52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

Learning to Coordinate in a Beauty Contest Game

Pooya Molavi, Ceyhun Eksin, Alejandro Ribeiro, and Ali Jadbabaie

Abstract— **We study a dynamic game in which a group of players attempt to coordinate on a desired, but only partially known, outcome. The desired outcome is represented by an unknown state of the world. Agents' stage payoffs are represented by a quadratic utility function that captures the kind of trade-off exemplified by the Keynesian beauty contest: each agent's stage payoff is decreasing in the distance between her action and the unknown state; it is also decreasing in the distance between her action and the average action taken by other agents. The agents thus have the incentive to correctly *estimate* the state while trying to *coordinate* with and learn from others. We show that myopic, but Bayesian, agents who repeatedly play this game and observe the actions of their neighbors in a connected network eventually succeed in coordinating on a single action. However, as we show through an example, the consensus action is not necessarily optimal given all the available information.**

I. Introduction

Consider a group of agents that wish to coordinate on a desired outcome that is not fully known to any one of them. Agents choose actions which are close to what they consider to be the desired outcome; but they also need to coordinate with other agents by choosing actions that are similar to what they expect others to choose. There is a trade-off between acting according to one's best estimate of the desired outcome and trying to coordinate with other agents. Such trade-offs are important in trade decision in financial markets [1], consumption decisions [2], and in problems in cooperative robotics [3] or organizational coordination [4]. The decisions of traders in stock market, for example, depend on their beliefs about the fundamental stock values. Nonetheless, traders also tend to consider how other traders will behave as their decisions could directly affect the gains from trade. When choosing between substitute products that exhibit network externality, consumers tend to consider the products that are expected to be chosen by other consumers, in addition to the alternatives with the highest perceived quality. In a case of cooperative robotic movement, the robots' goal is to rendezvous at a point whose location is known to the robots only through noisy private observations while also maintaining the initial formation. In all of these examples, agents make decisions by attempting to second-guess the decisions of others while also guessing the value of an unknown (stock value, product quality, or the location of a goal). The other complicating factor is that oftentimes agents can only communicate with a handful of other agents, while at the same time, trying to coordinate with and learn from everybody else.

We use the framework of dynamic games of incomplete information to model the agents' coordination problem. Agents play a game with payoffs that have two components: an estimation term and a coordination term. The estimation term serves to capture the agents' desire to make decisions that are optimal given their private information about an unknown parameter. The coordination term captures the payoffs agents receive by taking actions that are close to the average action taken by the rest of the population. The game is played over multiple stages. At each stage of the game, agents observe the previous choices made by a subset of other agents, called their neighbors. For an individual the neighbors represent other individuals in her social clique, whereas for a mobile robot the neighbors are other robots in its proximity. An agent's action may reveal some information to her neighbors that was previously unknown to them. The neighbors can use this information to reevaluate their beliefs about the underlying parameter and their predictions of others' future behavior. These reevaluations may, in turn, lead agents to revise their actions over time.

Given this dynamic environment, different behavioral assumptions lead to different outcomes. In particular, the way agents revise their views in face of new information and the actions they choose given these views determine the long-run outcome of the game. In this paper, we assume that agents are *Bayesian* and *myopic*. Bayesian agents use Bayes' rule to incorporate new observations in their beliefs. Myopic agents choose actions at each stage of the game which maximize their stage payoffs, without regard for the effect of these actions on their future payoffs. The assumption on myopic agent behavior simplifies the analysis significantly and results in an essentially unique equilibrium—which is unlikely with forward-looking agents.[1] We use this behavioral assumption to define an equilibrium, and prove formal results regarding the agents' asymptotic equilibrium behavior, assuming a quadratic utility function.

Our analysis yields several interesting results. First, we show that an equilibrium exists and that it is unique up to sets of measure zero. Second, we show that the agents' actions asymptotically converge for almost all realizations of the game. Furthermore, given a connected observation network, agents' actions converge to the same value. In other words, agents eventually coordinate on the same action. We also show that the agents reach consensus in their best estimates of the underlying parameter. These results suggest that in a coordination game—where the agents' interests are aligned—repeated interactions between agents who are selfish and myopic could eventually lead them to coordinate on the same outcome. However, as we show through an example, the agents do not necessarily coordinate on the action which is optimal given the information dispersed among them. The results extend our previous work in [7] and [8] on consensus in beauty contest games, and complement our work in [9] wherein we present a tractable and decentralized algorithm for computing the equilibrium actions.

This paper is related to two major lines of research in game theory. The first one is on learning in games. This literature

Research supported in parts by AFOSR MURI FA9550-10-1-0567 and ARO MURI W911NF-12-1-0509.

The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104-6228 {pooya, ceksin, aribeiro, jadbabai}@seas.upenn.edu

[1] A series of results in game theory, all of them known by the name "folk theorem", establishes that in games played by sufficiently patient forward-looking agents, any individually rational payoff can be obtained as an equilibrium payoff. We are not aware of any such theorem that directly applies to our model. However, based on the results proved in the literature, a unique equilibrium is unlikely to obtain in our setting if the agents are forward-looking. For two examples of a folk theorem, a classic result and a more recent result proved for games played on networks, see [5], [6].

goes back to the seminal work of Aumann and Maschler [10]. Other works in the same spirit include [11]–[15]. The central question in this literature is whether and how agents can learn to play a Nash (or Bayesian Nash) equilibrium. In the current work, in contrast, we assume that the agents always behave as prescribed by an equilibrium. Said differently, agents in our model learn *in equilibrium* rather than learning about the equilibrium. Our work is also related to the literature on social learning and distributed estimation where a canonical model consists of a set of agents connected via a network and exchanging their estimates of an unknown state. The focus of the social learning literature is on modeling the way agents use their observations to update their beliefs (or estimates) and characterizing the outcomes of the learning process. There are two distinct families of social learning models: In Bayesian models, sophisticated agents incorporate the information about the unknown parameter using Bayes' rule and discard the redundant information [16]–[18]. The focus in this family of models is on asymptotic outcomes. In non-Bayesian models, a heuristic update rule is employed by naïve agents [19]–[22]. These simple rules make a more complete characterization of the learning process possible, but they are also often harder to motivate. In this paper, we extend the Bayesian social learning framework to an environment with payoff externalities, i.e., one where an agent's stage payoff is a function of other agents' actions, in addition to the realization of an unknown parameter.

