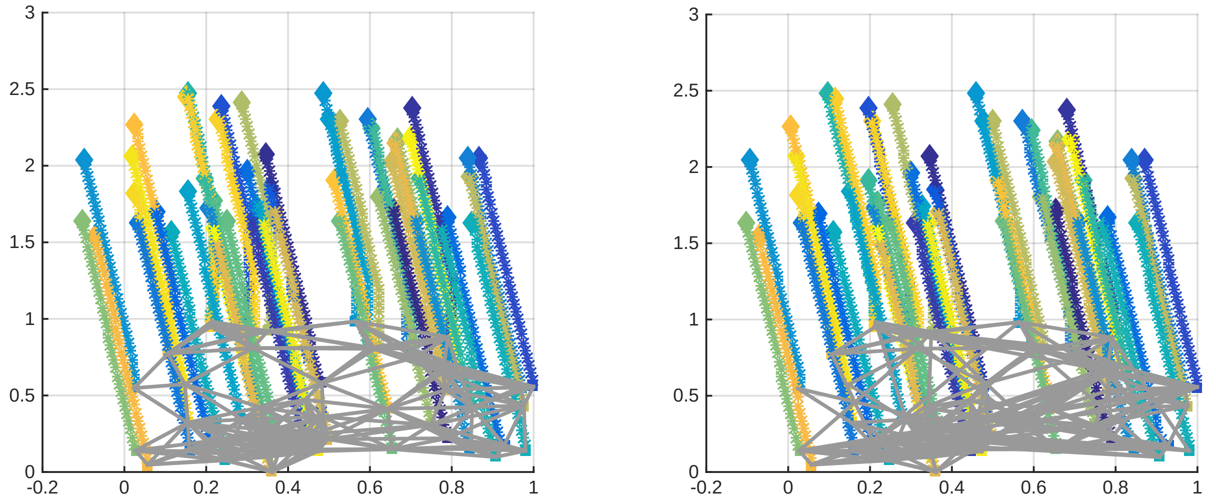
In Figs. 1 and 2, we plot robot positions and chosen actions, respectively, when robots use averaging to update their beliefs on the state θ based on receiving a single initial private signal with signal generating function in (35). That is, robots share their mean beliefs on the state and average their observations to obtain their beliefs on θ for the next time step. Figs. 1(a) and 2(a) correspond to the behavior in a geometric network when robots are placed on a 1 meter × 1 meter square randomly and connecting pairs with distance less than 0.3 meter between them. Figs. 1(b) and 2(b) correspond to the behavior in a small-world network when the edges in the geometric network are rewired with random nodes with probability 0.2. The geometric network illustrated in Fig. 1(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. 1(b) has a diameter of $\Delta_r = 4$ with an average length among users equal to 2. In figs. 2 (a)-(b), solid lines denote agents' actions over time, the dashed line marks the optimal estimate of the state θ given all of the signals which is equal to $96.1°$, the dotted dashed line is the actual value of the state $\theta = 90°$. 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 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.

V. Concluding Remarks

This paper introduced the distributed fictitious play algorithm as a bounded rational behavior model in potential games of incomplete information. Before presenting the algorithm, we established that a potential game of incomplete information with identical beliefs is equal to a potential game of complete information where the payoff is obtained by taking expectation of the payoff with respect to the state parameter. In the distributed fictitious play algorithm, each agent keeps an empirical distribution of the others based on the information received from their neighbors and incorrectly assumes that other agents are going to play with respect to this empirical distribution in the next time. Agents observe past actions of their neighbors and infer about their future behavior by keeping an empirical distribution. In addition, each agent makes observations about the unknown state or share information with each other regarding the state that allows it to learn about the state parameter through a learning process. Given the assumption that the learning process is fast enough to reach a belief agreement among agents, we showed that the empirical distributions converge to a consensus NE strategy of a symmetric potential game. In other words, empirical distribution of everyone converges to the same distribution and each agent knows that this

