Bayesian Learning in Social Networks[*]

Daron Acemoglu[†], Munther A. Dahleh[‡], Ilan Lobel[§] and Asuman Ozdaglar[¶]

May 11, 2008

Abstract

We study the (perfect Bayesian) equilibrium of a model of learning over a general social network. Each individual receives a signal about the underlying state of the world, observes the past actions of a stochastically-generated neighborhood of individuals, and chooses one of two possible actions. The stochastic process generating the neighborhoods defines the network topology (social network). The special case where each individual observes all past actions has been widely studied in the literature. We characterize pure-strategy equilibria for arbitrary stochastic and deterministic social networks and characterize the conditions under which there will be *asymptotic learning*—that is, the conditions under which, as the social network becomes large, individuals converge (in probability) to taking the right action. We show that when private beliefs are unbounded (meaning that the implied likelihood ratios are unbounded), there will be asymptotic learning as long as there is some minimal amount of "expansion in observations". Our main theorem shows that when the probability that each individual observes some other individual from the recent past converges to one as the social network becomes large, unbounded private beliefs are sufficient to ensure asymptotic learning. This theorem therefore establishes that, with unbounded private beliefs, there will be asymptotic learning in almost all reasonable social networks. We also show that for most network topologies, when private beliefs are bounded, there will not be asymptotic learning. In addition, in contrast to the special case where all past actions are observed, asymptotic learning is possible even with bounded beliefs in certain stochastic network topologies.

Keywords: information aggregation, learning, social networks, herding, information cascades.

JEL Classification: C72, D83.

[†]Department of Economics, MIT, Cambridge MA, 02139 (e-mail: daron@mit.edu)

[‡]Laboratory for Information and Decision Systems, Electrical Engineering and Computer Science Department, MIT, Cambridge MA, 02139 (e-mail: dahleh@mit.edu)

[§]Operations Research Center, MIT, Cambridge MA, 02139 (e-mail: lobel@mit.edu)

[¶]Laboratory for Information and Decision Systems, Electrical Engineering and Computer Science Department, MIT, Cambridge MA, 02139 (e-mail: asuman@mit.edu)

1 Introduction

How is dispersed and decentralized information held by a large number of individuals aggregated? Imagine a situation in which each of a large number of individuals has a noisy signal about an underlying state of the world. This state of the world might concern, among other things, earning opportunities in a certain occupation, the quality of a new product, the suitability of a particular political candidate for office or payoff-relevant actions taken by the government. If signals are unbiased, the combination—*aggregation*— of the information of the individuals will be sufficient for the society to "learn" the true underlying state. The above question can be formulated as the investigation of what types of behaviors and communication structures will lead to this type of information aggregation.

Condorcet's Jury Theorem provides a natural benchmark, where sincere (truthful) reporting of their information by each individual is sufficient for aggregation of information by a law of large numbers argument (Condorcet, 1788). Against this background, a number of papers, most notably Bikchandani, Hirshleifer and Welch (1992), Banerjee (1992) and Smith and Sorensen (2000), show how this type of aggregation might fail in the context of the (perfect) *Bayesian equilibrium* of a dynamic game: when individuals act sequentially, after observing the actions of all previous individuals (agents), many reasonable situations will lead to the wrong conclusion with positive probability.

An important modeling assumption in these papers is that each individual observes all past actions. In practice, individuals are situated in complex social networks, which provide their main source of information. For example, Granovetter (1973), Montgomery (1991), Munshi (2003) and Iaonnides and Loury (2004) document the importance of information obtained from the social network of an individual for employment outcomes. Besley and Case (1994), Foster and Rosenzweig (1995), Munshi (2004), and Udry and Conley (2001) show the importance of the information obtained from social networks for technology adoption. Jackson (2006, 2007) provide excellent surveys of the work on the importance of social networks in many diverse situations. In this paper, we address how the structure of social networks, which determines the information that individuals receive, affects equilibrium information aggregation.

