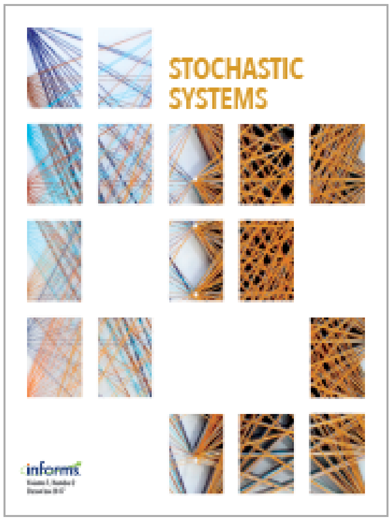

Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Framework for Studying Decentralized Bayesian Learning with Strategic Agents

Deepanshu Vasal, Achilleas Anastasopoulos

To cite this article:
Deepanshu Vasal, Achilleas Anastasopoulos (2022) A Framework for Studying Decentralized Bayesian Learning with Strategic Agents. Stochastic Systems 12(3):274-292. https://doi.org/10.1287/stsy.2021.0092

STOCHASTIC SYSTEMS
Vol. 12, No. 3, September 2022, pp. 274–292
ISSN 1946-5238 (online)
https://pubsonline.informs.org/journal/stsy

A Framework for Studying Decentralized Bayesian Learning with Strategic Agents

Deepanshu Vasal,[a,*] Achilleas Anastasopoulos[b]

[a] Department of Electrical and Computer Engineering, University of Texas, Austin, Texas 78712; [b] Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48105
*Corresponding author
Contact: dvasal@utexas.edu, https://orcid.org/0000-0003-1089-8080 (DV); anastas@umich.edu, https://orcid.org/0000-0003-3795-1654 (AA)

Received: July 30, 2021
Revised: November 24, 2021
Accepted: December 3, 2021
Published Online in Articles in Advance: March 22, 2022

https://doi.org/10.1287/stsy.2021.0092

Abstract. We study the problem of 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 investigated under a simplifying model where selfish players appear sequentially and act once in the game. It has been 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 just described. In particular, our model incorporates cases where players can act repeatedly and there is strategic interaction in that each agent's payoff may also depend on other players' actions. The proposed framework hinges on a sequential decomposition methodology for finding structured perfect Bayesian equilibria of a general class of dynamic games with asymmetric information. Using this methodology, we study a specific dynamic learning model where players make decisions about public investment based on their estimates of everyone's states. We characterize a set of informational cascades for this problem where learning stops for the team as a whole. Moreover, we show that such cascades occur almost surely.

Funding: This work was funded by NSF [Grant ECCS-1608361].

Keywords: Bayesian learning • social networks • informational cascades • dynamic games with asymmetric information • perfect Bayesian equilibrium

1. Introduction

The problem of how information spreads in a social network is of profound importance in understanding how learning occurs in a group of people or in a society, and it is important even more so today, with the ubiquitous presence of the internet and social media. Some scenarios of interest include how people vote for a candidate; the process of deciding to buy competitors' products; or dynamics of mass protests and movements, fads, trends, or cult behavior. In these examples there exists a group of people who have access to certain private information available through their peers or their own experience and certain publicly available information, such as the actions of others, available through mass media. Using this information, people make decisions that affect their reward and further spread of information in the system.

Such problems have been addressed in various disciplines such as behavioral economics, statistics, engineering, and computer science (DeGroot 1974; Banerjee 1992; Bikhchandani et al. 1992; Ellison and Fudenberg 1993, 1995; Bala and Goyal 1998; Smith and Sörensen 2000; Gale and Kariv 2003; Acemoglu et al. 2011; Guarino et al. 2011; Jadbabaie et al. 2012; Guarino and Jehiel 2013; Herrera and Hörner 2013; Mossel and Tamuz 2013; Harel et al. 2014; Mossel et al. 2014, 2015, 2020; Nedić et al. 2016; Le et al. 2017; Molavi et al. 2017). These problems have the following key features: (a) there are *multiple decision makers* (henceforth referred to as players) who can be cooperative or strategic, based on whether they have the same or different objectives; (b) there is *asymmetry of information* such that players have private and common information; and (c) there is *dynamic evolution* of the system. From a mathematical perspective, analysis of such problems entails two challenges: (i) one that is *decision theoretic*: finding the optimum or equilibrium or heuristic strategies of players, and (ii) one that is *statistical/probabilistic/analytic*: understanding the evolution and limiting behavior of the system dynamics under those strategies.

In two seminal papers (Banerjee 1992, Bikhchandani et al. 1992), the authors investigate the occurrence of fads in a social network, which was later generalized in Smith and Sörensen (2000). In particular, these works study a problem of learning over a social network with pure informational externalities (i.e., where a player's reward does not directly depend on other players' actions; however, those actions provide useful information about the state of the system). In this model, there is a product that is either good or bad, and there are countably many buyers (i.e., *different decision makers*) who are chosen exogenously and act exactly once in the process. Players make a 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 scenario, where the publicly available information becomes powerful enough that a user discards her own private information and follows the majority action of her predecessors. As a result, the user's 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. Although a good cascade is desirable, there is a positive probability of a bad cascade that hurts all future users in the community. Alternative learning models that study cascades have also appeared in the literature, such as in Acemoglu et al. (2011) and Le et al. (2017). Inspired by social networks, Acemoglu et al. consider a model where players observe only a random set of past actions. They show that under sufficient conditions of *expanding observations* and unbounded private belief log-likelihood ratios, players learn the true state asymptotically, and thus cascading does not occur. Le et al. study a model where agents observe the past actions through a noisy process where, again, they show that cascading does not occur. The common assumption in all of these models is that players act only once in the game, and there are informational externalities only, which allows for easy computation of game equilibrium strategies.

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 (with dynamic zero-sum games in the extreme), or more generally, there is strategic interaction in that each player's payoff may also depend on other players' actions. Studying such scenarios may reveal more interesting and richer equilibrium behaviors, including cascading phenomena not manifested in the models considered in the current literature. Mossel et al. (2020, p. 2) introduce a *static* notion of equilibrium called social learning equilibrium (SLE) to study social learning. They argue that "the analysis of asymptotic equilibrium behavior in dynamic games is not straightforward, resulting in a limited range of tractable models and a focus on extremely stylized settings; in particular, the literature has largely avoided studying models of repeated actions by rational non-myopic players." They also cite Gale and Kariv (2003, p. 20), saying that "the computational difficulty of solving the model is massive even in the case of three persons .... This is an important subject for future research." SLE is an asymptotic concept (which is also myopic in nature, as it is defined at the "end of the game") that takes away all the nuances that come about in the dynamics of a game partly because, as stated, analyzing the dynamics is quite challenging. The aforementioned difficulties are precisely what our work is addressing. Specifically, in this paper, we provide a sequential decomposition methodology to study all such sequential models of repeated actions by rational nonmyopic players. We provide a methodology that allows for the study such games in a more nuanced way and reveals exactly what kind of information will aggregate.

An indispensable tool for studying cascades and, generally speaking, Bayesian learning with strategic agents in such complex settings 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 the perfect Bayesian equilibrium (PBE), sequential equilibrium, and trembling hand equilibrium (Fudenberg and Tirole 1991, Osborne and Rubinstein 1994). Each of these notions of equilibrium consists of a strategy and a belief profile of all players where the equilibrium strategies satisfy sequential rationality (i.e., no player has an incentive to unilaterally deviate at equilibrium) given the equilibrium beliefs, and the equilibrium beliefs are derived from the equilibrium strategy profile using Bayes' rule (whenever possible). For the games considered in the current literature, including Bikhchandani et al. (1992), Smith and Sörensen (2000), Acemoglu et al. (2011), and Le et al. (2014), because every buyer participates only for one time period, finding the PBE reduces to solving a straightforward, one-shot optimization problem. However, for general dynamic games with asymmetric information, finding the PBE is hard, because 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 on the difficulty of finding PBEs, see Fudenberg and Tirole (1991), chap. 8). In general, there is no known sequential decomposition methodology for finding the general PBE for such games.

1.1. Contributions

In this paper, we consider a general model appropriate to the study of Bayesian learning where a finite number of players have different states associated with them that evolve as conditionally independent Markov processes.

Players do not perfectly observe their states; rather, they make independent, noisy observations of those states. The important new ingredient in this model is that players act repeatedly throughout the game, and there is strategic interaction in that each player's payoff may also depend on other players' actions. As a result, players have to strategize over the entire time horizon of the game. This model extends in a significant way the models considered in Ouyang et al. (2017) and Vasal et al. (2019), where players observe their state perfectly. Perfectly knowing one's state is a simplistic assumption that may not hold in the real world, and we argue that not perfectly knowing one's state is a more realistic assumption. The investment example we present at the end of this paper is one practical case where players may not know their states. Other motivating scenarios include, for instance, a wind-producing firm that may not, from the beginning, know the true cost of its wind production or a university student who may not know the true worth of his or her talents. They make noisy observations of their values over the course of time and eventually learn them. In these examples, the wind producer would observe the costs of wind produced every day, and a candidate may observe his or her grades at the university. Our model provides a general framework for studying such scenarios. Our contributions are as follows.

a. We first present in Theorem 1 a backward/forward algorithm for finding the *structured* PBE (SPBE) of the asymmetric information dynamic game. The term "structured" refers to the fact that equilibrium strategies in the SPBE depend on appropriately defined belief states instead of the whole observed history of the player. These equilibria are analogous to Markov perfect equilibria (MPE) defined in Maskin and Tirole (2001) but for the case of asymmetric information. The results in Vasal et al. (2019) vis-à-vis Theorem 1 in this paper can be interpreted with the analogy of dynamic programming methodology for Markov decision processes (MDP) versus that for partially observed Markov decision processes (POMDP), where in the former, the state of the system is perfectly observed by the controller, and in the latter, the state is imperfectly observed, and thus a new belief state is introduced that then behaves as an MDP.

b. We then utilize the aforementioned framework to study Bayesian learning dynamics and, specifically, informational cascades in dynamic games with asymmetric information. In general, an informational cascade at time t is the set of those public histories for which players' actions from that point onward stop depending on their private information. As a result, once a cascade is entered, the system dynamics are governed only through the common information, and any private information is discarded. By focusing on structured equilibrium strategies, we propose a concise characterization of such cascades as sets of appropriately defined public beliefs with the above-mentioned property. Unlike other settings 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 scenarios where players may be adversarial to each others' learning.

c. Finally, we consider a specific dynamic learning model with pure informational externalities where each player makes a decision to invest (or not invest) in the team, depending on her estimate of the average of all players' states. Players' states relate to their cost for investing. In this setting, learning players' states is an important aspect of the problem, although players are not adversarial to each others' learning. Using the methodology presented earlier, we characterize (in Theorem 2) a set of informational cascades for this model where, once in a cascade, players' estimates of others' states freeze, and learning stops for the team. This occurs despite the fact that, asymptotically, players learn their own states perfectly. This example serves as motivation for exploring the vast landscape of scenarios that can be studied through the proposed methodology.

The rest of this paper is structured as follows. In Section 2, we describe the model and problem statement. In Section 3, we provide a general methodology to find SPBEs for such games. In Section 4, we provide an application of the model and methodology described in Sections 2 and 3, where we formally define informational cascades and specialize our methodology to study a specific Bayesian learning game, for which we characterize a class of informational cascades. We conclude in Section 5. Part of the paper, without all the proofs and Theorem 3, was presented as Vasal and Anastasopoulos (2016a).

1.2. 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 the notation $-i$ to represent all players other than player i (i.e., $-i = \{1, 2, \ldots, i-1, i+1, \ldots, N\}$). We use the notation $a_{t:t'}$ to represent the 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 $\mathcal{S}$, $\Delta(\mathcal{S})$ represents the space of probability measures on $\mathcal{S}$, and $|\mathcal{S}|$ represents its cardinality. We denote by $\mathbb{P}^g$ (or $\mathbb{E}^g$) the probability measure generated by (or expectation with respect to) strategy profile g. For random variables X, Y, we use the notation $\mathbb{P}(x|y)$ to indicate $\mathbb{P}(X=x|Y=y)$. 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 the 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 shorthand 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 an almost sure (a.s.) sense.

2. General Model

2.1. 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

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

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

where $a_t = (a_t^1, \ldots, a_t^N)$, and a_t^i is the action taken by player i at time t. We follow the convention that for $t=1$, the above-mentioned kernels reduce to $\mathbb{P}(x_1) = \prod_{i=1}^N Q_x^i(x_1^i)$, and we use the same notation for the kernels Q_x^i for any t with the distinction being obvious from the context. Player i does not observe her state perfectly; rather, she 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, $\forall t \in \{1,\ldots T\}$,

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

Again, for $t=1$, the meaning of the preceding is that $\mathbb{P}(w_1|x_1) = \prod_{i=1}^N Q_w^i(w_t^i|x_t^i)$, and we use the same notation for the kernels Q_w^i for any t with the distinction being obvious from the context. 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, and we also assume that both kernels Q_x and Q_w have full support.[1] We further assume that all quantities $Q_x^i(x_t^i|x_{t-1}^i, a_{t-1})$ and $Q_w^i(w_t^i|x_t^i, a_{t-1})$ are rational numbers. The first assumption will simplify the discussion about off-equilibrium strategies and beliefs. The second is a mild assumption that is sufficient to resolve some technical issues with defining beliefs on beliefs as is done in Section 3. Let $g^i = (g_t^i)_{t\in\mathcal{T}}$ be a probabilistic strategy of player i, where $g_t^i : \mathcal{A}^{t-1}\times(\mathcal{W}^i)^t \to \Delta(\mathcal{A}^i)$ 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 time t, player i receives an instantaneous reward $R^i(x_t, a_t)$. The objective of player i is to maximize her 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 asymmetric information and with simultaneous moves. Although this model considers all N players acting in all periods of the game, 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 her state process $(X_t^0)_t$, has reward function zero, and plays actions $a_t^0 = w_t^0 = x_t^0$. For instance, let $\mathcal{X}^0 = \mathcal{A}^0 = \mathcal{N}$. Once the quantity a_t^0 is publicly observed, all players can determine that the acting player (at time t) will be the one indicated by a_t^0. Thus $R_t^i(x_t, a_t) = 0$ if $i \neq a_t^0$, and $\forall i$, $Q_x(x_{t+1}^i|x_t^i, a_t) = Q_x(x_{t+1}^i|x_t^i, a_t^{a_t^0})$. Here, in each period, only one player (player $a_t^0 = w_t^0 = x_t^0$) acts in the game while all other nonacting players receive zero rewards during that period.