II. The Model

A. Agents and payoffs

Consider n agents indexed by $i \in \{1, \dots, n\}$ who repeatedly play a game with uncertain payoffs. The payoff relevant uncertainty is captured by a *common* unknown parameter θ (also known as the state of the world) that takes values in $\Theta = \mathbb{R}$. Despite having incomplete information about θ, agents start with a common prior belief about the unknown parameter, denoted by $\mathbb{P}$. We make the following technical assumption about $\mathbb{P}$.

Assumption 1: θ is square integrable with respect to $\mathbb{P}$, i.e.,

$$\int_{\Theta} \theta^2 d\mathbb{P}(\theta) < \infty.$$

The game is played over a countable number of time periods that are indexed by the positive integers. At every stage of the game, each agent privately observes a signal, takes an action simultaneously with other agents, and receives a payoff. We use $s_{it} \in S_i$ to denote the private signal observed by agent i at time t, where S_i is a complete separable metric space. We also let $s_t = (s_{1t}, \dots, s_{nt}) \in S = \times_{i=1}^{n} S_i$ denote the signal profile observed by agents at time t. The action taken by agent i at time t is denoted by $a_{it} \in A_i = \mathbb{R}$. Finally, we use $u_i(a_t, \theta)$ to denote the payoff received by agent i at time period t when the action profile $a_t = (a_{1t}, \dots, a_{nt}) \in A = \mathbb{R}^n$ is chosen and the realized parameter is θ.

B. Information structure

The space of plays is the measurable set $\Omega = \Theta \times (S \times A)^{\mathbb{N}}$ with the generic element ω called the *path of play*. The set of all possible histories at time t is defined as $H_t = \Theta \times S^{t-1} \times A^{t-1}$ with the generic element denoted by h_t. The history h_t is a complete description of the realization of the unknown parameter θ in addition to the signals observed by the agents and actions taken by them up to time period t.

Agents' private signals are functions of the realized state as well as the actions previously taken by the agents. Given the history $h_t \in H_t$, signal profile s_t is generated according to some probability distribution $\pi_t(h_t)[\cdot]$ over S. More formally, the signaling function π_t is a transition probability from H_t to S that maps histories to probability distributions over S.

Agents do not observe the realized state; neither do they observe the realized histories. Rather, at time t, agent i's information is limited to the private observations she has made so far. The information available to agent i at time t is denoted by $h_{it} = (s_{i1}, \dots, s_{it-1})$ and called the time t *private history* of agent i.[2] We let $H_{it} = S_i^{t-1}$ denote the set of all possible time t private histories for agent i and let $H_i = \cup_{t=1}^{\infty} H_{it}$.

The information content of histories can be expressed as σ-algebras over the measurable space $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is the Borel σ-algebra. The time t history h_t as well as the time t private histories h_{it} are uniquely determined given the path of play ω. In other words, h_t and h_{it} are H_t-valued and H_{it}-valued random variables, respectively. We can therefore define $\mathcal{H}_t$ and $\mathcal{H}_{it}$ to be the σ-algebra of subsets of Ω generated by h_t and h_{it}, respectively. Likewise, we can define $\mathcal{H}_{i\infty}$ to be the σ-algebra generated by the union of $\mathcal{H}_{it}$ over all t. It represents agent i's information at the end of the game.

Although previous action profiles might remain unknown to every agent, we assume that each agent observes the previous actions chosen by a subset of the population. The observability of agents' actions is captured by the *observation network* $G \in \{0,1\}^{n \times n}$, where $G_{ij} = 1$ if and only if i can observe the actions previously taken by j. We impose the following restrictions on the observation network.

Assumption 2:

1) $G_{ii} = 1$ for all i.
2) $G_{ij} = G_{ji}$ for all i, j.
3) The network G is connected.

We let $N_i = \{j : G_{ij} = 1\}$ denote the set of *neighbors* of agent i in the observation network whose actions she can observe. a_{jt-1} is measurable with respect to $\mathcal{H}_{it}$ for all i and $j \in N_i$.

C. Strategies

Each agent's actions at any given time period can only depend on the information available to her at that time period. A strategy is a function that captures this dependence. More formally, a pure *behavior strategy* for agent i is a function $\sigma_i : H_i \to A_i$. This is a complete contingency plan determining the action to be taken by agent i at *all* time periods and given *any* private history. Similarly, the behavior of the agents is completely characterized by the strategy profile $\sigma = (\sigma_1, \dots, \sigma_n)$, where σ_i is the strategy of agent i. We also use $\sigma(h_t)$ to mean $(\sigma_1(h_{1t}), \dots, \sigma_n(h_{nt}))$, where h_{it} is the private history observed by agent i up to time t when the time t history is given by h_t. Given a strategy profile σ, the action chosen by agent i at time t is given by $a_{it} = \sigma_i(h_{it})$. This is a random variable over $(\Omega, \mathcal{F})$ which is measurable with respect to $\mathcal{H}_{it}$.