We start with the canonical sequential learning problem, except that instead of full observation of past actions, we allow for a general *social network* connecting individuals. More specifically, a large number of agents sequentially choose between two actions. An underlying state determines the payoffs of these two actions. Each agent receives a signal on which of these two actions yields a higher payoff. Preferences of all agents are aligned in the sense that, given the underlying state of the world, they all prefer the same action. The game is characterized by two features: (i) the *signal structure*, which determines how informative the signals received by the individuals are; (ii) the *social network structure*, which determines the observations of each individual in the game. We model the social network structure as a stochastic process that determines each individual's *neighborhood*. Each individual only observes the (past) actions of agents in his neighborhood. Motivated by the social network interpretation, throughout it is assumed that each individual knows the identity of the agents in his neighborhood (e.g., he can distinguish whether the action observed is by a friend or neighbor or by some

outside party). Nevertheless, the realized neighborhood of each individual as well as his private signal are private information.

We also refer to the stochastic process generating neighborhoods as the *network topology* of this social network. For some of our results, it will be useful to distinguish between deterministic and stochastic network topologies. With deterministic network topologies, there is no uncertainty concerning the neighborhood of each individual and these neighborhoods are common knowledge. With stochastic network topologies, there is uncertainty about these neighborhoods.

The environment most commonly studied in the previous literature is the *full observation network topology*, which is the special case where all past actions are observed. Another deterministic special case is the network topology where each individual observes the actions of the most recent $M \geq 1$ individuals. Other relevant social networks include stochastic topologies in which each individual observes a random subset of past actions, as well as those in which, with a high probability, each individual observes the actions of some "influential" group of agents, who may be thought of as "leaders" or the media.

We provide a systematic characterization of the conditions under which there will be equilibrium information aggregation in social networks. We say that there is *information aggregation* or equivalently *asymptotic learning*, when, in the limit as the size of the social network becomes arbitrarily large, individual actions converge (in probability) to the action that yields the higher payoff. We say that asymptotic learning fails if, as the social network becomes large, the correct action is not chosen (or more formally, the lim inf of the probability that the right action is chosen is strictly less than 1).

Two concepts turn out to be crucial in the study of information aggregation in social networks. The first is whether the *likelihood ratio* implied by individual signals is always finite and bounded away from 0.[1] Smith and Sorensen (2000) refer to beliefs that satisfy this property as *bounded (private) beliefs*. With bounded beliefs, there is a maximum amount of information in any individual signal. In contrast, when there exist signals with arbitrarily high and low likelihood ratios, *(private) beliefs are unbounded*. Whether bounded or unbounded beliefs provide a better approximation to reality is partly an interpretational and partly an empirical question. Smith and Sorensen's main result is that when each individual observes all past actions and private beliefs are unbounded, information will be aggregated and the correct action will be chosen asymptotically. In contrast, the results in Bikchandani, Hirshleifer and Welch (1992), Banerjee (1992) and Smith and Sorensen (2000) indicate that with bounded beliefs, there will not be asymptotic learning (or information aggregation). Instead, as emphasized by Bikchandani, Hirshleifer and Welch (1992) and Banerjee (1992), there will be "herding" or "informational cascades," where individuals copy past actions and/or completely ignore their own signals.

The second key concept is that of a network topology with expanding observations. To describe this concept, let us first introduce another notion: a finite group of agents is *excessively influential* if there exists an infinite number of agents who, with probability

---

[1]The likelihood ratio is the ratio of the probabilities or the densities of a signal in one state relative to the other.

uniformly bounded away from 0, observe only the actions of a subset of this group. For example, a group is excessively influential if it is the source of *all* information (except individual signals) for an infinitely large component of the social network. If there exists an excessively influential group of individuals, then the social network has *nonexpanding observations*, and conversely, if there exists no excessively influential group, the network has *expanding observations*. This definition implies that most reasonable social networks have expanding observations, and in particular, a minimum amount of "arrival of new information " in the social network is sufficient for the expanding observations property.[2] For example, the environment studied in most of the previous work in this area, where all past actions are observed, has expanding observations. Similarly, a social network in which each individual observes one uniformly drawn individual from those who have taken decisions in the past or a network in which each individual observes his immediate neighbor all feature expanding observations. Note also that a social network with expanding observations need not be *connected*. For example, the network in which even-numbered [odd-numbered] individuals only observe the past actions of even-numbered [odd-numbered] individuals has expanding observations, but is *not* connected. A simple, but typical, example of a network with nonexpanding observations is the one in which all future individuals only observe the actions of the first $K < \infty$ agents.