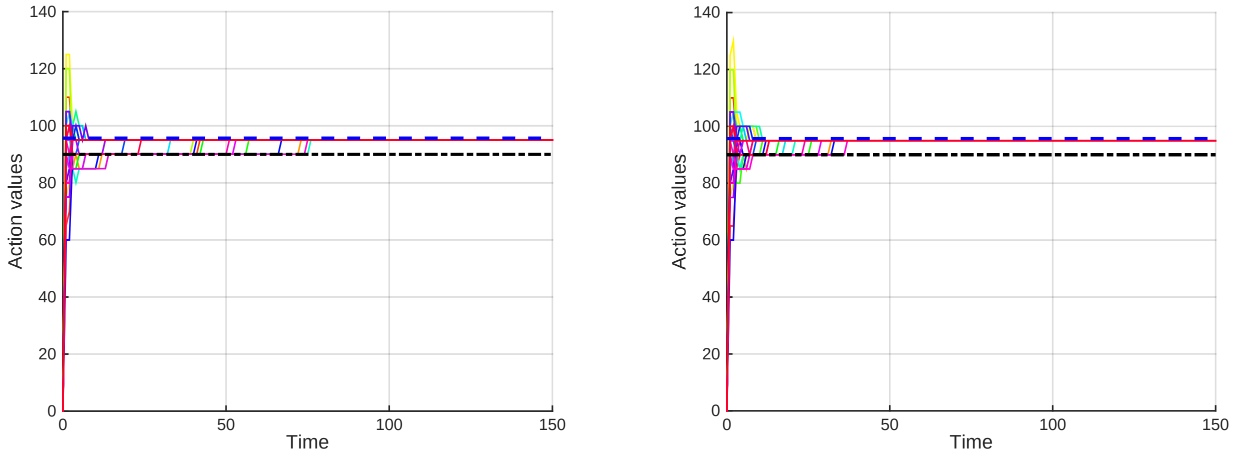
[1] Diameter is the longest shortest path among all pairs of nodes in the network. The average length is the average number of steps along the shortest path for all pairs of nodes in the network.



(a)

(b)

Fig. 1. 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 (36). 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 $\theta = 90°$.



(a)

(b)

Fig. 2. 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 θ and the dashed line is the optimal estimate of the state given all of the signals. Agents reach consensus in the movement direction 95° faster in the small-world network than the geometric network.

is the distribution that others are playing with respect to. We exemplified the algorithm in a coordination game and observed that the diameter of the network is influential in convergence rate where the shorter the diameter is, the faster is the convergence.

Appendix I
Intermediate convergence results

The following intermediate results are equivalent to derivations of the results stated in Appendix B in [4]. They are stated here for completeness.

Lemma 2 *If the processes $g_t \in \triangle^N$ and $h_t \in \triangle^N$ are such that for all $i \in \mathcal{N}$ $||g_{-it} - h_{-it}|| = O(\log t/t)$ and the state learning processes SL_i for all $i \in \mathcal{N}$ that generates estimate beliefs $\{\{\hat{\mu}^i\}_{t=0}^{\infty}\}_{i\in\mathcal{N}}$ satisfy Assumption 2, then for the potential utility function defined in Section II and the expected utility for best response behavior defined in* (5)*, the following holds*

$$||v(g_{-it}; \hat{\mu}_t^i) - v(h_{-it}; \hat{\mu}^*)|| = O(\frac{\log t}{t}). \qquad (37)$$

Proof: The proof is detailed in Lemma 4 in [4]. The proof follows by first making the observation that the expected utility defined in (3) 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 (5) for $g_{-it}, \hat{\mu}_t^i$ and $h_{-it}, \hat{\mu}^*$ by the distance between $g_{-it}, \hat{\mu}_t^i$ and $h_{-it}, \hat{\mu}^*$ multiplied by the Lipschitz constant. ■

Lemma 3 *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 [4]. ∎

Lemma 4 *If for any* $\epsilon > 0$ *the following holds*

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : \bar{f}_t^N \notin C_\epsilon(\hat{\mu}^*)\}}{T} = 0 \tag{38}$$

then $\lim_{t\to\infty} d(\bar{f}_t^N, C(\hat{\mu}^*)) = 0$.

Proof: By Lemma 7 in [4], (38) implies that for a given $\delta > 0$ there exists an $\epsilon > 0$ such that

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : \bar{f}_t^N \notin B_\delta(C(\hat{\mu}^*))\}}{T} = 0 \tag{39}$$

Using above equation, the result follows by Lemma 1 in [31]. ∎

Lemma 5 *For the potential game with function* $u(\cdot)$ *in* (1) *and expected best response utility* (5)*, consider a sequence of distributions* $f_t \in \triangle^N$ *for* $t = 1, 2, \ldots$ *and a common belief on the state* $\hat{\mu}^* \in \mathbf{P}$. *Define the process* $\beta_t := \sum_{i=1}^{N} v(f_{-it}; \hat{\mu}^*) - u(f_{it}, f_{-it}; \hat{\mu}^*)$ *for* $t = 1, 2, \ldots$. *If*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{\beta_t}{t} = 0 \tag{40}$$

then $\lim_{t\to\infty} d(f_t, K(\hat{\mu}^*)) = 0$.