2.2. Solution Concept: PBE

In this section, we introduce the PBE as an appropriate equilibrium concept for the game considered. Any history of this game at which players take action is of the form $h_t = (a_{1:t-1}, x_{1:t}, w_{1:t}) \in \mathcal{H}_t$. At any time t, player i observes $h_t^i = (a_{1:t-1}, w_{1:t}^i) \in \mathcal{H}_t^i$, and all players together observe $h_t^c = a_{1:t-1} \in \mathcal{H}_t^c$. An appropriate concept of equilibrium for such games is the PBE (Fudenberg and Tirole 1991), which consists of a pair (β,μ): a strategy profile $\beta = (\beta_t^i)_{t\in\mathcal{T}, i\in\mathcal{N}}$, where $\beta_t^i : \mathcal{H}_t^i \to \Delta(\mathcal{A}^i)$, and a belief profile[2] $\mu = ({}^i\mu_t)_{t\in\mathcal{T}, i\in\mathcal{N}}$, where ${}^i\mu_t : \mathcal{H}_t^i \to \Delta(\mathcal{X}\times(\mathcal{W}^{-i})^t)$, that satisfy the following two conditions.

1. *Sequential Rationality:* For all $\forall i \in \mathcal{N}, t \in \mathcal{T}, h_t^i \in \mathcal{H}_t^i, \tilde{\beta}^i$,

$$\mathbb{E}^{\beta_{t:T}^i \beta_{t:T}^{-i},\, {}^i\mu_t}\left\{\sum_{n=t}^{T} R^i(X_n, A_n) | h_t^i\right\} \geq \mathbb{E}^{\tilde{\beta}_{t:T}^i \beta_{t:T}^{-i},\, {}^i\mu_t}\left\{\sum_{n=t}^{T} R^i(X_n, A_n) | h_t^i\right\}. \quad (4)$$

2. *Belief Consistency:* For all $\forall i \in \mathcal{N}, t \in \mathcal{T} \setminus \{1\}, h_{t-1}^i \in \mathcal{H}_{t-1}^i, h_t^i = (h_{t-1}^i, a_{t-1}, w_t^i) \in \mathcal{H}_t^i$,
   - if $\mathbb{P}^{\beta}(h_t^i | h_{t-1}^i) > 0$, we must have[3] (Bayes' rule update)

$$ {}^i\mu_t[h_t^i](x_t, w_{1:t}^{-i}) = \frac{Q_w(w_t | x_t, a_{t-1}) \sum_{x_{t-1}} Q_x(x_t | x_{t-1}, a_{t-1}) \beta_{t-1}^{-i}(a_{t-1}^{-i} | a_{1:t-2}, w_{1:t-1}^{-i})\, {}^i\mu_{t-1}[h_{t-1}^i](x_{t-1}, w_{1:t-1}^{-i})}{\sum_{x_t, w_{1:t}^{-i}} Q_w(w_t | x_t, a_{t-1}) \sum_{x_{t-1}} Q_x(x_t | x_{t-1}, a_{t-1}) \beta_{t-1}^{-i}(a_{t-1}^{-i} | a_{1:t-2}, w_{1:t-1}^{-i})\, {}^i\mu_{t-1}[h_{t-1}^i](x_{t-1}, w_{1:t-1}^{-i})}; \quad (5)$$

   - if $\mathbb{P}^{\beta}(h_t^i | h_{t-1}^i) = 0$, then set ${}^i\mu_t[h_t^i](x_t, w_{1:t}^{-i})$ arbitrarily, with the only restriction being that ${}^i\mu_t[h_t^i](x_t, w_{1:t}^{-i}) = 0$ for those values of x_t, w_t^{-i} that are not in the support of $\mathbb{P}\tilde{\beta}(x_t, w_{1:t}^{-i} | h_{t-1}^i)$ for any strategy profile $\tilde{\beta}$.

In the following, we will define belief states $\xi_t^i \in \Delta(\mathcal{X}^i)$ and $\pi_t^i \in \Delta(\Delta(\mathcal{X}^i))$ that act as summaries of the histories h_t^i and h_t^c, respectively, and we will consider strategies that are defined on these belief states.

3. A Methodology for Characterizing the Structured PBE of the Game $\mathfrak{D}$

In this section, we provide a methodology to find the PBE of the game $\mathfrak{D}$ that consists of strategies whose domain is time invariant (although there may exist other equilibria that cannot be found using this methodology). Specifically, we seek equilibrium strategies that are structured in the sense that they depend on players' common and private observations through belief states. In order to achieve this, at any time t, we summarize player i's information set, $(a_{1:t-1}, w_{1:t}^i)$, in the belief $\xi_t^i \in \Delta(\mathcal{X}^i)$, and the common information, $a_{1:t-1}$, in the common belief $\pi_t \in \Delta(\times_{i\in\mathcal{N}} \Delta(\mathcal{X}^i))$, where ξ_t^i and π_t are defined as follows. For a strategy profile g, let $\xi_t^i(x_t^i) := \mathbb{P}^g(x_t^i | a_{1:t-1}, w_{1:t}^i)$ be the belief of player i on her current state conditioned on her information set. Similarly, we define $\pi_t(\xi_t) := \mathbb{P}^g(\xi_t | a_{1:t-1})$ as common joint belief on ξ_t based on the players' common information, $a_{1:t-1}$, and the corresponding marginals $\pi_t^i \in \Delta(\Delta(\mathcal{X}^i))$ as $\pi_t^i(\xi_t^i) := \mathbb{P}^g(\xi_t^i | a_{1:t-1})$. As will be shown later, because of 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 a 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 $\underline{\pi}_t \in \times_{i\in\mathcal{N}} \Delta(\Delta(\mathcal{X}^i))$ as the vector of marginal beliefs where $\underline{\pi}_t := (\pi_t^i)_{i\in\mathcal{N}}$. In addition, because of the above-mentioned conditional independence structure, it can be shown (in a similar way, conditional independence was shown in a simpler model in claim 1 of Vasal et al. (2019) that agent i's private information does not provide any new information on x_t^{-i} than what is already obtained from the common belief. Thus agent i only puts a belief on her own state x_t^i given her information set, and the belief on x_t^{-i} can be derived from π_t (i.e., $\mathbb{P}^g(x_t^i, x_t^{-i} | a_{t-1}, w_{1:t}^i) = \xi_t^i(x_t^i) \sum_{\xi_t^{-i}} \xi_t^{-i}(x_t^{-i}) \pi_t^{-i}(\xi_t^{-i})$). Finally, we point out that because of the assumption of rational kernels Q_x and Q_w, and because of the update equation of ξ_t^i (in Lemma 1), the private beliefs $\xi_t^i(x_t^i)$ are rational, and thus the beliefs π_t are well defined.

Inspired by the "common agent" approach in decentralized team problems (Nayyar et al. 2013), we now generate players' structured strategies as follows: player i at time t observes the common belief vector $\underline{\pi}_t$ and takes action γ_t^i, where $\gamma_t^i : \Delta(\mathcal{X}^i) \to \Delta(\mathcal{A}^i)$ is a partial (stochastic) function from her private belief ξ_t^i such that $A_t^i \sim \gamma_t^i(\cdot | \xi_t^i)$. These actions are generated through some policy $\theta^i = (\theta_t^i)_{t\in\mathcal{T}}$, $\theta_t^i : \times_{i\in\mathcal{N}} \Delta(\Delta(\mathcal{X}^i)) \to \{\Delta(\mathcal{X}^i) \to \Delta(\mathcal{A}^i)\}$ 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.

In the following, we will show a systematic approach to characterize equilibria of this game with structured strategies of the form $A_t^i \sim \theta_t^i[\underline{\pi}_t](\cdot | \xi_t^i)$. The justification for seeking such structured equilibria is threefold. First, through a "focusing argument," agents are convinced to play such structured strategies because there is a way to evaluate them (offline) instead of trying to find general equilibria with strategies of the form $A_t^i \sim g_t^i(\cdot | a_{1:t-1}, w_{1:t}^i)$ having ever-expanding domains. There is, however, a second, stronger justification for restricting attention to such equilibria. It can be shown (similar to section III of Vasal et al. (2019)) that these structured policies form a sufficiently rich set of policies. Specifically, it can be shown that policies g are outcome equivalent to policies θ—that is, 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. Finally, a third justification is that it can be shown (similar to section III of Vasal et al. (2019), and following the standard common agent methodology proposed by Nayyar et al. (2013)) that the optimal policies for the corresponding "dynamic team" problem have exactly this structure.

In the following lemma, we present update functions for the private beliefs ξ_t^i and the public beliefs π_t^i.

Lemma 1. *There exist update functions G^i, independent of players' strategies g, such that*

$$\xi^i_{t+1} = G^i(\xi^i_t, w^i_{t+1}, a_t) \tag{6}$$

and update functions F^i, independent of θ, such that

$$\pi^i_{t+1} = F^i(\pi^i_t, \gamma^i_t, a_t). \tag{7}$$

Thus $\underline{\pi}_{t+1} = \underline{F}(\underline{\pi}_t, \underline{\gamma}_t, a_t)$, where $\underline{F}$ is appropriately defined through (7).

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

We now define an SPBE as follows.

Definition 1 (SPBE). A structured perfect Bayesian equilibrium of the dynamic game $\mathfrak{D}$ is a PBE (β^*, μ^*) for which at any time t, for any agent i, her equilibrium strategy $\beta^{*,i}_t$ depends on player i's information $h^i_t = (a_{1:t-1}, w^i_{1:t})$ only through the beliefs ξ^i_t and $\underline{\pi}_t$. Furthermore, the belief ${}^i\mu^*_t$ depends on player i's information $h^i_t = (a_{1:t-1}, w^i_{1:t})$ only through the public belief $\underline{\pi}_t$. Finally, because of the aforementioned special structure, ${}^i\mu^*_t$ need only be a belief on $\xi^{-i}_t \in \times_{j\in\mathcal{N}\setminus\{i\}}\Delta(\mathcal{X}^j)$.

We remark that although we seek structured equilibria of the above-mentioned form, deviations are not restricted to be structured; that is, users are sequentially rational assuming every possible unilateral deviation. We now present the backward/forward algorithm to find the SPBE of the game $\mathfrak{D}$. The algorithm resembles the one presented in Vasal et al. (2019) for perfectly observable states.

3.1. Backward and Forward Recursions

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

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

$$V^i_{T+1}(\underline{\pi}_{T+1}, \xi^i_{T+1}) := 0. \tag{8}$$

2. For $t = T, T-1, \ldots, 1$, $\forall \underline{\pi}_t \in \times_{i\in\mathcal{N}}\Delta(\Delta(\mathcal{X}^i))$, 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, of the following fixed-point equation: $\forall i \in \mathcal{N}, \xi^i_t \in \Delta(\mathcal{X}^i)$,

$$\tilde{\gamma}^i_t(\cdot|\xi^i_t) \in \arg\max_{\gamma^i_t(\cdot|\xi^i_t)} \mathbb{E}^{\gamma^i_t(\cdot|\xi^i_t)\tilde{\gamma}^{-i}_t, \pi_t}\{R^i(X_t, A_t) + V^i_{t+1}(\underline{F}(\underline{\pi}_t, \tilde{\gamma}_t, A_t), \Xi^i_{t+1})|\xi^i_t\}, \tag{9}$$

where the expectation in (9) is with respect to random variables (X_t, A_t, Ξ^i_{t+1}) through the measure $\xi_t(x_t)\pi^{-i}_t(\xi^{-i}_t)\gamma^i_t(a^i_t|\xi^i_t)\tilde{\gamma}^{-i}_t(a^{-i}_t|\xi^{-i}_t)Q^i(\xi^i_{t+1}|\xi^i_t, a_t)$, $\underline{F}$ and Q^i are defined in Lemma 1, and Q^i is defined in appendix A of Vasal and Anastasopoulos (2016b). Furthermore, set

$$V^i_t(\underline{\pi}_t, \xi^i_t) := \mathbb{E}^{\tilde{\gamma}^i_t(\cdot|\xi^i_t)\tilde{\gamma}^{-i}_t, \pi_t}\{R^i(X_t, A_t) + V^i_{t+1}(\underline{F}(\underline{\pi}_t, \tilde{\gamma}_t, A_t), \Xi^i_{t+1})|\xi^i_t\}. \tag{10}$$

It should be noted that (9) 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, this is not to be confused with a best response type of a fixed-point equation as in a Bayesian Nash equilibrium. This distinct construction is pivotal in the proof of Theorem 1, and its roots can be traced back to the PBE construction in Vasal et al. (2019).

