Games and Economic Behavior 76 (2012) 210–225

Contents lists available at SciVerse ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Non-Bayesian social learning ☆

Ali Jadbabaie [a], Pooya Molavi [a], Alvaro Sandroni [b,*], Alireza Tahbaz-Salehi [c]

[a] *Department of Electrical and Systems Engineering, University of Pennsylvania, United States*
[b] *Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, United States*
[c] *Decisions, Risk, and Operations Division, Columbia Business School, Columbia University, United States*

ARTICLE INFO

Article history:
Received 10 February 2010
Available online 13 June 2012

JEL classification:
D83
L14

Keywords:
Social networks
Learning
Information aggregation

ABSTRACT

We develop a dynamic model of opinion formation in social networks when the information required for learning a parameter may not be at the disposal of any single agent. Individuals engage in communication with their neighbors in order to learn from their experiences. However, instead of incorporating the views of their neighbors in a fully Bayesian manner, agents use a simple updating rule which linearly combines their personal experience and the views of their neighbors. We show that, as long as individuals take their *personal* signals into account in a Bayesian way, repeated interactions lead them to successfully aggregate information and learn the true parameter. This result holds in spite of the apparent naïveté of agents' updating rule, the agents' need for information from sources the existence of which they may not be aware of, worst prior views, and the assumption that no agent can tell whether her own views or those of her neighbors are more accurate.

1. Introduction

In everyday life, people form opinions over various economic, political, and social issues — such as how to educate their children or whether to invest in a certain asset — which do not have an obvious solution. These issues allow for a great variety of opinions because even if a satisfactory solution exists, it is not easily recognizable. In addition, the relevant information for such problems is often not concentrated in any source or body of sufficient knowledge. Instead, the data are dispersed throughout a vast network, where each individual observes only a small fraction, consisting of his/her personal experience. This motivates an individual to engage in communication with others in order to learn from other people's experiences. For example, Hagerstrand (1969) and Rogers (1983) document such a phenomenon in the choice of new agricultural techniques by various farmers, while Kotler (1986) shows the importance of learning from others in the purchase of consumer products.

In many scenarios, however, the information available to an individual is not directly observable by others. At most, each individual only knows the opinions of few individuals (such as colleagues, family members, and maybe a few news organizations), will never know the opinions of everyone in the society, and might not even know the full personal experience of anyone but herself. This limited observability, coupled with the complex interactions of opinions arising from dispersed

☆ We thank the editors, Vincent Crawford and Matthew Jackson, and two anonymous referees for very helpful remarks and suggestions. We also thank Ilan Lobel and numerous seminar and conference participants for useful feedback and comments. Jadbabaie, Molavi and Tahbaz-Salehi gratefully acknowledge financial support from the Air Force Office of Scientific Research (Complex Networks Program) and the Office of Naval Research. Sandroni gratefully acknowledges the financial support of the National Science Foundation (Grants SES-0820472 and SES-0922404).

* Corresponding author.
E-mail address: sandroni@kellogg.northwestern.edu (A. Sandroni).

0899-8256/$ – see front matter
http://dx.doi.org/10.1016/j.geb.2012.06.001

information over the network, makes it highly impractical for agents to incorporate other people's views in a Bayesian fashion.

The difficulties with Bayesian updating are further intensified if agents do not have complete information about the structure of the social network or the probability distribution of signals observed by other individuals. Such incomplete information means that they would need to form and update opinions not only on the states of the world, but also on the network topology as well as other individuals' signal structures. This significantly complicates the required calculations for Bayesian updating of beliefs, well beyond agents' regular computational capabilities. Nevertheless, the complications with Bayesian learning persist even when individuals have complete information about the network structure, as they still need to perform deductions about the information of every other individual in the network, while only observing the evolution of opinions of their neighbors.[1] The necessary information and the computational burden of these calculations are simply prohibitive for adopting Bayesian learning, even in relatively simple networks.[2]

In this paper, we study the evolution of opinions in a society where agents, instead of performing Bayesian updates, apply a simple learning rule to incorporate the views of individuals in their social clique. We assume that at every time period, each individual receives a private signal, and observes the opinions (i.e., the beliefs) held by her neighbors at the previous period. The individual updates her belief as a convex combination of: (i) the Bayesian posterior belief conditioned on her private signal; and (ii) the opinions of her neighbors. The weight an individual assigns to the opinion of a neighbor represents the influence (or persuasion power) of that neighbor on her. At the end of the period, agents report their opinions truthfully to their neighbors. The influence that agents exert on one another can be large or small, and may depend on each pair of agents. Moreover, this persuasion power may be independent of the informativeness of their signals. In particular, more persuasive agents may not be better informed or hold more accurate views. In such cases, in initial periods, agents' views may move towards the views of the most persuasive agents and, hence, away from the data generating process.

We analyze the flow of opinions as new observations accumulate. First, we show that agents eventually make correct forecasts, provided that the social network is strongly connected; that is, there exists either a direct or an indirect information path between any two agents. Hence, the seemingly naïve updating rule will eventually transform the existing data into a near perfect guide for the future even though the truth is not recognizable, agents do not know if their views are more or less accurate than the views of their neighbors, and the most persuasive agents may have the least accurate views. By the means of an example we show that the assumption of strong connectivity cannot be disposed of.

We further show that in strongly connected networks, the non-Bayesian learning rule also enables agents to successfully aggregate dispersed information. Each agent eventually learns the truth even though no agent and her neighbors, by themselves, may have enough information to infer the underlying parameter. Eventually, each agent learns as if she were completely informed of all observations of *all* agents *and* updated her beliefs according to Bayes' rule. This aggregation of information is achieved while agents avoid the computational complexity involved in Bayesian updating. Thus, with a constant flow of new information, a sufficient condition for social learning in strongly connected networks is that individuals simply take their *personal* signals into account in a Bayesian manner. If such a condition is satisfied, then repeated interactions over the social network guarantee that the viewpoints of different individuals will eventually coincide, leading to complete aggregation of information.

Our results also highlight the role of social networks in information propagation and aggregation. An agent can learn from individuals with whom she is not in direct contact, and even from the ones of whose existence she is unaware. In other words, the indirect communication path in the social network guarantees that she will eventually incorporate the information initially revealed to agents in distant corners of the network into her beliefs. Thus, agents can learn the true state of the world even if they all face an identification problem.[3]

Our basic learning results hold in a wide spectrum of networks and under conditions that are seemingly not conducive to learning. For example, assume that one agent receives uninformative signals and has strong persuasive powers over all agents, including the only individual with informative signals (but who may not know that her signals are more informative than the signals of others). The agent with informative signals cannot directly influence the more persuasive agents and only has a small, direct persuasive power over a few other agents. We show that all agents' views will eventually be as if they were based on informative signals, despite the fact that most agents will never see these informative signals and will not know where they come from. Thus, the paper also establishes that whenever agents take their own information into account in a Bayesian way, neither the fine details of the network structure (beyond strong connectivity) nor the prior beliefs can prevent them from learning, as the effects of both are eventually "washed away" by the constant flow of new information.

[1] Gale and Kariv (2003) illustrate the complications that can arise due to repeated Bayesian deductions in a simple network. Also, as DeMarzo et al. (2003) point out, in order to disentangle old information from new, a Bayesian agent needs to recall the information she received from her neighbors in the previous communication rounds, and therefore, "[w]ith multiple communication rounds, such calculations would become quite laborious, even if the agent knew the entire social network."

[2] An exception, as shown recently by Mossel and Tamuz (2010), is the case in which agents' signal structures, their prior beliefs, and the social network are common knowledge and all signals and priors are normally distributed.

[3] It is important that not all agents face the same identification problem. We formalize this statement in the following sections.

The paper is organized as follows. The next section discusses the related literature. Section 3 contains our model. Our main results are presented in Section 4. In Section 5 we show that, under suitable conditions, Bayesian and non-Bayesian learning procedures asymptotically coincide. Section 6 concludes. All proofs can be found in Appendix A.

2. Related literature

There exists a large body of works on learning over social networks, both boundedly and fully rational. The Bayesian social learning literature focuses on formulating the problem as a dynamic game with incomplete information and characterizing its equilibria. However, since characterizing the equilibria in complex networks is generally intractable, the literature for the most part studies relatively simple and stylized environments. More specifically, rather than considering repeated interactions over the network, it focuses on models where agents interact sequentially and communicate with their neighbors only *once*. Examples include Banerjee (1992), Bikhchandani et al. (1992), Smith and Sørensen (2000), Banerjee and Fudenberg (2004), and more recently, Acemoglu et al. (2011). An exception is Rosenberg et al. (2009), who study a model of dynamic games with purely informational externalities and provide conditions under which players eventually reach a consensus.

Our work is also related to the social learning literature that focuses on non-Bayesian learning models, such as Ellison and Fudenberg (1993, 1995) and Bala and Goyal (1998, 2001), in which agents use simple rule-of-thumb methods to update their beliefs. In the same spirit are DeMarzo et al. (2003), Golub and Jackson (2010), and Acemoglu et al. (2010), which are based on the opinion formation model of DeGroot (1974). In DeGroot-style models, each individual initially receives *one* signal about the state of the world and the focus is on conditions under which individuals in the connected components of the social network converge to similar opinions. Golub and Jackson further show that if the size of the network grows unboundedly, this asymptotic consensus opinion converges to the true state of the world, provided that there are not overly influential agents in the society.

A feature that distinguishes our model from the works that are based on DeGroot's model, such as Golub and Jackson (2010), is the constant arrival of new information over time. Whereas in DeGroot's model each agent has only a single observation, the individuals in our model receive information in small bits over time. This feature of our model can potentially lead to learning in finite networks, a feature absent in other DeGroot-style models, where learning can only occur when the number of agents increases unboundedly. Thus, as long as there is a constant flow of new information and agents take their personal signals into account in a Bayesian manner, the resulting learning process asymptotically coincides with Bayesian learning, despite the fact that agents use a DeGroot-style update to incorporate the views of their neighbors.