Proof: By Lemma 6 in [4], the condition (40) implies that for all $\epsilon > 0$

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : f_t \notin K_\epsilon(\hat{\mu}^*)\}}{T} = 0. \tag{41}$$

By Lemma 7 in [4], (41) implies that for all $\delta > 0$ the following is true

$$\lim_{T\to\infty} \frac{\#\{1 \leq t \leq T : f_t \notin B_\delta(K(\hat{\mu}^*))\}}{T} = 0 \tag{42}$$

The above convergence result yields desired convergence result by Lemma 1 in [31]. ∎

References

[1] G. W. Brown. Iterative solution of games by fictitious play. *Activity analysis of production and allocation*, 13(1):374–376, 1951.

[2] D. Monderer and L.S. Shapley. Fictitious play property for games with identical interests. *Journal of economic theory*, 68(1):258–265, 1996.

[3] D. Monderer and L.S. Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996.

[4] B. Swenson, S. Kar, and J. Xavier. Empirical centroid fictitious play: An approach for distributed learning in multi-agent games, 2013, arXiv preprint arXiv:1304.4577.

[5] S. Shahrampour, A. Rakhlin, and A. Jadbabaie. Distributed detection: Finite-time analysis and impact of network topology, 2014, arXiv preprint arXiv:1409.8606.

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

[7] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[8] P.M. Djuric and Y. Wang. Distributed bayesian learning in multiagent systems. *IEEE Signal Process. Mag.*, 29:65–76, March, 2012.

[9] D. Fudenberg and D.K. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge, MA, 1. edition, 1998.

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

[11] D. Fudenberg and D.M. Kreps. Learning mixed equilibria. *Games and Economic Behavior*, 5(3):320–367, 1993.

[12] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie. Bayesian quadratic network game filters. *IEEE Trans. Signal Process.*, 62(9):2250 – 2264, May 2014.

[13] E. Dekel, D. Fudenberg, and D.K. Levine. Learning to play bayesian games. *Games and Economic Behavior*, 46(2):282–303, 2004.

[14] J.S. Shamma and G. Arslan. Dynamic fictitious play, dynamic gradient play, and distributed convergence to nash equilibria. *IEEE Trans. Automatic Control*, 50(3):312–327, 2005.

[15] D. Fudenberg and S. Takahashi. Heterogeneous beliefs and local information in stochastic fictitious play. *Games and Economic Behavior*, 71(1):100–120, 2011.

[16] J.R. Marden and J.S. Shamma. Revisiting log-linear learning: Asynchrony, completeness and a payoff-based implementation. *Games and Economic Behavior*, 75(2):788–808, 2012.

[17] J.R. Marden, G. Arslan, and J.S. Shamma. Joint strategy fictitious play with inertia for potential games. *IEEE Trans. Automatic Control*, 54(2):208–220, 2009.

[18] G. Arslan, J.R. Marden, and J.S. Shamma. Autonomous vehicle-target assignment: A game-theoretical formulation. *Journal of Dynamic Systems, Measurement, and Control*, 129(5):584–596, 2007.

[19] 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.*, 56(1):1650–1666, January 2008.

[20] 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)*, pages 1535–1540, New Orleans, LA, USA, Dec. 2007.

[21] R. Olfati-Saber. Distributed kalman filtering for sensor networks. In *46th IEEE Conference on Decision and Control, 2007*, pages 5492–5498. IEEE, 2007.

[22] 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*, pages 169–182. Springer Berlin Heidelberg, 2006.

[23] S. Kar, J. M. Moura, and K. Ramanan. Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication. *Unpublished Manuscript*, 2008.

[24] J. Chen and A.H. Sayed. Diffusion adaptation strategies for distributed optimization and learning over networks. *Signal Processing, IEEE Transactions on*, 60(8):4289–4305, 2012.

[25] A. Jadbabaie, J. Lin, and A.S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Trans. Autom. Control*, 48(6):988–1001, 2003.

[26] A. Kashyap, T. Basar, and R. Srikant. Quantized consensus. *Automatica*, 43(7):1192–1203, 2007.

[27] A. Nedic, A. Olshevsky, A. Ozdaglar, and J.N. Tsitsiklis. On distributed averaging algorithms and quantization effects. *IEEE Trans. Autom. Control*, 54(11), 2009.

[28] X. Vives. Learning from others: a welfare analysis. *Games Econ. Behav.*, 20(2):177–200, 1997.

[29] D. Gale and S. Kariv. Bayesian learning in social networks. *Games Econ. Behav.*, 45(2):329–346, 2003.

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

[31] D. Monderer and L.S. Shapley. Fictitious play property for games with identical interests. *Journal of economic theory*, 68(1):258–265, 1996.

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