Our main results in this paper are presented in four theorems.

1. Theorem 1 shows that there is no asymptotic learning in networks with nonexpanding observations. This result is not surprising, since information aggregation is not possible when the set of observations on which (an infinite subset of) individuals can build their decisions remains limited forever.

2. Our most substantive result, Theorem 2, shows that when (private) beliefs are unbounded and the network topology is expanding, there will be asymptotic learning. This is a very strong result (particularly if we consider unbounded beliefs to be a better approximation to reality than bounded beliefs), since almost all reasonable social networks have the expanding observations property. This theorem, for example, implies that when some individuals, such as "informational leaders," are overrepresented in the neighborhoods of future agents (and are thus "influential," though not excessively so), learning may slow down, but asymptotic learning will still obtain as long as private beliefs are unbounded.

   The idea of the proof of Theorem 2 is as follows. We first establish a *strong improvement principle* under unbounded beliefs, whereby in a network where each individual has a single agent in his neighborhood, he can receive a strictly higher payoff than this agent and this improvement remains bounded away from zero as long as asymptotic learning has not been achieved. We then show that the same insight applies when individuals stochastically observe one or multiple agents (in particular, with multiple agents, the improvement is no less than the case in which the individual observes a single agent from the past). Finally, the property that the

---

[2]Here, "arrival of new information" refers to the property that the probability of each individual observing the action of some individual from the recent past converges to one as the social network becomes arbitrarily large.

network topology has expanding observations is sufficient for these improvements to accumulate to asymptotic learning.

3. Theorem 3 presents a partial converse to Theorem 2. It shows that for the most common deterministic and stochastic networks, bounded private beliefs are incompatible with asymptotic learning. It therefore generalizes existing results on asymptotic learning, for example, those in Bikchandani, Hirshleifer and Welch (1992), Banerjee (1992), and Smith and Sorensen (2000) to general networks.

4. Our final main result, Theorem 4, establishes that asymptotic learning is possible with bounded private beliefs for certain stochastic network topologies. In these cases, there is sufficient arrival of new information (incorporated into the "social belief") because some agents make decisions on the basis of limited observations. As a consequence, even bounded private beliefs may aggregate and lead to asymptotic learning. This finding is particularly important, since it shows how moving away from simple network structures has major implications for equilibrium learning dynamics.

The rest of the paper is organized as follows. The next section discusses the related literature and clarifies the contribution of our paper. Section 3 introduces the model. Section 4 formally introduces the concepts of bounded and unbounded beliefs, and network topologies with expanding and nonexpanding observations. This section then presents our main results, Theorems 1-4, and discusses some of their implications (as well as presenting a number of corollaries to facilitate interpretation). The rest of the paper characterizes the (pure-strategy) perfect Bayesian equilibria of the model presented in Section 3 and provides proofs of these theorems. Section 5 presents a number of important results on the characterization of pure-strategy equilibria. Section 6 occupies the bulk of the paper and provides a detailed proof of Theorem 2, which involves the statement and proof of several lemmas. Section 7 provides a proof of Theorem 3, while Section 8 shows how asymptotic learning is possible with bounded private beliefs. Section 9 concludes. Appendices A and B contain proofs omitted from the main text, including the proof of Theorem 1.

2   Related Literature

The literature on social learning is vast. Roughly speaking, the literature can be separated according to two criteria: whether learning is Bayesian or myopic, and whether individuals learn from communication of exact signals or from the payoffs of others, or simply from observing others' actions. Typically, Bayesian models focus on learning from past actions, while most, but not all, myopic learning models focus on learning from communication.