The crucial difference in results between Golub and Jackson (2010) and Acemoglu et al. (2010), on the one hand, and our model, on the other, is the role played by the social network in successful information aggregation. These papers show that presence of "influential" individuals — those who are connected to a large number of people or affect their opinions disproportionally — may lead to disagreements or spread of misinformation. In contrast, in our environment, strong connectivity is the only requirement on the network for successful learning, and neither the network topology nor the influence level of different agents can prevent learning. In fact, social learning is achieved even if the most influential agents (both in terms of their persuasion power and in terms of their location in the network) are the ones with the least informative signals.

Finally, our work is also related to Epstein et al. (2008), who provide choice-theoretic foundations for non-Bayesian opinion formation dynamics of a *single* agent. However, the focus of our analysis is on the process of information aggregation over a network comprising of many agents.

3. The model

3.1. Agents and observations

Let Θ denote a finite set of possible states of the world. We consider a set $\mathcal{N} = \{1, 2, \ldots, n\}$ of agents interacting over a social network. Each agent i starts with a prior belief $\mu_{i,0} \in \Delta\Theta$ which is a probability distribution over the set Θ. More generally, we denote the opinion of agent i at time period $t \in \{0, 1, 2 \ldots\}$ by $\mu_{i,t} \in \Delta\Theta$.

Conditional on the state of the world θ, at each time period $t \geqslant 1$, an observation profile $\omega_t = (\omega_{1,t}, \ldots, \omega_{n,t}) \in S_1 \times \cdots \times S_n \equiv S$ is generated by the likelihood function $\ell(\cdot|\theta)$. We let $\omega_{i,t} \in S_i$ denote the signal privately observed by agent i at period t and S_i denote agent i's signal space, which we assume to be finite. The privately observed signals are independent over time, but might be correlated among agents at the same time period. We assume that $\ell(s|\theta) > 0$ for all $(s, \theta) \in S \times \Theta$ and use $\ell_i(\cdot|\theta)$ to denote the i-th marginal of $\ell(\cdot|\theta)$. We further assume that every agent i knows the conditional likelihood function $\ell_i(\cdot|\theta)$, known as her *signal structure*.

We do not require the observations to be informative about the state. In fact, each agent may face an identification problem, in the sense that she might not be able to distinguish between two states. We say two states are *observationally equivalent from the point of view of an agent* if the conditional distributions of her signals under the two states coincide. More specifically, the elements of the set $\Theta_i^\theta = \{\tilde{\theta} \in \Theta\colon\ \ell_i(s_i|\tilde{\theta}) = \ell_i(s_i|\theta) \text{ for all } s_i \in S_i\}$ are observationally equivalent to state θ from the point of view of agent i.

Finally, for a fixed $\theta \in \Theta$, we define a probability triple $(\Omega, \mathcal{F}, \mathbb{P}^\theta)$, where Ω is the space containing sequences of realizations of the signals $\omega_t \in S$ over time, $\mathcal{F}$ is the σ-field generated by the sequence of signal profiles, and $\mathbb{P}^\theta$ is the probability measure induced over sample paths in Ω. In other words, $\mathbb{P}^\theta = \bigotimes_{t=1}^{\infty} \ell(\cdot|\theta)$. We use $\mathbb{E}^\theta[\cdot]$ to denote the expectation operator associated with measure $\mathbb{P}^\theta$. Define $\mathcal{F}_{i,t}$ as the σ-field generated by the past history of agent i's observations up to time period t, and let $\mathcal{F}_t$ be the smallest σ-field containing all $\mathcal{F}_{i,t}$ for $1 \leqslant i \leqslant n$.

3.2. Social structure

We assume that when updating their views about the underlying state of the world, agents communicate their beliefs with individuals in their social clique. An advantage of communicating beliefs over signals is that all agents share the same space of beliefs, whereas their signal spaces may differ, making it difficult for them to interpret the signals observed by others. Moreover, in many scenarios, private signals of an individual are in the form of personal experiences which may not be easily communicable to other agents.

We capture the social interaction structure between agents by a directed graph $G = (V, E)$, where each vertex in V corresponds to an agent, and an edge connecting vertex i to vertex j, denoted by the ordered pair $(i, j) \in E$, captures the fact that agent j has access to the opinion held by agent i. Note that opinion of agent i might be accessible to agent j, but not the other way around.

For each agent i, define $\mathcal{N}_i = \{j \in V\colon (j, i) \in E\}$, called the set of *neighbors* of agent i. The elements of this set are agents whose opinions are available to agent i at each time period. We assume that individuals report their opinions truthfully to their neighbors.

A directed path in $G = (V, E)$ from vertex i to vertex j, is a sequence of vertices starting with i and ending with j such that each vertex is a neighbor of the next vertex in the sequence. We say the social network is *strongly connected* if there exists a directed path from each vertex to any other vertex.

3.3. Belief updates

Before the beginning of each period, agents observe the opinions of their neighbors. At the beginning of period t, signal profile $\omega_t = (\omega_{1,t}, \ldots, \omega_{n,t})$ is realized and signal $\omega_{i,t}$ is privately observed by agent i. Following the realization of the private signals, each agent computes her Bayesian posterior belief conditional on the signal observed, and then sets her final belief to be a linear combination of the Bayesian posterior and the opinions of her neighbors, observed right before the beginning of the period. At the end of the period, agents report their opinions to their neighbors. More precisely, if we denote the belief that agent i assigns to state $\theta \in \Theta$ at time period t by $\mu_{i,t}(\theta)$, then

$$\mu_{i,t+1} = a_{ii}\,\mathrm{BU}(\mu_{i,t}; \omega_{i,t+1}) + \sum_{j \in \mathcal{N}_i} a_{ij}\mu_{j,t}, \tag{1}$$

where $a_{ij} > 0$ captures the weight that agent i assigns to the opinion of agent j in her neighborhood, $\mathrm{BU}(\mu_{i,t}; \omega_{i,t+1})(\cdot)$ is the Bayesian update of $\mu_{i,t}$ when signal $\omega_{i,t+1}$ is observed, and a_{ii} is the weight that the agent assigns to her Bayesian posterior conditional on her private signal, which we refer to as the measure of *self-reliance* of agent i.[4] Note that weights a_{ij} must satisfy $\sum_{j \in \mathcal{N}_i \cup \{i\}} a_{ij} = 1$, in order for the period $t+1$ beliefs to form a well-defined probability distribution.

Even though agents incorporate their private signals into their beliefs using Bayes' rule, their belief updating is *non-Bayesian*: rather than conditioning their beliefs on all the information available to them, agents treat the beliefs generated through linear interactions with their neighbors as Bayesian priors when incorporating their private signals.

The beliefs of agent i on Θ at any time period induce forecasts about the future events. We define the *k-step-ahead forecasts* of agent i at a given time period as the probability measure induced by her beliefs over the realizations of her private signals in the next k consecutive time periods. More specifically, we denote the period t belief of agent i that signals $s_{i,1}, \ldots, s_{i,k} \in S_i$ will be realized in time periods $t+1$ through $t+k$, respectively, by $m_{i,t}^{(k)}(s_{i,1}, \ldots, s_{i,k})$. Thus, the k-step-ahead forecasts of agent i at time t are given by

$$m_{i,t}^{(k)}(s_{i,1}, \ldots, s_{i,k}) = \int_{\Theta} \left[\ell_i(s_{i,1}|\theta)\ell_i(s_{i,2}|\theta)\ldots\ell_i(s_{i,k}|\theta)\right] d\mu_{i,t}(\theta), \tag{2}$$

and therefore, the law of motion for the beliefs about the parameters can be written as

$$\mu_{i,t+1}(\theta) = a_{ii}\mu_{i,t}(\theta)\frac{\ell_i(\omega_{i,t+1}|\theta)}{m_{i,t}(\omega_{i,t+1})} + \sum_{j \in \mathcal{N}_i} a_{ij}\mu_{j,t}(\theta), \tag{3}$$

[4] One can generalize this belief update model and assume that agent i's belief depends linearly on her own beliefs at the previous time period as well. Such an assumption is equivalent to adding a prior-bias to the model, as stated in Epstein et al. (2010). Since this added generality does not change the results, we assume that agents have no prior bias.

for all $\theta \in \Theta$.[5] The dynamics for belief updating in our model are *local*, in the sense that each individual only uses the beliefs of her immediate neighbors to form her opinions, ignores the structure of the network, and does not make any inferences about the beliefs of other individuals. The above dynamics for opinion formation, compared to the Bayesian case, impose a significantly smaller computational burden on the individuals. Moreover, individuals do not need to keep track of the identities of their neighbors and the exact information provided by them. They only need to know the "average belief" held in their neighborhood, given by the term $\sum_{j\in\mathcal{N}_i} a_{ij}\mu_{j,t}(\cdot)$. In the special case that the signals observed by an agent are uninformative (or equivalently, there are no signals) after time $t=0$, Eq. (3) reduces to the belief update model of DeGroot (1974), used by Golub and Jackson (2010).

When analyzing the asymptotic behavior of the beliefs, sometimes it is more convenient to use matrix notation. Define A to be a real $n \times n$ matrix which captures the social interaction of the agents as well as the weight that each agent assigns to her neighbors. More specifically, we let the ij element of the matrix A be a_{ij} when agent j is a neighbor of agent i, and zero otherwise. Thus, Eq. (3) can be rewritten as

$$\mu_{t+1}(\theta) = A\mu_t(\theta) + \operatorname{diag}\left(a_{11}\left[\frac{\ell_1(\omega_{1,t+1}|\theta)}{m_{1,t}(\omega_{1,t+1})} - 1\right], \ldots, a_{nn}\left[\frac{\ell_n(\omega_{n,t+1}|\theta)}{m_{n,t}(\omega_{n,t+1})} - 1\right]\right)\mu_t(\theta) \tag{4}$$

where $\mu_t(\cdot) = [\mu_{1,t}, \ldots, \mu_{n,t}]'(\cdot)$, and diag of a vector is a diagonal matrix which has the entries of the vector as its diagonal. In the special case that A is the identity matrix, our model reduces to the standard Bayesian case, in which the society consists of n Bayesian agents who do not have access to the beliefs of other members of the society, and only observe their own private signals.