Fixing a strategy profile for the agents rules out the possibility of observing certain histories that are off the path of play. Given σ and for $t \geq 1$, the history $h_{t+1} = (h_t, s_t, a_t)$ is said to be on the path of play if h_t is on the path of play, s_t is in the support of $\pi_t(h_t)$, and $a_t = \sigma(h_t)$. The history $h_1 = (\theta)$ is on the path of play if θ is in the support of $\mathbb{P}$.

Any strategy profile, together with the agents' common prior and the signaling functions, induces a probability distribution over the measurable space $(H_t, \mathcal{H}_t)$ for any $t \in \mathbb{N}$.

[2] Throughout, we use the convention that $h_{i1} = \emptyset$ and that a function with domain $\emptyset$ is a constant.

We use $\mathbb{P}_{t\sigma}$ to denote this induced probability distribution defined as follows: For all h_{t+1} off the path of play let $d\mathbb{P}_{t+1\sigma}(h_{t+1}) = 0$; for $h_{t+1} = (h_t, s_t, a_t)$ on the path of play, on the other hand, define $\mathbb{P}_{t+1\sigma}$ recursively as

$$d\mathbb{P}_{t+1\sigma}(h_{t+1}) = d\mathbb{P}_{t\sigma}(h_t) d\pi_t(h_t)[s_t],$$

with

$$d\mathbb{P}_{1\sigma}(h_1) = d\mathbb{P}_{1\sigma}(\theta) = d\mathbb{P}(\theta).$$

Since the sequence of probability measures $\{\mathbb{P}_{t\sigma}\}_{t\in\mathbb{N}}$ is consistent, by Kolmogorov's extension theorem, there exists a probability measure $\mathbb{P}_\sigma$ on $(\Omega, \mathcal{F})$ whose marginal on $(H_t, \mathcal{H}_t)$ agrees with $\mathbb{P}_{t\sigma}$. Finally, we use $\mathbb{E}_\sigma$ to denote the expectation operator corresponding to $\mathbb{P}_\sigma$.

D. Equilibrium

So far, we have not constrained the behavior of the agents in any way. As mentioned in the introduction, however, we assume that agents behave selfishly and myopically optimal. To make this restriction precise, we first need to define the agents' expected utilities given a strategy profile. Given that other agents follow strategy profile σ and the agents' beliefs are induced by σ, the expected utility to agent i at time t of following strategy $\tilde{\sigma}_i$ is any random variable satisfying

$$U_{it}(\tilde{\sigma}_i; \sigma) = \mathbb{E}_\sigma \left[u_i(\tilde{\sigma}_i(h_{it}), \sigma_{-i}(h_{-it}), \theta) | \mathcal{H}_{it} \right],$$

where h_t and θ are understood to be random variables over $(\Omega, \mathcal{F})$.

Definition 1: Strategy profile σ^* is a *myopic weak perfect Bayesian equilibrium* if for all i and t the random equilibrium action $\sigma_i^*(h_{it})$ is $\mathbb{P}_{\sigma^*}$-square integrable and with $\mathbb{P}_{\sigma^*}$-probability one,

$$U_{it}(\sigma_i^*; \sigma^*) \geq U_{it}(\sigma_i; \sigma^*), \tag{1}$$

for any strategy σ_i.

The square integrability of equilibrium actions is a technical condition that is imposed to rule out the equilibria where each agent's expected payoff is $-\infty$ regardless of her own strategy.

Agents who play according to a myopic Bayesian equilibrium are selfish in that they choose actions that maximize their own expected utilities. They are myopic in that they do not account for the effect of their current actions on their future payoffs. An alternative equilibrium notion is obtained by assuming that agents choose actions that maximize the average (or discounted sum) of their payoffs over their lifetime. However, imposing this requirement will significantly complicate the calculations agents need to perform in order to find their optimal actions.

E. Quadratic coordination games

We restrict our attention to a model for agents' payoffs that is presented in [23] and induces strategic behavior in the spirit of the "beauty contest" example in Keynes's General Theory [24].

Assumption 3: Agents' stage payoffs have the following form

$$u_i(a_t, \theta) = -(1-\lambda)(a_{it} - \theta)^2 - \lambda(a_{it} - \bar{a}_t^i)^2, \tag{2}$$

where $\lambda \in (0, 1)$ is a constant and

$$\bar{a}_t^i = \frac{1}{n-1} \sum_{j \neq i} a_{jt}.$$

The first term in the payoff is a standard quadratic loss in the distance between the realized parameter and agent's action, whereas the second term is the beauty contest term that measures the distance between the action of agent i and the average action taken by the rest of the population.

Given an equilibrium σ^*, the action profiles $a_t^* = \sigma^*(h_t)$ prescribed by the equilibrium maximize the agents' payoffs over the path of play, that is,

$$\mathbb{E}_{\sigma^*} \left[u_i(a_{it}^*, a_{-it}^*, \theta) | \mathcal{H}_{it} \right] \geq \mathbb{E}_{\sigma^*} \left[u_i(a_{it}, a_{-it}^*, \theta) | \mathcal{H}_{it} \right],$$

for all i and any other $\mathcal{H}_{it}$-measurable random variable a_{it}. When agents play the quadratic coordination game with payoffs as in (1) and Assumption 1 is satisfied, in particular, the equilibrium is characterized by the following first-order condition

$$\frac{\partial}{\partial a_{it}} \mathbb{E}_{\sigma^*}[u_i(a_{it}, a_{-it}^*, \theta) | \mathcal{H}_{it}] \Big|_{a_{it} = a_{it}^*} = 0 \qquad \text{for all } i,$$

which can be written more explicitly as

$$a_{it}^* = (1-\lambda)\mathbb{E}_{\sigma^*}[\theta | \mathcal{H}_{it}] + \frac{\lambda}{n-1} \sum_{j \neq i} \mathbb{E}_{\sigma^*}[a_{jt}^* | \mathcal{H}_{it}] \quad \text{for all } i. \tag{3}$$

Any a_t^* that satisfies the fixed-point equation (3) can be the action profile chosen at time t given some equilibrium; that is, any such $\{a_t^*\}_{t\in\mathbb{N}}$ defines an equilibrium.

