Decentralized Bayesian learning in dynamic games

Deepanshu Vasal and Achilleas Anastasopoulos

Abstract

We study the problem of decentralized Bayesian learning in a dynamical system involving strategic agents with asymmetric information. In a series of seminal papers in the literature, this problem has been studied under a simplifying model where selfish players appear sequentially and act once in the game, based on private noisy observations of the system state and public observation of past players' actions. It is shown that there exist information cascades where users discard their private information and mimic the action of their predecessor. In this paper, we provide a framework for studying Bayesian learning dynamics in a more general setting than the one described above. In particular, our model incorporates cases where players participate for the whole duration of the game, and cases where an endogenous process selects which subset of players will act at each time instance. The proposed methodology hinges on a sequential decomposition for finding perfect Bayesian equilibria (PBE) of a general class of dynamic games with asymmetric information, where user-specific states evolve as conditionally independent Markov process and users make independent noisy observations of their states. Using our methodology, we study a specific dynamic learning model where players make decisions about investing in the team, based on their estimates of everyone's types. We characterize a set of informational cascades for this problem where learning stops for the team as a whole.

I. Introduction

In today's world, there are many scenarios where strategic agents with different observations (i.e. information sets) interact among themselves to learn about each other, and take actions that affect their reward and further spread of information in the system. One such scenario was

Both authors are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, 48105 USA e-mail: {dvasal, anastas} at umich.edu.

The authors would like to thank Vijay Subramanian for his contribution to the paper.

This work is supported in part by NSF grant CIF-1111061.

studied in the seminal papers [1], [2] where the authors investigated the occurrence of fads in a social network, which was later generalized in [3]. The authors in [1], [2] and [3] study the problem of learning over a social network where there is a product which is either good or bad. There are countably many buyers, i.e. *different decision makers*, that are chosen exogenously and act exactly once in the process. They make noisy observation about the value of the product and sequentially act *strategically* to either buy or not buy the product. Their actions are based on their own private observation and the actions of the previous users. It is shown that herding can occur in such a case where a user discards its own private information and follows the majority action of its predecessors (characterizing fads in social networks). As a result, the users' action does not reveal any new information and all future users repeat this behavior. This phenomenon is defined as an informational cascade where learning stops for the group as a whole. While a good cascade is desirable, there's a positive probability of a bad cascade that hurts all future users in the community. There is a growing body of literature on alternative learning models that aim at avoiding bad cascades (for example see [4], [5]).

There are however more general scenarios, such as cases where players participate in the game more than once, deterministically or randomly, through an exogenous or even an endogenous process. Furthermore, there are practical scenarios where players may be adversarial to each others' learning (e.g. dynamic zero-sum games). Studying such scenarios may reveal more interesting and richer equilibrium behaviors including cascading phenomena, not manifested in the models considered in the current literature. An indispensable tool for studying cascades is a framework for finding equilibria for these dynamical systems involving strategic players with different information sets, which are modeled as dynamic games with asymmetric information. Appropriate equilibrium concepts for such games include perfect Bayesian equilibrium (PBE), sequential equilibrium, trembling hand equilibrium [6], [7]. Each of these notions of equilibrium consists of a strategy and a belief profile of all players where the equilibrium strategies are optimal given the beliefs and the beliefs are derived from the equilibrium strategy profile using Bayes' rule (whenever possible). For the games considered in the current literature including [2]–[5], since every buyer participates only for one time period and thus acts myopically, finding PBE reduces to solving a straightforward, one-shot optimization problem. However, for general dynamic games with asymmetric information, finding PBE is hard, since it requires solving a fixed point equation in the space of strategy and belief profiles across all users and all time periods (for a more elaborate discussion, see [7, Ch. 8]). There is no known sequential decomposition methodology for finding PBE for such games.

Recently we presented a methodology in [8] for finding PBE for a general class of dynamic games where a finite number of players have different states associated to them that evolve as conditionally independent Markov processes, and are observed perfectly by the corresponding players. In this paper, we start by generalizing that model to the case when players' do not perfectly observe their states; rather they make independent, noisy observations. Unlike other scenarios in the cascades literature discussed before, the proposed general framework can incorporate, as special cases, scenarios where players participate in the game more than once, deterministically or randomly through an exogenous or endogenous process, and/or scenarios where players may be adversarial to each others' learning. For a dynamic game with asymmetric information and a given PBE, we define informational cascades as those histories of the game where players' actions do not depend on their private information from that point on, and thus the system dynamics are governed only through the common information. We then consider a specific dynamic learning model where each player makes a decision to invest (or not invest) in the team, depending on its estimate of the average of all players' types. Thus learning players types is important aspect of the problem. Using the methodology presented, we characterize a set of informational cascades for this model where learning stops for the team. Limited as it is, this example provides analysis and intuition on the learning dynamics in decentralized games, and also serves as motivation for exploring a vast landscape of the scenarios that can be studied through the proposed methodology.

The paper is structured as follows. In section III, we provide a general methodology to find a class of PBEs for such games. In Section IV, we formally define informational cascades and specialize our methodology to study a specific Bayesian learning game, for which we characterize its informational cascades. We conclude in Section V.

II. NOTATION

We use uppercase letters for random variables and lowercase for their realizations. For any variable, subscripts represent time indices and superscripts represent player identities. We use notation $-i$ to represent all players other than player i i.e. $-i = \{1, 2, \ldots i-1, i+1, \ldots, N\}$. We use notation $a_{t:t'}$ to represent vector $(a_t, a_{t+1}, \ldots a_{t'})$ when $t' \geq t$ or an empty vector if $t' < t$.

We use a_t^{-i} to mean $(a_t^1, a_t^2, \ldots, a_t^{i-1}, a_t^{i+1} \ldots, a_t^N)$. We remove superscripts or subscripts if we want to represent the whole vector, for example a_t represents $(a_t^1, \ldots, a_t^N)$. In a similar vein, for any collection of finite sets $(\mathcal{X}^i)_{i\in\mathcal{N}}$, we denote $\times_{i=1}^N \mathcal{X}^i$ by $\mathcal{X}$. We denote the indicator function of any set A by $I_A(\cdot)$. For any finite set $\mathcal{S}$, $\mathcal{P}(\mathcal{S})$ represents space of probability measures on $\mathcal{S}$ and $|\mathcal{S}|$ represents its cardinality. We denote by P^g (or E^g) the probability measure generated by (or expectation with respect to) strategy profile g. We denote the set of real numbers by $\mathbb{R}$. For a probabilistic strategy profile of players $(\beta_t^i)_{i\in\mathcal{N}}$ where probability of action a_t^i conditioned on $a_{1:t-1}x_{1:t}^i$ is given by $\beta_t^i(a_t^i|a_{1:t-1}, x_{1:t}^i)$, we use the short hand notation $\beta_t^{-i}(a_t^{-i}|a_{1:t-1}, x_{1:t}^{-i})$ to represent $\prod_{j\neq i} \beta_t^j(a_t^j|a_{1:t-1}, x_{1:t}^j)$. All equalities and inequalities involving random variables are to be interpreted in *a.s.* sense.

III. General model

A. Model

We consider a discrete-time dynamical system with N strategic players in the set $\mathcal{N} := \{1, 2, \ldots N\}$, over a finite time horizon $\mathcal{T} := \{1, 2, \ldots T\}$ and with perfect recall. The system state is $x_t := (x_t^1, x_t^2, \ldots x_t^N)$, where $x_t^i \in \mathcal{X}^i$ is the state of player i at time t. Players' states evolve as conditionally independent, controlled Markov processes such that

$$P(x_t|x_{1:t-1}, a_{1:t-1}) = P(x_t|x_{t-1}, a_{t-1}) \tag{1a}$$

$$= \prod_{i=1}^N Q_x^i(x_t^i|x_{t-1}^i, a_{t-1}), \tag{1b}$$

where $a_t = (a_t^1, \ldots, a_t^N)$ and a_t^i is the action taken by player i at time t. Player i does not observe its state perfectly, rather it makes a private observation $w_t^i \in \mathcal{W}^i$ at time t, where all observations are conditionally independent across time and across players given x_t and a_{t-1}, in the following way, $\forall t \in 1, \ldots T$,

$$P(w_{1:t}|x_{1:t}, a_{1:t-1}) = \prod_{n=1}^t \prod_{i=1}^N Q_w^i(w_n^i|x_n^i, a_{n-1}). \tag{2}$$

Player i takes action $a_t^i \in \mathcal{A}^i$ at time t upon observing $a_{1:t-1}$, which is common information among players, and $w_{1:t}^i$, which is player i's private information. The sets $\mathcal{A}^i, \mathcal{X}^i, \mathcal{W}^i$ are assumed to be finite. Let $g^i = (g_t^i)_t$ be a probabilistic strategy of player i where $g_t^i : (\times_{j=1}^N \mathcal{A}^j)^{t-1} \times (\mathcal{W}^i)^t \to \mathcal{P}(\mathcal{A}^i)$ such that player i plays action a_t^i according to $A_t^i \sim g_t^i(\cdot|a_{1:t-1}, w_{1:t}^i)$. Let $g := (g^i)_{i\in\mathcal{N}}$ be a strategy profile of all players. At the end of interval t, player i gets an instantaneous reward $R^i(x_t, a_t)$. The objective of player i is to maximize its total expected reward

$$J^{i,g} := \mathbb{E}^g \left[\sum_{t=1}^{T} R^i(X_t, A_t) \right]. \tag{3}$$

With all players being strategic, this problem is modeled as a dynamic game $\mathfrak{D}$ with imperfect and asymmetric information, and with simultaneous moves. Although this model considers all N players acting at all times, it can accommodate cases where at each time t, players are chosen through an endogenously defined (controlled) Markov process. This can be done by introducing a nature player 0, who perfectly observes its state process $(X_t^0)_t$, has reward function zero, and plays actions $a_t^0 = w_t^0 = x_t^0$. Equivalently, all players publicly observe a controlled Markov process $(X_{t-1}^0)_t$, and a player selection process could be defined through this process. For instance, let $\mathcal{X}^0 = \mathcal{A}^0 = \mathcal{N}$, $\forall i$, $R_t^i(x_t, a_t) = 0$ if $i \neq a_t^0$, and $Q(x_{t+1}^i | x_t^i, a_t) = Q(x_{t+1}^i | x_t^i, a_t^{a_t^0})$. Here, in each period only one player acts in the game who is selected through an internal, controlled Markov process.