4. Social learning

Given the model described above, we are interested in the evolution of opinions in the network, and whether this evolution can lead to learning in the long run. Learning may either signify uncovering the true parameter or learning to forecast future outcomes. These two notions of learning are distinct and may not occur simultaneously. We start this section by specifying the exact meaning of both types of learning. Throughout, we denote the true parameter by $\theta^* \in \Theta$ and the probability law generating signals $(\omega_1, \omega_2, \ldots) \in \Omega$ by $\mathbb{P}^* = \bigotimes_{t=1}^{\infty} \ell(\cdot|\theta^*)$.

Definition 1. The *k-step-ahead forecasts* of agent i are *eventually correct* on a path $\omega = (\omega_1, \omega_2, \ldots)$ if, along that path,

$$m_{i,t}^{(k)}(s_{i,1}, s_{i,2}, \ldots, s_{i,k}) \to \ell_i(s_{i,1}|\theta^*)\ell_i(s_{i,2}|\theta^*)\ldots\ell_i(s_{i,k}|\theta^*) \quad \text{as } t \to \infty$$

for all $(s_{i,1}, \ldots, s_{i,k}) \in S_i^k$. Moreover, we say the beliefs of agent i *weakly merge to the truth* on some path if, along that path, her k-step-ahead forecasts are eventually correct for all natural numbers k.

This notion of learning captures the ability of agents to correctly forecast events in the near future.[6] It is well-known that, under suitable assumptions, repeated applications of Bayes' rule lead to eventually correct forecasts with probability one. The key condition is absolute continuity of the true measure with respect to initial beliefs. In the presence of absolute continuity, the mere repetition of Bayes' rule eventually transforms the historical record into a near perfect guide for the future. However, predicting events in near future accurately is not the same as learning the underlying state of the world. In fact, depending on the signal structure of each agent, there might be an "identification problem" which can potentially prevent the agent from learning the true parameter θ^*. We define an alternative notion of asymptotic learning according to which agents uncover the underlying parameter:

Definition 2. Agent $i \in \mathcal{N}$ *asymptotically learns* the true parameter θ^* on a path $\omega = (\omega_1, \omega_2, \ldots)$ if, along that path,

$$\mu_{i,t}(\theta^*) \to 1 \quad \text{as } t \to \infty.$$

Asymptotic learning occurs when the agent assigns probability one to the true parameter. As mentioned earlier, making correct forecasts about future events does not necessarily guarantee learning the true state. In general, the converse is not true either.[7] However, it is straightforward to show that in our model, asymptotically learning θ^* implies eventually correct forecasts.

[5] To simplify notation and where no confusion arises, we denote the one-step-ahead forecasts of agent i by $m_{i,t}(\cdot)$ rather than $m_{i,t}^{(1)}(\cdot)$.

[6] Note that our notion of weak merging is weaker than what Kalai and Lehrer (1994) refer to as weak merging of opinions. Whereas Definition 1 only requires that k-step-ahead forecasts of agent i are eventually correct for all $k < \infty$, Kalai and Lehrer's notion of weak merging of opinions requires that she assign asymptotically correct probabilities to all events that are measurable with respect to $\mathcal{F}_{i,t}$ for all $t < \infty$.

[7] For an example, see Lehrer and Smorodinsky (1996).

4.1. Correct one-step-ahead forecasts

We now turn to the main question of this paper: under what circumstances does learning occur over the social network?

Our first result shows that under mild assumptions, in spite of local interactions, limited observability, and the non-Bayesian belief update, agents will eventually have correct one-step-ahead forecasts. The proof is provided in Appendix A.

Proposition 1. *Suppose that the social network is strongly connected, all agents have strictly positive self-reliances, and there exists an agent with positive prior belief on the true parameter θ^*. Then, the one-step-ahead forecasts of all agents are eventually correct with $\mathbb{P}^*$-probability one.*

This proposition states that, when agents use non-Bayesian updating rule (3) to form and update their opinions, they will eventually make accurate predictions about the realization of their private signals in the next period. Note that as long as the social network remains strongly connected, neither the topology of the network nor the influence levels of different individuals prevent agents from making correct predictions.

One of the features of Proposition 1 is the absence of absolute continuity of the true measure with respect to the prior beliefs of all agents in the society: as long as some agent assigns a positive prior belief to the true parameter θ^* all agents will eventually be able to correctly predict the next period realizations of their private signals in the sense of Definition 1. In fact, eventually correct one-step-ahead forecasts arise even if the only agent for whom absolute continuity holds is located at the fringe of the society, has very small persuasive power over her neighbors, and almost everyone in the network is unaware of her existence. The main reason for this phenomenon is that, due to the naïve part of the updating, there is a "contagion" of beliefs to *all agents*, which eventually leads to absolute continuity of their beliefs with respect to the truth.

Besides the existence of an agent with a positive prior belief on the true state, the above proposition requires the existence of positive self-reliances to guarantee correct forecasts. This requirement is intuitive: it prohibits agents from completely discarding information provided to them through their observations. Clearly, if all agents discard their private signals, no new information is incorporated into their opinions, and (3) simply turns into a diffusion of prior beliefs.

The final requirement for accurate predictions is strong connectivity of the social network. The following example illustrates that this assumption cannot be disposed of.

Example 1. Consider a society consisting of two agents, $\mathcal{N} = \{1, 2\}$, and assume that $\Theta = \{\theta_1, \theta_2\}$ with the true state being $\theta^* = \theta_1$. Both agents have non-degenerate prior beliefs over Θ. Assume that signals observed by the agents are conditionally independent, and belong to the set $S_1 = S_2 = \{H, T\}$. We further assume that Agent 2's signals are non-informative, while Agent 1's observations are perfectly informative about the state; that is, $\ell_1(H|\theta_1) = \ell_1(T|\theta_2) = 1$, and $\ell_2(s|\theta_1) = \ell_2(s|\theta_2)$ for $s \in \{H, T\}$. As for the social structure, we assume that Agent 1 has access to the opinion of Agent 2, while Agent 2 cannot observe the opinion of Agent 1. Clearly, the social network is not strongly connected. We let the social interaction matrix be

$$A = \begin{bmatrix} 1-\alpha & \alpha \\ 0 & 1 \end{bmatrix},$$

where $\alpha \in (0, 1)$ is the weight that Agent 1 assigns to the opinion of Agent 2, when updating her beliefs using Eq. (3). Since the private signals observed by the latter are non-informative, her beliefs, at all times, remain equal to her prior. Clearly, she makes correct forecasts at all times. Agent 1's forecasts, on the other hand, will always remain incorrect. Notice that since her signals are perfectly informative, Agent 1's one-step-ahead predictions are eventually correct if and only if she eventually assigns probability 1 to the true state, θ_1. However, the belief she assigns to θ_2 follows the law of motion

$$\mu_{1,t+1}(\theta_2) = (1-\alpha)\mu_{1,t}(\theta_2)\frac{\ell_1(\omega_{1,t+1}|\theta_2)}{m_{1,t}(\omega_{1,t+1})} + \alpha\mu_{2,t}(\theta_2)$$

which cannot converge to zero, as $\mu_{2,t}(\theta_2) = \mu_{2,0}(\theta_2)$ is strictly positive.

The intuition for failure of learning in this example is simple. Given the same observations, the two agents make different interpretations about the state, even if they have equal prior beliefs. Moreover, Agent 1 follows the beliefs of the less informed Agent 2 but is unable to influence her back. This one-way persuasion and non-identical interpretations of signals (due to non-identical signal structures) result in incorrect one-step-ahead forecasts on the part of Agent 1. Finally, note that had Agent 1 discarded the opinions of Agent 2 and updated her beliefs according to Bayes' rule, she would have learned the truth after a single observation.

4.2. Weak merging to the truth

The key implication of Proposition 1 is that as long as the social network is strongly connected, the one-step-ahead forecasts of all agents will eventually be correct. Our next result establishes that not only agents make accurate predictions about their private observations in the next period, but also, under the same set of assumptions, make correct predictions about any finite time horizon in the future.

Proposition 2. *Suppose that the social network is strongly connected, all agents have strictly positive self-reliances, and there exists an agent with positive prior belief on the true parameter θ^*. Then, the beliefs of all agents weakly merge to the truth with $\mathbb{P}^*$-probability one.*

The above proposition states that in strongly connected societies, having eventually correct one-step-ahead forecasts is equivalent to weak merging of agents' opinions to the truth. This equivalence has already been established by Kalai and Lehrer (1994) for Bayesian agents. Note that in the purely Bayesian case, an agent's k-step-ahead forecasts are simply products of her one-step-ahead forecasts, making the equivalence between one-step-ahead correct forecasts and weak merging of opinions immediate. However, due to the non-Bayesian updating of the beliefs, the k-step-ahead forecasts of agents in our model do not have such a multiplicative decomposition, making this implication significantly less straightforward.

4.3. Social learning

Proposition 2 shows that in strongly connected social networks, the predictions of all individuals about the realizations of their signals in any finite time horizon will eventually be correct, implying that their asymptotic opinions cannot be arbitrary. The following proposition, which is our main result, establishes that strong connectivity of the social network not only leads to correct forecasts, but also guarantees successful aggregation of information: all individuals eventually learn the true state.

Proposition 3. *Suppose that*:

(a) *The social network is strongly connected.*
(b) *All agents have strictly positive self-reliances.*
(c) *There exists an agent with positive prior belief on the true parameter θ^*.*
(d) *There is no state $\theta \neq \theta^*$ that is observationally equivalent to θ^* from the point of view of all agents in the network.*

Then, all agents in the social network learn the true state of the world $\mathbb{P}^*$-almost surely; that is, $\mu_{i,t}(\theta^*) \to 1$ with $\mathbb{P}^*$-probability one for all $i \in \mathcal{N}$, as $t \to \infty$.