Proposition 1: There exists a myopic Bayesian equilibrium σ^* which is unique up to histories of $\mathbb{P}_{\sigma^*}$-measure zero.

Proof: Proof in the Appendix.

In the sequel, we repeatedly employ the characterization of the equilibrium given in (3). We also use $\mathbb{P}^*$ and $\mathbb{E}^*$ to denote the probability distribution and expectation operator over $(\Omega, \mathcal{F})$, respectively, given an arbitrary (but fixed) equilibrium.

III. Consensus in Actions

In this section we show that in a connected network agents eventually reach consensus in their actions. To prove this result, we first prove that the agents' actions converge.

Lemma 1: Let $a_{it}^* = \sigma_i^*(h_{it})$ for some equilibrium σ^*. Then, for all i,

$$a_{it}^* \xrightarrow{L^1} a_{i\infty}^* \qquad \text{as} \qquad t \to \infty,$$

where $a_{i\infty}^*$ is a square integrable $\mathcal{H}_{i\infty}$-measurable random variable satisfying

$$\mathbb{E}^* \left[u_i(a_{i\infty}^*, a_{-i\infty}^*, \theta) | \mathcal{H}_{i\infty} \right] \geq \mathbb{E}^* \left[u_i(a_{i\infty}, a_{-i\infty}^*, \theta) | \mathcal{H}_{i\infty} \right], \tag{4}$$

for any square integrable $\mathcal{H}_{i\infty}$-measurable random variable $a_{i\infty}$.

Proof: First, define the L^2-norm of a_i as

$$\|a_i\|_2 = \mathbb{E}^*[|a_i|].$$

Also, define the L^2-norm of $a = (a_1, \ldots, a_n)$ as

$$\|a\|_2 = \sum_{i=1}^n \|a_i\|_2.$$

The condition expressed in (4) can be characterized (up to sets of measure zero) by the following first-order condition:

$$a_{i\infty}^* = (1-\lambda)\mathbb{E}^*[\theta | \mathcal{H}_{i\infty}] + \frac{\lambda}{n-1} \sum_{j \neq i} \mathbb{E}^*[a_{j\infty}^* | \mathcal{H}_{i\infty}].$$

By Lemma 3, the above set of equations have a solution which is unique up to sets of measure zero. Let $a^*_{i\infty}$ be a solution. We have that

$$a^*_{it} - a^*_{i\infty} = (1-\lambda)\left(\mathbb{E}^*[\theta|\mathcal{H}_{it}] - \mathbb{E}^*[\theta|\mathcal{H}_{i\infty}]\right) + \frac{\lambda}{n-1}\sum_{j\neq i}\left(\mathbb{E}^*[a^*_{jt}|\mathcal{H}_{it}] - \mathbb{E}^*[a^*_{j\infty}|\mathcal{H}_{i\infty}]\right).$$

Adding and subtracting $\frac{\lambda}{n-1}\sum_{j\neq i}\mathbb{E}^*[a^*_{j\infty}|\mathcal{H}_{it}]$ and using the triangle inequality,

$$\begin{aligned}\|a^*_{it} - a^*_{i\infty}\|_2 \leq{}& (1-\lambda)\left\|\mathbb{E}^*[\theta|\mathcal{H}_{it}] - \mathbb{E}^*[\theta|\mathcal{H}_{i\infty}]\right\|_2 \\ &+ \frac{\lambda}{n-1}\sum_{j\neq i}\left\|\mathbb{E}^*[a^*_{jt} - a^*_{j\infty}|\mathcal{H}_{it}]\right\|_2 \\ &+ \frac{\lambda}{n-1}\sum_{j\neq i}\left\|\mathbb{E}^*[a^*_{j\infty}|\mathcal{H}_{it}] - \mathbb{E}^*[a^*_{j\infty}|\mathcal{H}_{i\infty}]\right\|_2.\end{aligned} \tag{5}$$

Since conditional expectation is a contraction in L^2,

$$\sum_{j\neq i}\left\|\mathbb{E}^*[a^*_{jt} - a^*_{j\infty}|\mathcal{H}_{it}]\right\|_2 \leq \sum_{j\neq i}\left\|a^*_{jt} - a^*_{j\infty}\right\|_2.$$

Summing (5) over i and using the above inequality results in

$$\begin{aligned}\|a^*_t - a^*_\infty\|_2 \leq{}& (1-\lambda)\sum_{i=1}^n\left\|\mathbb{E}^*[\theta|\mathcal{H}_{it}] - \mathbb{E}^*[\theta|\mathcal{H}_{i\infty}]\right\|_2 \\ &+ \lambda\|a^*_t - a^*_\infty\|_2 \\ &+ \frac{\lambda}{n-1}\sum_{i=1}^n\sum_{j\neq i}\left\|\mathbb{E}^*[a^*_{j\infty}|\mathcal{H}_{it}] - \mathbb{E}^*[a^*_{j\infty}|\mathcal{H}_{i\infty}]\right\|_2,\end{aligned}$$

which implies that

$$\begin{aligned}&\|a^*_t - a^*_\infty\|_2 \\ &\quad\leq \sum_{i=1}^n\left\|\mathbb{E}^*[\theta|\mathcal{H}_{it}] - \mathbb{E}^*[\theta|\mathcal{H}_{i\infty}]\right\|_2 \\ &\qquad + \frac{\lambda}{1-\lambda}\frac{1}{n-1}\sum_{i=1}^n\sum_{j\neq i}\left\|\mathbb{E}^*[a^*_{j\infty}|\mathcal{H}_{it}] - \mathbb{E}^*[a^*_{j\infty}|\mathcal{H}_{i\infty}]\right\|_2.\end{aligned}$$