Bikchandani, Hirshleifer and Welch (1992) and Banerjee (1992) started the literature on learning in situations in which individuals are Bayesian and observe past actions. Smith and Sorensen (2000) provide the most comprehensive and complete analysis of

this environment. Their results and the importance of the concepts of bounded and unbounded beliefs, which they introduced, have already been discussed in the introduction and will play an important role in our analysis in the rest of the paper. Other important contributions in this area include, among others, Welch (1992), Lee (1993), Chamley and Gale (1994), and Vives (1997). An excellent general discussion is contained in Bikchandani, Hirshleifer and Welch (1998). These papers typically focus on the special case of full observation network topology in terms of our general model.

The two papers most closely related to ours are Banerjee and Fudenberg (2004) and Smith and Sorensen (1998). Both of these papers study social learning with sampling of past actions. In Banerjee and Fudenberg, there is a continuum of agents and the focus is on proportional sampling (whereby individuals observe a "representative" sample of the overall population). They establish that asymptotic learning is achieved under mild assumptions as long as the sample size is no smaller than two. The existence of a continuum of agents is important for this result since it ensures that the fraction of individuals with different posteriors evolves deterministically. Smith and Sorensen, on the other hand, consider a related model with a countable number of agents. In their model, as in ours, the evolution of beliefs is stochastic. Smith and Sorensen provide conditions under which asymptotic learning takes place.

A crucial difference between Banerjee and Fudenberg and Smith and Sorensen, on the one hand, and our work, on the other, is the information structure. These papers assume that "samples are unordered" in the sense that individuals do not know the identity of the agents they have observed. In contrast, as mentioned above, our setup is motivated by a social network and assumes that individuals have stochastic neighborhoods, but know the identity of the agents in their realized neighborhood. We view this as a better approximation to learning in social networks. In addition to its descriptive realism, this assumption leads to a sharper characterization of the conditions under which asymptotic learning occurs. For example, in Smith and Sorensen's environment, asymptotic learning fails whenever an individual is "oversampled," in the sense of being overrepresented in the samples of future agents. In contrast, in our environment, asymptotic learning occurs when the network topology features expanding observations (and private beliefs are unbounded). Expanding observations is a much weaker requirement than "non-oversampling." For example, when each individual observes agent 1 and a randomly chosen agent from his predecessors, the network topology satisfies expanding observations, but there is oversampling.[3]

Other recent work on social learning includes Celen and Kariv (2004) who study Bayesian learning when each individual observes his immediate predecessor, Gale and Kariv (2003) who generalize the payoff equalization result of Bala and Goyal (1998) in connected social networks (discussed below) to Bayesian learning, and Callander and Horner (2006), who show that it may be optimal to follow the actions of agents that deviate from past average behavior.

The second branch of the literature focuses on non-Bayesian learning, typically with agents using some reasonable rules of thumb. This literature considers both learning from

---

[3]This also implies that, in the terminology of Bala and Goyal, a "royal family" precludes learning in Smith and Sorensen's model, but not in ours, see below.

past actions and from payoffs (or directly from beliefs). Early papers in this literature include Ellison and Fudenberg (1993, 1995), which show how rule-of-thumb learning can converge to the true underlying state in some simple environments. The papers most closely related to our work in this genre are Bala and Goyal (1998, 2001), DeMarzo, Vayanos and Zwiebel (2003) and Golub and Jackson (2007). These papers study non-Bayesian learning over an arbitrary, connected social network. Bala and Goyal (1998) establish the important and intuitive payoff equalization result that, asymptotically, each individual must receive a payoff equal to that of an arbitrary individual in his "social network," since otherwise he could copy the behavior of this other individual. Our paper can be viewed as extending Bala and Goyal's results to a situation with Bayesian learning. A similar "imitation" intuition plays an important role in our proof of asymptotic learning with unbounded beliefs and unbounded observations.