Proposition 3 states that under regularity assumptions on the social network's topology and the individuals' signal structures, all agents will eventually learn the true underlying state of the world. Notice that agents only interact with their neighbors and perform no deductions about the world beyond their immediate neighbors. Nonetheless, the non-Bayesian updating rule eventually enables them to obtain relevant information from others, without exactly knowing where it comes from. In fact, they can be completely oblivious to important features of the social network — such as the number of individuals in the society, the topology of the network, other people's signal structures, the existence of some agent who considers the truth plausible, or the influence level of any agent in the network — and still learn the parameter. Moreover, all these results are achieved with a significantly smaller computational burden than what is required for Bayesian updating.[8]

The other significant feature of Proposition 3 is that neither the topology of the network, the signal structures, nor the influence levels of different agents prevent learning. For instance, even if the agents with the least informative signals are the most persuasive ones and are located at the bottlenecks of the network, everyone will eventually learn the true state. Social learning is achieved despite the fact that the truth is not recognizable to any individual, and she would not have learned it by herself in isolation.

The intuition behind Proposition 3 is simple. Recall that Proposition 2 implies that the vector of beliefs of individual i, i.e., $\mu_{i,t}(\cdot)$, cannot vary arbitrarily forever, and instead, will eventually be restricted to the subspace that guarantees correct forecasts. Asymptotic convergence to such a subspace requires that she assigns an asymptotic belief of zero to any state θ which is not observationally equivalent to the truth from her point of view; otherwise, she would not be able to form correct forecasts about future realizations of her signals. However, due to the non-Bayesian part of the update corresponding to the social interactions of agent i with her neighbors, an asymptotic belief of zero is possible only if all her neighbors also consider θ asymptotically unlikely. This means that the information available to agent i must be eventually incorporated into every other individuals' beliefs.

The role of the assumptions of Proposition 3 can be summarized as follows: the strong connectivity assumption creates the possibility of information flow between any pair of agents in the social network. The assumption on positive self-reliances guarantees that agents do not discard the information provided to them through their private observations. The third assumption states that some agent assigns a positive prior belief to the truth. This agent may be at the fringe of the society, may have a very small influence on her neighbors, and almost no one may be aware of her existence. Hence, the ultimate source of learning may remain unknown to almost everyone. Clearly, if the prior beliefs of all agents assigned to the truth is equal to zero, then they will never learn.

[8] As we show in the next section, the asymptotic beliefs of all individuals do indeed coincide with those of a set of agents who make full Bayesian deductions.

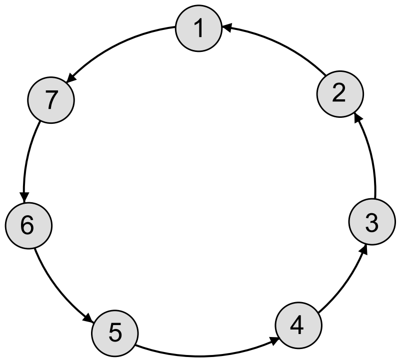

Fig. 1. A strongly connected social network of 7 agents.

The last assumption indicates that the collection of observations of all agents is informative enough about the true state; that is, $\Theta_1^* \cap \cdots \cap \Theta_n^* = \{\theta^*\}$, where Θ_i^* is the set of states that are observationally equivalent to θ^* from the point of view of agent i.[9] This assumption guarantees that it is possible to learn the truth if one has access to the observations of all agents. In the absence of this assumption, even highly sophisticated Bayesian agents with access to all relevant information (such as the topology of the network and the signal structures), would not be able to completely learn the state, due to an identification problem. Finally, note that when agents have identical signal structures (and therefore, $\Theta_i^* = \Theta_j^*$ for all i and j), they do not benefit from the information provided by their neighbors as they would be able to asymptotically learn just as much through their private observations.

The next examples show the power and limitations of Proposition 3.

Example 2. Consider the collection of agents $\mathcal{N} = \{1, 2, \dots, 7\}$ who are located in a social network as depicted in Fig. 1: at every time period, agent $i \leqslant 6$ can observe the opinion of agent $i+1$ and agent 7 has access to the opinion held by agent 1. Clearly, this is a strongly connected social network.

Assume that the set of possible states of the world is given by $\Theta = \{\theta^*, \theta_1, \theta_2, \dots, \theta_7\}$, where θ^* is the true underlying state of the world. We also assume that the signals observed by the agents belong to the set $S_i = \{H, T\}$ for all i, are conditionally independent, and have conditional distributions given by

$$\ell_i(H|\theta) = \begin{cases} \frac{i}{i+1} & \text{if } \theta = \theta_i \\ \frac{1}{(i+1)^2} & \text{otherwise} \end{cases}$$

for all $i \in \mathcal{N}$.

The signal structures are such that each agent suffers from some identification problem; i.e., the information in the observations of any agent is not sufficient for learning the true state of the world in isolation. More precisely, $\Theta_i^* = \Theta/\{\theta_i\}$ for all i, which means that from the point of view of agent i, all states except for θ_i are observationally equivalent to the true state θ^*. Nevertheless, for any given state $\theta \neq \theta^*$, there exists an agent whose signals are informative enough to distinguish the two; that is, $\bigcap_{i=1}^7 \Theta_i^* = \{\theta^*\}$. Therefore, Proposition 3 implies that as long as one agent assigns a positive prior belief on the true state θ^* and all agents have strictly positive self-reliances, then $\mu_{i,t}(\theta^*) \to 1$, as $t \to \infty$ for all agents i, with $\mathbb{P}^*$-probability one. In other words, all agents will asymptotically learn the true underlying state of the world. Clearly, if agents discard the information provided to them by their neighbors, they have no means of learning the true state.

Example 3. Consider a strongly connected social network consisting of two individuals $\mathcal{N} = \{1, 2\}$. Assume that $\Theta = \{\theta_1, \theta_2\}$, and $S_1 = S_2 = \{H, T\}$. Also assume that the distribution function describing the random private observations of the agents conditional on the underlying state of the world is given by the following tables:

$\ell(s_1 s_2|\theta_1)$:

	H	T
H	1/2	0
T	0	1/2

$\ell(s_1 s_2|\theta_2)$:

	H	T
H	0	1/2
T	1/2	0

In other words, under state θ_1, the private observations of the two agents are perfectly correlated, while when the underlying state of the world is θ_2, their observations are perfectly negatively correlated. Even though the joint distributions of the signals generated by θ_1 and θ_2 are different, we have $\ell_i(H|\theta_1) = \ell_i(H|\theta_2) = \frac{1}{2}$ for $i = 1, 2$; i.e., the local signal structure of each agent is the same under either state. As a result, despite the fact that agents will eventually agree on their opinions and make correct forecasts, they do not uncover the underlying parameter, as θ_1 and θ_2 are observationally equivalent from the point of view of both agents. That is, in this example, assumption (d) of Proposition 3 does not hold.

[9] This is a stronger restriction than requiring $\ell(\cdot|\theta) \neq \ell(\cdot|\theta^*)$ for all $\theta \neq \theta^*$. See also Example 3.

Finally, we remark that Proposition 3 is silent on the rate at which information is aggregated. In particular, the rate of learning may depend on the fine details of the structure of the social network as well as agents' signal structures. In addition, while non-Bayesian agents learn the truth as if they have access to all relevant signals and update their beliefs according to Bayes' rule, this aggregation of information may not happen at the same rate that Bayesian agents learn.

5. Bayesian learning

In the Introduction, we claimed that with a constant flow of new information, Bayesian and non-Bayesian learning rules asymptotically coincide. In this section, we make this claim precise. In particular, we consider the problem of aggregation of information on a strongly connected social network when individuals incorporate the views of their neighbors in a fully Bayesian manner, and show that under assumptions similar to those of Proposition 3, Bayesian agents learn the true parameter. In the remainder of the paper, we assume that individuals have full information about the topology of the social network and the signal structures of all other agents.[10] Our results remain valid even if agents have prior beliefs about the network structure and the distributions from which others' signals are drawn.

Suppose that agents have a common prior belief ν_0 over the finite set of states of the world Θ. Let $(X, \mathcal{B}, \mathbb{Q})$ be a probability space, where $X = \Theta \times \Omega$ is the space containing the realizations of the underlying state of the world and the sequence of signal profiles over time, $\mathcal{B} = 2^{\Theta} \times \mathcal{F}$ is the product σ-algebra, and $\mathbb{Q}$ is the probability measure defined as

$$\mathbb{Q}(\theta, \omega) = \nu_0(\theta) \cdot \mathbb{P}^{\theta}(\omega).$$

We use $\mathbb{E}[\cdot]$ to denote the expectation operator associated with measure $\mathbb{Q}$.

Proposition 4. *Suppose that all agents incorporate their private signals and views of their neighbors in a fully Bayesian manner. Also suppose that*

(a) *The social network is strongly connected.*
(b) *The agents' common prior has full support over Θ.*
(c) *There are no two states that are observationally equivalent from the point of view of all agents.*

Then, all agents in the social network learn the realized state of the world with $\mathbb{Q}$-probability one.

Thus, Bayesian agents will eventually learn the true parameter as long as the network is strongly connected and no two states are observationally equivalent from the point of view of all agents. However, in contrast to Proposition 3, this result requires all agents to assign a non-zero prior belief on all states. Clearly, if a Bayesian agent considers a state θ to be impossible, then no new information would convince her that θ is the underlying state of the world. So, apart from the case where agents assign zero prior probability to the realized state of the world, our non-Bayesian learning procedure asymptotically coincides with Bayesian learning. Thus, even if agents incorporate the views of their neighbors using a DeGroot-style update, their beliefs are asymptotically as if they performed full Bayesian deductions. This result holds provided that agents take their private signals into account in a Bayesian manner and that new information continuously arrives over time.