It is easy to verify that $\mathbb{E}^*[\theta|\mathcal{H}_{it}]$ is a martingale with respect to the filtration $\mathcal{H}_{it} \uparrow \mathcal{H}_{i\infty}$. Furthermore,

$$\sup_t \|\mathbb{E}^*[\theta|\mathcal{H}_{it}]\|_2 \leq \|\theta\|_2 < \infty,$$

where the first inequality is a consequence of the fact that conditional expectation is a contraction and the second one is due to Assumption 1. Thus, by the L^p convergence theorem, $\mathbb{E}^*[\theta|\mathcal{H}_{it}]$ converges in the L^2 sense to $\mathbb{E}^*[\theta|\mathcal{H}_{i\infty}]$.[3] That is,

$$\lim_{t\to\infty}\|\mathbb{E}^*[\theta|\mathcal{H}_{it}] - \mathbb{E}^*[\theta|\mathcal{H}_{i\infty}]\|_2 = 0.$$

By a similar argument, relying on the fact that $a^*_{j\infty}$ is square integrable, for all j,

$$\lim_{t\to\infty}\left\|\mathbb{E}^*[a^*_{j\infty}|\mathcal{H}_{it}] - \mathbb{E}^*[a^*_{j\infty}|\mathcal{H}_{i\infty}]\right\|_2 = 0.$$

Therefore,

$$\|a^*_t - a^*_\infty\|_2 \longrightarrow 0 \qquad \text{as} \qquad t\to\infty,$$

which proves the desired result. ■

So far we have shown that each agent's action converges in the L^2 sense to some limit action. The next result asserts that agents can identify the limit actions of all their neighbors.

Lemma 2: If $j \in N_i$, then agent i can asymptotically identify the limit action of agent j, i.e., $a^*_{j\infty} \in \mathcal{H}_{i\infty}$.

Proof: Since a^*_{jt} converges to $a^*_{j\infty}$ in L^2 (and hence in L^1) and $\mathcal{H}_{it} \uparrow \mathcal{H}_{i\infty}$,

$$\mathbb{E}^*\left[a^*_{jt-1}|\mathcal{H}_{it}\right] \longrightarrow \mathbb{E}^*\left[a^*_{j\infty}|\mathcal{H}_{i\infty}\right] \qquad \text{as} \qquad t\to\infty,$$

in the L^1 sense.[4] On the other hand, since $a^*_{jt-1} \in \mathcal{H}_{it}$,

$$\mathbb{E}^*\left[a^*_{jt-1}|\mathcal{H}_{it}\right] = a^*_{jt-1} \xrightarrow{L^1} a^*_{j\infty} \qquad \text{as} \qquad t\to\infty,$$

which implies that $\mathbb{E}^*\left[a^*_{j\infty}|\mathcal{H}_{i\infty}\right] = a^*_{j\infty}$. Therefore, $a^*_{j\infty} \in \mathcal{H}_{i\infty}$. ■

Agents' strategies at any stage of the game are mappings from their private histories to their action spaces. Consequently, agent i's action at time t is constrained to be measurable with respect to $\mathcal{H}_{it}$, her information at time t. The previous lemma shows that if agent j is a neighbor of i, her actions are measurable with respect to agent i's information at infinity. Therefore, agent i can asymptotically *imitate* the actions of agent j. Because the observation graph is assumed to be undirected, agent j can imitate the actions of agent i as well. Agents i and j must, therefore, each asymptotically believe that their actions are better than the ones taken by the other. The following proposition shows that this is only possible if any two neighbors asymptotically play the same action, regardless of the realization of the state of the world.

Proposition 2: For any two neighboring agents i and j, $a^*_{i\infty} = a^*_{j\infty}$ except on a set of $\mathbb{P}^*$-probability zero.

Proof: By construction,

$$\mathbb{E}^*\left[u_i(a^*_{i\infty}, a^*_{-i\infty}, \theta)|\mathcal{H}_{i\infty}\right] \geq \mathbb{E}^*\left[u_i(a_{i\infty}, a^*_{-i\infty}, \theta)|\mathcal{H}_{i\infty}\right],$$

for any square integrable $\mathcal{H}_{i\infty}$-measurable random variable $a_{i\infty}$. By Lemma 2, $a^*_{j\infty} \in \mathcal{H}_{i\infty}$. Therefore,

$$\mathbb{E}^*\left[u_i(a^*_{i\infty}, a^*_{-i\infty}, \theta)|\mathcal{H}_{i\infty}\right] \geq \mathbb{E}^*\left[u_i(a^*_{j\infty}, a^*_{-i\infty}, \theta)|\mathcal{H}_{i\infty}\right].$$

Taking expectations of the above equation with respect to $\mathbb{P}^*$,

$$\mathbb{E}^*\left[u_i(a^*_{i\infty}, a^*_{-i\infty}, \theta)\right] \geq \mathbb{E}^*\left[u_i(a^*_{j\infty}, a^*_{-i\infty}, \theta)\right]. \tag{6}$$

By a similar argument,

$$\mathbb{E}^*\left[u_j(a^*_{j\infty}, a^*_{-j\infty}, \theta)\right] \geq \mathbb{E}^*\left[u_j(a^*_{i\infty}, a^*_{-j\infty}, \theta)\right]. \tag{7}$$

Summing (6) and (7) and simplifying the result, we can conclude that

$$\mathbb{E}^*\left[\left(a^*_{i\infty} - a^*_{j\infty}\right)^2\right] \leq 0,$$

which proves the proposition. ■

The proposition implies that in a connected network all the agents asymptotically take the same action.