B. Solution concept: PBE

In this section, we introduce PBE as an appropriate equilibrium concept for the game considered. Any history of this game at which players take action is of the form $h_t = (a_{1:t-1}, x_{1:t}, w_{1:t})$. Let $\mathcal{H}_t$ be the set of such histories at time t and $\mathcal{H} := \cup_{t=0}^{T} \mathcal{H}_t$ be the set of all possible such histories. At any time t player i observes $h_t^i = (a_{1:t-1}, w_{1:t}^i)$ and all players together observe $h_t^c = a_{1:t-1}$ as common history. Let $\mathcal{H}_t^i$ be the set of observed histories of player i at time t and $\mathcal{H}_t^c$ be the set of common histories at time t. An appropriate concept of equilibrium for such games is the PBE [7] which consists of a pair (β^*, μ^*) of strategy profile $\beta^* = (\beta_t^{*,i})_{t \in \mathcal{T}, i \in \mathcal{N}}$ where $\beta_t^{*,i} : \mathcal{H}_t^i \to \mathcal{P}(\mathcal{A}^i)$ and a belief profile $\mu^* = ({}^i\mu_t^*)_{t \in \mathcal{T}, i \in \mathcal{N}}$ where ${}^i\mu_t^* : \mathcal{H}_t^i \to \mathcal{P}(\mathcal{H}_t)$ that satisfy sequential rationality so that $\forall i \in \mathcal{N}, t \in \mathcal{T}, h_t^i \in \mathcal{H}_t^i, \beta^i$

$$\mathbb{E}^{(\beta^{*,i}\beta^{*,-i}, \mu^*)} \left\{ \sum_{n=t}^{T} R^i(X_n, A_n) \Big| h_t^i \right\} \geq \mathbb{E}^{(\beta^i\beta^{*,-i}, \mu^*)} \left\{ \sum_{n=t}^{T} R^i(X_n, A_n) \Big| h_t^i \right\}, \tag{4}$$

and the beliefs satisfy consistency conditions as described in [7, p. 331].

In general, ${}^i\mu_t^*$ is defined as the belief of player i at time t on the history $h_t = (a_{1:t-1}, x_{1:t}, w_{1:t})$, conditioned on its private history $h_t^i = (a_{1:t-1}, w_{1:t}^i)$. In our model, due to independence of types and observations, player i's private observations $w_{1:t}^i$ do not provide any information about

$(x_{1:t}^{-i}, w_{1:t}^{-i})$, as will be shown later. For this reason we consider beliefs that are functions of each agent's history h_t^i only through the common history $h_t^c = a_{1:t-1}$. Moreover, player i's relevant uncertainty, as required to compute its expected reward-to-go, can be sufficiently represented by beliefs on players' private beliefs ξ_t (which are defined later). Hence, instead of ${}^i\mu_t^*[h_t^i](h_t)$, every agent uses a common belief $\mu_t^*[a_{1:t-1}](\xi_t)$ derived from the common history $h_t^c = a_{1:t-1}$, where $\mu_t^*[a_{1:t-1}](\xi_t)$ itself factorizes into a product of marginals $\prod_{j\in\mathcal{N}} \mu_t^{*,j}[a_{1:t-1}](\xi_t^i)$.

C. PBE of the game $\mathfrak{D}$

In this section, we provide a methodology to find PBE of the game $\mathfrak{D}$ that consists of strategies whose domain is time-invariant (while there may exist other equilibria that can not be found using this methodology). Specifically, we seek equilibrium strategies that are structured in the sense that they depend on players' common and private information through belief states. In order to achieve this, at any time t, we summarize player i's private information, $w_{1:t}^i$, in the belief ξ_t^i, and its common information, $a_{1:t-1}$, in the belief π_t, where ξ_t^i and π_t are defined as follows. For a strategy profile g, let $\xi_t^i(x_t^i) := P^g(x_t^i|a_{1:t-1}, w_{1:t}^i)$ be the belief of player i on its current state conditioned on its information, where $\xi_t^i \in \mathcal{P}(\mathcal{X}^i)$. Also we define $\pi_t^i(\xi_t^i) := P^g(\xi_t^i|a_{1:t-1})$ as common belief on ξ_t^i based on the common information of the players, $a_{1:t-1}$, where $\pi_t^i \in \mathcal{P}(\mathcal{P}(\mathcal{X}^i))$. As it will be shown later, due to the independence of states and their evolution as independent controlled Markov processes, for any strategy profile of the players, joint beliefs on states can be factorized as product of their marginals i.e. $\pi_t(\xi_t) = \prod_{i=1}^N \pi_t^i(\xi_t^i)$. To accentuate this independence structure, we define $\underline{\pi}_t \in \times_{i\in\mathcal{N}}\mathcal{P}(\mathcal{X}^i)$ as vector of marginal beliefs where $\underline{\pi}_t := (\pi_t^i)_{i\in\mathcal{N}}$.

Inspired by the common agent approach in decentralized team problems [9], we now generate players' structured strategies as follows: player i at time t observes a common belief vector $\underline{\pi}_t$ and takes action γ_t^i, where $\gamma_t^i : \mathcal{P}(\mathcal{X}^i) \to \mathcal{P}(\mathcal{A}^i)$ is a partial (stochastic) function from its private belief ξ_t^i to a_t^i of the form $\gamma_t^i(a_t^i|\xi_t^i)$. These actions are generated through some policy $\theta^i = (\theta_t^i)_{t\in\mathcal{T}}$, $\theta_t^i : \times_{i\in\mathcal{N}}\mathcal{P}(\mathcal{P}(\mathcal{X}^i)) \to \{\mathcal{P}(\mathcal{X}^i) \to \mathcal{P}(\mathcal{A}^i)\}$, that operates on the common belief vector $\underline{\pi}_t$ so that $\gamma_t^i = \theta_t^i[\underline{\pi}_t]$. Then, the generated policy of the form $A_t^i \sim \theta_t^i[\underline{\pi}_t](\cdot|\xi_t^i)$ is also a policy of the form $A_t^i \sim g_t^i(\cdot|a_{1:t-1}, w_{1:t}^i)$ for an appropriately defined g. Although this is not relevant to our proofs, it can be shown that these structured policies form a sufficiently large, rich set of policies, which provides a good motivation for restricting attention to such equilibria. Specifically, it can be shown that policies g are outcome equivalent to policies of state θ, i.e., any expected total reward profile of the players that can be generated through a general policy profile g can also be generated through some policy profile θ. In the following lemma, we present the update functions of the private belief ξ_t^i and the public belief π_t^i.

Lemma 1: There exist update functions F^i, independent of players' strategies g, such that

$$\xi_{t+1}^i = F^i(\xi_t^i, w_{t+1}^i, a_t) \tag{5}$$

and update functions $\bar{F}^i$, independent of θ, such that

$$\pi_{t+1}^i = \bar{F}^i(\pi_t^i, \gamma_t^i, a_t). \tag{6}$$

Thus $\underline{\pi}_{t+1} = \bar{F}(\underline{\pi}_t, \gamma_t, a_t)$ where $\bar{F}$ is appropriately defined through (6).

Proof: The proofs are straightforward using Bayes' rule and the fact that players' state and observation histories, $X_{1:t}^i, W_{1:t}^i$, are conditionally independent across players given the action history $a_{1:t-1}$, and are provided in Appendix A. ■

Based on (5), we define an update kernel of ξ_t^i in (34) as $Q^i(\xi_{t+1}^i|\xi_t^i, a_t) := P(\xi_{t+1}^i|\xi_t^i, a_t)$. We now present the backward-forward algorithm to find PBE of the game $\mathfrak{D}$, where strategies of the players are of state θ. The algorithm resembles the one presented in [8] by the same authors for perfectly observable states.

1) Backward Recursion: In this section, we define an equilibrium generating function $\theta = (\theta_t^i)_{i\in\mathcal{N}, t\in\mathcal{T}}$ and a sequence of functions
$(V_t^i)_{i\in\mathcal{N}, t\in\{1,2,\ldots T+1\}}$, where $V_t^i : \times_{i\in\mathcal{N}}\mathcal{P}(\mathcal{P}(\mathcal{X}^i)) \times \mathcal{P}(\mathcal{X}^i) \to \mathbb{R}$, in a backward recursive way, as follows.

1. Initialize $\forall \underline{\pi}_{T+1} \in \times_{i\in\mathcal{N}}\mathcal{P}(\mathcal{P}(\mathcal{X}^i)), \xi_{T+1}^i \in \mathcal{P}(\mathcal{X}^i)$,

$$V_{T+1}^i(\underline{\pi}_{T+1}, \xi_{T+1}^i) := 0. \tag{7}$$

2. For $t = T, T-1, \ldots 1$, $\forall \underline{\pi}_t \in \times_{i\in\mathcal{N}}\mathcal{P}(\mathcal{P}(\mathcal{X}^i))$, let $\theta_t[\underline{\pi}_t]$ be generated as follows. Set $\tilde{\gamma}_t = \theta_t[\underline{\pi}_t]$, where $\tilde{\gamma}_t$ is the solution, if it exists[1],, of the following fixed point equation, $\forall i \in \mathcal{N}, \xi_t^i \in \mathcal{P}(\mathcal{X}^i)$,

$$\tilde{\gamma}_t^i(\cdot|\xi_t^i) \in \arg\max_{\gamma_t^i(\cdot|\xi_t^i)} \mathbb{E}^{\gamma_t^i(\cdot|\xi_t^i)\tilde{\gamma}_t^{-i}, \pi_t} \left\{ R^i(X_t, A_t) + V_{t+1}^i(\bar{F}(\underline{\pi}_t, \tilde{\gamma}_t, A_t), \Xi_{t+1}^i) \middle| \xi_t^i \right\}, \tag{8}$$

[1]Similar to the existence results shown in [10], it can be shown that in the special case where agent i's instantaneous reward does not depend on its private state x_t^i, and for uncontrolled states and observations, the fixed point equation always has a state-independent, myopic solution $\tilde{\gamma}_t^i(\cdot)$, since it degenerates to a Bayesian-Nash like best-response equation.

where expectation in (8) is with respect to random variables (X_t, A_t, Ξ^i_{t+1}) through the measure
$\xi_t(x_t)\pi_t^{-i}(\xi_t^{-i})\gamma^i_t(a^i_t|\xi^i_t)\tilde{\gamma}_t^{-i}(a_t^{-i}|\xi_t^{-i})Q^i(\xi^i_{t+1}|\xi^i_t, a_t)$, F is defined in Lemma 3 and Q^i is defined in (34). Furthermore, set

$$V^i_t(\underline{\pi}_t, \xi^i_t) := \mathbb{E}^{\tilde{\gamma}^i_t(\cdot|\xi^i_t)\tilde{\gamma}_t^{-i}, \pi_t}\left\{R^i(X_t, A_t) + V^i_{t+1}(\bar{F}(\underline{\pi}_t, \tilde{\gamma}_t, A_t), \Xi^i_{t+1})\middle|\xi^i_t\right\}. \tag{9}$$

It should be noted that (8) is a fixed point equation where the maximizer $\tilde{\gamma}^i_t$ appears in both, the left-hand-side and the right-hand-side of the equation. However, it is not the outcome of the maximization operation as in a best response equation, similar to that of a Bayesian Nash equilibrium.

2) Forward Recursion: Based on θ defined above in (7)–(9), we now construct a set of strategies β^* and beliefs μ^* for the game $\mathfrak{D}$ in a forward recursive way, as follows. As before, we will use the notation $\underline{\mu}^*_t[a_{1:t-1}] := (\mu^{*,i}_t[a_{1:t-1}])_{i\in\mathcal{N}}$ and $\mu^*_t[a_{1:t-1}]$ can be constructed from $\underline{\mu}^*_t[a_{1:t-1}]$ as $\mu^*_t[a_{1:t-1}](\xi_t) = \prod_{i=1}^N \mu^{*,i}_t[a_{1:t-1}](\xi^i_t)$, where $\mu^{*,i}_t[a_{1:t-1}]$ is a belief on ξ^i_t.