6. Conclusions

In this paper, we develop a model of dynamic opinion formation in social networks, bridging the gap between Bayesian and DeGroot-style non-Bayesian models of social learning. Agents fail to incorporate the views of their neighbors in a fully Bayesian manner, and instead, use a local updating rule. More specifically, at every time period, the belief of each individual is a convex combination of her Bayesian posterior belief and her neighbors' expressed beliefs. Our results show that agents eventually make correct forecasts, as long as the social network is strongly connected. In addition, agents successfully aggregate all information over the entire social network: they eventually learn the true underlying state of the world as if they were completely informed of *all* signals *and* updated their beliefs according to Bayes' rule.

The main insight suggested by our results is that, with a constant flow of new information, the key condition for social learning is that individuals take their *personal* signals into account in a Bayesian way. Repeated interactions over the social network guarantee that the differences of opinions eventually vanish and learning is obtained. The aggregation of information is achieved even if individuals are unaware of important features of the environment. In particular, agents do not need to have any information (or form beliefs) about the structure of the social network nor the views or characteristics of most agents, as they only update their opinions *locally* and do not make any deductions about the world beyond their immediate neighbors. Moreover, the individuals do not need to know the signal structure of any other agent in the network, besides their own. Thus, individuals eventually achieve full learning through a simple local updating rule and avoid the highly complex computations that are essential for full Bayesian updating over the network.

[10] Similar results are shown by Rosenberg et al. (2009) and Mueller-Frank (forthcoming), albeit under different sets of assumptions.

Even though our results establish that asymptotic learning is achieved in all strongly connected social networks, the rate at which information is aggregated depends on the topology of the network as well as agents' signal structures. Relatedly, the fine details of the social network structure would also affect asymptotic learning if agents' influences vary over time. In particular, if the influences of individuals on their neighbors vanish over time, disagreements may persist even if the social network is strongly connected. Another feature of the model studied in this paper is the assumption that agents can communicate their beliefs with their neighbors; a potentially unrealistic assumption when the size of the state space is large. This leads to the open questions of whether there are more efficient modes of communication and whether asymptotic social learning can be sustained when agents communicate some sufficient statistics of their beliefs with one another. We intend to investigate these questions in future work.

Appendix A. Proofs

A.1. Two auxiliary lemmas

Before presenting the proofs of the results in the paper, we state and prove two lemmas, both of which are consequences of the martingale convergence theorem.

Lemma 1. *Let A denote the matrix of social interactions. The sequence $\sum_{i=1}^n v_i\mu_{i,t}(\theta^*)$ converges $\mathbb{P}^*$-almost surely as $t \to \infty$, where v is any non-negative left eigenvector of A corresponding to its unit eigenvalue.*

Proof. First, note that since A is stochastic, it always has at least one eigenvalue equal to 1.[11] Moreover, there exists a non-negative left eigenvector corresponding to this eigenvalue.[12] We denote such a vector by v.

Evaluate Eq. (4) at the true parameter θ^* and multiply both sides by v' from the left

$$v'\mu_{t+1}(\theta^*) = v'A\mu_t(\theta^*) + \sum_{i=1}^n v_i\mu_{i,t}(\theta^*)a_{ii}\left[\frac{\ell_i(\omega_{i,t+1}|\theta^*)}{m_{i,t}(\omega_{i,t+1})} - 1\right].$$

Thus,

$$\mathbb{E}^*\left[\sum_{i=1}^n v_i\mu_{i,t+1}(\theta^*)\big|\mathcal{F}_t\right] = \sum_{i=1}^n v_i\mu_{i,t}(\theta^*) + \sum_{i=1}^n v_i a_{ii}\mu_{i,t}(\theta^*)\mathbb{E}^*\left[\frac{\ell_i(\omega_{i,t+1}|\theta^*)}{m_{i,t}(\omega_{i,t+1})} - 1\big|\mathcal{F}_t\right], \tag{5}$$

where $\mathbb{E}^*$ denotes the expectation operator associated with measure $\mathbb{P}^*$. Since $f(x) = 1/x$ is a convex function, Jensen's inequality implies that

$$\mathbb{E}^*\left[\frac{\ell_i(\omega_{i,t+1}|\theta^*)}{m_{i,t}(\omega_{i,t+1})}\big|\mathcal{F}_t\right] \geqslant \left(\mathbb{E}^*\left[\frac{m_{i,t}(\omega_{i,t+1})}{\ell_i(\omega_{i,t+1}|\theta^*)}\big|\mathcal{F}_t\right]\right)^{-1} = 1,$$

and therefore,

$$\mathbb{E}^*\left[\sum_{i=1}^n v_i\mu_{i,t+1}(\theta^*)\big|\mathcal{F}_t\right] \geqslant \sum_{i=1}^n v_i\mu_{i,t}(\theta^*).$$

The last inequality is due to the fact that v is element-wise non-negative. As a result, $\sum_{i=1}^n v_i\mu_{i,t}(\theta^*)$ is a submartingale with respect to the filtration $\mathcal{F}_t$, which is also bounded above by $\|v\|_1$. Hence, it converges $\mathbb{P}^*$-almost surely. □

Lemma 2. *Suppose that there exists an agent i such that $\mu_{i,0}(\theta^*) > 0$. Also suppose that the social network is strongly connected. Then, the sequence $\sum_{i=1}^n v_i \log\mu_{i,t}(\theta^*)$ converges $\mathbb{P}^*$-almost surely as $t \to \infty$, where v is any non-negative left eigenvector of A corresponding to its unit eigenvalue.*

Proof. Similar to the proof of the previous lemma, we show that $\sum_{i=1}^n v_i \log\mu_{i,t}(\theta^*)$ is a bounded submartingale and invoke the martingale convergence theorem to obtain almost sure convergence.

By evaluating the law of motion at θ^*, taking log from both sides, and using the fact that the row sums of A are equal to one, we obtain

$$\log\mu_{i,t+1}(\theta^*) \geqslant a_{ii}\log\mu_{i,t}(\theta^*) + a_{ii}\log\left(\frac{\ell_i(\omega_{i,t+1}|\theta^*)}{m_{i,t}(\omega_{i,t+1})}\right) + \sum_{j\in\mathcal{N}_i} a_{ij}\log\mu_{j,t}(\theta^*),$$

11 A matrix is said to be *stochastic* if it is entry-wise non-negative and all its row sums are equal to one.

12 This is a consequence of the Perron–Frobenius theorem. For more on the properties of non-negative and stochastic matrices, see Berman and Plemmons (1979).

where we have used the concavity of the logarithm function. Note that since the social network is strongly connected, the existence of one agent with a positive prior on θ^* guarantees that after at most n periods all agents assign a strictly positive probability to the true parameter, which means that $\log \mu_{i,t}(\theta^*)$ is well-defined for large enough t and all i.

Our next step is to show that $\mathbb{E}^*[\log \frac{\ell_i(\omega_{i,t+1}|\theta^*)}{m_{i,t}(\omega_{i,t+1})}|\mathcal{F}_t] \geqslant 0$. To obtain this, note that

$$\begin{aligned}\mathbb{E}^*\left[\log \frac{\ell_i(\omega_{i,t+1}|\theta^*)}{m_{i,t}(\omega_{i,t+1})}\middle|\mathcal{F}_t\right] &= -\mathbb{E}^*\left[\log \frac{m_{i,t}(\omega_{i,t+1})}{\ell_i(\omega_{i,t+1}|\theta^*)}\middle|\mathcal{F}_t\right]\\ &\geqslant -\log\left(\mathbb{E}^*\left[\frac{m_{i,t}(\omega_{i,t+1})}{\ell_i(\omega_{i,t+1}|\theta^*)}\middle|\mathcal{F}_t\right]\right)\\ &= 0.\end{aligned}$$

Thus,

$$\mathbb{E}^*\left[\log \mu_{i,t+1}\left(\theta^*\right)\middle|\mathcal{F}_t\right] \geqslant a_{ii} \log \mu_{i,t}\left(\theta^*\right) + \sum_{j\in\mathcal{N}_i} a_{ij} \log \mu_{j,t}\left(\theta^*\right),$$

which can be rewritten in matrix form as $\mathbb{E}^*[\log \mu_{t+1}(\theta^*)|\mathcal{F}_t] \geqslant A \log \mu_t(\theta^*)$, where by the logarithm of a vector, we mean its entry-wise logarithm. Multiplying both sides by A's non-negative left eigenvector v' leads to

$$\mathbb{E}^*\left[\sum_{i=1}^n v_i \log \mu_{i,t+1}\left(\theta^*\right)\middle|\mathcal{F}_t\right] \geqslant \sum_{i=1}^n v_i \log \mu_{i,t}\left(\theta^*\right).$$

Thus, the non-positive sequence $\sum_{i=1}^n v_i \log \mu_{i,t}(\theta^*)$ is a submartingale with respect to filtration $\mathcal{F}_t$, and therefore, converges with $\mathbb{P}^*$-probability one. $\square$

With these lemmas in hand, we can prove Proposition 1.