Corollary 1: For any two agents i and j, $a^*_{i\infty} = a^*_{j\infty}$ with $\mathbb{P}^*$-probability one.

[3] For a statement and proof of the L^p convergence theorem, see, for instance, p. 215 of Durrett [25].

[4] See for instance Exercise 5.5.8. in Durrett [25].

IV. Asymptotic Efficiency

In this section we explore the question of whether agents' consensus action is optimal given all the information available to the whole population. First, we have to characterize this optimal action. Let $\mathcal{H}_\infty$ be the σ-algebra generated by the union of $\mathcal{H}_{i\infty}$ over all i. This is all the information available to the agents at the end of the game. Given $\mathcal{H}_\infty$, the socially optimal action profile is defined as the action profile that maximizes the expected social welfare

$$\mathbb{E}^*\left[\sum_{i=1}^{n} u_i(a,\theta)\middle|\mathcal{H}_\infty\right].$$

It is easy to see that $a^{**} = (a_1^{**}, \ldots, a_n^{**})$ defined below is the maximizer of the above expression.

$$a^{**} = (\mathbb{E}^*[\theta|\mathcal{H}_\infty], \ldots, \mathbb{E}^*[\theta|\mathcal{H}_\infty]) \tag{8}$$

The optimal action profile requires the agents to coordinate on playing the expectation of θ given the information collectively available to them.

Corollary 1 suggests that, as agents reach consensus in their actions, their coordination motive would disappear and their realized utilities would only depend on how close their actions are to the realized state of the world. The following proposition formalizes this observation.

Proposition 3: For all i,

$$a_{it}^* - \mathbb{E}^*\left[\theta|\mathcal{H}_{it}\right] \xrightarrow{L^1} 0 \qquad \text{as} \qquad t \to \infty.$$

Proof: By construction,

$$a_{i\infty}^* = (1-\lambda)\mathbb{E}^*[\theta|\mathcal{H}_{i\infty}] + \frac{\lambda}{n-1}\sum_{j\neq i}\mathbb{E}^*[a_{j\infty}^*|\mathcal{H}_{i\infty}],$$

which since $a_{j\infty}^* = a_{i\infty}^*$ for all j with $\mathbb{P}^*$-probability one, results in

$$a_{i\infty}^* = \mathbb{E}^*[\theta|\mathcal{H}_{i\infty}].$$

On the other hand, since $\mathcal{H}_{it} \uparrow \mathcal{H}_{i\infty}$,

$$\mathbb{E}^*\left[\theta|\mathcal{H}_{it}\right] \xrightarrow{L^1} \mathbb{E}^*\left[\theta|\mathcal{H}_{i\infty}\right] \qquad \text{as} \qquad t \to \infty.$$

The last two equations together with Lemma 1 complete the proof. ■

This result, together with Corollary 1, implies that $\mathbb{E}^*[\theta|\mathcal{H}_{i\infty}] = \mathbb{E}^*[\theta|\mathcal{H}_{j\infty}]$ with $\mathbb{P}^*$-probability one for all i, j. That is, agents eventually also reach consensus in their best estimate of θ. However, $\mathbb{E}^*\left[\theta|\mathcal{H}_{i\infty}\right]$ is generally different than $\mathbb{E}^*\left[\theta|\mathcal{H}_\infty\right]$. The following example shows that agents might asymptotically reach consensus on an action that is different from the optimal action a^{**}, even if the observation network is the complete network.

Example 1: Consider two agents who are endowed with the common prior $\mathbb{P}$ with $\operatorname{supp}\mathbb{P} = \{-1, 1\}$ and $\mathbb{P}(1) = \mathbb{P}(-1) = 1/2$. Agents' private signals belong to the sets $S_1 = S_2 = \{\mathrm{H}, \mathrm{T}\}$, and the signaling functions π_t are given by

$$\pi_t(h_t) = \begin{cases} \frac{1}{2}\delta_{(\mathrm{H},\mathrm{H})} + \frac{1}{2}\delta_{(\mathrm{T},\mathrm{T})} & \text{if} \quad \theta = 1, \\ \frac{1}{2}\delta_{(\mathrm{H},\mathrm{T})} + \frac{1}{2}\delta_{(\mathrm{T},\mathrm{H})} & \text{if} \quad \theta = -1, \end{cases}$$

where δ_{s_t} is the degenerate probability distribution with unit mass on the signal profile s_t. At each stage of the game, each agent receives a signal that is Heads (Tails) with probability one half, regardless of the realization of θ. Agents' private signals are thus completely uninformative about the realized state. A single observation of the *signal profile*, on the other hand, completely reveals the realized state: agents' signals are perfectly correlated if the state is $\theta = 1$, whereas they are perfectly negatively correlated if the state is $\theta = -1$.

Since the agents' signals are completely uninformative and given their prior, it is optimal for them both to choose $a_{i1} = 0$ in the first stage of the game. At time period $t = 2$, each agent observes the other agent's action. These observations, however, contain no information regarding the agents' private signals. Therefore, agents play $a_{it} = 0$ at time $t = 2$, and in all subsequent stages of the game. Agents choose the same action ($a_{1t} = a_{2t} = 1$) and have the same best estimate of the state ($\mathbb{E}^*[\theta|\mathcal{H}_{1t}] = \mathbb{E}^*[\theta|\mathcal{H}_{2t}] = 0$) starting from the first stage of the game. However, the information available to agents is not fully aggregated in this example: $a^{**} = \mathbb{E}^*[\theta|\mathcal{H}_\infty] = \theta \neq a_{i\infty} = 0$. Agents fail to reach consensus on the socially optimal action for all realizations of the game.[5]

V. Conclusion

This paper studies a repeated game in which agents attempt to coordinate on an outcome about which they have incomplete and asymmetric information. Any agent's actions reveal information which is used by other agents to revise their beliefs, and hence, their actions. We prove formal results regarding the asymptotic outcomes obtained when myopic agents play the actions prescribed by the Bayesian Nash equilibrium. In particular, we show that agents reach consensus in their actions if the observation network is connected.