3.1.2. Forward Recursion. On the basis of θ defined in (8)–(10), 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 $\underline{\mu}^*_t[a_{1:t-1}] := (\mu^{*,i}_t[a_{1:t-1}])_{i\in\mathcal{N}}$ for the collection of marginal beliefs, and the joint belief $\mu^*_t[a_{1:t-1}]$ 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 $t = 0$,

$$\mu^{*,i}_0[\phi](\xi_0) := \delta_{Q^i_x}(\xi^i_0). \tag{11}$$

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

$$\beta^{*,i}_t(a^i_t|a_{1:t-1}, w^i_{1:t}) := \theta^i_t[\underline{\mu}^*_t[a_{1:t-1}]](a^i_t|\xi^i_t) \tag{12a}$$

$$\mu^{*,i}_{t+1}[a_{1:t}] := F^i(\mu^{*,i}_t[a_{1:t-1}], \theta^i_t[\underline{\mu}^*_t[a_{1:t-1}]], a_t). \tag{12b}$$

We conclude the construction by noting that the required beliefs ${}^i\mu^*_t : \mathcal{H}^i_t \to \times_{j\in\mathcal{N}\setminus\{i\}}\Delta(\mathcal{X}^j)$ can now be generated directly from $\underline{\mu}^*_t$ as ${}^i\mu^*_t[h^i_t](\xi^{-i}_t) = \prod_{j\neq i}\mu^{*,j}_t[a_{1:t-1}](\xi^j_t)$ with the understanding that under structured strategies, a belief on ξ^{-i}_t is part of the *information state or sufficient statistic* for user i to compute her future expected reward conditioned on the history h^i_t.

The main result of this section is summarized in the following theorem.

Theorem 1. *A strategy and belief profile* (β^*, μ^*), *constructed through the backward/forward recursive algorithm, is a PBE of the game.*

Proof. The proof relies on the specific fixed-point construction in (8)–(10) and the conditional independence structure of states and observations, and it is provided in Appendix B in Vasal and Anastasopoulos (2016b). □

3.2. Remarks

Several remarks are in order with regard to the preceding methodology and the result.

Remark 1. When players observe their types perfectly (i.e., when $\mathcal{W}^i = \mathcal{X}^i$ and $Q^i_w(w^i_t|x^i_t, a_{t-1}) = \delta_{x^i_t}(w^i_t)$, $\forall i, w^i_t, x^i_t, a_{t-1}$), then $\xi^i_t(\cdot) = \delta_{x^i_t}(\cdot)$, $\forall x^i_t$, and the results in this section reduce to the results in Vasal et al. (2019), as expected.

Remark 2. In the preceding special case with players perfectly observing their own types, it was shown in Vasal et al. (2019, theorem 2) that all SPBEs of the game can be found using this methodology. Using a similar argument, it can also be shown that all SPBEs of the game considered in this paper (i.e., with noisy types) can be found using this methodology.

Remark 3. The second subcase in the definition of F in appendix A in Vasal and Anastasopoulos (2016b) dictates how beliefs are updated for histories with zero probability. The particular expression used is only one of many possible updates than can be used here. Dynamics that govern the evolution of public beliefs at histories with zero probability of occurrence affect equilibrium strategies. Thus, the construction proposed for calculating SPBEs in this paper will produce a different set of equilibria if one changes the second subcase. The most well-known example of another such update is the *intuitive criterion* proposed in Cho and Kreps (1987) for Nash equilibria, later generalized to sequential equilibria in Cho (1987). The intuitive criterion assigns zero probability to states that can be excluded based on data available to all players (in our case, action profile history $a_{1:t-1}$). Another example of belief update is *universal divinity*, proposed in Banks and Sobel (1987).

Remark 4. To highlight the significance of the unique structure of (9), one can think as follows. When all players other than player i play structured strategies (i.e., strategies of the form $A^j_t \sim \theta^j_t[\underline{\pi}_t](\cdot|\xi^j_t)$), one may want to study the optimization problem from the viewpoint of the ith player in order to characterize her best response. In particular, one may want to show that although player i can play general strategies of the form $A^i_t \sim g^i_t(\cdot|w^i_{1:t}, a_{1:t-1})$, it is sufficient to best respond with structured strategies of the form $A^i_t \sim \theta^i_t[\underline{\pi}_t](\cdot|\xi^i_t)$ as well. To show that, one may entertain the thought that player i faces an MDP with state $(\Xi^i_t, \underline{\Pi}_t)$ and action A^i_t at time t. If that were true, then player i's optimal action could be characterized (using standard MDP results) by a dynamic programming equation similar to (9) of the form

$$\tilde{\gamma}^i_t(\cdot|\xi^i_t) \in \arg\max_{\gamma^i_t(\cdot|\xi^i_t)} \mathbb{E}^{\gamma^i_t(\cdot|\xi^i_t)\tilde{\gamma}^{-i}_t, \pi_t}\{R^i_t(X_t, A_t) + V^i_{t+1}(\underline{F}(\underline{\pi}_t, \gamma^i_t(\cdot|\xi^i_t), \tilde{\gamma}^i_t(\cdot|\cdot), \tilde{\gamma}^{-i}_t, A_t), \Xi^i_{t+1})|\xi^i_t\}, \tag{13}$$

where, unlike (9), in the belief update equation, the partial strategy $\gamma^i_t(\cdot|\xi^i_t)$ is also optimized over. However, as it turns out, user i does not face such an MDP problem: the reason is that the update equation $\underline{\pi}_{t+1} = \underline{F}(\underline{\pi}_t, \gamma_t, a_t)$ also depends on γ^i_t, which is the partial strategy of player i, and this has not been fixed in the aforementioned setting. If, however, the update equation is first fixed (so it is updated as $\underline{\pi}_{t+1} = \underline{F}(\underline{\pi}_t, \tilde{\gamma}^i_t, \tilde{\gamma}^{-i}_t, a_t) = \underline{F}(\underline{\pi}_t, \theta_t[\underline{\pi}_t], a_t)$—i.e., using the equilibrium strategies even for player i), then, indeed, the problem faced by user i is the MDP previously defined. It is now clear why (9) has the flavor of a fixed-point equation: the update of beliefs needs to be fixed beforehand with the equilibrium action $\tilde{\gamma}^i_t$ even for user i, and only then can user i's best response depend only on the MDP state $(\Xi^i_t, \underline{\Pi}_t)$ thus being a structured strategy as well. This implies that her optimal action $\tilde{\gamma}^i_t$ appears both on the left-hand and right-hand sides of this equation, giving rise to (9).

Remark 5. Nayyar et al. (2014) consider a general model of dynamic game of asymmetric information and provide a similar methodology to ours to solve for *nonsignaling* equilibria for that model. The corresponding equation in Nayyar et al. (2014) is Equation (9), which solves for a Bayesian Nash–like fixed-point equation in the space of γ_t functions, which occurs as a result of the assumption that the beliefs' updates are independent of the strategies, which leads to no signaling. As one would expect, their optimization problem is a special case of our development when there is no signaling (i.e., when the update of the common belief π_t does not depend on γ_t)—in other words, when $\pi_{t+1} = F(\pi_t, a_t)$, in which case (9) reduces to the same Bayesian Nash equation as in Nayyar et al. By contrast, in our work, (9) is a unique fixed-point equation not considered in the literature of games before, where the equilibrium strategy $\tilde{\gamma}^i_t(\cdot|x^i_t)$ in (9) appears both on the left-hand side and on the right-hand side, where the common belief is updated using the solution of the fixed-point equation.

Remark 6. In this paper, we find a class of PBEs of the game, although there may exist other equilibria that are not structured and cannot be found by directly using the proposed methodology. The rationale for using structured equilibria over others is the same as that for using Markov perfect equilibria over subgame perfect equilibria for a symmetric information game—a focusing argument (Osborne and Rubinstein 1994) for using simpler strategies being one of them.

Remark 7. Vasal (2019) extends the methodology of Vasal et al. (2019) to arbitrary correlated though *static* types, arguing that the same methodology could not be used for dynamic correlated types as a result of the lack of an information state. From the observation of Vasal (2019), in this remark, we argue (without proof) that the machinery developed in this paper can be used to extend the results to correlated, though *static*, types with conditionally independent observations.

4. Application: Informational Cascades

In the simple herding model introduced in the seminal papers of Banerjee (1992) and Bikhchandani et al. (1992), where selfish players sequentially acted once in the game, the authors introduce the notion of informational cascades as those histories where all future players' actions do not depend on their private information and they repeat the same action. In this section, we define a more general notion of informational cascades as those histories of the game where the dynamic game of asymmetric information collapses into a dynamic game of symmetric information, and the system dynamics from that point on depend only on the common information. We define two notions of information cascades, one based on common history and the other based on common belief, and in Lemma 2, we show the connection between the two definitions.

Definition 2. For a given strategy and belief profile (β^*, μ^*) that constitutes a PBE of the game, and for any time t and sequence of action profiles $a_{t:T}$, an informational cascade is defined as the set of public histories h^c_t of the game such that at h^c_t and under (β^*, μ^*), actions $a_{t:T}$ are played almost surely, irrespective of players' future private history realizations. More precisely,

$$\mathcal{C}^{a_{t:T}}_t := \{h^c_t \in \mathcal{H}^c_t \,|\, \forall i,\ \forall n \geq t,\ \forall h^i_n \text{ that are consistent with } h^c_t \text{ and occur with nonzero probability}, \beta^{*,i}_n(a^i_n|h^i_n) = 1\}. \tag{14}$$

We can also specialize the definition to a constant informational cascade if action profiles in the cascade are constant across time; that is, for time t and action profile a, constant cascades are defined by

$$\mathcal{C}^a_t := \mathcal{C}^{a_{t:T}}_t, \text{ where } a_n = a \text{ for } n = t, \ldots, T. \tag{15}$$

In this definition, we define cascades for a general model using action sets that may not be very useful in characterizing cascades using the SPBE methodology defined before. In the following, we provide an alternative definition that, because of its recursive nature, is well suited for characterizing information cascades associated with structured strategies.

Definition 3. For a given equilibrium generating function θ, and for any time t and sequence of action profiles $a_{t:T}$, an informational cascade for the game $\mathfrak{D}$ is defined recursively through the sets $\{\tilde{\mathcal{C}}^{a_{t:T}}_t\}_{t=1,\ldots,T+1}$ as follows. For $t = T, T-1, \ldots, 1$,

$$\tilde{\mathcal{C}}_{T+1} := \{\text{All possible common beliefs } \underline{\pi}_{T+1}\}, \tag{16a}$$

$$\tilde{\mathcal{C}}^{a_{t:T}}_t := \{\underline{\pi}_t \,|\, \forall i,\ \forall \xi^i_t \in \text{supp}(\pi^i_t), \theta^i_t[\underline{\pi}_t](a^i_t|\xi^i_t) = 1 \text{ and } \underline{F}(\underline{\pi}_t, \theta_t[\underline{\pi}_t], a_t) \in \tilde{\mathcal{C}}^{a_{t+1:T}}_{t+1}\}. \tag{16b}$$

Similar to the previous definition, a constant informational cascade for time t and action profile a is defined as

$$\tilde{\mathcal{C}}_t^a := \tilde{\mathcal{C}}_t^{a_{t:T}}, \text{ where } a_n = a \text{ for } n = t, \ldots, T. \tag{17}$$

This backward recursive definition characterizes informational cascades as those subsets of beliefs $\underline{\pi}_t$ that result in players taking certain predefined actions a_t almost surely regardless of their private information ξ_t. In addition, and because of the preceding, the belief updates F in appendix A in Vasal and Anastasopoulos (2016b) are simplified as

$$\pi_{t+1}^i(\xi_{t+1}^i) = \sum_{\substack{\xi_t^i, x_t^i,\\ x_{t+1}^i, w_{t+1}^i}} \pi_t^i(\xi_t^i)\xi_t^i(x_t^i)Q_x^i(x_{t+1}^i|x_t^i, a_t)Q_w^i(w_{t+1}^i|x_{t+1}^i, a_t)I_{G^i(\xi_t^i, w_{t+1}^i, a_t)}(\xi_{t+1}^i), \tag{18}$$

and these beliefs further result in players ignoring their private information ξ_{t+1}^i in forming their future actions. Thus we can correspondingly represent the update of π_t under nonsignaling policies as $\pi_{t+1} = F(\pi_t, a_t)$. In this situation, actions control the evolution of the state x_t and the spread of information, but the action of player i does not reveal any information about x_t^i. In other words, there is control but no signaling.

The following lemma establishes the connection between the preceding two definitions of cascades, one through the action sets and the other through the beliefs.

Lemma 2. *Let (β^*, μ^*) be an SPBE of the game $\mathfrak{D}$ generated by an equilibrium generating function θ through the backward/forward algorithm presented in Section 3. Then $\forall t, a_{t:T}$,*

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

Similarly, for a constant informational cascade, $\forall t, a$,

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

Proof. See Appendix D here or appendix E in Vasal and Anastasopoulos (2016b).

The preceding lemma makes precise the equivalence between the two definitions of informational cascades, Definitions 2 and 3, which are defined on two different objects—namely, the space of common histories and the space of common beliefs, respectively. Lemma 2 connects these two definitions such that if one finds a common history in a cascading set, so then using the lemma, one can find a corresponding common belief that is cascading, and vice versa, so long as such a belief corresponds to some common history.

4.1. Example with Nonadversarial Learning

We now consider a specific model that captures the learning aspect in a dynamic setting with strategic agents and decentralized information. The model is inspired by the model considered in Bikhchandani et al. (1992) and Smith and Sörensen (2000), where now we consider a finite number of players who take action in every time and participate during the entire duration of the game.

Assume there are N players, each having a constant static type that represents her talent, capabilities, or popularity, and the player makes a decision to either invest (action = 1) or not invest (action = 0) in the team as a whole, where her instantaneous reward depends on some combination of the capabilities of all the players (including herself). 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.