1. Initialize at time $t = 0$,

$$\mu^{*,i}_0[\phi](\xi_0) := \delta_{Q^i_x}(\xi^i_0). \tag{10}$$

2. For $t = 1, 2 \dots T, i \in \mathcal{N}, \forall a_{1:t}, w^i_{1:t}$

$$\beta^{*,i}_t(a^i_t|a_{1:t-1}, w^i_{1:t}) := \theta^i_t[\underline{\mu}^*_t[a_{1:t-1}]](a^i_t|\xi^i_t) \tag{11a}$$

$$\mu^{*,i}_{t+1}[a_{1:t}] := \bar{F}(\mu^{*,i}_t[a_{1:t-1}], \theta^i_t[\underline{\mu}^*_t[a_{1:t-1}]], a_t) \tag{11b}$$

where $\bar{F}$ is defined in Lemma 3.

Theorem 1: A strategy and belief profile (β^*, μ^*), constructed through backward/forward recursive algorithm is a PBE of the game, i.e. $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t-1}, w^i_{1:t}), \beta^i$,

$$\mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]}\left\{\sum_{n=t}^T R^i(X_n, A_n)\middle|a_{1:t-1}, w^i_{1:t}\right\} \geq$$
$$\mathbb{E}^{\beta^{i}_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]}\left\{\sum_{n=t}^T R^i(X_n, A_n)\middle|a_{1:t-1}, w^i_{1:t}\right\}. \tag{12}$$

Proof: The proof relies crucially on the specific fixed point construction in (8) and the conditional independence structure of states and observations, and is provided in Appendix B. ■

IV. Informational Cascades

In the following definition, we define informational cascades for a dynamic game with asymmetric information, and for a given PBE of that game, as those public histories of the game at which the future actions of the players are predictable[2].

Definition 1: For a given strategy and belief profile (β^*, μ^*) that constitute a PBE of the game[3], and for any time t and a sequence of action profile $a_{t:T}$, informational cascades can be defined as set of public histories h_t^c of the game such that at h_t^c and under (β^*, μ^*), actions $a_{t:T}$ are played almost surely, irrespective of players' future private history realizations, i.e. for a PBE (β^*, μ^*) and time t and actions $a_{t:T}$, cascades are defined by

$$\mathcal{C}_t^{a_{t:T}} := \{h_t^c \in \mathcal{H}_t^c \mid \forall i, \forall n \geq t, \forall h_n^i \text{ that are consistent with } h_t^c, \text{ and occur with non-zero probability, } \beta_n^{*,i}(a_n^i | h_n^i) = 1\}. \quad (13)$$

We also call an informational cascade a constant informational cascade if action profiles in the cascade are constant across time, i.e. for time t and action profile a, constant cascades are defined by

$$\mathcal{C}_t^{a} := \{h_t^c \in \mathcal{H}_t^c \mid \forall i, \forall n \geq t, \forall h_n^i \text{ that are consistent with } h_t^c, \text{ and occur with non-zero probability, } \beta_n^{*,i}(a^i | h_n^i) = 1\}. \quad (14)$$

For the general games considered in this section, which are dynamic game with asymmetric information and independent states, a more useful definition of cascades is the following.

Definition 2: For a given equilibrium generating function θ, and for time t and actions $a_{t:T}$, informational cascades are defined by the sets $\{\tilde{\mathcal{C}}_t^{a_{t:T}}\}_{t=1,\ldots T+1}$, which are defined as follows. For $t = T, T-1, \ldots 1$,

$$\tilde{\mathcal{C}}_{T+1} := \left\{ \text{ All possible common beliefs } \underline{\pi}_{T+1} \right\} \quad (15)$$

$$\tilde{\mathcal{C}}_t^{a_{t:T}} := \left\{ \underline{\pi}_t \mid \forall i, \ \forall \xi_t^i \in supp(\pi_t^i), \theta_t^i[\underline{\pi}_t](a_t^i | \xi_t^i) = 1 \text{ and } \bar{F}(\underline{\pi}_t, \theta_t[\underline{\pi}_t], a_t) \in \tilde{\mathcal{C}}_{t+1}^{a_{t+1:T}} \right\}. \quad (16)$$

[2]More generally, informational cascades can be thought as those histories of the game at which the system dynamics, from that point on, only depend on the common information.

[3]A stronger notion of informational cascade could be defined for *all* PBEs of the game.

A constant informational cascade for time t and actions profile a is defined as,

$$\tilde{\mathcal{C}}_{T+1} := \left\{ \text{ All possible common beliefs } \underline{\pi}_{T+1} \right\} \tag{17}$$

$$\begin{aligned}\tilde{\mathcal{C}}_t^a := \Big\{ & \underline{\pi}_t \mid \forall i,\ \forall \xi_t^i \in supp(\pi_t^i), \theta_t^i[\underline{\pi}_t](a^i|\xi_t^i) = 1 \\ & \text{and } \bar{F}(\underline{\pi}_t, \theta_t[\underline{\pi}_t], a) \in \tilde{\mathcal{C}}_{t+1}^a \Big\}.\end{aligned} \tag{18}$$

In the following lemma, we show the connection between the two definitions.

Lemma 2: Let (β^*, μ^*) be an SPBE of a dynamic game with asymmetric information and independent states, generated by an equilibrium generating function θ. Then $\forall t, a_{t:T}$,

$$(\mu_t^*)^{-1}(\tilde{\mathcal{C}}_t^{a_{t:T}}) = \mathcal{C}_t^{a_{t:T}}. \tag{19}$$

Proof: See Appendix D. ■

Corollary 1: Let (β^*, μ^*) be an SPBE of a dynamic game with asymmetric information and independent states, generated by an equilibrium generating function θ. Then $\forall t, a$,

$$(\mu_t^*)^{-1}(\tilde{\mathcal{C}}_t^{a}) = \mathcal{C}_t^{a}. \tag{20}$$

A. *Specific learning model*

We now consider a specific model that captures the learning aspect in a dynamic setting with strategic agents and decentralized information. The model is similar in spirit to the model considered in [2], [3] except we consider a finite number of players who take action in every epoch and participate during the entire duration of the game. We assume that players' states are uncontrollable and static i.e. $Q_x^i(x_{t+1}^i|x_t^i, a_t) = \delta_{x_t^i}(x_{t+1}^i)$, where $\mathcal{X}^i = \{-1, 1\}$. Since the set of states, $\mathcal{X}^i$ is has cardinality 2, the measure ξ_t^i can be sufficiently described by $\xi_t^i(1)$. Henceforth, in this section and in Appendix E, with slight abuse of notation, we also denote $\xi_t^i(1)$ by $\xi_t^i \in [0, 1]$, and reference is clear from context. In each epoch t, player i makes independent observation w_t^i about its state where $\mathcal{W}^i = \{-1, 1\}$, through an observation kernel of the form $Q_w^i(w_t^i|x_t^i, a_{t-1}^i)$ which does not depend on a_{t-1}^{-i}. Based on its information, it takes action a_t^i, where $\mathcal{A}^i = \{0, 1\}$, and earns an instantaneous reward given by

$$R^i(x, a_t^i) = a_t^i \left(\lambda x^i + \bar{\lambda} \frac{\sum_{j \neq i} x^j}{N-1} \right), \tag{21}$$

where $\lambda \in [0, 1], \bar{\lambda} = 1 - \lambda$. This scenario can thought of the case when players' states represent their talent, capabilities or popularity, and a player makes a decision to either invest (action =

1) or not invest (action = 0) in these players, where its instantaneous reward depends on some combination of the capabilities of all the players. We note that the instantaneous reward does not depend on other players' actions but on their states, and thus learning players' states is an important aspect of the problem.

1) Partially controlled observations: We consider the case where observations of the player i do depend on other players' actions, i.e. the observation kernel is of the form $Q_w^i(w_t^i|x_t^i, a_{t-1}^i)$. These observations are made through a binary symmetric channel such that $Q_w^i(-1|1,a^i) = Q_w^i(1|-1,a^i) = p_{a^i}$, where $p_1 \leq p_0 < 1/2$. This model implies that taking action 1 can improve the quality of a player's future private belief. In this case, the update functions of ξ_t^i and π_t^i in (5), (6) reduce to

$$\xi_{t+1}^i = F^i(\xi_t^i, w_{t+1}^i, a_t^i) \tag{22a}$$

$$\pi_{t+1}^i = \bar{F}^i(\pi_t^i, \gamma_t^i, a_t^i). \tag{22b}$$

and (8) in the backward recursion reduces to

$$\begin{aligned}\tilde{\gamma}_t^i(\cdot|\xi_t^i) \in \arg\max_{\gamma_t^i(\cdot|\xi_t^i)} \sum_{a_t^i} a_t^i \gamma_t^i(a_t^i|\xi_t^i)(\lambda(2\xi_t^i - 1) + \bar{\lambda}(2\hat{\xi}_t^{-i} - 1))\\ + \mathbb{E}^{\gamma_t^i(\cdot|\xi_t^i)\tilde{\gamma}_t^{-i},\pi_t}\left\{V_{t+1}^i(\bar{F}(\underline{\pi}_t, \tilde{\gamma}_t, A_t), \Xi_{t+1}^i)\middle|\xi_t^i\right\},\end{aligned} \tag{23}$$

For the learning model considered in Section IV-A, we characterize constant informational cascades through a time invariant set $\hat{\mathcal{C}}^a$ of common beliefs $\underline{\pi}$, defined as follows. Let

$$\begin{aligned}\hat{\mathcal{C}}^a := \Big\{\underline{\pi} \mid \forall i, \frac{1}{2} - \frac{\bar{\lambda}}{\lambda}(\hat{\xi}^{-i} - \frac{1}{2}) \geq 1 \text{ if } a^i = 0,\\ \frac{1}{2} - \frac{\bar{\lambda}}{\lambda}(\hat{\xi}^{-i} - \frac{1}{2}) \leq 0 \text{ if } a^i = 1\Big\},\end{aligned} \tag{24}$$

where

$$\hat{\xi}^{-i} := \frac{1}{N-1}\sum_{j\neq i}\mathbb{E}^{\pi^j}[\Xi^j]. \tag{25}$$

In the following theorem we show that the set $\hat{\mathcal{C}}^a$ defined in (24) characterizes a set of constant informational cascades for this problem. Specifically, we show that $\hat{\mathcal{C}}^a \subset \tilde{\mathcal{C}}^a$.

Theorem 2: If for some time t_0 and action profile a, $\underline{\pi}_{t_0} \in \hat{\mathcal{C}}^a$, then $\forall t \geq t_0, \underline{\pi}_t \in \hat{\mathcal{C}}^a$ and solutions of (23) satisfy $\tilde{\gamma}_t^i(a^i|\xi_t^i) = 1 \ \forall \xi_t^i \in [0,1]$. Moreover, for $t_0 \leq t \leq T$, V_t^i is given by,

$\forall \underline{\pi}_t \in \hat{\mathcal{C}}^a$,

$$V_t^i(\underline{\pi}_t, \xi_t^i) = (T - t + 1)(\lambda(2\xi_t^i - 1) + \bar{\lambda}(2\hat{\xi}_t^{-i} - 1))a^i. \tag{26}$$

Proof: See Appendix E. ■

As is the case in (26), for any π_t in a cascading set $\tilde{\mathcal{C}}_t^{a_{t:T}}, V_t^i(\pi_t, \cdot)$ represents reward to go for open loop control policy $a_{t:T}$.