We proved these results assuming that the agents' payoffs are represented by a quadratic utility function. However, the insights of our analysis do not seem to hinge on the particular utility function used. In fact, similar results can be proved for more general coordination games with payoffs that satisfy some symmetry, concavity, and supermodularity conditions. We intend to investigate this extension in future work.

Example 1 showed that agents do not necessarily coordinate on the optimal action. However, our extensive simulations suggest that "generically" agents reach consensus on the optimal action—at least when their private observations are independent of the history of the game.[6] We intend to formalize and investigate this conjecture in future research.

Appendix

We first prove a technical lemma.

Lemma 3: Let $(X, \mathcal{F}, P)$ be a probability triple, and let E be the expectation operator corresponding to P. Let θ be a square integrable random variable measurable with respect to $\mathcal{F}$. Also let $\mathcal{G}_i \subset \mathcal{F}$ be σ-algebras for $i = 1, \ldots, n$. Then, there exist square integrable random variables $a_1, \ldots, a_n$ such that a_i is measurable with respect to $\mathcal{G}_i$ and $a = (a_1, \ldots, a_n)$ is an essentially unique fixed-point of the equation

$$a_i = (1-\lambda)E[\theta|\mathcal{G}_i] + \frac{\lambda}{n-1}\sum_{j\neq i}E[a_j|\mathcal{G}_i].$$

[5] If agents were forward-looking, they could coordinate on the optimal action by signaling their private signals through their actions. For instance, by following the strategy that requires each agent to choose action ϵ (action -ϵ) at the first stage of the game if her private signal is Heads (Tails), agents learn the realized state in the second stage. As a result of following this strategy, agents' expected payoffs are lower in the first period but higher in all subsequent period compared to when playing according to the myopic equilibrium. When agents are sufficiently patient and ϵ is sufficiently small, the strategy described above is a (non-myopic) weak Bayesian perfect equilibrium.

[6] See the complementary paper by the authors for numerical examples [9].

Proof: Let $L_i^2(X)$ be the set of P-almost everywhere equivalent class of $\mathcal{G}_i$-measurable random variables with the norm

$$\|a_i\|_2 = \left(\int_X a_i^2 dP \right)^{\frac{1}{2}}.$$

By the Riesz-Fischer theorem, $L_i^2(X)$ is a Banach space. Let $L^2(X) = \times_{i=1}^n L_i^2(X)$ with the norm $\|a\|_2 = \sum_{i=1}^n \|a_i\|_2$. Define $T : L^2(X) \to L^2(X)$ as

$$T_i(a) = (1-\lambda) E[\theta|\mathcal{G}_i] + \frac{\lambda}{n-1} \sum_{j \neq i} E[a_j|\mathcal{G}_i].$$

Note that

$$\begin{aligned} \|T_i(a) - T_i(b)\|_2 &= \frac{\lambda}{n-1} \left\| \sum_{j \neq i} E[a_j - b_j|\mathcal{G}_i] \right\|_2 \\ &\leq \frac{\lambda}{n-1} \sum_{j \neq i} \|E[a_j - b_j|\mathcal{G}_i]\|_2 \\ &\leq \frac{\lambda}{n-1} \sum_{j \neq i} \|a_j - b_j\|_2, \end{aligned}$$

where the first inequality is the triangle inequality and the second one is due to the fact that conditional expectation is a contraction in L^2. Therefore,

$$\begin{aligned} \|T(a) - T(b)\|_2 &= \sum_{i=1}^n \|T_i(a) - T_i(b)\|_2 \\ &\leq \frac{\lambda}{n-1} \sum_{i=1}^n \sum_{j \neq i} \|a_j - b_j\|_2 \\ &= \lambda \|a - b\|_2. \end{aligned}$$

That is, T is a contraction mapping with Lipschitz constant $\lambda < 1$. Hence, by the Banach fixed-point theorem, T has a fixed-point $a \in L^1(X)$ which is unique—up to sets of P-measure zero. ■

Proof of Proposition 1: The proof is constructive. We start at $t = 1$ and inductively construct some functions σ_{it}^*. The equilibrium strategy is given by $\sigma_i^*(H_{it}) = \sigma_{it}^*(H_{it})$. For all t, let $\Omega^t = \Theta \times S^{t-1} \times A^{t-1}$ and let $\mathcal{F}^t$ be the product σ-algebra over Ω^t.

Let P^1 be the probability distribution over $(\Omega^1, \mathcal{F}^1)$ induced by $\mathbb{P}$ and π_1, and let E^1 be the corresponding expectation operator. Note that the marginal of $\mathbb{P}_\sigma$ over $(\Omega^1, \mathcal{F}^1)$ is equal to P^1 for any strategy profile σ. Furthermore, θ is measurable with respect to $\mathcal{F}^1$ and agent i's time 1 action need to be measurable with respect to $\mathcal{H}_{i1} \subset \mathcal{F}^1$. Therefore, the first-order equilibrium condition at time $t = 1$ can be written as

$$a_{i1} = (1-\lambda) E^1[\theta|\mathcal{H}_{i1}] + \frac{\lambda}{n-1} \sum_{j \neq i} E^1 \left[a_{j1} | \mathcal{H}_{i1} \right].$$

By Lemma 3, the above equation has an essentially unique square integrable solution $a_1^* = (a_{11}^*, \ldots, a_{n1}^*)$. Since a_{i1}^* is $\mathcal{H}_{i1}$-measurable, there exists σ_{i1}^* such that $a_{i1}^* = \sigma_{i1}^*(h_{i1})$.