To simplify the exposition, we assume that players' states are uncontrollable and static; that is, $Q_x^i(x_{t+1}^i|x_t^i, a_t) = \delta_{x_t^i}(x_{t+1}^i)$, where $\mathcal{X}^i = \{-1, 1\}$ and $\mathbb{P}(X^i = -1) = \mathbb{P}(X^i = 1) = 1/2$.[4] Because the set of states, $\mathcal{X}^i$, has cardinality 2, the measure ξ_t^i can be sufficiently described by $\xi_t^i(1)$. Henceforth, in this section and in appendix F in Vasal and Anastasopoulos (2016b), with slight abuse of notation, we also denote $\xi_t^i(1)$ by $\xi_t^i \in [0, 1]$, and the reference is clear from the context. In each time t, player i makes an independent observation w_t^i about her 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}. 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 $p_1 \le p_0 < 1/2$. This model implies that taking action 1 can improve the quality of a player's future private belief. From her information, agent i takes action a_t^i, where $\mathcal{A}^i = \{0, 1\}$ and earns an instantaneous reward given by

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

where $\lambda \in [0,1], \bar{\lambda} = 1-\lambda$. To study asymptotic learning, we assume a discounted infinite horizon model with discount factor $\delta < 1$.

In this case, the update functions of ξ_t^i and π_t^i in (6) and (7) reduce to

$$\xi_{t+1}^i = G^i(\xi_t^i, w_{t+1}^i, a_t^i), \tag{22a}$$
$$\pi_{t+1}^i = F^i(\pi_t^i, \gamma_t^i, a_t^i). \tag{22b}$$

And an infinite horizon version of (9) in the backward recursion reduces to

$$\tilde{\gamma}_t^i(\cdot|\xi_t^i) \in \arg\max_{\gamma_t^i(\xi_t^i)} \gamma_t^i(1|\xi_t^i)(\lambda(2\xi_t^i - 1) + \bar{\lambda}(2\hat{\xi}_t^{-i} - 1)) + \delta \mathbb{E}^{\gamma_t^i(\xi_t^i)\tilde{\gamma}_t^{-i},\pi_t}\{V^i(\underline{F}(\underline{\pi}_t, \tilde{\gamma}_t, A_t), \Xi_{t+1}^i)\,|\,\xi_t^i\}, \tag{23}$$

$$V(\pi_t, \xi_t^i) = \tilde{\gamma}_t^i(1|\xi_t^i)(\lambda(2\xi_t^i - 1) + \bar{\lambda}(2\hat{\xi}_t^{-i} - 1)) + \delta \mathbb{E}^{\tilde{\gamma}_t^i(\xi_t^i)\tilde{\gamma}_t^{-i},\pi_t}\{V^i(\underline{F}(\underline{\pi}_t, \tilde{\gamma}_t, A_t), \Xi_{t+1}^i)|\xi_t^i\}, \tag{24}$$

where

$$\hat{\xi}_t^{-i} = \hat{\xi}^{-i}(\underline{\pi}^{-i}) := \frac{1}{N-1}\sum_{j\neq i} \mathbb{E}^{\pi_t^j}[\Xi_t^j] = \frac{1}{N-1}\sum_{j\neq i}\int \xi_t^j \pi_t^j(d\xi_t^j). \tag{25}$$

The intuition behind this equation should be clear. The instantaneous reward of player i is proportional to the probability of investing, $\gamma_t^i(1|\xi_t^i)$, as well as the perceived talent of the entire team formed by the combination of his perceived talent, $2\xi_t^i - 1$, and his perceived talent of the rest of the team, $2\hat{\xi}_t^{-i} - 1$. Furthermore, the estimate of user i on player's j talent is the same for all players i and is a result of their common belief π_t^j.

In the following, we show that for the specific learning model considered in this section, the players learn their true state asymptotically. We note that the result is true independent of the equilibrium (because the update of ξ_t^i does not depend on strategy θ).

Fact 1. The private beliefs Ξ_t^i converge to delta functions on the true value of the state, i.e.,

$$\Xi_t^i \xrightarrow[t\to\infty]{a.s.} \delta_{x^i}. \tag{26}$$

Proof. This is a classical Bayesian learning problem, and there are many techniques that can be used to prove Fact 1 (e.g., see Cover and Thomas (2012), pp. 314–316). □

This result is quite intuitive because players make independent observations about their respective states (independent of their strategies) and thus eventually learn their (static) state.

As it turns out, although players eventually learn their private states, almost surely, the system exhibits informational cascades. This surprising result implies that for a group of players who have sufficiently good talents, it may happen that they eventually do not invest in the team. This may happen because initially—because of a player's own atypical bad observations w_t^i—the player may take actions that may lead the game into a cascade of everybody not investing. Now, even though the player eventually learns perfectly her true type and realizes that her actions could have initially signaled the wrong information (which led the system into a wrong cascade), it is still rational for her to not invest and keep the same belief about others as before.

We now define a time-invariant set $\hat{\mathcal{C}}^a$ of common beliefs $\underline{\pi}$. This set for $a^i = 1$ includes those public beliefs for which player i believes that the other players have high enough types (on average) such that action $a^i = 1$ is taken irrespective of her private belief, ξ_t^i, on her own type, x^i, and similarly for $a^i = 0$. Let

$$\hat{\mathcal{C}}^a := \{\underline{\pi}\ |\forall i, \lambda + \bar{\lambda}(2\hat{\xi}^{-i} - 1) \le 0 \text{ if } a^i = 0, -\lambda + \bar{\lambda}(2\hat{\xi}^{-i} - 1) \ge 0 \text{ if } a^i = 1\}, \tag{27}$$

where $\hat{\xi}^{-i} = \hat{\xi}^{-i}(\underline{\pi}^{-i})$ is defined in (25). The intuition behind defining this set is clear if we compare with the instantaneous reward in (23). Regardless of how good or bad the estimate of the private state is, the estimate of other players' talent is so bad (or good) that player i does not (or does) invest. This is also implied by the fact that private beliefs form a sub- or super-martingale as shown in Appendix E and from the Martingale convergence theorem.

In the following theorem we show that the set $\hat{\mathcal{C}}^a$ defined in (27) characterizes a set of constant informational cascades for this problem. Specifically, we show that $\hat{\mathcal{C}}^a \subseteq \tilde{\mathcal{C}}_t^a$ for any t.

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

$$V^i(\underline{\pi}_t, \xi_t^i) = \frac{(\lambda(2\xi_t^i - 1) + \bar{\lambda}(2\hat{\xi}_t^{-i} - 1))a^i}{1-\delta}. \tag{28}$$

Proof. See Appendix F here and in Vasal and Anastasopoulos (2016b).

Several remarks are in order regarding this result.

Remark 8. In addition to proving that $\hat{\mathcal{C}}^a$ is a cascade, Theorem 2 provides an explicit expression for the reward-to-go of each player inside this cascade. Although it is, in general, difficult to solve the fixed-point equation (23), the special structure of players' actions and the special belief update inside a cascade makes this possible. Equation (28) implies that for those players who do not invest, their expected reward is 0, and for those who invest, their expected reward at the time t_0 they enter the cascade is $\frac{(\lambda(2\xi_t^i-1)+\bar{\lambda}(2\hat{\xi}_t^{-i}-1))}{1-\delta} \geq \frac{(-\lambda+\bar{\lambda}(2\hat{\xi}_t^{-i}-1))}{1-\delta} \geq 0$.

Remark 9. For the simplified problem considered in Bikhchandani et al. (1992), cascades can be characterized as the fixed points of the common belief update function so that the common belief gets "stuck" once it reaches that state. It was shown in Bikhchandani et al. that cascades eventually occur with probability 1 for that model. For the general model considered in this paper, common beliefs π_t still evolve in a cascade governed by the uninformative, nonsignaling update of the common belief π_t by $F(\pi_t, a_t)$, 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.

Remark 10. Conceptually, informational cascades can be thought of as absorbing states of the system. Indeed, given an equilibrium strategy profile, the common belief $(\underline{\Pi}_t)_{t\geq 1}$ is a Markov chain. It is thus natural to ask questions regarding the dynamics of the process that could lead to those states—for example, hitting times of such sets and absorption probabilities. We remark that this is a rather difficult task because it involves finding the equilibrium strategies—that is, solving the fixed-point equation (23) for all values of $\underline{\pi}_t$ and not only for those values of $\underline{\pi}_t$ inside the cascade as done in Theorem 2. One trivial case when cascades could occur for this model is if the system was born in a cascade; that is, the initial common belief, based on the prior distributions, is $\pi_1 \in \hat{C}^a$. More generally, a cascade could occur as in the following case. Suppose all players have low states (i.e., $x^i = -1$ for all $i \in \mathcal{N}$), 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 the players eventually learn their true states, they remain in a (bad) cascade, with each player believing that others have high states on average.

Theorem 2 gives a characterization of informational cascades. However, it is not clear if players always reach any of the informational cascades. In the following theorem, we show that in an infinite horizon model, this is true almost surely.

Theorem 3. *Under an infinite time horizon, the system eventually falls into one of the cascades.*

Proof. We prove this by contradiction. Suppose there exists a nonempty set of trajectories Ω such that the system never falls into any of the cascading sets $\hat{\mathcal{C}}^a$. Consider the process π_t conditioned on the set Ω and the true state x. Then, by definition of information cascades, there exists $\exists i$ such that π_t^i is evolving on this set such that the corresponding sequence of γ_t^i is informative.

Then for all such i, π_t^i converges to x^i on this set using the same arguments as in Fact 1 (i.e., $\Pi_t^i|\Omega \xrightarrow[t\to\infty]{a.s.} \delta(\delta_{x^i})$). Thus the system eventually falls into a cascade (note that out of all the cascading sets, there exists a cascading set that is optimal with respect to the players' type such that players would play that if everyone's types are known).

This, however, contradicts the fact that Ω is nonempty. □

Remark 11. In this paper, we considered the model where users observe the entire action history of the players. However, there exist many scenarios where users may observe a coarser version of the action history, as has been studied in Çelen and Kariv (2004), Guarino et al. (2011), Herrera and Hörner (2013), and Guarino and Jehiel (2013). We note that our model does not incorporate the flexibility of using such information structures, and we pose it as an open problem for future research if our methodology of finding an SPBE can be extended to different information structures as studied in the above-mentioned works.

5. Conclusion

In this paper we study Bayesian learning dynamics of a specific class of dynamic games with asymmetric information. In the literature, a simplifying model is considered where herding behavior by selfish players is shown in a sequential buyers' game where a countable number of strategic buyers buy a product exactly once in the game. In this paper, we consider a more general scenario where players could participate in the game throughout the duration of the game. Players' states evolve as conditionally independent controlled Markov processes, and players made noisy observations of their states. We first present a sequential decomposition methodology to find the SPBE of the game. We then study a specific learning model with forward-looking agents and characterize information cascades using the general methodology described before. We show a corresponding result as that in Banerjee (1992) and Bikhchandani et al. (1992) although for forward-looking agents, that for infinite horizon all players fall into one of the cascades almost surely. In general, the methodology presented in the paper serves as a framework for studying learning dynamics of decentralized systems with strategic agents. Although the presented model considers all N players acting in all periods of the game, it can be readily extended to 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 her state process $(X_t^0)_t$, has reward function 0, and plays actions $a_t^0 = w_t^0 = x_t^0$. For instance, let $\mathcal{X}^0 = \mathcal{A}^0 = \mathcal{N}$. Once the quantity $a_{t-1}^0 = w_{t-1}^0 = x_{t-1}^0$ is publicly observed, all players can determine that the acting player (at time t) will be the one indicated by a_{t-1}^0.

Some important future 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 information design to avoid bad cascades.

Appendix A. Proof of Lemma 1

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

Lemma A.1. *For any policy profile g and $\forall t$,*

$$\mathbb{P}^g(x_{1:t}, w_{1:t} | a_{1:t-1}) = \prod_{i=1}^N \mathbb{P}^{g^i}(x_{1:t}^i, w_{1:t}^i | a_{1:t-1}). \tag{A.1}$$

Proof. Note that

$$\mathbb{P}^g(x_{1:t}, w_{1:t} | a_{1:t-1}) = \frac{\mathbb{P}^g(x_{1:t}, w_{1:t}, a_{1:t-1})}{\sum_{x_{1:t}, w_{1:t}} \mathbb{P}^g(x_{1:t}, w_{1:t}, a_{1:t-1})} \tag{A.2a}$$

$$= \frac{\prod_{i=1}^N Q_x^i(x_1^i) Q_w^i(w_1^i | x_1^i) \prod_{n=1}^{t-1} g_n^i(a_n^i | a_{1:n-1}, w_{1:n-1}^i) Q_x^i(x_{n+1}^i | a_n, x_n^i) Q_w^i(w_{n+1}^i | x_{n+1}^i, a_n)}{\sum_{x_{1:t}, w_{1:t}} \prod_{i=1}^N Q_x^i(x_1^i) Q_w^i(w_1^i | x_1^i) \prod_{n=1}^{t-1} g_n^i(a_n^i | a_{1:n-1}, w_{1:n-1}^i) Q_x^i(x_{n+1}^i | a_n, x_n^i) Q_w^i(w_{n+1}^i | x_{n+1}^i, a_n)} \tag{A.2b}$$

$$= \frac{\prod_{i=1}^N Q_x^i(x_1^i) Q_w^i(w_1^i | x_1^i) \prod_{n=1}^{t-1} g_n^i(a_n^i | a_{1:n-1}, w_{1:n-1}^i) Q_x^i(x_{n+1}^i | a_n, x_n^i) Q_w^i(w_{n+1}^i | x_{n+1}^i, a_n)}{\prod_{i=1}^N \sum_{x_{1:t}^i, w_{1:t}^i} Q_x^i(x_1^i) Q_w^i(w_1^i | x_1^i) \prod_{n=1}^{t-1} g_n^i(a_n^i | a_{1:n-1}, w_{1:n-1}^i) Q_x^i(x_{n+1}^i | a_n, x_n^i) Q_w^i(w_{n+1}^i | x_{n+1}^i, a_n)}, \tag{A.2c}$$

and thus

$$\mathbb{P}^g(x_{1:t}, w_{1:t} | a_{1:t-1}) = \prod_{i=1}^N \mathbb{P}^{g^i}(x_{1:t}^i, w_{1:t}^i | a_{1:t-1}). \tag{A.2d}$$