DeMarzo, Vayanos and Zwiebel and Golub and Jackson also study similar environments and derive consensus-type results, whereby individuals in the connected components of the social network will converge to similar beliefs. They provide characterization results on which individuals in the social network will be influential and investigate the likelihood that the consensus opinion will coincide with the true underlying state. Golub and Jackson, in particular, show that social networks where some individuals are "influential" in the sense of being connected to a large number of people make learning more difficult or impossible. A similar result is also established in Bala and Goyal, where they show that the presence of a *royal family*, i.e., a small set of individuals observed by everyone, precludes learning. This both complements and contrasts with our results. In our environment, an excessively influential group of individuals prevents learning, but influential agents in Golub and Jackson's sense or Bala and Goyal's royal family are *not excessively influential* and still allow asymptotic learning. This is because with Bayesian updating over a social network, individuals recognize who the oversampled individuals or the royal family are and accordingly adjust the weight they give to their action/information.

The literature on the information aggregation role of elections is also related, since it revisits the original context of Condorcet's Jury Theorem. This literature includes, among others, the papers by Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1996, 1997), McLennan (1998), Myerson (1998, 2000), and Young (1988). Most of these papers investigate whether dispersed information will be accurately aggregated in large elections. Although the focus on information aggregation is common, the set of issues and the methods of analysis are very different, particularly since, in these models, there are no sequential decisions.

Finally, there is also a literature in engineering, which studies related problems, especially motivated by aggregation of information collected by decentralized sensors. These include Cover (1969), Papastavrou and Athans (1990), Lorenz, Marciniszyn and, Steger (2007), and Tay, Tsitsiklis and Win (2007). The work by Papastavrou and Athans contains a result that is equivalent to the characterization of asymptotic learning with the observation of the immediate neighbor.

3   Model

A countably infinite number of agents (individuals), indexed by $n \in \mathbb{N}$, sequentially make a single decision each. The payoff of agent n depends on an underlying state of the world θ and his decision. To simplify the notation and the exposition, we assume that both the underlying state and decisions are binary. In particular, the decision of agent n is denoted by $x_n \in \{0, 1\}$ and the underlying state is $\theta \in \{0, 1\}$. The payoff of agent n is

$$u_n (x_n, \theta) = \begin{cases} 1 & \text{if } x_n = \theta \\ 0 & \text{if } x_n \neq \theta. \end{cases}$$

Again to simplify notation, we assume that both values of the underlying state are equally likely, so that $\mathbb{P}(\theta = 0) = \mathbb{P}(\theta = 1) = 1/2$.

The state θ is unknown. Each agent $n \in \mathbb{N}$ forms beliefs about this state from a *private signal* $s_n \in S$ (where S is a metric space or simply a Euclidean space) and from his observation of the actions of other agents. Conditional on the state of the world θ, the signals are independently generated according to a probability measure $\mathbb{F}_\theta$. We refer to the pair of measures $(\mathbb{F}_0, \mathbb{F}_1)$ as the *signal structure* of the model. We assume that $\mathbb{F}_0$ and $\mathbb{F}_1$ are *absolutely continuous* with respect to each other, which immediately implies that no signal is fully revealing about the underlying state. We also assume that $\mathbb{F}_0$ and $\mathbb{F}_1$ are not identical, so that some signals are *informative*. These two assumptions on the signal structure are maintained throughout the paper and will not be stated in the theorems explicitly.