Let P^2 be the probability distribution over $(\Omega^2, \mathcal{F}^2)$ induced by $\mathbb{P}$, π_1, π_2, and $(\sigma_{11}^*, \ldots, \sigma_{n1}^*)$, and let E^2 be the expectation operator corresponding to P^2. The marginal of $\mathbb{P}_{\sigma^1}$ over $(\Omega^2, \mathcal{F}^2)$ is equal to P^2 for any σ^1 with $\sigma_i^1(H_{i1}) = \sigma_{i1}^*(H_{i1})$, that is, for all beliefs which are consistent with σ_{i1}^*. Therefore, the time 2 first-order equilibrium condition can be written as

$$a_{i2} = (1-\lambda) E^2[\theta|\mathcal{H}_{i2}] + \frac{\lambda}{n-1} \sum_{j \neq i} E^2 \left[a_{j2} | \mathcal{H}_{i2} \right],$$

which by Lemma 3 has an essentially unique square integrable solution $a_2^* = (a_{12}^*, \ldots, a_{n2}^*)$. Since a_{i2}^* is $\mathcal{H}_{i2}$-measurable, there exists σ_{i2}^* such that $a_{i2}^* = \sigma_{i2}^*(h_{i2})$.

Proceeding inductively, one can construct an equilibrium strategy profile σ^* with $\sigma_i^*(H_{it}) = \sigma_{it}^*(H_{it})$ for all i and t. Lemma 3 guarantees that the resulting random equilibrium actions $a_{it}^* = \sigma_i^*(h_{it})$ are square integrable and unique up to sets of measure zero. ■

References

[1] G. Angeletos and A. Pavan, "Efficient use of information and social value of information," *Econometrica*, vol. 75, no. 4, pp. 1103–1142, 2007.
[2] Y. Bramoullé, R. Kranton, and M. D'Amours, "Strategic interaction and networks," *Working Paper*, 2009.
[3] J. Marden, G. Arslan, and J. Shamma, "Cooperative control and potential games," *IEEE Trans. Syst., Man, Cybern. B*, vol. 39, no. 6, pp. 1393–1407, 2009.
[4] A. Calvó-Armengol and J. d. M. Beltran, "Information gathering in organizations: equilibrium, welfare, and optimal network structure," *Journal of the European Economic Association*, vol. 7, no. 1, pp. 116–161, 2009.
[5] D. Abreu, D. Pearce, and E. Stacchetti, "Toward a theory of discounted repeated games with imperfect monitoring," *Econometrica*, vol. 58, no. 5, pp. 1041–1063, 1990.
[6] M. Laclau, "A folk theorem for repeated games played on a network," *Games and Economic Behavior*, vol. 76, no. 2, pp. 711–737, 2012.
[7] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie, "Learning in linear games over networks," in *Proceedings of the 50th Annual Allerton Conference on Communications, Control, and Computing*, Allerton, IL, 2012.
[8] ——, "Games with side information in economic networks," in *Proceedings of the 46th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2012.
[9] ——. (2013, Feb.) Bayesian Quadratic Network Game Filters. arXiv e-prints. arXiv:1302.0249. [Online]. Available: http://arxiv.org/abs/1302.0249
[10] R. J. Aumann and M. Maschler, "Repeated games with incomplete information: A survey of recent results," *Mathematica*, vol. 116, pp. 287–403, 1967.
[11] J. S. Jordan, "Bayesian learning in normal form games," *Games and Economic Behavior*, vol. 3, no. 1, pp. 60–81, 1991.
[12] ——, "Bayesian learning in repeated games," *Games and Economic Behavior*, vol. 9, no. 1, pp. 8–20, 1995.
[13] E. Kalai and E. Lehrer, "Rational learning leads to nash equilibrium," *Econometrica*, vol. 61, no. 5, pp. 1019–1045, 1993.
[14] J. H. Nachbar, "Prediction, optimization, and learning in repeated games," *Econometrica*, vol. 65, no. 2, pp. 275–309, 1997.
[15] D. P. Foster and H. P. Young, "Learning, hypothesis testing, and Nash equilibrium," *Games and Economic Behavior*, vol. 45, no. 1, pp. 73–96, 2003.
[16] D. Gale and S. Kariv, "Bayesian learning in social networks," *Games and Economic Behavior*, vol. 45, no. 2, pp. 329–346, 2003.
[17] D. Rosenberg, E. Solan, and N. Vieille, "Informational externalities and emergence of consensus," *Games and Economic Behavior*, vol. 66, no. 2, pp. 979–994, 2009.
[18] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *The Review of Economic Studies*, vol. 78, no. 4, pp. 1201–1236, 2011.
[19] V. Bala and S. Goyal, "Learning from neighbours," *The Review of Economic Studies*, vol. 65, no. 3, pp. 595–621, 1998.
[20] P. M. DeMarzo, D. Vayanos, and J. Zwiebel, "Persuasion bias, social influence, and unidimensional opinions," *The Quarterly Journal of Economics*, vol. 118, no. 3, pp. 909–968, 2003.
[21] B. Golub and M. Jackson, "Naïve learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, vol. 2, no. 1, pp. 112–149, 2010.
[22] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games and Economic Behavior*, vol. 76, no. 1, pp. 210–225, 2012.
[23] S. Morris and H. Shin, "Social value of public information," *The American Economic Review*, vol. 92, no. 5, pp. 1521–1534, 2002.
[24] J. M. Keynes, *The General Theory of Interest, Employment and Money*. London: Macmillan, 1936.
[25] R. Durrett, *Probability: Theory and Examples*. Cambridge University Press, 2010.