Now for any g we have

$$\xi_{t+1}^i(x_{t+1}^i) \triangleq \mathbb{P}^g(x_{t+1}^i | a_{1:t}, w_{1:t+1}^i) \tag{A.3a}$$

$$= \frac{\sum_{x_t^i} \mathbb{P}^g(x_t^i, a_t, x_{t+1}^i, w_{t+1}^i | a_{1:t-1}, w_{1:t}^i)}{\sum_{\tilde{x}_{t+1}^i \tilde{x}_t^i} \mathbb{P}^g(\tilde{x}_t^i, a_t, w_{t+1}^i, \tilde{x}_{t+1}^i | a_{1:t-1}, w_{1:t}^i)} \tag{A.3b}$$

$$= \frac{\sum_{x_t^i} \xi_t^i(x_t^i) \mathbb{P}^g(a_t^{-i} | a_{1:t-1}, w_{1:t}^i, x_t^i) Q_x^i(x_{t+1}^i | a_t, x_t^i) Q_w^i(w_{t+1}^i | x_{t+1}^i, a_t)}{\sum_{\tilde{x}_{t+1}^i, \tilde{x}_t^i} \xi_t^i(\tilde{x}_t^i) \mathbb{P}^g(a_t^{-i} | a_{1:t-1}, w_{1:t}^i, \tilde{x}_t^i) Q_x^i(\tilde{x}_{t+1}^i | a_t, \tilde{x}_t^i) Q_w^i(w_{t+1}^i | \tilde{x}_{t+1}^i, a_t)}, \tag{A.3c}$$

where (A.3c) is true because a_t^i is a function of $(a_{1:t-1}, w_{1:t}^i)$, and thus the term involving a_t^i can be cancelled in the numerator and denominator. We now consider the quantity $\mathbb{P}^g(a_t^{-i} | a_{1:t-1}, w_{1:t}^i, x_t^i)$:

$$\mathbb{P}^g(a_t^{-i} | a_{1:t-1}, w_{1:t}^i, x_t^i) = \sum_{w_{1:t}^{-i}} \mathbb{P}^g(a_t^{-i}, w_{1:t}^{-i} | a_{1:t-1}, w_{1:t}^i, x_t^i) \tag{A.4a}$$

$$= \sum_{w_{1:t}^{-i}} \mathbb{P}^g(w_{1:t}^{-i} | a_{1:t-1}, w_{1:t}^i, x_t^i) \prod_{j \neq i} g_t^j(a_t^j | a_{1:t-1}, w_{1:t}^j) \tag{A.4b}$$

$$= \sum_{w_{1:t}^{-i}} \mathbb{P}^{g^{-i}}(w_{1:t}^{-i} | a_{1:t-1}) \prod_{j \neq i} g_t^j(a_t^j | a_{1:t-1}, w_{1:t}^j) \tag{A.4c}$$

$$= \mathbb{P}^{g^{-i}}(a_t^{-i} | a_{1:t-1}), \tag{A.4d}$$

where (A.4c) follows from lemma 1 in Vasal and Anastasopoulos (2016b) because $w_{1:t}^{-i}$ is conditionally independent of $(w_{1:t}^i, x_t^i)$ given $a_{1:t-1}$ and is only a function of g^{-i}. Because this term does not depend on x_t^i, it gets cancelled in the final expression of ξ^i_{t+1}:

$$\xi^i_{t+1}(x^i_{t+1}) = \frac{\sum_{x^i_t} \xi^i_t(x^i_t) Q^i_x(x^i_{t+1}|x^i_t, a_t) Q^i_w(w^i_{t+1}|x^i_{t+1}, a_t)}{\sum_{\tilde{x}^i_{t+1}} \sum_{x^i_t} \xi^i_t(x^i_t) Q^i_x(\tilde{x}^i_{t+1}|x^i_t, a_t) Q^i_w(w^i_{t+1}|\tilde{x}^i_{t+1}, a_t)}. \tag{A.5}$$

Note that because of the full support of kernels Q_x and Q_w, the preceding recursion is always well defined. In addition, because of the kernels being rationals, the beliefs $\xi^i_t(x_t)$ will always be rationals as well. Thus the claim of the lemma follows. On the basis of this claim, we can conclude that

$$\xi^i_t(x^i_t) = \mathbb{P}^g(x^i_t|a_{1:t-1}, w^i_{1:t}) = \mathbb{P}(x^i_t|a_{1:t-1}, w^i_{1:t}). \tag{A.6}$$

Also, from the update of ξ^i_t in (6), we define an update kernel,

$$Q^i(\xi^i_{t+1}|\xi^i_t, a_t) := \mathbb{P}(\xi^i_{t+1}|\xi^i_t, a_t) \tag{A.7}$$

$$= \sum_{x^i_t, x^i_{t+1}, w^i_{t+1}} \xi^i_t(x^i_t) Q^i_x(x^i_{t+1}|x^i_t, a_t) Q^i_w(w^i_{t+1}|x^i_{t+1}, a_t) I_{G^i(\xi^i_t, w^i_{t+1}, a^i_t)}(\xi^i_{t+1}). \tag{A.8}$$

We now turn to the second part of the lemma regarding the update of the common belief π_t. We have

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

$$= \mathbb{P}^\psi(\xi_{t+1}|a_{1:t}, \gamma_{1:t+1}) \tag{A.9a}$$

$$= \mathbb{P}^\psi(\xi_{t+1}|a_{1:t}, \gamma_{1:t}) \tag{A.9b}$$

$$= \frac{\sum_{\substack{\xi_t, x_t, \\ x_{t+1}, w_{t+1}}} \mathbb{P}^\psi(\xi_t, x_t, a_t, x_{t+1}, w_{t+1}, \xi_{t+1}|a_{1:t-1}, \gamma_{1:t})}{\sum_{\xi_t} \mathbb{P}^\psi(\xi_t, a_t|a_{1:t-1}, \gamma_{1:t})} \tag{A.9c}$$

$$= \frac{\sum_{\substack{\xi_t, x_t, \\ x_{t+1}, w_{t+1}}} \prod_{i=1}^N \pi^i_t(\xi^i_t)\xi^i_t(x^i_t)\gamma^i_t(a^i_t|\xi^i_t) Q^i_x(x^i_{t+1}|x^i_t, a_t) Q^i_w(w^i_{t+1}|x^i_{t+1}, a_t) I_{G^i(\xi^i_t, w^i_{t+1}, a_t)}(\xi^i_{t+1})}{\sum_{\xi_t} \prod_{i=1}^N \pi^i_t(\xi^i_t)\gamma^i_t(a^i_t|\xi^i_t)} \tag{A.9d}$$

$$= \frac{\prod_{i=1}^N \sum_{\substack{\xi^i_t, x^i_t, \\ x^i_{t+1}, w^i_{t+1}}} \pi^i_t(\xi^i_t)\xi^i_t(x^i_t)\gamma^i_t(a^i_t|\xi^i_t) Q^i_x(x^i_{t+1}|x^i_t, a_t) Q^i_w(w^i_{t+1}|x^i_{t+1}, a_t) I_{G^i(\xi^i_t, w^i_{t+1}, a_t)}(\xi^i_{t+1})}{\sum_{\xi_t} \prod_{i=1}^N \pi^i_t(\xi^i_t)\gamma^i_t(a^i_t|\xi^i_t)} \tag{A.9e}$$

$$= \frac{\prod_{i=1}^N \sum_{\substack{\xi^i_t, x^i_t, \\ x^i_{t+1}, w^i_{t+1}}} \pi^i_t(\xi^i_t)\xi^i_t(x^i_t)\gamma^i_t(a^i_t|\xi^i_t) Q^i_x(x^i_{t+1}|x^i_t, a_t) Q^i_w(w^i_{t+1}|x^i_{t+1}, a_t) I_{G^i(\xi^i_t, w^i_{t+1}, a_t)}(\xi^i_{t+1})}{\prod_{i=1}^N \sum_{\xi^i_t} \pi^i_t(\xi^i_t)\gamma^i_t(a^i_t|\xi^i_t)}. \tag{A.9f}$$

When the denominator in the aforementioned equation is 0, we define

$$\pi_{t+1}(\xi_{t+1}) = \prod_{i=1}^N \sum_{\substack{\xi^i_t, x^i_t, \\ x^i_{t+1}, w^i_{t+1}}} \pi^i_t(\xi^i_t)\xi^i_t(x^i_t) Q^i_x(x^i_{t+1}|x^i_t, a_t) Q^i_w(w^i_{t+1}|x^i_{t+1}, a_t) I_{G^i(\xi^i_t, w^i_{t+1}, a_t)}(\xi^i_{t+1}). \tag{A.9g}$$

We remark here that the necessity of defining a belief update for the case when the denominator is 0 stems from the requirement of defining appropriate beliefs on the off-equilibrium paths in a PBE. Indeed, when an off-equilibrium action a_t is observed, the belief update is defined as in (A.9g). Furthermore, because of the presence of the indicator function $I_{G^i(\xi^i_t, w^i_{t+1}, a_t)}(\xi^i_{t+1})$ in the update, the support of π_{t+1} is not consistent with the prior belief π_t and the observations w^i_{t+1} and a_t, thus guaranteeing belief consistency as required in the definition of the PBE. Thus we have

$$\pi_{t+1} = \prod_{i=1}^N F^i(\pi^i_t, \gamma^i_t, a_t). \tag{A.9h}$$

Appendix B. Proof of Theorem 1

We prove sequential rationality using induction and from the results of Lemmas C.1–C.3 proved in Appendix C. For the base case at $t = T$, $\forall i \in \mathcal{N}, (a_{1:T-1}, w^i_{1:T}) \in \mathcal{H}^i_T, \beta^i$:

$$\mathbb{E}^{\beta^{*,i}_T \beta^{*,-i}_T, \mu^*_T[a_{1:T-1}]}\{R^i(X_T, A_T)|a_{1:T-1}, w^i_{1:T}\}$$

$$= V^i_T(\underline{\mu}^*_T[a_{1:T-1}], \xi^i_T) \tag{B.1a}$$

$$\geq \mathbb{E}^{\beta^i_T \beta^{*,-i}_T, \mu^*_T[a_{1:T-1}]}\{R^i(X_T, A_T)|\, a_{1:T-1}, w^i_{1:T}\}, \tag{B.1b}$$

where (B.1a) follows from Lemma C.3 and (B.1b) from Lemma C.1 in Appendix C.

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

$$\mathbb{E}^{\beta^{*,i}_{t+1:T}\beta^{*,-i}_{t+1:T},\mu^*_{t+1}[a_{1:t}]}\left\{\sum_{n=t+1}^{T} R^i(X_n, A_n) | a_{1:t}, w^i_{1:t+1}\right\} \geq$$
$$\mathbb{E}^{\beta^{i}_{t+1:T}\beta^{*,-i}_{t+1:T},\mu^*_{t+1}[a_{1:t}]}\left\{\sum_{n=t+1}^{T} R^i(X_n, A_n) | a_{1:t}, w^i_{1:t+1}\right\}. \tag{B.2}$$

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

$$\mathbb{E}^{\beta^{*,i}_{t:T}\beta^{*,-i}_{t:T},\mu^*_t[a_{1:t-1}]}\left\{\sum_{n=t}^{T} R^i(X_n, A_n) | a_{1:t-1}, w^i_{1:t}\right\} \tag{B.3a}$$
$$= V^i_t(\underline{\mu}^*_t[a_{1:t-1}], \xi^i_t)$$
$$\geq \mathbb{E}^{\beta^i_t\beta^{*,-i}_t,\mu^*_t[a_{1:t-1}]}\{R^i(X_t, A_t) + V^i_{t+1}(\underline{\mu}^*_{t+1}[a_{1:t-1}A_t], \Xi^i_{t+1}) | a_{1:t-1}, w^i_{1:t}\} \tag{B.3b}$$
$$= \mathbb{E}^{\beta^i_t\beta^{*,-i}_t,\mu^*_t[a_{1:t-1}]}\{R^i(X_t, A_t) + \mathbb{E}^{\beta^{*,i}_{t+1:T}\beta^{*,-i}_{t+1:T},\mu^*_{t+1}[a_{1:t-1},A_t]}\left\{\sum_{n=t+1}^{T} R^i(X_n, A_n) | a_{1:t-1}, A_t, w^i_{1:t}W^i_{t+1}\right\} | a_{1:t-1}, w^i_{1:t}\right\} \tag{B.3c}$$
$$\geq \mathbb{E}^{\beta^i_t\beta^{*,-i}_t,\mu^*_t[a_{1:t-1}]}\{R^i(X_t, A_t) + \mathbb{E}^{\beta^{i}_{t+1:T}\beta^{*,-i}_{t+1:T}\mu^*_{t+1}[a_{1:t-1},A_t]}\left\{\sum_{n=t+1}^{T} R^i(X_n, A_n) \; | a_{1:t-1}, A_t, w^i_{1:t}, W^i_{t+1}\right\} | a_{1:t-1}, w^i_{1:t}\right\} \tag{B.3d}$$
$$= \mathbb{E}^{\beta^i_t\beta^{*,-i}_t,\mu^*_t[a_{1:t-1}]}\{R^i(X_t, A_t) + \mathbb{E}^{\beta^{i}_{t:T}\beta^{*,-i}_{t:T}\mu^*_{t}[a_{1:t-1}]}\left\{\sum_{n=t+1}^{T} R^i(X_n, A_n) \; | a_{1:t-1}, A_t, w^i_{1:t}, W^i_{t+1}\right\} | a_{1:t-1}, w^i_{1:t}\right\} \tag{B.3e}$$
$$= \mathbb{E}^{\beta^{i}_{t:T}\beta^{*,-i}_{t:T},\mu^*_{t}[a_{1:t-1}]}\left\{\sum_{n=t}^{T} R^i(X_n, A_n) \; | a_{1:t-1}, w^i_{1:t}\right\}, \tag{B.3f}$$

where (B.3a) follows from Lemma C.3, (B.3b) follows from Lemma C.1, (B.3c) follows from Lemma C.3, (B.3d) follows from induction hypothesis in (B.2), and (B.3e) follows from Lemma C.2. Moreover, the construction of θ in (9)—and consequently, the definition of β^* in (12a)—are pivotal for (B.3e) to follow from (B.3d).