B. Discussion

We characterize informational cascades by those histories of the game where learning stops for the players as a whole. Conceptually, they could be thought of as absorbing states of the system. It begets questions regarding the dynamics of the process that could lead to those states, for example hitting times of such sets and absorption probabilities. For the simplified problem considered in [2], cascades can be characterized as the fixed points of common belief update function, so that the common belief gets "stuck" once it reaches that state. It was shown that cascades eventually occur with probability 1 for that model. For the learning model considered in this section, common beliefs π_t still evolve in a cascade, although uninformatively, i.e., their evolution is directed by the primitives of the process and not on the new random variables being generated, namely, players' private observations. Also, if players' observations are informative, they asymptotically learn their true states, i.e., their private beliefs converge to dirac delta function on their true states. One trivial case when cascades could occur for this model is if the system was born in a cascade, i.e., the initial common belief, based on the prior distributions, is in cascades, $\pi_1 \in \hat{C}^a$. In general, a cascade could occur as in the following case. Suppose all players have low states (i.e. $x^i = -1$), but they get atypical observations initially, which lead them into believing that their states are high ($x^i = 1$). This information is conveyed through their actions, which leads the public belief into a cascade. Now, even though players eventually learn their true states, yet they remain in a (bad) cascade, each player believing that others have high states on average.

V. Conclusion

In this chapter we studied Bayesian learning dynamics of a specific class of dynamic games with asymmetric information. In the literature, as simplifying model is considered where herding behavior by selfish players is shown in a ergodic sequential buyers' game where a countable number of strategic buyers buy a product exactly once in the game. In this paper, we considered a more general scenario where players could participate in the game throughout the duration of the game. Players' states evolved as conditionally independent controlled Markov processes and players made noisy observations of their states. We first presented a sequential decomposition methodology to find SPBE of the game. We then studied a specific learning model and characterized information cascades using the general methodology described before. In general, the methodology presented serves as a framework for studying learning dynamics of decentralized systems with strategic agents. Some important research directions include characterization of cascades for specific classes of models, studying convergent learning behavior in such games including the probability and the rate of "falling" into a cascade, and incentive or mechanism design to avoid bad cascades.

Appendix A

Proof: We first prove the following claim on conditional independence of $x_{1:t}, w_{1:t}$ given $a_{1:t-1}$.

Claim 1: For any policy profile g and $\forall t$,

$$P^g(x_{1:t}, w_{1:t}|a_{1:t-1}) = \prod_{i=1}^{N} P^{g^i}(x^i_{1:t}, w^i_{1:t}|a_{1:t-1}) \tag{27}$$

Proof:

$$P^g(x_{1:t}, w_{1:t}|a_{1:t-1})$$

$$= \frac{P^g(x_{1:t}, w_{1:t}, a_{1:t-1})}{\sum_{x_{1:t},w_{1:t}} P^g(x_{1:t}, w_{1:t}, a_{1:t-1})} \tag{28a}$$

$$= \frac{\prod_{i=1}^N Q^i_x(x^i_1)Q^i_w(w^i_1|x^i_1)\prod_{n=1}^{t-1} g^i_n(a^i_n|a_{1:n-1}, w^i_{1:n-1})Q^i_x(x^i_{n+1}|a_n, x^i_n)Q^i_w(w^i_{n+1}|x^i_{n+1}, a_n)}{\sum_{x_{1:t},w_{1:t}} \prod_{i=1}^N Q^i_x(x^i_1)Q^i_w(w^i_1|x^i_1)\prod_{n=1}^{t-1} g^i_n(a^i_n|a_{1:n-1}, w^i_{1:n-1})Q^i_x(x^i_{n+1}|a_n, x^i_n)Q^i_w(w^i_{n+1}|x^i_{n+1}, a_n)} \tag{28b}$$

$$= \frac{\prod_{i=1}^N Q^i_x(x^i_1)Q^i_w(w^i_1|x^i_1)\prod_{n=1}^{t-1} g^i_n(a^i_n|a_{1:n-1}, w^i_{1:n-1})Q^i_x(x^i_{n+1}|a_n, x^i_n)Q^i_w(w^i_{n+1}|x^i_{n+1}, a_n)}{\prod_{i=1}^N \sum_{x^i_{1:t},w^i_{1:t}} Q^i_x(x^i_1)Q^i_w(w^i_1|x^i_1)\prod_{n=1}^{t-1} g^i_n(a^i_n|a_{1:n-1}, w^i_{1:n-1})Q^i_x(x^i_{n+1}|a_n, x^i_n)Q^i_w(w^i_{n+1}|x^i_{n+1}, a_n)} \tag{28c}$$

and thus

$$P^g(x_{1:t}, w_{1:t}|a_{1:t-1}) = \prod_{i=1}^{N} P^{g^i}(x^i_{1:t}, w^i_{1:t}|a_{1:t-1}) \tag{28d}$$

■

Now for any g we have,

$$\xi^i_{t+1}(x^i_{t+1}) \stackrel{\triangle}{=} P^g(x^i_{t+1}|a_{1:t}, w^i_{1:t+1}) \tag{29a}$$

$$= \frac{\sum_{x^i_t} P^g(x^i_t, a_t, x^i_{t+1}, w^i_{t+1}|a_{1:t-1}, w^i_{1:t})}{\sum_{\tilde{x}^i_{t+1}\tilde{x}^i_t} P^g(\tilde{x}^i_t, a_t, w^i_{t+1}, \tilde{x}^i_{t+1}|a_{1:t-1}, w^i_{1:t})} \tag{29b}$$

$$= \frac{\sum_{x^i_t} \xi^i_t(x^i_t) P^g(a^{-i}_t|a_{1:t-1}, w^i_{1:t}, x^i_t) Q^i_x(x^i_{t+1}|a_t, x^i_t) Q^i_w(w^i_{t+1}|x^i_{t+1}, a_t)}{\sum_{\substack{\tilde{x}^i_{t+1}\\ \tilde{x}^i_t}} \xi^i_t(\tilde{x}^i_t) P^g(a^{-i}_t|a_{1:t-1}, w^i_{1:t}, \tilde{x}^i_t) Q^i_x(\tilde{x}^i_{t+1}|a_t, \tilde{x}^i_t) Q^i_w(w^i_{t+1}|\tilde{x}^i_{t+1}, a_t)}, \tag{29c}$$

where (29c) is true because a^i_t is a function of $(a_{1:t-1}, w^i_{1:t})$ and thus term involving can be cancelled in numerator and denominator. We now consider the quantity $P^g(a^{-i}_t|a_{1:t-1}, w^i_{1:t}, x^i_t)$

$$P^g(a^{-i}_t|a_{1:t-1}, w^i_{1:t}, x^i_t) = \sum_{w^{-i}_{1:t}} P^g(a^{-i}_t, w^{-i}_{1:t}|a_{1:t-1}, w^i_{1:t}, x^i_t) \tag{30a}$$

$$= \sum_{w^{-i}_{1:t}} P^g(w^{-i}_{1:t}|a_{1:t-1}, w^i_{1:t}, x^i_t) \prod_{j\neq i} g^j_t(a^j_t|a_{1:t-1}, w^j_{1:t}) \tag{30b}$$

$$= \sum_{w^{-i}_{1:t}} P^{g^{-i}}(w^{-i}_{1:t}|a_{1:t-1}) \prod_{j\neq i} g^j_t(a^j_t|a_{1:t-1}, w^j_{1:t}) \tag{30c}$$

$$= P^{g^{-i}}(a^{-i}_t|a_{1:t-1}) \tag{30d}$$

where (30c) follows from Claim 1 in Appendix A since $w^{-i}_{1:t}$ is conditionally independent of $(w^i_{1:t}, x^i_t)$ given $a_{1:t-1}$ and is only a function of g^{-i}. Since this term does not depend on x^i_t, it gets cancelled in the final expression of ξ^i_{t+1}

$$\xi^i_{t+1}(x^i_{t+1}) = \frac{\sum_{x^i_t} \xi^i_t(x^i_t) Q^i_x(x^i_{t+1}|x^i_t, a_t) Q^i_w(w^i_{t+1}|x^i_{t+1}, a_t)}{\sum_{\tilde{x}^i_{t+1}} \sum_{x^i_t} \xi^i_t(x^i_t) Q^i_x(\tilde{x}^i_{t+1}|x^i_t, a_t) Q^i_w(w^i_{t+1}|\tilde{x}^i_{t+1}, a_t)}. \tag{31}$$

Thus the claim of the Lemma follows. Based on this claim, we can conclude that

$$\xi^i_t(x^i_t) \;=\; P^g(x^i_t|a_{1:t-1}, w^i_{1:t}) \;=\; P(x^i_t|a_{1:t-1}, w^i_{1:t}). \tag{32}$$

Also, based on the update of ξ_t^i in (5), we define an update kernel

$$Q^i(\xi_{t+1}^i|\xi_t^i, a_t) := P(\xi_{t+1}^i|\xi_t^i, a_t) \tag{33}$$

$$= \sum_{x_t^i, x_{t+1}^i, w_{t+1}^i} \xi_t^i(x_t^i) Q_x^i(x_{t+1}^i|x_t^i, a_t) Q_w^i(w_{t+1}^i|x_{t+1}^i, a_t) I_{F(\xi_t^i, w_{t+1}^i, a_t^i)}(\xi_{t+1}^i) \tag{34}$$

■

Lemma 3: There exists an update function $\bar{F}$ of π_t, independent of ψ

$$\pi_{t+1}^i = \bar{F}(\pi_t^i, \gamma_t^i, a_t) \tag{35}$$

Proof:

$$\pi_{t+1}(\xi_{t+1})$$

$$= P^\psi(\xi_{t+1}|a_{1:t}, \gamma_{1:t+1}) \tag{36a}$$

$$= P^\psi(\xi_{t+1}|a_{1:t}, \gamma_{1:t}) \tag{36b}$$

$$= \frac{\sum_{\substack{\xi_t, x_t, \\ x_{t+1}, w_{t+1}}} P^\psi(\xi_t, x_t, a_t, x_{t+1}, w_{t+1}, \xi_{t+1}|a_{1:t-1}, \gamma_{1:t})}{\sum_{\xi_t} P^\psi(\xi_t, a_t|a_{1:t-1}, \gamma_{1:t})} \tag{36c}$$

$$= \frac{\sum_{\substack{\xi_t, x_t, \\ x_{t+1}, w_{t+1}}} \prod_{i=1}^N \pi_t^i(\xi_t^i) \xi_t^i(x_t^i) \gamma_t^i(a_t^i|\xi_t^i) Q_x^i(x_{t+1}^i|x_t^i, a_t) Q_w^i(w_{t+1}^i|x_{t+1}^i, a_t) I_{F^i(\xi_t^i, w_{t+1}^i, a_t)}(\xi_{t+1}^i)}{\sum_{\xi_t} \prod_{i=1}^N \pi_t^i(\xi_t^i) \gamma_t^i(a_t^i|\xi_t^i)} \tag{36d}$$

$$= \frac{\prod_{i=1}^N \sum_{\substack{\xi_t^i, x_t^i, \\ x_{t+1}^i, w_{t+1}^i}} \pi_t^i(\xi_t^i) \xi_t^i(x_t^i) \gamma_t^i(a_t^i|\xi_t^i) Q_x^i(x_{t+1}^i|x_t^i, a_t) Q_w^i(w_{t+1}^i|x_{t+1}^i, a_t) I_{F^i(\xi_t^i, w_{t+1}^i, a_t)}(\xi_{t+1}^i)}{\sum_{\xi_t} \prod_{i=1}^N \pi_t^i(\xi_t^i) \gamma_t^i(a_t^i|\xi_t^i)} \tag{36e}$$

$$= \frac{\prod_{i=1}^N \sum_{\substack{\xi_t^i, x_t^i, \\ x_{t+1}^i, w_{t+1}^i}} \pi_t^i(\xi_t^i) \xi_t^i(x_t^i) \gamma_t^i(a_t^i|\xi_t^i) Q_x^i(x_{t+1}^i|x_t^i, a_t) Q_w^i(w_{t+1}^i|x_{t+1}^i, a_t) I_{F^i(\xi_t^i, w_{t+1}^i, a_t)}(\xi_{t+1}^i)}{\prod_{i=1}^N \sum_{\xi_t^i} \pi_t^i(\xi_t^i) \gamma_t^i(a_t^i|\xi_t^i)} \tag{36f}$$