A.2. Proof of Proposition 1

First, note that since the social network is strongly connected, the social interaction matrix A is an irreducible stochastic matrix, and therefore its left eigenvector corresponding to the unit eigenvalue is strictly positive.[13]

According to Lemma 1, $\sum_{i=1}^n v_i \mu_{i,t}(\theta^*)$ converges with $\mathbb{P}^*$-probability one, where v is the positive left eigenvector of A corresponding to its unit eigenvalue. Thus, evaluating Eq. (4) at θ^* and multiplying both sides by v' from the left imply

$$\sum_{i=1}^n v_i a_{ii} \mu_{i,t}\left(\theta^*\right)\left(\frac{\ell_i(\omega_{i,t+1}|\theta^*)}{m_{i,t}(\omega_{i,t+1})} - 1\right) \to 0 \quad \mathbb{P}^*\text{-a.s.}$$

and therefore, by the dominated convergence theorem for conditional expectations,[14]

$$\sum_{i=1}^n v_i a_{ii} \mu_{i,t}\left(\theta^*\right)\left(\mathbb{E}^*\left[\frac{\ell_i(\omega_{i,t+1}|\theta^*)}{m_{i,t}(\omega_{i,t+1})}\middle|\mathcal{F}_t\right] - 1\right) \to 0 \quad \mathbb{P}^*\text{-a.s.}$$

Since the term $v_i a_{ii} \mu_{i,t}(\theta^*)\mathbb{E}^*[\ell_i(\omega_{i,t+1}|\theta^*)/m_{i,t}(\omega_{i,t+1}) - 1|\mathcal{F}_t]$ is non-negative for all i, each such term converges to zero with $\mathbb{P}^*$-probability one. Moreover, the assumptions that all diagonal entries of A are strictly positive and that of its irreducibility (which means that v is entry-wise positive) lead to

$$\mu_{i,t}\left(\theta^*\right)\left(\mathbb{E}^*\left[\frac{\ell_i(\omega_{i,t+1}|\theta^*)}{m_{i,t}(\omega_{i,t+1})}\middle|\mathcal{F}_t\right] - 1\right) \to 0 \quad \text{for all } i \quad \mathbb{P}^*\text{-a.s.} \tag{6}$$

Furthermore, Lemma 2 guarantees that $\sum_{i=1}^n v_i \log \mu_{i,t}(\theta^*)$ converges almost surely, implying that $\mu_{i,t}(\theta^*)$ converges to a strictly positive number with probability one for all i. Note that, once again we are using the fact that v is a strictly positive vector. Hence, $\mathbb{E}^*[\frac{\ell_i(\omega_{i,t+1}|\theta^*)}{m_{i,t}(\omega_{i,t+1})}|\mathcal{F}_t] \to 1$ almost surely. Thus,

[13] An $n \times n$ matrix A is said to be *reducible*, if for some permutation matrix P, the matrix $P'AP$ is block upper triangular. If a square matrix is not reducible, it is said to be *irreducible*. For more on this, see e.g., Berman and Plemmons (1979).

[14] For the statement and a proof of the theorem, see page 262 of Durrett (2005).

$$\mathbb{E}^*\left[\frac{\ell_i(\omega_{i,t+1}|\theta^*)}{m_{i,t}(\omega_{i,t+1})}\middle|\mathcal{F}_t\right]-1=\sum_{s_i\in S_i}\ell_i\big(s_i|\theta^*\big)\left(\frac{\ell_i(s_i|\theta^*)}{m_{i,t}(s_i)}-1\right)$$
$$=\sum_{s_i\in S_i}\left(\ell_i\big(s_i|\theta^*\big)\frac{\ell_i(s_i|\theta^*)-m_{i,t}(s_i)}{m_{i,t}(s_i)}+m_{i,t}(s_i)-\ell_i\big(s_i|\theta^*\big)\right)$$
$$=\sum_{s_i\in S_i}\frac{[\ell_i(s_i|\theta^*)-m_{i,t}(s_i)]^2}{m_{i,t}(s_i)}\to 0\quad \mathbb{P}^*\text{-a.s.},$$

where the second equality is due to the fact that both $\ell_i(\cdot|\theta^*)$ and $m_{i,t}(\cdot)$ are probability measures on S_i, and therefore, $\sum_{s_i\in S_i}\ell_i(s_i|\theta^*)=\sum_{s_i\in S_i}m_{i,t}(s_i)=1$.

In the last expression, the term in the brackets and the denominator are always non-negative and therefore,

$$m_{i,t}(s_i)\to\ell_i\big(s_i|\theta^*\big)\quad \mathbb{P}^*\text{-a.s.}$$

for all $s_i\in S_i$ and all $i\in\mathcal{N}$. □

A.3. Proof of Proposition 2

We first present and prove a simple lemma which is later used in the proof of the proposition.

Lemma 3. *Suppose that the social network is strongly connected, all agents have strictly positive self-reliances, and there exists an agent i such that $\mu_{i,t}(\theta^*)>0$. Then, for all $\theta\in\Theta$,*

$$\mathbb{E}^*\big[\mu_{t+1}(\theta)|\mathcal{F}_t\big]-A\mu_t(\theta)\to 0$$

with $\mathbb{P}^*$-probability one, as $t\to\infty$.

Proof. Taking conditional expectations from both sides of Eq. (4) implies

$$\mathbb{E}^*\big[\mu_{t+1}(\theta)|\mathcal{F}_t\big]-A\mu_t(\theta)=\mathrm{diag}\left(a_{11}\mathbb{E}^*\left[\frac{\ell_1(\omega_{1,t+1}|\theta)}{m_{1,t}(\omega_{1,t+1})}-1\middle|\mathcal{F}_t\right],\ldots,a_{nn}\mathbb{E}^*\left[\frac{\ell_n(\omega_{n,t+1}|\theta)}{m_{n,t}(\omega_{n,t+1})}-1\middle|\mathcal{F}_t\right]\right)\mu_t(\theta).$$

On the other hand, we have

$$\mathbb{E}^*\left[\frac{\ell_i(\omega_{i,t+1}|\theta)}{m_{i,t}(\omega_{i,t+1})}\middle|\mathcal{F}_t\right]=\sum_{s_i\in S_i}\ell_i(s_i|\theta)\frac{\ell_i(s_i|\theta^*)}{m_{i,t}(s_i)}\to\sum_{s_i\in S_i}\ell_i(s_i|\theta)\quad \mathbb{P}^*\text{-a.s.}$$

where the convergence is a consequence of Proposition 1. The fact that $\ell_i(\cdot|\theta)$ is a probability measure on S_i implies $\sum_{s_i\in S_i}\ell_i(s_i|\theta)=1$, completing the proof. □

We now present the proof of Proposition 2.

Proof of Proposition 2.[15] We prove this proposition by induction. Note that by definition, agent i's beliefs weakly merge to the truth if her k-step-ahead forecasts are eventually correct for all natural numbers k. In Proposition 1, we established that the claim is true for $k=1$. For the rest of the proof, we assume that the claim is true for $k-1$ and show that $m_{i,t}^{(k)}(s_{i,1},\ldots,s_{i,k})$ converges to $\prod_{r=1}^{k}\ell_i(s_{i,r}|\theta^*)$ for any arbitrary sequence of signals $(s_{i,1},\ldots,s_{i,k})\in S_i^k$.[16]

First, note that Lemma 3 and Eq. (3) imply that for all $\theta\in\Theta$,

$$\mathbb{E}^*\big[\mu_{i,t+1}(\theta)|\mathcal{F}_t\big]-\mu_{i,t+1}(\theta)+a_{ii}\left[\frac{\ell_i(\omega_{i,t+1}|\theta)}{m_{i,t}(\omega_{i,t+1})}-1\right]\mu_{i,t}(\theta)\to 0\quad \mathbb{P}^*\text{-a.s.}$$

Multiplying both sides by $\prod_{r=2}^{k}\ell_i(s_{i,r}|\theta)$ for an arbitrary sequence of signals $(s_{i,2},\ldots,s_{i,k})\in S_i^{k-1}$ and summing up over all $\theta\in\Theta$ lead to

$$\sum_{\theta\in\Theta}\left(\prod_{r=2}^{k}\ell_i(s_{i,r}|\theta)\right)\big(\mathbb{E}^*\big[\mu_{i,t+1}(\theta)|\mathcal{F}_t\big]-\mu_{i,t+1}(\theta)\big)+a_{ii}\sum_{\theta\in\Theta}\left(\prod_{r=2}^{k}\ell_i(s_{i,r}|\theta)\right)\left[\frac{\ell_i(\omega_{i,t+1}|\theta)}{m_{i,t}(\omega_{i,t+1})}-1\right]\mu_{i,t}(\theta)\to 0$$

[15] We would like to thank a referee for bringing a technical mistake in an earlier draft of the paper to our attention.

[16] Recall that we use $s_i\in S_i$ to denote a generic element of the signal space of agent i, whereas $\omega_{i,t}$ denotes the random variable corresponding to i's observation at period t.

with $\mathbb{P}^*$-probability one. On the other hand,

$$\sum_{\theta\in\Theta}\left(\prod_{r=2}^{k}\ell_i(s_{i,r}|\theta)\right)\left(\mathbb{E}^*\left[\mu_{i,t+1}(\theta)|\mathcal{F}_t\right]-\mu_{i,t+1}(\theta)\right)=\mathbb{E}^*\left[m_{i,t+1}^{(k-1)}(s_{i,2},\dots,s_{i,k})|\mathcal{F}_t\right]-m_{i,t+1}^{(k-1)}(s_{i,2},\dots,s_{i,k}),$$

where we have used the definition of $(k-1)$-step-ahead forecasts of agent i. The induction hypothesis and the dominated convergence theorem for conditional expectations imply that the right-hand side of the above equation converges to zero with $\mathbb{P}^*$-probability one. Therefore,

$$\sum_{\theta\in\Theta}\left(\prod_{r=2}^{k}\ell_i(s_{i,r}|\theta)\right)\left[\frac{\ell_i(\omega_{i,t+1}|\theta)}{m_{i,t}(\omega_{i,t+1})}-1\right]\mu_{i,t}(\theta)\to 0\quad \mathbb{P}^*\text{-a.s.}$$

for any arbitrary sequence of signals $(s_{i,2},\dots,s_{i,k})\in S_i^{k-1}$, which is equivalent to

$$\frac{1}{m_{i,t}(\omega_{i,t+1})}m_{i,t}^{(k)}(\omega_{i,t+1},s_{i,2},\dots,s_{i,k})-m_{i,t}^{(k-1)}(s_{i,2},\dots,s_{i,k})\to 0\quad \mathbb{P}^*\text{-a.s.}$$

Thus, once again by the induction hypothesis,

$$m_{i,t}^{(k)}(\omega_{i,t+1},s_{i,2},\dots,s_{i,k})-m_{i,t}(\omega_{i,t+1})\prod_{r=2}^{k}\ell_i\left(s_{i,r}|\theta^*\right)\to 0$$

with $\mathbb{P}^*$-probability one. The dominated convergence theorem for conditional expectations implies

$$\mathbb{E}^*\left[\left|m_{i,t}^{(k)}(\omega_{i,t+1},s_{i,2},\dots,s_{i,k})-m_{i,t}(\omega_{i,t+1})\prod_{r=2}^{k}\ell_i\left(s_{i,r}|\theta^*\right)\right|\middle|\mathcal{F}_t\right]\to 0\quad \mathbb{P}^*\text{-a.s.}$$