We conclude by noting that the constructed μ^* satisfies the belief consistency conditions specified in the PBE definition by the very construction in (A.9g) (see the comment in lemma 1 in Vasal and Anastasopoulos (2016b) and after (A.9g)).

Appendix C. Lemmas

Lemma C.1. *For all* $\forall t \in \mathcal{T}, i \in \mathcal{N}, (a_{1:t-1}, w^i_{1:t}) \in \mathcal{H}^i_t, \beta^i_t$,

$$V^i_t(\underline{\mu}^*_t[a_{1:t-1}], \xi^i_t) \geq \mathbb{E}^{\beta^i_t\beta^{*,-i}_t,\mu^*_t[a_{1:t-1}]}\{R^i(X_t, A_t) + \tag{C.1}$$
$$V^i_{t+1}(\underline{F}(\underline{\mu}^*_t[a_{1:t-1}], \beta^*_t(\cdot|a_{1:t-1},\cdot), A_t), \Xi^i_{t+1}) \; | a_{1:t-1}, w^i_{1:t}\}. \tag{C.2}$$

Proof. We prove this lemma by contradiction.

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

$$\mathbb{E}^{\hat{\beta}^i_t\beta^{*,-i}_t,\mu^*_t[\hat{a}_{1:t-1}]}\left\{R^i(X_t, A_t) + V^i_{t+1}(\underline{F}(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \beta^*_t(\cdot|\hat{a}_{1:t-1},\cdot), A_t), \Xi^i_{t+1}) | \hat{a}_{1:t-1}, \hat{w}^i_{1:t}\right\} > V^i_t(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \hat{\xi}^i_t). \tag{C.3}$$

We will show that this contradicts the definition of V^i_t in (10).

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

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

$$V^i_t(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \hat{\xi}^i_t) = \max_{\gamma^i_t(\hat{\xi}^i_t)} \mathbb{E}^{\gamma^i_t(\hat{\xi}^i_t)\beta^{*,-i}_t,\mu^*_t[\hat{a}_{1:t-1}]}\left\{R^i(X_t, A_t) + V^i_{t+1}(\underline{F}(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \beta^*_t(\cdot|\hat{a}_{1:t-1},\cdot), A_t), \Xi^i_{t+1}) \; | \hat{\xi}^i_t\right\} \tag{C.4a}$$
$$\geq \mathbb{E}^{\hat{\gamma}^i_t(\hat{\xi}^i_t)\beta^{*,-i}_t,\mu^*_t[\hat{a}_{1:t-1}]}\left\{R^i(X_t, A_t) + V^i_{t+1}(\underline{F}(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \beta^*_t(\cdot|\hat{a}_{1:t-1},\cdot), A_t), \Xi^i_{t+1}) \; | \hat{\xi}^i_t\right\} \tag{C.4b}$$
$$= \sum_{x_t, \xi^{-i}_t, a_t, \xi_{t+1}} \left\{R^i(x_t, a_t) + V^i_{t+1}(\underline{F}(\underline{\mu}^*_t[\hat{a}_{1:t-1}], \beta^*_t(\cdot|\hat{a}_{1:t-1},\cdot), a_t), \xi^i_{t+1})\right\} \times$$
$$\hat{\xi}^i_t(x^i_t)\xi^{-i}_t(x^{-i}_t)\mu^{*,-i}_t[\hat{a}_{1:t-1}](\xi^{-i}_t)\hat{\gamma}^i_t(a^i_t|\hat{\xi}^i_t)\beta^{*,-i}_t(a^{-i}_t|\hat{a}_{1:t-1}, \xi^{-i}_t)Q^i(\xi^i_{t+1}|\hat{\xi}^i_t, a_t) \tag{C.4c}$$

$$= \sum_{\substack{x_t, \xi_t^{-i}, \\ a_t, \xi_{t+1}}} \{R^i(x_t, a_t) + V_{t+1}^i(\underline{F}(\underline{\mu}_t^*[\hat{a}_{1:t-1}], \beta_t^*(\cdot|\hat{a}_{1:t-1}, \cdot), a_t), \xi_{t+1}^i)\} \times \hat{\xi}_t^i(x_t^i)\xi_t^{-i}(x_t^{-i})\mu_t^{*,-i}[\hat{a}_{1:t-1}](\xi_t^{-i})\hat{\beta}_t^i(a_t^i|\hat{a}_{1:t-1}, \hat{w}_{1:t}^i)\beta_t^{*,-i}(a_t^{-i}|\hat{a}_{1:t-1}, \xi_t^{-i})Q^i(\xi_{t+1}^i|\hat{\xi}_t^i, a_t) \tag{C.4d}$$

$$= \mathbb{E}^{\hat{\beta}_t^i \beta_t^{*,-i}, \mu_t^*[\hat{a}_{1:t-1}]}\{R^i(X_t, A_t) + V_{t+1}^i(\underline{F}(\underline{\mu}_t^*[\hat{a}_{1:t-1}], \beta_t^*(\cdot|\hat{a}_{1:t-1}, \cdot), A_t), X_{t+1}^i)|\hat{a}_{1:t-1}, \hat{w}_{1:t}^i\} \tag{C.4e}$$

$$> V_t^i(\underline{\mu}_t^*[\hat{a}_{1:t-1}], \hat{\xi}_t^i), \tag{C.4f}$$

where (C.4a) follows from the definition of V_t^i in (10), (C.4d) follows from definition of $\hat{\gamma}_t^i$, and (C.4f) follows from (C.3). However, this leads to a contradiction. □

Lemma C.2. *For all* $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t}, w_{1:t+1}^i) \in \mathcal{H}_{t+1}^i$, *and* β_t^i,

$$\mathbb{E}^{\beta_{t:T}^i \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]}\left\{\sum_{n=t+1}^{T} R^i(X_n, A_n)|a_{1:t}, w_{1:t+1}^i\right\} = \mathbb{E}^{\beta_{t+1:T}^i \beta_{t+1:T}^{*,-i}, \mu_{t+1}^*[a_{1:t}]}\left\{\sum_{n=t+1}^{T} R^i(X_n, A_n)|a_{1:t}, w_{1:t+1}^i\right\}. \tag{C.5}$$

Thus these quantities do not depend on β_t^i.

Proof. Essentially this claim stands on the fact that $\mu_{t+1}^{*,-i}[a_{1:t}]$ can be updated from $\mu_t^{*,-i}[a_{1:t-1}], \beta_t^{*,-i}$ and a_t, as $\mu_{t+1}^{*,-i}[a_{1:t}] = \prod_{j \neq i} F^{-i}(\mu_t^{*,-i}[a_{1:t-1}], \beta_t^{*,-i}, a_t)$, as in Lemma 1. Because the above-mentioned expectations involve random variables $X_{t+1:T}, A_{t+1:T}$, we consider $\mathbb{P}^{\beta_{t:T}^i \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]}(x_{t+1:T}, a_{t+1:T}|a_{1:t}, w_{1:t+1}^i)$:

$$\mathbb{P}^{\beta_{t:T}^i \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]}(x_{t+1:T}, a_{t+1:T}|a_{1:t}, w_{1:t+1}^i) = \frac{\mathbb{P}^{\beta_{t:T}^i \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]}(a_t, x_{t+1}, w_{t+1}^i|a_{1:t-1}, w_{1:t}^i)}{\mathbb{P}^{\beta_{t:T}^i \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]}(a_t, w_{t+1}^i|a_{1:t-1}, w_{1:t}^i)} \mathbb{P}^{\beta_{t:T}^i \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]}(a_{t+1:T}, x_{t+2:T}|a_{1:t-1}, w_{1:t}^i, x_{t+1}) \tag{C.6a}$$

We note that

$$\mathbb{P}^{\beta_{t:T}^i \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]}(a_{t+1:T}, x_{t+2:T}|a_{1:t-1}, w_{1:t}^i, x_{t+1}) = \beta_{t+1}^i(a_{t+1}^i|a_{1:t-1}, w_{1:t}^i)\beta_{t+1}^{-i}(a_{t+1}^{-i}|a_{1:t-1}, w_{1:t}^i)\xi_{t+1}^i(x_{t+1}^i)\sum_{\xi_{t+1}^{-i}} \mu_{t+1}^{*,-i}[a_{1:t}](\xi_{t+1}^{-i})\xi_{t+1}^{-i}(x_{t+1}^{-i}) \tag{C.6b}$$

$$\mathbb{P}^{\beta_{t+1:T}^i \beta_{t+1:T}^{*,-i}, \mu_{t+1}^*[a_{1:t}]}(a_{t+2:T}, x_{t+3:T}|a_{1:t-1}, w_{1:t}^i, x_{t+3}), = \mathbb{P}^{\beta_{t+1:T}^i \beta_{t+1:T}^{*,-i}, \mu_{t+1}^*[a_{1:t}]}(a_{t+1:T}, x_{t+2:T}|a_{1:t}, w_{1:t}^i, x_{t+1}). \tag{C.6c}$$

We consider the numerator and the denominator on the left-hand side of the preceding equation separately. The numerator in (C.6a) is given by

$$Nr = \sum_{x_t, \xi_t^{-i}} \mathbb{P}^{\beta_{t:T}^i \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]}(x_t, \xi_t^{-i}|a_{1:t-1}, w_{1:t}^i)\beta_t^i(a_t^i|a_{1:t-1}, w_{1:t}^i)\beta_t^{*,-i}(a_t^{-i}|a_{1:t-1}, \xi_t^{-i})Q_x(x_{t+1}|x_t, a_t) \cdot Q_w^i(w_{t+1}^i|x_{t+1}^i, a_t) \tag{C.6d}$$

$$= \left(\sum_{x_t^i} \xi_t^i(x_t^i)Q_x^i(x_{t+1}^i|x_t, a_t)Q_w^i(w_{t+1}^i|x_{t+1}^i, a_t)\right) \cdot \left(\sum_{x_t^{-i}, \xi_t^{-i}} \xi_t^{-i}(x_t^{-i})\mu_t^{*,-i}[a_{1:t-1}](\xi_t^{-i})\beta_t^i(a_t^i|a_{1:t-1}, w_{1:t}^i)\beta_t^{*,-i}(a_t^{-i}|a_{1:t-1}, \xi_t^{-i})Q_x^{-i}(x_{t+1}^{-i}|x_t, a_t)\right), \tag{C.6e}$$

where (C.6e) follows from the fact that the probability on $(a_{t+1:T}, x_{2+t:T})$, given $a_{1:t}, w_{1:t+1}^i, x_{t:t+1}, \mu_t^*[a_{1:t-1}]$, depends on $a_{1:t}, w_{1:t+1}^i, x_{t+1}, \mu_{t+1}^*[a_{1:t}]$ through $\beta_{t+1:T}^i \beta_{t+1:T}^{*,-i}$. Similarly, the denominator in (C.6a) is given by

$$Dr = \sum_{\tilde{x}_t, \tilde{\xi}_t^{-i}, \tilde{x}_{t+1}^i} \mathbb{P}^{\beta_{t:T}^i \beta_{t:T}^{*,-i}, \mu_t^*}(\tilde{x}_t, \xi_t^{-i}|a_{1:t-1}, w_{1:t}^i)\beta_t^i(a_t^i|a_{1:t-1}, w_{1:t}^i)\beta_t^{*,-i}(a_t^{-i}|a_{1:t-1}, \tilde{\xi}_t^{-i})Q_x^i(\tilde{x}_{t+1}^i|\tilde{x}_t^i, a_t) \cdot Q_w^i(w_{t+1}^i|\tilde{x}_{t+1}^i, a_t) \tag{C.6f}$$

$$= \left(\sum_{\tilde{x}_t^i, \tilde{x}_{t+1}^i} \xi_t^i(\tilde{x}_t^i)Q_x^i(\tilde{x}_{t+1}^i|\tilde{x}_t^i, a_t)Q_w^i(w_{t+1}^i|\tilde{x}_{t+1}^i, a_t)\right) \cdot \sum_{\tilde{x}_t, \tilde{\xi}_t^{-i}, \tilde{x}_{t+1}^i} \tilde{\xi}_t^{-i}(\tilde{x}_t^{-i})\mu_t^{*,-i}[a_{1:t-1}](\tilde{\xi}_t^{-i})\beta_t^i(a_t^i|a_{1:t-1}, w_{1:t}^i)\beta_t^{*,-i}(a_t^{-i}|a_{1:t-1}, \tilde{\xi}_t^{-i}). \tag{C.6g}$$