Thus we have,

$$\pi_{t+1} = \prod_{i=1}^N \bar{F}(\pi_t^i, \gamma_t^i, a_t) \tag{36g}$$

■

APPENDIX B
(PROOF OF THEOREM 1)

Proof: We prove (12) using induction and from results in Lemma 4, 5 and 6 proved in Appendix C. For base case at $t = T$, $\forall i \in \mathcal{N}, (a_{1:T-1}, w^i_{1:T}) \in \mathcal{H}^i_T, \beta^i$

$$\mathbb{E}^{\beta^{*,i}_T \beta^{*,-i}_T, \mu^*_T[a_{1:T-1}]} \left\{ R^i(X_T, A_T) \middle| a_{1:T-1}, w^i_{1:T} \right\}$$
$$= V^i_T(\underline{\mu}^*_T[a_{1:T-1}], \xi^i_T) \tag{37a}$$
$$\geq \mathbb{E}^{\beta^i_T \beta^{*,-i}_T, \mu^*_T[a_{1:T-1}]} \left\{ R^i(X_T, A_T) \middle| a_{1:T-1}, w^i_{1:T} \right\} \tag{37b}$$

where (37a) follows from Lemma 6 and (37b) follows from Lemma 4 in Appendix C.

Let the induction hypothesis be that for $t+1$, $\forall i \in \mathcal{N}, (a_{1:t}, w^i_{1:t+1}) \in \mathcal{H}^i_{t+1}, \beta^i$,

$$\mathbb{E}^{\beta^{*,i}_{t+1:T} \beta^{*,-i}_{t+1:T}, \mu^*_{t+1}[a_{1:t}]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \middle| a_{1:t}, w^i_{1:t+1} \right\} \geq$$
$$\mathbb{E}^{\beta^i_{t+1:T} \beta^{*,-i}_{t+1:T}, \mu^*_{t+1}[a_{1:t}]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \middle| a_{1:t}, w^i_{1:t+1} \right\}. \tag{38a}$$

Then $\forall i \in \mathcal{N}, (a_{1:t-1}, w^i_{1:t}) \in \mathcal{H}^i_t, \beta^i$, we have

$$\mathbb{E}^{\beta^{*,i}_{t:T} \beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]} \left\{ \sum_{n=t}^{T} R^i(X_n, A_n) \middle| a_{1:t-1}, w^i_{1:t} \right\}$$
$$= V^i_t(\underline{\mu}^*_t[a_{1:t-1}], \xi^i_t) \tag{39a}$$
$$\geq \mathbb{E}^{\beta^i_t \beta^{*,-i}_t, \mu^*_t[a_{1:t-1}]} \left\{ R^i(X_t, A_t) + V^i_{t+1}(\underline{\mu}^*_{t+1}[a_{1:t-1}A_t], \Xi^i_{t+1}) \middle| a_{1:t-1}, w^i_{1:t} \right\} \tag{39b}$$
$$= \mathbb{E}^{\beta^i_t \beta^{*,-i}_t, \mu^*_t[a_{1:t-1}]} \Big\{ R^i(X_t, A_t) +$$
$$\mathbb{E}^{\beta^{*,i}_{t+1:T} \beta^{*,-i}_{t+1:T}, \mu^*_{t+1}[a_{1:t-1}, A_t]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \middle| a_{1:t-1}, A_t, w^i_{1:t} W^i_{t+1} \right\} \Big| a_{1:t-1}, w^i_{1:t} \Big\} \tag{39c}$$
$$\geq \mathbb{E}^{\beta^i_t \beta^{*,-i}_t, \mu^*_t[a_{1:t-1}]} \Big\{ R^i(X_t, A_t) +$$
$$\mathbb{E}^{\beta^i_{t+1:T} \beta^{*,-i}_{t+1:T} \mu^*_{t+1}[a_{1:t-1}, A_t]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \middle| a_{1:t-1}, A_t, w^i_{1:t}, W^i_{t+1} \right\} \Big| a_{1:t-1}, w^i_{1:t} \Big\} \tag{39d}$$

$$= \mathbb{E}^{\beta_t^i \beta_t^{*,-i}, \mu_t^*[a_{1:t-1}]} \Big\{ R^i(X_t, A_t)$$

$$+\mathbb{E}^{\beta_{t:T}^i \beta_{t:T}^{*,-i} \mu_t^*[a_{1:t-1}]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \Big| a_{1:t-1}, A_t, w_{1:t}^i, W_{t+1}^i \right\} \Big| a_{1:t-1}, w_{1:t}^i \Big\} \tag{39e}$$

$$= \mathbb{E}^{\beta_{t:T}^i \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]} \left\{ \sum_{n=t}^{T} R^i(X_n, A_n) \Big| a_{1:t-1}, w_{1:t}^i \right\}, \tag{39f}$$

where (39a) follows from Lemma 6, (39b) follows from Lemma 4, (39c) follows from Lemma 6, (39d) follows from induction hypothesis in (38a) and (39e) follows from Lemma 5. Moreover, construction of θ in (8), and consequently definition of β^* in (11a) are pivotal for (39e) to follow from (39d).

We note that μ^* satisfies the consistency condition of [7, p. 331] from the fact that (a) for all t and for every common history $a_{1:t-1}$, all players use the same belief $\mu_t^*[a_{1:t-1}]$ on x_t and (b) the belief μ_t^* can be factorized as $\mu_t^*[a_{1:t-1}] = \prod_{i=1}^N \mu_t^{*,i}[a_{1:t-1}]$ $\forall a_{1:t-1} \in \mathcal{H}_t^c$ where $\mu_t^{*,i}$ is updated through Bayes' rule ($\bar{F}$) as in Lemma 3 in Appendix A.

■

APPENDIX C

Lemma 4: $\forall t \in \mathcal{T}, i \in \mathcal{N}, (a_{1:t-1}, w_{1:t}^i) \in \mathcal{H}_t^i, \beta_t^i$

$$V_t^i(\underline{\mu}_t^*[a_{1:t-1}], \xi_t^i) \geq \mathbb{E}^{\beta_t^i \beta_t^{*,-i}, \mu_t^*[a_{1:t-1}]} \Big\{ R^i(X_t, A_t) + \tag{40}$$

$$V_{t+1}^i(F(\underline{\mu}_t^*[a_{1:t-1}], \beta_t^*(\cdot|a_{1:t-1}, \cdot), A_t), \Xi_{t+1}^i) \Big| a_{1:t-1}, w_{1:t}^i \Big\}. \tag{41}$$

Proof: We prove this Lemma by contradiction.

Suppose the claim is not true for t. This implies $\exists i, \hat{\beta}_t^i, \hat{a}_{1:t-1}, \hat{w}_{1:t}^i$ such that

$$\mathbb{E}^{\hat{\beta}_t^i \beta_t^{*,-i}, \mu_t^*[\hat{a}_{1:t-1}]} \Big\{ R^i(X_t, A_t) + V_{t+1}^i(F(\underline{\mu}_t^*[\hat{a}_{1:t-1}], \beta_t^*(\cdot|\hat{a}_{1:t-1}, \cdot), A_t), \Xi_{t+1}^i) \Big| \hat{a}_{1:t-1}, \hat{w}_{1:t}^i \Big\}$$
$$> V_t^i(\underline{\mu}_t^*[\hat{a}_{1:t-1}], \hat{\xi}_t^i). \tag{42}$$

We will show that this contradicts the definition of V_t^i in (9).

Construct $\hat{\gamma}_t^i(a_t^i|\xi_t^i) = \begin{cases} \hat{\beta}_t^i(a_t^i|\hat{a}_{1:t-1}, \hat{w}_{1:t}^i) & \xi_t^i = \hat{\xi}_t^i \\ \text{arbitrary} & \text{otherwise.} \end{cases}$

Then for $\hat{a}_{1:t-1}, \hat{w}^i_{1:t}$, we have

$$
\begin{aligned}
&V^i_t(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \hat{\xi}^i_t) \\
&= \max_{\gamma^i_t(\cdot|\hat{\xi}^i_t)} \mathbb{E}^{\gamma^i_t(\cdot|\hat{\xi}^i_t)\beta^{*,-i}_t, \mu^*_t[\hat{a}_{1:t-1}]} \left\{ R^i(X_t, A_t) + V^i_{t+1}(F(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \beta^*_t(\cdot|\hat{a}_{1:t-1}, \cdot), A_t), \Xi^i_{t+1}) \middle| \hat{\xi}^i_t \right\}, \quad (43a) \\
&\geq \mathbb{E}^{\hat{\gamma}^i_t(\cdot|\hat{\xi}^i_t)\beta^{*,-i}_t, \mu^*_t[\hat{a}_{1:t-1}]} \left\{ R^i(X_t, A_t) + V^i_{t+1}(F(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \beta^*_t(\cdot|\hat{a}_{1:t-1}, \cdot), A_t), \Xi^i_{t+1}) \middle| \hat{\xi}^i_t \right\} \quad (43b) \\
&= \sum_{x_t, \xi^{-i}_t, a_t, \xi_{t+1}} \left\{ R^i(x_t, a_t) + V^i_{t+1}(F(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \beta^*_t(\cdot|\hat{a}_{1:t-1}, \cdot), a_t), \xi^i_{t+1}) \right\} \times \\
&\hat{\xi}^i_t(x^i_t)\xi^{-i}_t(x^{-i}_t)\mu^{*,-i}_t[\hat{a}_{1:t-1}](\xi^{-i}_t)\hat{\gamma}^i_t(a^i_t|\hat{\xi}^i_t)\beta^{*,-i}_t(a^{-i}_t|\hat{a}_{1:t-1}, \xi^{-i}_t)Q^i(\xi^i_{t+1}|\hat{\xi}^i_t, a_t) \quad (43c) \\
&= \sum_{\substack{x_t, \xi^{-i}_t, \\ a_t, \xi_{t+1}}} \left\{ R^i(x_t, a_t) + V^i_{t+1}(F(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \beta^*_t(\cdot|\hat{a}_{1:t-1}, \cdot), a_t), \xi^i_{t+1}) \right\} \times \\
&\hat{\xi}^i_t(x^i_t)\xi^{-i}_t(x^{-i}_t)\mu^{*,-i}_t[\hat{a}_{1:t-1}](\xi^{-i}_t)\hat{\beta}^i_t(a^i_t|\hat{a}_{1:t-1}, \hat{w}^i_{1:t})\beta^{*,-i}_t(a^{-i}_t|\hat{a}_{1:t-1}, \xi^{-i}_t)Q^i(\xi^i_{t+1}|\hat{\xi}^i_t, a_t) \quad (43d) \\
&= \mathbb{E}^{\hat{\beta}^i_t\beta^{*,-i}_t, \mu^*_t[\hat{a}_{1:t-1}]} \left\{ R^i(X_t, A_t) + V^i_{t+1}(F(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \beta^*_t(\cdot|\hat{a}_{1:t-1}, \cdot), A_t), X^i_{t+1}) \middle| \hat{a}_{1:t-1}, \hat{w}^i_{1:t} \right\} \quad (43e) \\
&> V^i_t(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \hat{\xi}^i_t) \quad (43f)
\end{aligned}
$$

where (43a) follows from the definition of V^i_t in (9), (43d) follows from definition of $\hat{\gamma}^i_t$ and (43f) follows from (42). However this leads to a contradiction. ■