In contrast to much of the literature on social learning, we assume that agents do not necessarily observe all previous actions. Instead, they observe the actions of other agents according to the structure of the *social network*. To introduce the notion of a social network, let us first define a *neighborhood*. Each agent n observes the decisions of the agents in his (stochastically-generated) neighborhood, denoted by $B(n)$.[4] Since agents can only observe actions taken previously, $B(n) \subseteq \{1, 2, ..., n-1\}$. Each neighborhood $B(n)$ is generated according to an arbitrary probability distribution $\mathbb{Q}_n$ over the set of all subsets of $\{1, 2, ..., n-1\}$. We impose no special assumptions on the sequence of distributions $\{\mathbb{Q}_n\}_{n \in \mathbb{N}}$ except that the draws from each $\mathbb{Q}_n$ are independent from each other for all n and from the realizations of private signals. The sequence $\{\mathbb{Q}_n\}_{n \in \mathbb{N}}$ is the *network topology* of the social network formed by the agents. The network topology is common knowledge, whereas the realized neighborhood $B(n)$ and the private signal s_n are the private information of agent n. We say that $\{\mathbb{Q}_n\}_{n \in \mathbb{N}}$ is a *deterministic* network topology if the probability distribution $\mathbb{Q}_n$ is a degenerate (Dirac) distribution for all n. Otherwise, that is, if $\{\mathbb{Q}_n\}$ for some n is nondegenerate, $\{\mathbb{Q}_n\}_{n \in \mathbb{N}}$ is a *stochastic* network topology.

A *social network* consists of a network topology $\{\mathbb{Q}_n\}_{n \in \mathbb{N}}$ and a signal structure $(\mathbb{F}_0, \mathbb{F}_1)$.

Example 1 Here are some examples of network topologies.

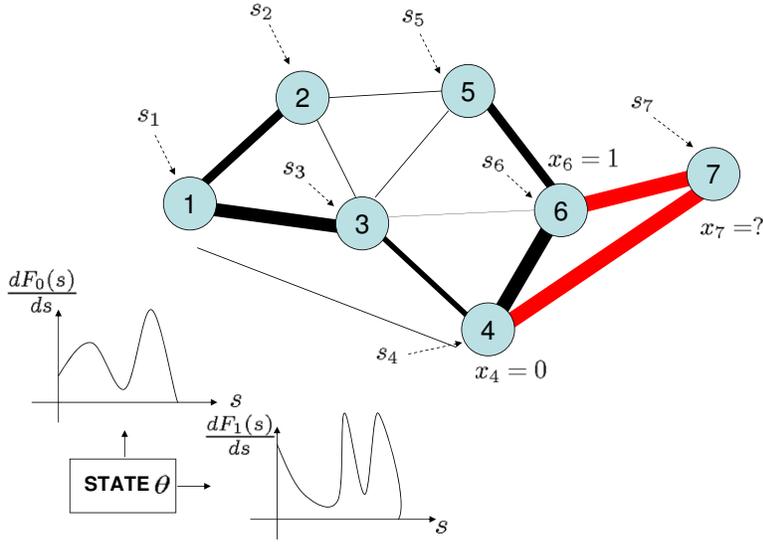[4]If $n' \in B(n)$, then agent n not only observes the action of n', but also knows the identity of this agent.

Figure 1: The figure illustrates the world from the perspective of agent 7. Agent 7 knows her private signal s_7, her realized neighborhood, $B(7) = \{4, 6\}$ and the decisions of agents 4 and 6, x_4 and x_6. She also knows the probabilistic model $\{\mathbb{Q}_n\}_{n<7}$ for neighborhoods of all agents $n < 7$.

1. If $\{\mathbb{Q}_n\}_{n\in\mathbb{N}}$ assigns probability 1 to neighborhood $\{1, 2..., n-1\}$ for each $n \in \mathbb{N}$, then the network topology is identical to the canonical one studied in the previous literature where each agent observes all previous actions (e.g., Banerjee (1992), Bikchandani, Hirshleifer and Welch (1992), Smith and Sorensen (2000)).

2. If $\{\mathbb{Q}_n\}_{n\in\mathbb{N}}$ assigns probability $1/(n-1)$ to each one of the subsets of size 1 of $\{1, 2..., n-1\}$ for each $n \in \mathbb{N}$, then we have a network topology of random sampling of one agent from the past.