By canceling the terms $\beta_t^i(\cdot)$ in the numerator and the denominator and using the update equation for ξ_t^i, (C.6a) is given by

$$Nr = \sum_{x_t^i} \xi_t^i(x_t^i) Q_x^i(x_{t+1}^i | x_t^i, a_t^i) Q_w^i(w_{t+1}^i | x_{t+1}^i, a_t) \cdot \sum_{x_t^{-i}, \xi_t^{-i}} \xi_t^{-i}(x_t^{-i}) \mu_t^{*,-i}[a_{1:t-1}](\xi_t^{-i}) \beta_t^{*,-i}(a_t^{-i} | a_{1:t-1}, \xi_t^{-i}) Q_x^{-i}(x_{t+1}^{-i} | x_t^{-i}, a_t), \tag{C.6h}$$

and

$$Dr = \sum_{\tilde{x}_t^i, \tilde{x}_{t+1}^i} \xi_t^i(\tilde{x}_t^i) Q_x^i(\tilde{x}_{t+1}^i | \tilde{x}_t^i, a_t) Q_w(w_{t+1}^i | \tilde{x}_{t+1}^i, a_t) \cdot \sum_{\tilde{x}_t^{-i}, \tilde{\xi}_t^{-i}} \tilde{\xi}_t^{-i}(\tilde{x}_t^{-i}) \mu_t^{*,-i}[a_{1:t-1}](\tilde{\xi}_t^{-i}) \beta_t^{*,-i}(a_t^{-i} | a_{1:t-1}, \tilde{\xi}_t^{-i}) \tag{C.6i}$$

$$= \xi_{t+1}^i(x_{t+1}^i) \sum_{\xi_{t+1}^{-i}} \mu_{t+1}^{*,-i}[a_{1:t}](\xi_{t+1}^{-i}) \xi_{t+1}^{-i}(x_{t+1}^{-i}) \mathbb{P}^{\beta_{t+1:T}^i \beta_{t+1:T}^{*,-i}, \mu_{t+1}^*[a_{1:t}]}(a_{t+1:T}, x_{t+2:T} | a_{1:t}, w_{1:t}^i, x_{t+1}) \tag{C.6j}$$

$$= \mathbb{P}^{\beta_{t+1:T}^i \beta_{t+1:T}^{*,-i}, \mu_{t+1}^*[a_{1:t}]}(x_{t+1}, a_{t+1:T}, x_{t+2:T} | a_{1:t}, w_{1:t+1}^i). \quad \square \tag{C.6k}$$

Lemma C.3. *For all* $\forall i \in \mathcal{N}, t \in \mathcal{T}, a_{1:t-1} \in \mathcal{H}_t^c, w_{1:t}^i \in (\mathcal{W}^i)^t$,

$$V_t^i(\underline{\mu}_t^*[a_{1:t-1}], \xi_t^i) = \mathbb{E}^{\beta_{t:T}^{*,i} \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]} \left\{ \sum_{n=t}^T R^i(X_n, A_n) \; | a_{1:t-1}, w_{1:t}^i \right\}. \tag{C.7}$$

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

$$\mathbb{E}^{\beta_T^{*,i} \beta_T^{*,-i}, \mu_T^*[a_{1:T-1}]}\{R^i(X_T, A_T) | a_{1:T-1}, w_{1:T}^i\} = \sum_{x_T^{-i} a_T} R^i(x_T, a_T) \xi_T(x_T) \mu_T^*[a_{1:T-1}](\xi_T^{-i}) \beta_T^{*,i}(a_T^i | a_{1:T-1}, \xi_T^i) \beta_T^{*,-i}(a_T^{-i} | a_{1:T-1}, \xi_T^{-i}) \tag{C.8a}$$

$$= V_T^i(\underline{\mu}_T^*[a_{1:T-1}], \xi_T^i), \tag{C.8b}$$

where (C.8b) follows from the definition of V_t^i in (10) and the definition of β_T^* in the forward recursion in (12a).

Suppose the claim is true for $t+1$; that is, $\forall i \in \mathcal{N}, t \in \mathcal{T}, (a_{1:t}, w_{1:t+1}^i) \in \mathcal{H}_{t+1}^i$,

$$V_{t+1}^i(\underline{\mu}_{t+1}^*[a_{1:t}], \xi_{t+1}^i) = \mathbb{E}^{\beta_{t+1:T}^{*,i} \beta_{t+1:T}^{*,-i}, \mu_{t+1}^*[a_{1:t}]} \left\{ \sum_{n=t+1}^T R^i(X_n, A_n) \; | a_{1:t}, w_{1:t+1}^i \right\}. \tag{C.9}$$

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

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

$$= \mathbb{E}^{\beta_{t:T}^{*,i} \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]}\{R^i(X_t, A_t) + \mathbb{E}^{\beta_{t:T}^{*,i} \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]}\left\{ \sum_{n=t+1}^T R^i(X_n, A_n) | a_{1:t-1}, A_t, w_{1:t}^i, W_{t+1}^i \right\} | a_{1:t-1}, w_{1:t}^i \} \tag{C.10a}$$

$$= \mathbb{E}^{\beta_{t:T}^{*,i} \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]}\{R^i(X_t, A_t) + \mathbb{E}^{\beta_{t+1:T}^{*,i} \beta_{t+1:T}^{*,-i}, \mu_{t+1}^*[a_{1:t-1}, A_t]} \left\{ \sum_{n=t+1}^T R^i(X_n, A_n) | a_{1:t-1}, A_t, w_{1:t}^i, W_{t+1}^i \right\} | a_{1:t-1}, w_{1:t}^i \} \tag{C.10b}$$

$$= \mathbb{E}^{\beta_{t:T}^{*,i} \beta_{t:T}^{*,-i}, \mu_t^*[a_{1:t-1}]} \left\{ R^i(X_t, A_t) + V_{t+1}^i(\underline{\mu}_{t+1}^*[a_{1:t-1} A_t], \Xi_{t+1}^i) \; | a_{1:t-1}, w_{1:t}^i \right\} \tag{C.10c}$$

$$= \mathbb{E}^{\beta_t^{*,i} \beta_t^{*,-i}, \mu_t^*[a_{1:t-1}]} \left\{ R^i(X_t, A_t) + V_{t+1}^i(\underline{\mu}_{t+1}^*[a_{1:t-1} A_t], \Xi_{t+1}^i) \; | a_{1:t-1}, w_{1:t}^i \right\} \tag{C.10d}$$

$$= V_t^i(\underline{\mu}_t^*[a_{1:t-1}], \xi_t^i), \tag{C.10e}$$

where (C.10b) follows from Lemma C.2; (C.10c) follows from the induction hypothesis in (C.9); (C.10d) follows because the random variables involved in expectation, X_t^{-i}, A_t, X_{t+1}^i, do not depend on $\beta_{t+1:T}^{*,i} \beta_{t+1:T}^{*,-i}$; and (C.10e) follows from the definition of β_t^* in the forward recursion in (12a), the definition of μ_{t+1}^* in (12b), and the definition of V_t^i in (10).

Appendix D. Proof of Lemma 2

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

$$(\mu_{t+1}^*)^{-1}(\tilde{C}_{t+1}^{a_{t+1:T}}) = C_{t+1}^{a_{t+1:T}}. \tag{D.1}$$

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

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

Because $h_t^c \in (\mu_t^*)^{-1}(\tilde{\mathcal{C}}_t^{a_{t:T}})$, this implies $\mu_t^*[h_t^c] \in \tilde{\mathcal{C}}_t^{a_{t:T}}$. Then by the definition of $\tilde{\mathcal{C}}_t^{a_{t:T}}$, $\forall i$, $\forall \xi_t^i \in \text{supp}(\mu_t^{*,i}[h_t^c])$, $\theta_t^i[\underline{\mu}_t^*[h_t^c]](a_t^i|\xi_t^i) = 1$. Because $\xi_t^i(x_t^i) = \mathbb{P}(x_t^i|h_t^i)\ \forall x_t^i$, $\mu_t^{*,i}[h_t^c](\xi_t^i) = \mathbb{P}^\theta(\xi_t^i|h_t^c)\ \forall \xi_t^i$ and $\beta_t^{*,i}(a_t^i|h_t^i) = \theta_t^i[\underline{\mu}_t^*[h_t^c]](a_t^i|\xi_t^i)$, 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 nonzero probability.

Also because $\mu_t^*[h_t^c] \in \tilde{\mathcal{C}}_t^{a_{t:T}}$, this implies $\underline{F}(\underline{\mu}_t^*[h_t^c], \theta_t[\underline{\mu}_t^*[h_t^c]], a_t) \in \tilde{\mathcal{C}}_{t+1}^{a_{t+1:T}}$ by definition of $\tilde{\mathcal{C}}_t^{a_{t:T}}$. Thus $\mu_{t+1}^*[h_t^c, a_t] \in \tilde{\mathcal{C}}_{t+1}^{a_{t+1:T}}$, because $\mu_{t+1}^*[h_t^c, a_t] = F(\mu_t^*[h_t^c], \theta_t[\underline{\mu}_t^*[h_t^c]], a_t)$ 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 nonzero probability.

These two facts suggest 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 nonzero 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 $\mu_t^*[h_t^c] \in \tilde{\mathcal{C}}_t^{a_{t:T}}$.

Because $h_t^c \in \mathcal{C}_t^{a_{t:T}}$, 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 nonzero probability. Because $\beta_t^{*,i}(a_t^i|h_t^i) = \theta_t^i[\underline{\mu}_t^*[h_t^c]](a_t^i|\xi_t^i)$, by the definition of β^*, where $\xi_t^i(x_t^i) = \mathbb{P}(x_t^i|h_t^i)\ \forall x_t^i$, this implies $\forall i, \theta_t^i[\underline{\mu}_t^*[h_t^c]](a_t^i|\xi_t^i) = 1$, $\forall \xi_t^i \in \text{supp}(\mu_t^{*,i}[h_t^c])$, where $\mu_t^{*,i}[h_t^c](\xi_t^i) = \mathbb{P}^\theta(\xi_t^i|h_t^c)\ \forall \xi_t^i$.

Also, because $h_t^c \in \mathcal{C}_t^{a_{t:T}}$, 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. Because, by definition, $\mu_{t+1}^*[h_t^c, a_t] = F([\mu_t^*[h_t^c], \theta_t[\underline{\mu}_t^*[h_t^c]], a_t)$, this implies $F(\mu_t^*[h_t^c], \theta_t[\underline{\mu}_t^*[h_t^c]], a_t) \in \tilde{\mathcal{C}}_{t+1}^{a_{t+1:T}}$.

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

These two cases complete the induction step. □

Appendix E.

In this appendix, we show that each private belief process conditioned on the true state of the world forms a sub- or super-martingale.

Lemma E.1. *Conditioned on $x^i = 1$, $\{\Xi_t^i\}_t$ is a submartingale.*

Proof. We know that

$$\xi_{t+1}^i = \begin{cases} G^i(\xi_t^i, w_{t+1}^i = 0, a_t^i) = \dfrac{\xi_t^i p_{a_t^i}}{\xi_t^i p_{a_t^i} + (1-\xi_t^i)(1-p_{a_t^i})} & \text{with probability } p_{a_t^i}, \\ G^i(\xi_t^i, w_{t+1}^i = 1, a_t^i) = \dfrac{\xi_t^i (1-p_{a_t^i})}{\xi_t^i(1-p_{a_t^i}) + (1-\xi_t^i)p_{a_t^i}} & \text{with probability } 1-p_{a_t^i}. \end{cases} \tag{E.1}$$

Thus,

$$\mathbb{E}[\xi_{t+1}^i|\xi_t^i, a_t^i] - \xi_t^i = \frac{\xi_t^i(p_{a_t^i})^2}{\xi_t^i p_{a_t^i} + (1-\xi_t^i)(1-p_{a_t^i})} + \frac{\xi_t^i(1-p_{a_t^i})^2}{\xi_t^i(1-p_{a_t^i}) + (1-\xi_t^i)p_{a_t^i}} - \xi_t^i \tag{E.2}$$

$$= \frac{\xi_t^i(1-\xi_t^i)^2(1-2p_{a_t^i})^2}{(\xi_t^i p_{a_t^i} + (1-\xi_t^i)(1-p_{a_t^i}))(\xi_t^i(1-p_{a_t^i}) + (1-\xi_t^i)p_{a_t^i})} \tag{E.3}$$

$$\geq 0, \tag{E.4}$$

with the inequality being strict for $p_{a_t^i} < \frac{1}{2}$ and $\xi_t^i \notin \{0,1\}$. □

Appendix F. Proof of Theorem 2

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

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

and because $\pi_T \in \hat{\mathcal{C}}^a$, 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 $V_T^i(\pi_T, \xi_T^i) = (\lambda(2\xi_T^i - 1) + \bar{\lambda}(2\hat{\xi}_T^{-i} - 1))a^i$. This establishes the base case.