Rewriting the conditional expectation operator as a sum over all possible realizations of $\omega_{i,t+1}$ leads to

$$\sum_{\tilde{s}_i\in S_i}\ell_i\left(\tilde{s}_i|\theta^*\right)\left|m_{i,t}^{(k)}(\tilde{s}_i,s_{i,2},\dots,s_{i,k})-m_{i,t}(\tilde{s}_i)\prod_{r=2}^{k}\ell_i\left(s_{i,r}|\theta^*\right)\right|\to 0$$

$\mathbb{P}^*$-almost surely, and therefore, guaranteeing

$$m_{i,t}^{(k)}(s_{i,1},s_{i,2},\dots,s_{i,k})-m_{i,t}(s_{i,1})\prod_{r=2}^{k}\ell_i\left(s_{i,r}|\theta^*\right)\to 0\quad \mathbb{P}^*\text{-a.s.}$$

for all $s_{i,1}\in S_i$.[17] Finally, the fact that $m_{i,t}(s_{i,1})\to\ell_i(s_{i,1}|\theta^*)$ with $\mathbb{P}^*$-probability one (Proposition 1) completes the proof. □

A.4. Proof of Proposition 3

We first show that for any agent i, there exists a finite sequence of private signals that is more likely to realize under the true state θ^* than any other state θ, unless θ is observationally equivalent to θ^* from the point of view of agent i.

Lemma 4. *For any agent i, there exists a positive integer $\hat{k}_i$, a sequence of signals $(\hat{s}_{i,1},\dots,\hat{s}_{i,\hat{k}_i})\in(S_i)^{\hat{k}_i}$, and constant $\delta_i\in(0,1)$ such that*

$$\prod_{r=1}^{\hat{k}_i}\frac{\ell_i(\hat{s}_{i,r}|\theta)}{\ell_i(\hat{s}_{i,r}|\theta^*)}\leqslant\delta_i\quad\forall\theta\notin\Theta_i^*\tag{7}$$

where $\Theta_i^*=\{\theta\in\Theta\colon \ell_i(s_i|\theta)=\ell_i(s_i|\theta^*)$ for all $s_i\in S_i\}$.

Proof. By definition, for any $\theta\notin\Theta_i^*$, the probability measures $\ell_i(\cdot|\theta)$ and $\ell_i(\cdot|\theta^*)$ are distinct. Therefore, by the Kullback–Leibler inequality, there exists some constant $\epsilon_i>0$ such that

$$\sum_{s_i\in S_i}\ell_i\left(s_i|\theta^*\right)\log\left[\frac{\ell_i(s_i|\theta^*)}{\ell_i(s_i|\theta)}\right]>\epsilon_i,$$

[17] Recall that, by assumption, $\ell_i(s_i|\theta)>0$ for all $(s_i,\theta)\in S_i\times\Theta$ and all i.

for all $\theta \notin \Theta_i^*$, which then implies

$$\prod_{s_i \in S_i} \left[\frac{\ell_i(s_i|\theta)}{\ell_i(s_i|\theta^*)} \right]^{\ell_i(s_i|\theta^*)} < \delta_i',$$

for $\delta_i' = \exp(-\epsilon_i)$. On the other hand, given the fact that rational numbers are dense on the real line, there exist strictly positive rational numbers $\{q(s_i)\}_{s_i \in S_i}$ — with $q(s_i)$ chosen arbitrarily close to $\ell_i(s_i|\theta^*)$ — satisfying $\sum_{s_i \in S_i} q(s_i) = 1$, such that[18]

$$\prod_{s_i \in S_i} \left[\frac{\ell_i(s_i|\theta)}{\ell_i(s_i|\theta^*)} \right]^{q(s_i)} < \delta_i' \quad \forall \theta \notin \Theta_i^*. \tag{8}$$

Therefore, the above inequality can be rewritten as

$$\prod_{s_i \in S_i} \left[\frac{\ell_i(s_i|\theta)}{\ell_i(s_i|\theta^*)} \right]^{k(s_i)} < \left(\delta_i'\right)^{\hat{k}_i} \quad \forall \theta \notin \Theta_i^*$$

for some positive integers $k(s_i)$ and $\hat{k}_i$, satisfying $\hat{k}_i = \sum_{s_i \in S_i} k(s_i)$. Picking the sequence of signals of length $\hat{k}_i$, $(\hat{s}_{i,1}, \ldots, \hat{s}_{i,\hat{k}_i})$, such that s_i appears $k(s_i)$ many times in the sequence and setting $\delta_i = (\delta_i')^{\hat{k}_i}$ proves the lemma. □

The above lemma shows that the sequence of private signals in which any signal $s_i \in S_i$ appears with a frequency close enough to $\ell_i(s_i|\theta^*)$ is more likely under the truth θ^* than any other state θ which is distinguishable from θ^*. We now proceed to the proof of Proposition 3.

Proof of Proposition 3. First, we prove that agent i assigns an asymptotic belief of zero to states that are not observationally equivalent to θ^* from her point of view.

Recall that according to Proposition 2, the k-step-ahead forecasts of agent i are eventually correct for all positive integers k, guaranteeing that $m_{i,t}^{(k)}(s_{i,1}, \ldots, s_{i,k}) \to \prod_{r=1}^{k} \ell_i(s_{i,k}|\theta^*)$ with $\mathbb{P}^*$-probability one for any sequence of signals $(s_{i,1}, \ldots, s_{i,k})$. In particular, the claim is true for the integer $\hat{k}_i$ and the sequence of signals $(\hat{s}_{i,1}, \ldots, \hat{s}_{i,\hat{k}_i})$ satisfying (7) in Lemma 4:

$$\sum_{\theta \in \Theta} \mu_{i,t}(\theta) \prod_{r=1}^{\hat{k}_i} \frac{\ell_i(\hat{s}_{i,r}|\theta)}{\ell_i(\hat{s}_{i,r}|\theta^*)} \to 1 \quad \mathbb{P}^*\text{-a.s.}$$

Therefore,

$$\sum_{\theta \notin \Theta_i^*} \mu_{i,t}(\theta) \prod_{r=1}^{\hat{k}_i} \frac{\ell_i(\hat{s}_{i,r}|\theta)}{\ell_i(\hat{s}_{i,r}|\theta^*)} + \sum_{\theta \in \Theta_i^*} \mu_{i,t}(\theta) - 1 \to 0 \quad \mathbb{P}^*\text{-a.s.}$$

leading to

$$\sum_{\theta \notin \Theta_i^*} \mu_{i,t}(\theta) \left(1 - \prod_{r=1}^{\hat{k}_i} \frac{\ell_i(\hat{s}_{i,r}|\theta)}{\ell_i(\hat{s}_{i,r}|\theta^*)} \right) \to 0$$

with $\mathbb{P}^*$-probability one. The fact that $\hat{k}_i$ and $(\hat{s}_{i,1}, \ldots, \hat{s}_{i,\hat{k}_i})$ were chosen to satisfy (7) implies that

$$1 - \prod_{r=1}^{\hat{k}_i} \frac{\ell_i(\hat{s}_{i,r}|\theta)}{\ell_i(\hat{s}_{i,r}|\theta^*)} > 1 - \delta_i > 0 \quad \forall \theta \notin \Theta_i^*,$$

and as a consequence, it must be the case that $\mu_{i,t}(\theta) \to 0$ as $t \to \infty$ for any $\theta \notin \Theta_i^*$. Therefore, with $\mathbb{P}^*$-probability one, agent i assigns an asymptotic belief of zero to any state θ that is not observationally equivalent to θ^* from her point of view.

[18] The fact that the rationals form a dense subset of the reals means that there are rational numbers $\{q'(s_i)\}_{s_i \in S_i}$ arbitrarily close to $\{\ell_i(s_i|\theta^*)\}_{s_i \in S_i}$. Setting $q(s_i) = q'(s_i)/\sum_{\tilde{s}_i \in S_i} q'(\tilde{s}_i)$ guarantees that one can always find strictly positive rational numbers $\{q(s_i)\}_{s_i \in S_i}$ adding up to one, while at the same time, (8) is satisfied.

Now consider the belief update rule for agent i given by Eq. (3), evaluated at some state $\theta \notin \Theta_i^*$:

$$\mu_{i,t+1}(\theta) = a_{ii}\mu_{i,t}(\theta)\frac{\ell_i(\omega_{i,t+1}|\theta)}{m_{i,t}(\omega_{i,t+1})} + \sum_{j\in\mathcal{N}_i} a_{ij}\mu_{j,t}(\theta).$$

We have already shown that $\mu_{i,t}(\theta) \to 0$, $\mathbb{P}^*$-almost surely. However, this is not possible unless $\sum_{j\in\mathcal{N}_i} a_{ij}\mu_{j,t}(\theta)$ converges to zero as well, which implies that $\mu_{j,t}(\theta) \to 0$ with $\mathbb{P}^*$-probability one for all $j \in \mathcal{N}_i$. Note that this happens even if θ is observationally equivalent to θ^* from the point of view of agent j; that is, even if $\theta \in \Theta_j^*$. As a result, all neighbors of agent i will assign an asymptotic belief of zero to parameter θ regardless of their signal structure. We can extend the same argument to the neighbors of neighbors of agent i, and by induction — since the social network is strongly connected — to all agents in the network. Thus, with $\mathbb{P}^*$-probability one,

$$\mu_{i,t}(\theta) \to 0 \quad \forall i \in \mathcal{N},\ \forall \theta \notin \Theta_1^* \cap \cdots \cap \Theta_n^*$$

implying that all agents assign an asymptotic belief of zero to states that are not observationally equivalent to θ^* from the point of view of all individuals in the society. Therefore, statement (d) in the assumptions of Proposition 3 implies that $\mu_{i,t}(\theta) \to 0$ for all $\theta \neq \theta^*$, with $\mathbb{P}^*$-probability one, guaranteeing complete learning by all agents. □

A.5. Proof of Proposition 4

We define $\mathcal{P}_{i,t}$ as the σ-field generated by the private signals observed by agent i up to period t,

$$\mathcal{P}_{i,t} = \sigma(\omega_{i,1}, \ldots, \omega_{i,t}),$$

and $\mathcal{P}_i$ as the smallest σ-field containing $\mathcal{P}_{i,t}$ for all t. We also define the posterior belief of agent i on Θ and the σ-field generated by all her observations up to a given time period — which consists of her private signals and the sequence of beliefs of her neighbors — recursively as follows:

$$\mathcal{I}_{i,t} = \sigma\big(\omega_{i,1}, \ldots, \omega_{i,t}, (\nu_{j,1}, \ldots, \nu_{j,t-1})_{j\in\mathcal{N}_i}\big),$$
$$\nu_{i,t}(\theta) = \mathbb{Q}(\theta|\mathcal{I}_{i,t}),$$

with $\nu_{i,0} = \nu_0$. We denote the smallest σ-field that contains $\mathcal{I}_{i,t}$ for all t by $\mathcal{I}_i$. Finally, we define $\mathcal{H}_{i,t}$ as the σ-field generated by the sequence of beliefs of agent i up to period t,

$$\mathcal{H}_{i,t} = \sigma(\nu_{i,1}, \ldots, \nu_{i,t}),$$

and $\mathcal{H}_i$ as the smallest σ-field containing $\mathcal{H}_{i,t}$ for all t.