Lemma 5: $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t}, w^i_{1:t+1}) \in \mathcal{H}^i_{t+1}$ and β^i_t

$$
\mathbb{E}^{\beta^i_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \middle| a_{1:t}, w^i_{1:t+1} \right\} =
$$

$$
\mathbb{E}^{\beta^i_{t+1:T}\beta^{*,-i}_{t+1:T}, \mu^*_{t+1}[a_{1:t}]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \middle| a_{1:t}, w^i_{1:t+1} \right\}. \quad (44)
$$

Thus the above quantities do not depend on β^i_t.

Proof: Essentially this claim stands on the fact that $\mu^{*,-i}_{t+1}[a_{1:t}]$ can be updated from $\mu^{*,-i}_t[a_{1:t-1}], \beta^{*,-i}_t$ and a_t, as $\mu^{*,-i}_{t+1}[a_{1:t}] = \prod_{j\neq i} \bar{F}(\mu^{*,-i}_t[a_{1:t-1}], \beta^{*,-i}_t, a_t)$ as in Lemma 3. Since the above expectations involve random variables $X_{t+1:T}, A_{t+1:T}$,, we consider $P^{\beta^i_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]}(x_{t+1:T}, a_{t+1:T} \middle| a_{1:t}, w^i_{1:t+1})$.

$$P^{\beta^i_{t:T}\beta^{*,-i}_{t:T},\,\mu^*_t[a_{1:t-1}]}(x_{t+1:T},a_{t+1:T}\Big|a_{1:t},w^i_{1:t+1}) =$$
$$\frac{P^{\beta^i_{t:T}\beta^{*,-i}_{t:T},\,\mu^*_t[a_{1:t-1}]}(a_t,x_{t+1},w^i_{t+1},a_{t+1:T},x_{t+2:T}\Big|a_{1:t-1},w^i_{1:t})}{P^{\beta^i_{t:T}\beta^{*,-i}_{t:T},\,\mu^*_t[a_{1:t-1}]}(a_t,w^i_{t+1}\Big|a_{1:t-1},w^i_{1:t})} \tag{45a}$$

We consider the numerator and the denominator separately. The numerator in (45a) is given by

$$Nr = \sum_{x_t,\xi_t^{-i}} P^{\beta^i_{t:T}\beta^{*,-i}_{t:T},\,\mu^*_t[a_{1:t-1}]}(x_t,\xi_t^{-i}\Big|a_{1:t-1},w^i_{1:t})\beta^i_t(a^i_t|a_{1:t-1},w^i_{1:t})\beta^{*,-i}_t(a^{-i}_t|a_{1:t-1},\xi^{-i}_t)Q_x(x_{t+1}|x_t,a_t)$$
$$Q^i_w(w^i_{t+1}|x^i_{t+1},a_t)P^{\beta^i_{t:T}\beta^{*,-i}_{t:T},\,\mu^*_t[a_{1:t-1}]}(a_{t+1:T},x_{t+2:T}|a_{1:t},w^i_{1:t+1},x_{t:t+1}) \tag{45b}$$
$$= \sum_{x_t,\xi_t^{-i}} \xi_t(x_t)\mu^{*,-i}_t[a_{1:t-1}](\xi^{-i}_t)\beta^i_t(a^i_t|a_{1:t-1},w^i_{1:t})\beta^{*,-i}_t(a^{-i}_t|a_{1:t-1},\xi^{-i}_t)Q_x(x_{t+1}|x_t,a_t)$$
$$Q^i_w(w^i_{t+1}|x^i_{t+1},a_t)P^{\beta^i_{t+1:T}\beta^{*,-i}_{t+1:T},\,\mu^*_{t+1}[a_{1:t}]}(a_{t+1:T},x_{t+2:T}|a_{1:t},w^i_{1:t+1},x_{t+1}) \tag{45c}$$

where (45c) follows from the fact that probability on $(a_{t+1:T},x_{2+t:T})$ given $a_{1:t},w^i_{1:t+1},x_{t:t+1},\mu^*_t[a_{1:t-1}]$ depends on $a_{1:t},w^i_{1:t+1},x_{t+1},\mu^*_{t+1}[a_{1:t}]$ through $\beta^i_{t+1:T}\beta^{*,-i}_{t+1:T}$. Similarly, the denominator in (45a) is given by

$$Dr = \sum_{\tilde{x}_t,\tilde{\xi}_t^{-i},\tilde{x}^i_{t+1}} P^{\beta^i_{t:T}\beta^{*,-i}_{t:T},\,\mu^*_t}(\tilde{x}_t,\xi_t^{-i}|a_{1:t-1},w^i_{1:t})\beta^i_t(a^i_t|a_{1:t-1},w^i_{1:t})\beta^{*,-i}_t(a^{-i}_t|a_{1:t-1},\tilde{\xi}^{-i}_t)Q^i_x(\tilde{x}^i_{t+1}|\tilde{x}^i_t,a_t)$$
$$Q^i_w(w^i_{t+1}|\tilde{x}^i_{t+1},a_t) \tag{45d}$$
$$= \sum_{\tilde{x}_t,\tilde{\xi}_t^{-i},\tilde{x}^i_{t+1}} \xi^i_t(\tilde{x}^i_t)\tilde{\xi}^{-i}_t(\tilde{x}^{-i}_t)\mu^{*,-i}_t[a_{1:t-1}](\tilde{\xi}^{-i}_t)\beta^i_t(a^i_t|a_{1:t-1},w^i_{1:t})\beta^{*,-i}_t(a^{-i}_t|a_{1:t-1},\tilde{\xi}^{-i}_t)Q^i_x(\tilde{x}^i_{t+1}|\tilde{x}^i_t,a_t)$$
$$Q^i_w(w^i_{t+1}|\tilde{x}^i_{t+1},a_t) \tag{45e}$$

By canceling the terms $\beta^i_t(\cdot)$ in the numerator and the denominator, (45a) is given by

$$\frac{Nr}{Dr}P^{\beta^i_{t+1:T}\beta^{*,-i}_{t+1:T},\,\mu^*_{t+1}[a_{1:t}]}(a_{t+1:T},x_{t+2:T}|a_{1:t},w^i_{1:t+1},x_{t+1}) \tag{45f}$$

where

$$Nr = \sum_{x_t, \xi_t^{-i}} \xi_t(x_t)\mu_t^{*,-i}[a_{1:t-1}](\xi_t^{-i})\beta_t^{*,-i}(a_t^{-i}|a_{1:t-1},\xi_t^{-i})Q_x(x_{t+1}|x_t,a_t)Q_w^i(w_{t+1}^i|x_{t+1}^i,a_t) \tag{45g}$$

$$= \sum_{x_t^i} \xi_t^i(x_t^i)Q_x^i(x_{t+1}^i|x_t^i,a_t^i)Q_w^i(w_{t+1}^i|x_{t+1}^i,a_t)$$

$$\sum_{x_t^{-i},\xi_t^{-i}} \xi_t^{-i}(x_t^{-i})\mu_t^{*,-i}[a_{1:t-1}](\xi_t^{-i})\beta_t^{*,-i}(a_t^{-i}|a_{1:t-1},\xi_t^{-i})Q_x^{-i}(x_{t+1}^{-i}|x_t^{-i},a_t) \tag{45h}$$

and

$$Dr = \sum_{\tilde{x}_t,\tilde{\xi}_t^{-i},\tilde{x}_{t+1}^i} \xi_t^i(\tilde{x}_t^i)\tilde{\xi}_t^{-i}(\tilde{x}_t^{-i})\mu_t^{*,-i}[a_{1:t-1}](\tilde{\xi}_t^{-i})\beta_t^{*,-i}(a_t^{-i}|a_{1:t-1},\tilde{\xi}_t^{-i})Q_x^i(\tilde{x}_{t+1}^i|\tilde{x}_t^i,a_t)Q_w(w_{t+1}^i|\tilde{x}_{t+1}^i,a_t) \tag{45i}$$

$$= \sum_{\tilde{x}_t^i,\tilde{x}_{t+1}^i} \xi_t^i(\tilde{x}_t^i)Q_x^i(\tilde{x}_{t+1}^i|\tilde{x}_t^i,a_t)Q_w(w_{t+1}^i|\tilde{x}_{t+1}^i,a_t) \sum_{\tilde{x}_t^{-i},\tilde{\xi}_t^{-i}} \tilde{\xi}_t^{-i}(\tilde{x}_t^{-i})\mu_t^{*,-i}[a_{1:t-1}](\tilde{\xi}_t^{-i})\beta_t^{*,-i}(a_t^{-i}|a_{1:t-1},\tilde{\xi}_t^{-i}) \tag{45j}$$

Thus (45a) is given by

$$= \xi_{t+1}(x_{t+1})\mu_{t+1}^{*,-i}[a_{1:t}](\xi_{t+1}^{-i})P^{\beta_{t+1:T}^i\beta_{t+1:T}^{*,-i},\,\mu_{t+1}^*[a_{1:t}]}(a_{t+1:T},x_{t+2:T}|a_{1:t},w_{1:t}^i,x_{t+1}) \tag{45k}$$

$$= P^{\beta_{t+1:T}^i\beta_{t+1:T}^{*,-i},\,\mu_{t+1}^*[a_{1:t}]}(x_{t+1},a_{t+1:T},x_{t+2:T}|a_{1:t},w_{1:t+1}^i). \tag{45l}$$

■

Lemma 6: $\forall i \in \mathcal{N}, t \in \mathcal{T}, a_{1:t-1} \in \mathcal{H}_t^c, w_{1:t}^i \in (\mathcal{W}^i)^t$

$$V_t^i(\underline{\mu}_t^*[a_{1:t-1}],\xi_t^i) = \mathbb{E}^{\beta_{t:T}^{*,i}\beta_{t:T}^{*,-i},\mu_t^*[a_{1:t-1}]}\left\{\sum_{n=t}^T R^i(X_n,A_n)\Big| a_{1:t-1},w_{1:t}^i\right\}. \tag{46}$$

Proof:

We prove the Lemma by induction. For $t = T$,

$$\mathbb{E}^{\beta_T^{*,i}\beta_T^{*,-i},\mu_T^*[a_{1:T-1}]}\left\{R^i(X_T,A_T)\Big| a_{1:T-1},w_{1:T}^i\right\}$$

$$= \sum_{x_T^{-i}a_T} R^i(x_T,a_T)\xi_T(x_T)\mu_T^*[a_{1:T-1}](\xi_T^{-i})\beta_T^{*,i}(a_T^i|a_{1:T-1},\xi_T^i)\beta_T^{*,-i}(a_T^{-i}|a_{1:T-1},\xi_T^{-i}) \tag{47a}$$

$$= V_T^i(\underline{\mu}_T^*[a_{1:T-1}],\xi_T^i), \tag{47b}$$

where (47b) follows from the definition of V_t^i in (9) and the definition of β_T^* in the forward recursion in (11a).