Now, suppose the claim is true for $t_0 = \tau + 1$; that is, 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, $\forall \pi_t \in \hat{\mathcal{C}}^a$,

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

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

$$\tilde{\gamma}^i_\tau(\cdot|\xi^i_\tau) \in \arg\max_{\gamma^i_\tau(\cdot|\xi^i_\tau)} \sum_{a^i_\tau} a^i_\tau \gamma^i_\tau(a^i_\tau|\xi^i_\tau)(\lambda(2\xi^i_\tau - 1) + \bar{\lambda}(2\hat{\xi}^{-i}_\tau - 1)) + \mathbb{E}^{\gamma^i_\tau(\cdot|\xi^i_\tau)\tilde{\gamma}^{-i}_\tau,\pi_\tau}\{V^i_{\tau+1}(\underline{F}(\underline{\pi}_\tau, \tilde{\gamma}_\tau, A_\tau), \Xi^i_{\tau+1})\,|\xi^i_\tau\} \tag{F.3}$$

$$= \arg\max_{\gamma^i_\tau(\cdot|\xi^i_\tau)} \sum_{a^i_\tau} a^i_\tau \gamma^i_\tau(a^i_\tau|\xi^i_\tau)(\lambda(2\xi^i_\tau - 1) + \bar{\lambda}(2\hat{\xi}^{-i}_\tau - 1)) + \mathbb{E}^{\gamma^i_\tau(\cdot|\xi^i_\tau)\tilde{\gamma}^{-i}_\tau,\pi_\tau}\{(T-\tau)(\lambda(2\Xi^i_{\tau+1} - 1) + \bar{\lambda}(2\hat{\Xi}^{-i}_{\tau+1} - 1))a^i|\xi^i_\tau\} \tag{F.4a}$$

$$= \arg\max_{\gamma^i_\tau(\cdot|\xi^i_\tau)} \sum_{a^i_\tau} a^i_\tau \gamma^i_\tau(a^i_\tau|\xi^i_\tau)(\lambda(2\xi^i_\tau - 1) + \bar{\lambda}(2\hat{\xi}^{-i}_\tau - 1)) + (T-\tau)(\lambda(2\xi^i_\tau - 1) + \bar{\lambda}(2\hat{\xi}^{-i}_\tau - 1))a^i \tag{F.4b}$$

$$= \arg\max_{\gamma^i_\tau(\cdot|\xi^i_\tau)} \sum_{a^i_\tau} a^i_\tau \gamma^i_\tau(a^i_\tau|\xi^i_\tau)(\lambda(2\xi^i_\tau - 1) + \bar{\lambda}(2\hat{\xi}^{-i}_\tau - 1)), \tag{F.4c}$$

where (F.4a) follows from the fact that $\underline{F}(\underline{\pi}_\tau, \tilde{\gamma}_\tau, a_\tau) \in C^a$, $\forall a_\tau$, as shown in Lemma F.1, and the induction hypothesis, (F.4b), follows from Lemmas F.1 and F.2, and (F.4c) follows from the fact that the second term does not depend on $\gamma^i_\tau(\cdot|\xi^i_\tau)$. This also shows that, $\forall \pi_t \in \hat{\mathcal{C}}^a$,

$$V^i_\tau(\pi_\tau, \xi^i_\tau) = (T - \tau + 1)(\lambda(2\xi^i_\tau - 1) + \bar{\lambda}(2\hat{\xi}^{-i}_\tau - 1))a^i, \tag{F.5}$$

which completes the induction step. □

Lemma F.1. *Expectation of π^i_{t+1} under noninformative $\tilde{\gamma}^i_t$ of the form $\tilde{\gamma}^i_t(a^i|\xi^i_t) = 1 \;\forall \xi^i_t \in [0,1]$ remains the same as the mean of π^i_t; that is,*

$$\mathbb{E}\{\Xi^i_{t+1}(1)|\pi^i_t, \tilde{\gamma}^i_t, a^i\} = \mathbb{E}\{\Xi^i_t(1)|\pi^i_t\}. \tag{F.6}$$

Proof. We have that

$$\mathbb{E}\{\Xi^i_{t+1}(1)|\pi^i_t, \tilde{\gamma}^i_t, a^i\} = \sum_{\xi^i_{t+1}(1)} \xi^i_{t+1}(1)F^i(\pi^i_t, \tilde{\gamma}^i_t, a^i)(\xi^i_{t+1}(1)) \tag{F.7a}$$

$$= \frac{\sum_{\xi^i_t, x^i, \xi^i_{t+1}(1)} \xi^i_{t+1}(1)\pi^i_t(\xi^i_t)\xi^i_t(x^i)\tilde{\gamma}^i_t(a^i_t|\xi^i_t)Q^i_w(w^i_{t+1}|x^i, a_t)I_{G^i(\xi^i_t, w^i_{t+1}, a_t)(1)}(\xi^i_{t+1}(1))}{\sum_{\xi^i_t, x^i, w^i_{t+1}} \pi^i_t(\xi^i_t)\xi^i_t(x^i)\tilde{\gamma}^i_t(a^i_t|\xi^i_t)} \tag{F.7b}$$

$$= \frac{\sum_{\xi^i_t, x^i, w^i_{t+1}, \xi^i_{t+1}(1)} \xi^i_{t+1}(1)\pi^i_t(\xi^i_t)\xi^i_t(x^i)Q^i_w(w^i_{t+1}|x^i, a^i)I_{G^i(\xi^i_t, w^i_{t+1}, a^i)(1)}(\xi^i_{t+1}(1))}{\sum_{\xi^i_t, x^i} \pi^i_t(\xi^i_t)\xi^i_t(x^i)} \tag{F.7c}$$

$$= \sum_{\xi^i_t, x^i, w^i_{t+1}} G^i(\xi^i_t, w^i_{t+1}, a^i)(1)\pi^i_t(\xi^i_t)\xi^i_t(x^i)Q^i_w(w^i_{t+1}|x^i, a^i) \tag{F.7d}$$

$$= \sum_{\xi^i_t, w^i_{t+1}} \frac{\xi^i_t(1)Q^i_w(w^i_{t+1}|1, a^i)}{\sum_{\tilde{x}^i} \xi^i_t(\tilde{x}^i)Q^i_w(w^i_{t+1}|\tilde{x}^i, a^i)} \pi^i_t(\xi^i_t) \sum_{x^i} \xi^i_t(x^i)Q^i_w(w^i_{t+1}|x^i, a^i) \tag{F.7e}$$

$$= \sum_{\xi^i_t} \xi^i_t(1)\pi^i_t(\xi^i_t(1)) \tag{F.7f}$$

$$= \mathbb{E}\{\Xi^i_t(1)|\pi^i_t\}. \quad \square \tag{F.7g}$$

Lemma F.2. *For any γ^i_t,*

$$\mathbb{E}\{\Xi^i_{t+1}(1)|\xi^i_t, \gamma^i_t\} = \xi^i_t(1). \tag{F.8}$$

Proof. We have that

$$\mathbb{E}\{\Xi^i_{t+1}(1)|\xi^i_t, \gamma^i_t\} = \sum_{x^i, w^i_{t+1}, a^i_t, \xi^i_{t+1}(1)} \xi^i_{t+1}(1)I_{F^i(\xi^i_t, w^i_{t+1}, a^i_t)(1)}(\xi^i_{t+1}(1))\xi^i_t(x^i)Q^i_w(w^i_{t+1}|x^i, a^i_t)\gamma^i_t(a^i_t|\xi^i_t) \tag{F.9a}$$

$$= \sum_{x^i, w^i_{t+1}, a^i_t} F^i(\xi^i_t, w^i_{t+1}, a^i_t)(1)\xi^i_t(x^i)Q^i_w(w^i_{t+1}|x^i, a^i_t)\gamma^i_t(a^i_t|\xi^i_t) \tag{F.9b}$$

$$= \sum_{a^i_t, w^i_{t+1}} \frac{\xi^i_t(1)Q^i_w(w^i_{t+1}|1, a^i_t)}{\sum_{\tilde{x}^i} \xi^i_t(\tilde{x}^i)Q^i_w(w^i_{t+1}|\tilde{x}^i, a^i_t)} \gamma^i_t(a^i_t|\xi^i_t) \sum_{x^i} \xi^i_t(x^i)Q^i_w(w^i_{t+1}|x^i, a^i_t) \tag{F.9c}$$

$$= \sum_{a^i_t, w^i_{t+1}} \xi^i_t(1)Q^i_w(w^i_{t+1}|1, a^i_t)\gamma^i_t(a^i_t|\xi^i_t) \tag{F.9d}$$

$$= \xi^i_t(1). \quad \square \tag{F.9e}$$

Endnotes

[1] As we will see in the example in Section 4, when types of the players are static (i.e., when $x^i_{t+1} = x^i_t = \cdots = x^i_1$), full support assumption on the priors $Q^i_x(x^i_1)$ is sufficient for our methodology to work.

[2] 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 her observed history $h^i_t = (a_{1:t-1}, w^i_{1:t})$ such that ${}^i\mu_t[h^i_t](h_t)$ is this conditional probability. However, because of the specific structure of our model, we only need to put a belief on the random variables $X_t, W^{-i}_{1:t}$.

[3] Because of the assumption of full support of kernel Q^i_w, the condition $\mathbb{P}^\beta(h^i_t|h^i_{t-1}) > 0$ simplifies to $\mathbb{P}^\beta(a_{t-1}|h^i_{t-1}) > 0$.

[4] We note that although the kernels Q^i_x do not have full support, in the case of static types that we consider in this example, the whole methodology for computing structured PBEs presented in the previous section goes through almost verbatim, assuming full support on the priors $Q^i_x(x^i_1)$. To emphasize this simplification, in the following we drop the time index when referring to state variables (i.e., $X^i_t = X^i$).

References

Acemoglu D, Dahleh MA, Lobel I, Ozdaglar A (2011) Bayesian learning in social networks. *Rev. Econom. Stud.* 78(4):1201–1236.
Bala V, Goyal S (1998) Learning from neighbours. *Rev. Econom. Stud.* 65(3):595–621.
Banerjee AV (1992) A simple model of herd behavior. *Quart. J. Econom.* 107(3):797–817.
Banks JS, Sobel J (1987) Equilibrium selection in signaling games. *Econometrica* 55(3):647–661.
Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Political Econom.* 100(5):992–1026.
Çelen B, Kariv S (2004) Distinguishing informational cascades from herd behavior in the laboratory. *Amer. Econom. Rev.* 94(3):484–498.
Cho I-K (1987) A refinement of sequential equilibria. *Econometrica* 55(6):1367–1389.
Cho I-K, Kreps DM (1987) Signaling games and stable equilibria. *Quart. J. Econom.* 102(2):179–221.
Cover TM, Thomas JA (2012) *Elements of Information Theory*, 2nd ed. (John Wiley & Sons, Hoboken, NJ).
DeGroot MH (1974) Reaching a consensus. *J. Amer. Statist. Assoc.* 69(345):118–121.
Ellison G, Fudenberg D (1993) Rules of thumb for social learning. *J. Political Econom.* 101(4):612–643.
Ellison G, Fudenberg D (1995) Word-of-mouth communication and social learning. *Quart. J. Econom.* 110(1):93–125.
Fudenberg D, Tirole J (1991) *Game Theory* (MIT Press, Cambridge, MA).
Gale D, Kariv S (2003) Bayesian learning in social networks. *Games Econom. Behav.* 45(2):329–346.
Guarino A, Jehiel P (2013) Social learning with coarse inference. *Amer. Econom. J. Microeconom.* 5(1):147–174.
Guarino A, Harmgart H, Huck S (2011) Aggregate information cascades. *Games Econom. Behav.* 73(1):167–185.
Harel M, Mossel E, Strack P, Tamuz O (2014) The speed of social learning. Preprint, submitted December 22, https://arxiv.org/abs/1412.7172v1.
Herrera H, Hörner J (2013) Biased social learning. *Games Econom. Behav.* 80:131–146.
Jadbabaie A, Molavi P, Sandroni A, Tahbaz-Salehi A (2012) Non-Bayesian social learning. *Games Econom. Behav.* 76(1):210–225.
Le TN, Subramanian VG, Berry RA (2014) The impact of observation and action errors on informational cascades. *53rd IEEE Conf. Decision Control* (IEEE, Piscataway, NJ), 1917–1922.
Le TN, Subramanian VG, Berry RA (2017) Information cascades with noise. *IEEE Trans. Signal Inform. Processing Networks* 3(2):239–251.
Maskin E, Tirole J (2001) Markov perfect equilibrium: I. Observable actions. *J. Econom. Theory* 100(2):191–219.
Molavi P, Tahbaz-Salehi A, Jadbabaie A (2017) Foundations of non-Bayesian social learning. Columbia Business School Research Paper 15-95, Columbia University, New York.
Mossel E, Tamuz O (2013) Making consensus tractable. *ACM Trans. Econom. Comput.* 1(4):1–19.
Mossel E, Sly A, Tamuz O (2014) Asymptotic learning on Bayesian social networks. *Probab. Theory Related Fields* 158(1–2):127–157.
Mossel E, Sly A, Tamuz O (2015) Strategic learning and the topology of social networks. *Econometrica* 83(5):1755–1794.
Mossel E, Mueller-Frank M, Sly A, Tamuz O (2020) Social learning equilibria. *Econometrica* 88(3):1235–1267.
Nayyar A, Mahajan A, Teneketzis D (2013) Decentralized stochastic control with partial history sharing: A common information approach. *IEEE Trans. Automatic Control* 58(7):1644–1658.
Nayyar A, Gupta A, Langbort C, Başar T (2014) Common information based Markov perfect equilibria for stochastic games with asymmetric information: Finite games. *IEEE Trans. Automatic Control* 59(3):555–570.
Nedić A, Olshevsky A, Uribe CA (2016) A tutorial on distributed (non-Bayesian) learning: Problem, algorithms and results. *2016 IEEE 55th Conf. Decision Control* (IEEE, Piscataway, NJ), 6795–6801.
Osborne MJ, Rubinstein A (1994) *A Course in Game Theory*, vol. 1 (MIT Press, Cambridge, MA).
Ouyang Y, Tavafoghi H, Teneketzis D (2017) Dynamic games with asymmetric information: Common information based perfect Bayesian equilibria and sequential decomposition. *IEEE Trans. Automatic Control* 62(1):222–237.
Smith L, Sörensen P (2000) Pathological outcomes of observational learning. *Econometrica* 68(2):371–398.
Vasal (2019) Sequential decomposition of repeated games with asymmetric information and dependent states. *2019 Amer. Control Conf. (ACC)*, 2309–2314. 10.23919/ACC.2019.8814782.
Vasal D, Anastasopoulos A (2016a) Decentralized Bayesian learning in dynamic games. *54th Allerton Conf. Comm. Control Comput*. (IEEE, Piscataway, NJ), 264–273.
Vasal D, Anastasopoulos A (2016b) Decentralized Bayesian learning in dynamic games: A framework to study informational cascades. Preprint, submitted July 22, https://arxiv.org/abs/1607.06847.
Vasal D, Sinha A, Anastasopoulos A (2019) A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information. *IEEE Trans. Automatic Control* 64(1):81–96.