By observing the sequence of her private signals, agent i can asymptotically distinguish between the states that are not observationally equivalent from her point of view[19]; that is, for any $\theta \in \Theta$,

$$\mathbb{E}[\mathbf{1}_{\Theta_i^\theta}|\mathcal{P}_{i,t}] \to \mathbf{1}_{\Theta_i^\theta} \quad \mathbb{Q}\text{-a.s.},$$

where $\mathbf{1}_{\Theta_i^\theta}$ is the indicator function of the event that the realized state is observationally equivalent to θ from the point of view of agent i. On the other hand, by the dominated convergence theorem for conditional expectations, $\mathbb{E}[\mathbf{1}_{\Theta_i^\theta}|\mathcal{P}_{i,t}]$ converges to $\mathbb{E}[\mathbf{1}_{\Theta_i^\theta}|\mathcal{P}_i]$ with $\mathbb{Q}$-probability one. Consequently,

$$\mathbb{E}[\mathbf{1}_{\Theta_i^\theta}|\mathcal{P}_i] = \mathbf{1}_{\Theta_i^\theta} \quad \mathbb{Q}\text{-a.s.},$$

which means that $\mathbf{1}_{\Theta_i^\theta}$ is measurable with respect to $\mathcal{P}_i$. The fact that $\mathcal{P}_i \subseteq \mathcal{I}_i$ guarantees that it is also measurable with respect to $\mathcal{I}_i$. Thus, for any $\theta \in \Theta$,

$$\nu_{i,t}\big(\Theta_i^\theta\big) = \mathbb{E}[\mathbf{1}_{\Theta_i^\theta}|\mathcal{I}_{i,t}] \to \mathbf{1}_{\Theta_i^\theta} \quad \mathbb{Q}\text{-a.s.} \tag{9}$$

In other words, agent i asymptotically rules out states that are not observationally equivalent to the underlying state of the world from her point of view.

Next, note that for any $\theta \in \Theta$,

$$\mathbb{E}[\mathbf{1}_{\Theta_i^\theta}|\mathcal{H}_{i,t}] = \mathbb{E}\big[\nu_{i,t}\big(\Theta_i^\theta\big)\big|\mathcal{H}_{i,t}\big] = \nu_{i,t}\big(\Theta_i^\theta\big),$$

[19] See, for example, Savage (1954).

where the first equality is a consequence of $\mathcal{H}_{i,t} \subseteq \mathcal{I}_{i,t}$,[20] and the second equality is due to the fact that $\nu_{i,t}$ is measurable with respect to $\mathcal{H}_{i,t}$. Therefore, by (9),

$$\mathbb{E}[\mathbf{1}_{\Theta_i^\theta} | \mathcal{H}_{i,t}] \to \mathbf{1}_{\Theta_i^\theta} \quad \mathbb{Q}\text{-a.s.}$$

On the other hand, by the dominated convergence theorem for conditional expectations, $\mathbb{E}[\mathbf{1}_{\Theta_i^\theta} | \mathcal{H}_{i,t}]$ converges to $\mathbb{E}[\mathbf{1}_{\Theta_i^\theta} | \mathcal{H}_i]$ with $\mathbb{Q}$-probability one, which implies that $\mathbf{1}_{\Theta_i^\theta}$ is measurable with respect to $\mathcal{H}_i$. Moreover, for any agent j such that $i \in \mathcal{N}_j$, we have $\mathcal{H}_i \subseteq \mathcal{I}_j$, which guarantees that $\mathbf{1}_{\Theta_i^\theta}$ is also measurable with respect to $\mathcal{I}_j$. Thus,

$$\nu_{j,t}\big(\Theta_i^\theta\big) = \mathbb{E}[\mathbf{1}_{\Theta_i^\theta} | \mathcal{I}_{j,t}] \to \mathbf{1}_{\Theta_i^\theta}$$

with $\mathbb{Q}$-probability one. Therefore, not only agent j would be able to eventually distinguish between any two states that are not observationally equivalent from her own point of view, but also between those that are not observationally equivalent from the points of view of any of her neighbors.

Finally, the fact that the social network is strongly connected guarantees that every individual would be able to eventually distinguish between any two states that are observationally equivalent from the point of view of some other agent. This observation coupled with the assumption that no two states are observationally equivalent from the point of view of all agents implies

$$\nu_{i,t}(\theta) \to \mathbf{1}_{\{\theta\}} \quad \mathbb{Q}\text{-a.s.},$$

completing the proof. □

References

Acemoglu, Daron, Dahleh, Munther, Lobel, Ilan, Ozdaglar, Asuman, 2011. Bayesian learning in social networks. The Review of Economic Studies 78, 1201–1236.

Acemoglu, Daron, Ozdaglar, Asuman, Parandeh-Gheibi, Ali, 2010. Spread of (mis)information in social networks. Games and Economic Behavior 70, 194–227.

Bala, Venkatesh, Goyal, Sanjeev, 1998. Learning from neighbors. The Review of Economic Studies 65, 595–621.

Bala, Venkatesh, Goyal, Sanjeev, 2001. Conformism and diversity under social learning. Economic Theory 17, 101–120.

Banerjee, Abhijit, 1992. A simple model of herd behavior. The Quarterly Journal of Economics 107, 797–817.

Banerjee, Abhijit, Fudenberg, Drew, 2004. Word-of-mouth learning. Games and Economic Behavior 46, 1–22.

Berman, Abraham, Plemmons, Robert J., 1979. Nonnegative Matrices in the Mathematical Sciences. Academic Press, New York.

Bikhchandani, Sushil, Hirshleifer, David, Welch, Ivo, 1992. A theory of fads, fashion, custom, and cultural change as information cascades. Journal of Political Economy 100, 992–1026.

DeGroot, Morris H., 1974. Reaching a consensus. Journal of American Statistical Association 69, 118–121.

DeMarzo, Peter M., Vayanos, Dimitri, Zwiebel, Jeffery, 2003. Persuasion bias, social influence, and unidimensional opinions. The Quarterly Journal of Economics 118, 909–968.

Durrett, Rick, 2005. Probability: Theory and Examples, third edition. Duxbury Press, Belmont, CA.

Ellison, Glenn, Fudenberg, Drew, 1993. Rules of thumb for social learning. The Journal of Political Economy 101, 612–643.

Ellison, Glenn, Fudenberg, Drew, 1995. Word-of-mouth communication and social learning. The Quarterly Journal of Economics 110, 93–126.

Epstein, Larry G., Noor, Jawwad, Sandroni, Alvaro, 2008. Non-Bayesian updating: a theoretical framework. Theoretical Economics 3, 193–229.

Epstein, Larry G., Noor, Jawwad, Sandroni, Alvaro, 2010. Non-Bayesian learning. The B.E. Journal of Theoretical Economics 10.

Gale, Douglas, Kariv, Shachar, 2003. Bayesian learning in social networks. Games and Economic Behavior 45, 329–346.

Golub, Benjamin, Jackson, Mathew O., 2010. Naïve learning in social networks and the wisdom of crowds. American Economic Journal: Microeconomics 2, 112–149.

Hagerstrand, Torsten, 1969. Innovation Diffusion as a Spatial Process. University of Chicago Press, Chicago, IL.

Kalai, Ehud, Lehrer, Ehud, 1994. Weak and strong merging of opinions. Journal of Mathematical Economics 23, 73–86.

Kotler, Philip, 1986. The Principles of Marketing, third edition. Prentice-Hall, New York, NY.

Lehrer, Ehud, Smorodinsky, Rann, 1996. Merging and learning. Statistics, Probability and Game Theory 30, 147–168.

Mossel, Elchanan, Tamuz, Omer, 2010. Efficient Bayesian learning in social networks with Gaussian estimators. Preprint, arXiv:1002.0747.

Mueller-Frank, Manuel, forthcoming. A general framework for rational learning in social networks. Theoretical Economics.

Rogers, Everett M., 1983. Diffusion of Innovations, third edition. Free Press, New York.

Rosenberg, Dinah, Solan, Eilon, Vieille, Nicolas, 2009. Informational externalities and emergence of consensus. Games and Economic Behavior 66, 979–994.

Savage, Leonard J., 1954. The Foundations of Statistics. Dover Reprint, New York. 1972.

Smith, Lones, Sørensen, Peter Norman, 2000. Pathological outcomes of observational learning. Econometrica 68, 371–398.

[20] Note that by definition, $\nu_{i,t} \in \mathcal{I}_{i,t}$, implying that $\mathcal{H}_{i,t} \subseteq \mathcal{I}_{i,t}$.