3. If $\{\mathbb{Q}_n\}_{n\in\mathbb{N}}$ assigns probability 1 to neighborhood $\{n-1\}$ for each $n \in \mathbb{N}$, then we have a network topology where each individual only observes his immediate neighbor (also considered in a different context in the engineering literature by Papastavrou and Athans (1990)).

4. If $\{\mathbb{Q}_n\}_{n\in\mathbb{N}}$ assigns probability 1 to neighborhoods that are subsets of $\{1, 2, ..., K\}$ for each $n \in \mathbb{N}$ for some $K \in \mathbb{N}$. In this case, all agents observe the actions of at most K agents.

5. Figure 1 depicts an arbitrary stochastic topology until agent 7. The thickness of the lines represents the probability with which a particular agent will observe the action of the corresponding preceding agent.

Given the description above, it is evident that the information set I_n of agent n is given by her signal s_n, her neighborhood $B(n)$, and all decisions of agents in $B(n)$, that

is,

$$I_n = \{s_n, B(n), x_k \text{ for all } k \in B(n)\}. \tag{1}$$

The set of all possible information sets of agent n is denoted by $\mathcal{I}_n$. A strategy for individual n is a mapping $\sigma_n : \mathcal{I}_n \to \{0, 1\}$ that selects a decision for each possible information set. A strategy profile is a sequence of strategies $\sigma = \{\sigma_n\}_{n \in \mathbb{N}}$. We use the standard notation $\sigma_{-n} = \{\sigma_1, \ldots, \sigma_{n-1}, \sigma_{n+1}, \ldots\}$ to denote the strategies of all agents other than n and also (σ_n, σ_{-n}) for any n to denote the strategy profile σ. Given a strategy profile σ, the sequence of decisions $\{x_n\}_{n \in \mathbb{N}}$ is a stochastic process and we denote the measure generated by this stochastic process by $\mathbb{P}_\sigma$.

Definition 1 *A strategy profile σ^* is a pure-strategy* **Perfect Bayesian Equilibrium** *of this game of social learning if for each $n \in \mathbb{N}$, σ_n^* maximizes the expected payoff of agent n given the strategies of other agents σ_{-n}^*.*

In the rest of the paper, we focus on pure-strategy Perfect Bayesian Equilibria, and simply refer to this as "equilibrium" (without the pure-strategy and the Perfect Bayesian qualifiers).

Given a strategy profile σ, the expected payoff of agent n from action $x_n = \sigma_n(I_n)$ is simply $\mathbb{P}_\sigma(x_n = \theta \mid I_n)$. Therefore, for any equilibrium σ^*, we have

$$\sigma_n^*(I_n) \in \arg\max_{y \in \{0,1\}} \mathbb{P}_{(y, \sigma_{-n}^*)}(y = \theta \mid I_n). \tag{2}$$

We denote the set of equilibria (pure-strategy Perfect Bayesian Equilibria) of the game by Σ^*. It is clear that Σ^* is nonempty. Given the sequence of strategies $\{\sigma_1^*, \ldots, \sigma_{n-1}^*\}$, the maximization problem in (2) has a solution for each agent n and each $I_n \in \mathcal{I}_n$. Proceeding inductively, and choosing either one of the actions in case of indifference determines an equilibrium. We note the existence of equilibrium here.

Proposition 1 *There exists a pure-strategy Perfect Bayesian Equilibrium.*

Our main focus is whether equilibrium behavior will lead to information aggregation. This is captured by the notion of *asymptotic learning*, which is introduced next.

Definition 2 *Given a signal structure $(\mathbb{F}_0, \mathbb{F}_1)$ and a network topology $\{\mathbb{Q}_n\}_{n \in \mathbb{N}}$, we say that **asymptotic learning occurs in equilibrium** σ if x_n converges to θ in probability (according to measure $\mathbb{P}_\sigma$), that is,*

$$\lim_{n \to \infty} \mathbb{P}_\sigma(x_n = \theta) = 1.$$

Notice that asymptotic learning requires that the probability of taking the correct action converges to 1. Therefore, asymptotic learning will fail when, as the network becomes large, the limit inferior of the probability of all individuals taking the correct action is strictly less than 1.