Suppose the claim is true for $t+1$, i.e., $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t}, w^i_{1:t+1}) \in \mathcal{H}^i_{t+1}$

$$V^i_{t+1}(\underline{\mu}^*_{t+1}[a_{1:t}], \xi^i_{t+1}) = \mathbb{E}^{\beta^{*,i}_{t+1:T}\beta^{*,-i}_{t+1:T}, \mu^*_{t+1}[a_{1:t}]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \Big| a_{1:t}, w^i_{1:t+1} \right\}. \tag{48}$$

Then $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t-1}, w^i_{1:t}) \in \mathcal{H}^i_t$, we have

$$\mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]} \left\{ \sum_{n=t}^{T} R^i(X_n, A_n) \Big| a_{1:t-1}, w^i_{1:t} \right\}$$

$$= \mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]} \big\{ R^i(X_t, A_t) +$$

$$\mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \Big| a_{1:t-1}, A_t, w^i_{1:t}, W^i_{t+1} \right\} \Big| a_{1:t-1}, w^i_{1:t} \Bigg\} \tag{49a}$$

$$= \mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]} \big\{ R^i(X_t, A_t) +$$

$$\mathbb{E}^{\beta^{*,i}_{t+1:T}\beta^{*,-i}_{t+1:T}, \mu^*_{t+1}[a_{1:t-1}, A_t]} \left\{ \sum_{n=t+1}^{T} R^i(X_n, A_n) \Big| a_{1:t-1}, A_t, w^i_{1:t}, W^i_{t+1} \right\} \Big| a_{1:t-1}, w^i_{1:t} \Bigg\} \tag{49b}$$

$$= \mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T}, \mu^*_t[a_{1:t-1}]} \left\{ R^i(X_t, A_t) + V^i_{t+1}(\underline{\mu}^*_{t+1}[a_{1:t-1}A_t], \Xi^i_{t+1}) \Big| a_{1:t-1}, w^i_{1:t} \right\} \tag{49c}$$

$$= \mathbb{E}^{\beta^{*,i}_{t}\beta^{*,-i}_{t}, \mu^*_t[a_{1:t-1}]} \left\{ R^i(X_t, A_t) + V^i_{t+1}(\underline{\mu}^*_{t+1}[a_{1:t-1}A_t], \Xi^i_{t+1}) \Big| a_{1:t-1}, w^i_{1:t} \right\} \tag{49d}$$

$$= V^i_t(\underline{\mu}^*_t[a_{1:t-1}], \xi^i_t), \tag{49e}$$

where (49b) follows from Lemma 5 in Appendix C, (49c) follows from the induction hypothesis in (48), (49d) follows because the random variables involved in expectation, X^{-i}_t, A_t, X^i_{t+1} do not depend on $\beta^{*,i}_{t+1:T}\beta^{*,-i}_{t+1:T}$ and (49e) follows from the definition of β^*_t in the forward recursion in (11a), the definition of μ^*_{t+1} in (11b) and the definition of V^i_t in (9). ■

APPENDIX D

Proof: We will prove the result by induction on t. The result is vacuously true for $T+1$. Suppose it is also true for $t+1$, i.e.

$$(\mu^*_{t+1})^{-1}(\tilde{\mathcal{C}}^{a_{t+1:T}}_{t+1}) = \mathcal{C}^{a_{t+1:T}}_{t+1}. \tag{50}$$

We show that the result holds true for t. In the following two cases, we show that if there exists an element in one set, it also belongs to the other. From the contrapositive of the statement, if one is empty, so is the other.

Case 1. We prove $(\mu^*_t)^{-1}(\tilde{\mathcal{C}}^{a_{t:T}}_t) \subset \mathcal{C}^{a_{t:T}}_t$

Let $h_t^c \in (\mu_t^*)^{-1}(\tilde{\mathcal{C}}_t^{a_{t:T}})$. We will show that $h_t^c \in \mathcal{C}_t^{a_{t:T}}$.

Since $h_t^c \in (\mu_t^*)^{-1}(\tilde{\mathcal{C}}_t^{a_{t:T}})$, this implies $\mu_t^*[h_t^c] \in \tilde{\mathcal{C}}_t^{a_{t:T}}$. Then by the definition of $\tilde{\mathcal{C}}_t^{a_{t:T}}, \forall i,\ \forall \xi_t^i \in supp(\mu_t^{*,i}[h_t^c]),\ \theta_t^i[\mu_t^*[h_t^c]](a_t^i|\xi_t^i) = 1$. Since $\xi_t^i(x_t^i) = P(x_t^i|h_t^i)\ \forall x_t^i$,
$\mu_t^{*,i}[h_t^c](\xi_t^i) = P^\theta(\xi_t^i|h_t^c)\ \forall \xi_t^i$ and $\beta_t^{*,i}(a_t^i|h_t^i) = \theta_t^i[\mu_t^*[h_t^c]](a_t^i|\xi_t^i)$ by the definition of β^*, this implies $\forall i, \beta_t^{*,i}(a_t^i|h_t^i) = 1$, $\forall h_t^i$ that are consistent with h_t^c and occur with non-zero probability.

Also since $\mu_t^*[h_t^c] \in \tilde{\mathcal{C}}_t^{a_{t:T}}$, this implies $\bar{F}([\mu_t^*[h_t^c], \theta_t[\mu_t^*[h_t^c]], a_t) \in \tilde{\mathcal{C}}_{t+1}^{a_{t+1:T}}$ by definition of $\tilde{\mathcal{C}}_t^{a_{t:T}}$. Thus $\mu_{t+1}^*[h_t^c, a_t] \in \tilde{\mathcal{C}}_{t+1}^{a_{t+1:T}}$, since $\mu_{t+1}^*[h_t^c, a_t] =$
$\bar{F}([\mu_t^*[h_t^c], \theta_t[\mu_t^*[h_t^c]], a_t)$ by definition. Using the induction hypothesis, $(h_t^c, a_t) \in \mathcal{C}_{t+1}^{a_{t+1:T}}$, which implies $\forall i, \beta_n^{*,i}(a_n^i|h_n^i) = 1,\ \forall n \geq t+1, \forall h_n^i$ that are consistent with (h_t^c, a_t) and occur with non-zero probability.

The above two facts conclude that $\forall i, \beta_n^{*,i}(a_n^i|h_n^i) = 1,\ \forall n \geq t, \forall h_n^i$ that are consistent with h_t^c and occur with non-zero probability, which implies $h_t^c \in \mathcal{C}_t^{a_{t:T}}$ by the definition of $\mathcal{C}_t^{a_{t:T}}$.

Case 2. We prove $(\mu_t^*)^{-1}(\tilde{\mathcal{C}}_t^{a_{t:T}}) \supset \mathcal{C}_t^{a_{t:T}}$.

Let $h_t^c \in \mathcal{C}_t^{a_{t:T}}$. We will show that $\mu_t^*[h_t^c] \in \tilde{\mathcal{C}}_t^{a_{t:T}}$.

Since $h_t^c \in \mathcal{C}_t^{a_{t:T}}$, this implies $\forall i, \beta_t^{*,i}(a_t^i|h_t^i) = 1,\ \forall h_t^i$ that are consistent with h_t^c and occur with non-zero probability. Since $\beta_t^{*,i}(a_t^i|h_t^i) = \theta_t^i[\mu_t^*[h_t^c]](a_t^i|\xi_t^i)$, by the definition of β^*, where $\xi_t^i(x_t^i) = P(x_t^i|h_t^i)\ \forall x_t^i$, this implies $\forall i, \theta_t^i[\mu_t^*[h_t^c]](a_t^i|\xi_t^i) = 1, \forall \xi_t^i \in supp(\mu_t^{*,i}[h_t^c])$, where $\mu_t^{*,i}[h_t^c](\xi_t^i) = P^\theta(\xi_t^i|h_t^c)\ \forall \xi_t^i$.

Also, since $h_t^c \in \mathcal{C}_t^{a_{t:T}}$, it is implied by the definition of $\mathcal{C}_t^{a_{t:T}}$ that $(h_t^c, a_t) \in \mathcal{C}_{t+1}^{a_{t+1:T}}$. This implies $\mu_{t+1}^*[h_t^c, a_t] \in \tilde{\mathcal{C}}_{t+1}^{a_{t+1:T}}$ by the induction hypothesis. Since, by definition, $\mu_{t+1}^*[h_t^c, a_t] = \bar{F}([\mu_t^*[h_t^c], \theta_t[\mu_t^*[h_t^c]], a_t)$, this implies $\bar{F}([\mu_t^*[h_t^c], \theta_t[\mu_t^*[h_t^c]], a_t) \in \tilde{\mathcal{C}}_{t+1}^{a_{t+1:T}}$.

Since we have shown that $\forall i, \theta_t^i[\mu_t^*[h_t^c]](a_t^i|\xi_t^i) = 1, \forall \xi_t^i \in supp(\mu_t^*[h_t^c])$ and
$\bar{F}([\mu_t^*[h_t^c], \theta_t[\mu_t^*[h_t^c]], a_t) \in \tilde{\mathcal{C}}_{t+1}^{a_{t+1:T}}$, this implies $\mu_t^*[h_t^c] \in \tilde{\mathcal{C}}_t^{a_{t:T}}$ by the definition of $\tilde{\mathcal{C}}_t^{a_{t:T}}$.

The above two cases complete the induction step.

■

APPENDIX E

Proof: We prove this by induction on t_0. For $t_0 = T$, (23) reduces to

$$\tilde{\gamma}_T^i(\cdot|\xi_T^i) \in \arg\max_{\gamma_T^i(\cdot|\xi_T^i)} \sum_{a_T^i} a_T^i \gamma_T^i(a_T^i|\xi_T^i)(\lambda(2\xi_T^i - 1) + \bar{\lambda}(2\hat{\xi}_T^{-i} - 1)), \tag{51}$$

and since $\pi_T \in \hat{\mathcal{C}}^a$, it is easy to verify that $\tilde{\gamma}_T^i(a^i|\xi_T^i) = 1,\ \forall \xi_T^i \in [0,1]$ and thus $V_T^i(\pi_T^{-i}, \xi_T^i) = (\lambda(2\xi_T^i - 1) + \bar{\lambda}(2\hat{\xi}_T^{-i} - 1))a^i$. This establishes the base case.

Now, suppose the claim is true for $t_0 = \tau + 1$ i.e. if $\pi_{\tau+1} \in \hat{\mathcal{C}}^a$, then $\forall t \geq \tau + 1, \pi_t \in \hat{\mathcal{C}}^a$ and $\tilde{\gamma}_t^i(a^i|\xi_t^i) = 1\ \forall \xi_t^i \in [0,1]$. Moreover, for $\tau + 1 \leq t \leq T$, V_t^i is given by, $\forall \pi_t \in \hat{\mathcal{C}}^a$,

$$V_t^i(\pi_t^{-i}, \xi_t^i) = (T - t + 1)(\lambda(2\xi_t^i - 1) + \bar{\lambda}(2\hat{\xi}_t^{-i} - 1))a^i. \tag{52}$$

Then if $\pi_\tau \in \hat{\mathcal{C}}^a$, then $\tilde{\gamma}_\tau^i(a^i|\xi_\tau^i) = 1\ \forall \xi_\tau^i \in [0,1]$ satisfies (23) since,

$$\begin{aligned}\tilde{\gamma}_\tau^i(\cdot|\xi_\tau^i) \in \arg\max_{\gamma_\tau^i(\cdot|\xi_\tau^i)} \sum_{a_\tau^i} a_\tau^i \gamma_\tau^i(a_\tau^i|\xi_\tau^i)(\lambda(2\xi_\tau^i - 1) + \bar{\lambda}(2\hat{\xi}_\tau^{-i} - 1))\\ + \mathbb{E}^{\gamma_\tau^i(\cdot|\xi_\tau^i)\tilde{\gamma}_\tau^{-i}, \pi_\tau} \left\{ V_{\tau+1}^i(F(\underline{\pi}_\tau^{-i}, \tilde{\gamma}_\tau^{-i}, A_\tau^{-i}), \Xi_{\tau+1}^i) \middle| \xi_\tau^i \right\}\end{aligned} \tag{53}$$

$$= \arg\max_{\gamma_\tau^i(\cdot|\xi_\tau^i)} \sum_{a_\tau^i} a_\tau^i \gamma_\tau^i(a_\tau^i|\xi_\tau^i)(\lambda(2\xi_\tau^i - 1) + \bar{\lambda}(2\hat{\xi}_\tau^{-i} - 1))$$

$$+ \mathbb{E}^{\gamma_\tau^i(\cdot|\xi_\tau^i)\tilde{\gamma}_\tau^{-i}, \pi_\tau} \left\{ (T - \tau)(\lambda(2\Xi_{\tau+1}^i - 1) + \bar{\lambda}(2\hat{\Xi}_{\tau+1}^{-i} - 1))a^i | \xi_\tau^i \right\} \tag{54}$$

$$= \arg\max_{\gamma_\tau^i(\cdot|\xi_\tau^i)} \sum_{a_\tau^i} a_\tau^i \gamma_\tau^i(a_\tau^i|\xi_\tau^i)(\lambda(2\xi_\tau^i - 1) + \bar{\lambda}(2\hat{\xi}_\tau^{-i} - 1))$$

$$+ (T - \tau)(\lambda(2\xi_\tau^i - 1) + \bar{\lambda}(2\hat{\xi}_\tau^{-i} - 1))a^i \tag{55}$$

$$= \arg\max_{\gamma_\tau^i(\cdot|\xi_\tau^i)} \sum_{a_\tau^i} a_\tau^i \gamma_\tau^i(a_\tau^i|\xi_\tau^i)(\lambda(2\xi_\tau^i - 1) + \bar{\lambda}(2\hat{\xi}_\tau^{-i} - 1)), \tag{56}$$

where (54) follows from the fact that $F(\underline{\pi}_\tau, \tilde{\gamma}_\tau, a_\tau) \in C^a,\ \forall a_\tau$, as shown in Claim 2, and induction hypothesis, (55) follows from Claim 2 and Claim 3 and (56) follows from the fact that the second term does not depend on $\gamma_\tau^i(\cdot|\xi_\tau^i)$. This also shows that, $\forall \pi_t \in \hat{\mathcal{C}}^a$,

$$V_\tau^i(\pi_\tau, \xi_\tau^i) = (T - \tau + 1)(\lambda(2\xi_\tau^i - 1) + \bar{\lambda}(2\hat{\xi}_\tau^{-i} - 1))a^i, \tag{57}$$

which completes the induction step. ■

Claim 2: Expectation of π_{t+1}^i under non-informative $\tilde{\gamma}_t^i$ of the form $\tilde{\gamma}_t^i(a^i|\xi_t^i) = 1\ \forall \xi_t^i \in [0,1]$, remains the same as mean of π_t^i, i.e.,

$$\mathbb{E}\{\Xi_{t+1}^i(1)|\pi_t^i, \tilde{\gamma}_t^i, a^i\} = \mathbb{E}\{\Xi_t^i(1)|\pi_t^i\} \tag{58}$$

Proof:

$$\begin{aligned}
&\mathbb{E}\{\Xi^i_{t+1}(1)|\pi^i_t,\tilde{\gamma}^i_t,a^i\} \\
&= \sum_{\xi^i_{t+1}(1)} \xi^i_{t+1}(1)\bar{F}^i(\pi^i_t,\tilde{\gamma}^i_t,a^i)(\xi^i_{t+1}(1)) && (59)\\
&= \frac{\sum_{\xi^i_t,x^i,\xi^i_{t+1}(1)} \xi^i_{t+1}(1)\pi^i_t(\xi^i_t)\xi^i_t(x^i)\tilde{\gamma}^i_t(a^i_t|\xi^i_t)Q^i_w(w^i_{t+1}|x^i,a_t)I_{F^i(\xi^i_t,w^i_{t+1},a_t)(1)}(\xi^i_{t+1}(1))}{\sum_{\xi^i_t,x^i,w^i_{t+1}} \pi^i_t(\xi^i_t)\xi^i_t(x^i)\tilde{\gamma}^i_t(a^i_t|\xi^i_t)} && (60)\\
&= \frac{\sum_{\xi^i_t,x^i,w^i_{t+1},\xi^i_{t+1}(1)} \xi^i_{t+1}(1)\pi^i_t(\xi^i_t)\xi^i_t(x^i)Q^i_w(w^i_{t+1}|x^i,a^i)I_{F^i(\xi^i_t,w^i_{t+1},a^i)(1)}(\xi^i_{t+1}(1))}{\sum_{\xi^i_t,x^i} \pi^i_t(\xi^i_t)\xi^i_t(x^i)} && (61)\\
&= \sum_{\xi^i_t,x^i,w^i_{t+1}} F^i(\xi^i_t,w^i_{t+1},a^i)(1)\pi^i_t(\xi^i_t)\xi^i_t(x^i)Q^i_w(w^i_{t+1}|x^i,a^i) && (62)\\
&= \sum_{\xi^i_t,w^i_{t+1}} \frac{\xi^i_t(1)Q^i_w(w^i_{t+1}|1,a^i)}{\sum_{\tilde{x}^i}\xi^i_t(\tilde{x}^i)Q^i_w(w^i_{t+1}|\tilde{x}^i,a^i)}\pi^i_t(\xi^i_t)\sum_{x^i}\xi^i_t(x^i)Q^i_w(w^i_{t+1}|x^i,a^i) && (63)\\
&= \sum_{\xi^i_t}\xi^i_t(1)\pi^i_t(\xi^i_t(1)) && (64)\\
&= \mathbb{E}\{\Xi^i_t(1)|\pi^i_t\} && (65)
\end{aligned}$$

■

Claim 3: For any γ^i_t,

$$\mathbb{E}\{\Xi^i_{t+1}(1)|\xi^i_t,\gamma^i_t\} = \xi^i_t(1) \quad (66)$$

Proof:

$$\begin{aligned}
&\mathbb{E}\{\Xi^i_{t+1}(1)|\xi^i_t,\gamma^i_t\} \\
&= \sum_{x^i,w^i_{t+1},a^i_t,\xi^i_{t+1}(1)} \xi^i_{t+1}(1)I_{\bar{F}^i(\xi^i_t,w^i_{t+1},a^i_t)(1)}(\xi^i_{t+1}(1))\xi^i_t(x^i)Q^i_w(w^i_{t+1}|x^i,a^i_t)\gamma^i_t(a^i_t|\xi^i_t) && (67)\\
&= \sum_{x^i,w^i_{t+1},a^i_t} \bar{F}^i(\xi^i_t,w^i_{t+1},a^i_t)(1)\xi^i_t(x^i)Q^i_w(w^i_{t+1}|x^i,a^i_t)\gamma^i_t(a^i_t|\xi^i_t) && (68)\\
&= \sum_{a^i_t,w^i_{t+1}} \frac{\xi^i_t(1)Q^i_w(w^i_{t+1}|1,a^i_t)}{\sum_{\tilde{x}^i}\xi^i_t(\tilde{x}^i)Q^i_w(w^i_{t+1}|\tilde{x}^i,a^i_t)}\gamma^i_t(a^i_t|\xi^i_t)\sum_{x^i}\xi^i_t(x^i)Q^i_w(w^i_{t+1}|x^i,a^i_t) && (69)\\
&= \sum_{a^i_t,w^i_{t+1}} \xi^i_t(1)Q^i_w(w^i_{t+1}|1,a^i_t)\gamma^i_t(a^i_t|\xi^i_t) && (70)\\
&= \xi^i_t(1) && (71)
\end{aligned}$$

■

REFERENCES

[1] A. V. Banerjee, "A simple model of herd behavior," *The Quarterly Journal of Economics*, pp. 797–817, 1992.

[2] S. Bikhchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy*, vol. 100, no. 5, pp. pp. 992–1026, 1992. [Online]. Available: http://www.jstor.org/stable/2138632

[3] L. Smith and P. Sörensen, "Pathological outcomes of observational learning," *Econometrica*, vol. 68, no. 2, pp. 371–398, 2000. [Online]. Available: http://dx.doi.org/10.1111/1468-0262.00113

[4] 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. [Online]. Available: http://restud.oxfordjournals.org/content/78/4/1201.abstract

[5] T. N. Le, V. Subramanian, and R. Berry, "The impact of observation and action errors on informational cascades," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, Dec 2014, pp. 1917–1922.

[6] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*, ser. MIT Press Books. The MIT Press, 1994, vol. 1.

[7] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.

[8] D. Vasal and A. Anastasopoulos, "A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information," in *American Control Conference*, Boston, US, 2016, available on arXiv.

[9] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," *Automatic Control, IEEE Transactions on*, vol. 58, no. 7, pp. 1644–1658, 2013.

[10] Y. Ouyang, H. Tavafoghi, and D. Teneketzis, "Dynamic oligopoly games with private Markovian dynamics," in *Proc. 54th IEEE Conf. Decision and Control (CDC)*, 2015.