Our goal in this paper is to characterize conditions on social networks—on signal structures and network topologies—that ensure asymptotic learning.

4 Main Results

In this section, we present our main results on asymptotic learning, in particular, Theorems 1-4, and we discuss some of the implications of these theorems. The proofs of the results stated in this section are provided in the rest of the paper.

We start by introducing the key properties of network topologies and signal structures that impact asymptotic learning. Intuitively, for asymptotic learning to occur, the information that each agent receives from other agents should not be confined to a bounded subset of agents. This property is established in the following definition. For this definition and throughout the paper, if the set $B(n)$ is empty, we set $\max_{b \in B(n)} b = 0$.

Definition 3 *The network topology has **expanding observations** if for all $K \in \mathbb{N}$, we have*

$$\lim_{n \to \infty} \mathbb{Q}_n \left(\max_{b \in B(n)} b < K \right) = 0.$$

*If the network topology does not satisfy this property, then we say it has **nonexpanding observations**.*

Recall that the neighborhood of agent n is a random variable $B(n)$ (with values in the set of subsets of $\{1, 2, ..., n-1\}$) and distributed according to $\mathbb{Q}_n$. Therefore, $\max_{b \in B(n)} b$ is a random variable that takes values in $\{0, 1, ..., n-1\}$. The expanding observations condition can be restated as the sequence of random variables $\{\max_{b \in B(n)} b\}_{n \in \mathbb{N}}$ converging to infinity in probability. Similarly, it follows from the preceding definition that the network topology has nonexpanding observations if and only if there exists some $K \in \mathbb{N}$ and some scalar $\epsilon > 0$ such that

$$\limsup_{n \to \infty} \mathbb{Q}_n \left(\max_{b \in B(n)} b < K \right) \geq \epsilon.$$

An alternative restatement of this definition might clarify its meaning. Let us refer to a finite set of individuals C as *excessively influential* if there exists a subsequence of agents who, with probability uniformly bounded away from zero, observe the actions of a subset of C. Then, the network topology has nonexpanding observations if and only if there exists an excessively influential group of agents. Note also that if there is a minimum amount of arrival of new information in the network, so that the probability of an individual observing some other individual from the recent past goes to one as the network becomes large, then the network topology will feature expanding observations. This discussion therefore highlights that the requirement that a network topology has expanding observations is quite mild and most social networks satisfy this requirement.

When the topology has nonexpanding observations, there is a subsequence of agents that draws information from the first K decisions with positive probability (uniformly bounded away from 0). It is then intuitive that network topologies with nonexpanding observations will preclude asymptotic learning. Our first theorem states this result. Though intuitive, the proof of this result is somewhat long and not essential for the rest of the argument and is thus provided in Appendix A.

Theorem 1 *Assume that the network topology $\{\mathbb{Q}_n\}_{n\in\mathbb{N}}$ has nonexpanding observations. Then, there exists no equilibrium $\sigma \in \Sigma^*$ with asymptotic learning.*

This theorem states the intuitive result that with nonexpanding observations, asymptotic learning will fail. This result is not surprising, since asymptotic learning requires the *aggregation* of the information of different individuals. But a network topology with nonexpanding observations does not allow such aggregation. Intuitively, nonexpanding observations, or equivalently the existence of an excessively influential group of agents, imply that infinitely many individuals will observe finitely many actions with positive probability and this will not enable them to aggregate the dispersed information collectively held by the entire social network.

The main question is then whether, once we exclude network topologies with nonexpanding observations, what other conditions need to be imposed to ensure asymptotic learning. To answer this question and state our main theorem, we need to introduce one more notion. Following Smith and Sorensen (2000), we define *private beliefs* as the posterior that the true state is $\theta = 1$ given individual signal s_n. We will see below that private beliefs will play a key role in the characterization of equilibrium behavior. For now, let $d\mathbb{F}_0/d\mathbb{F}_1$ denote the Radon-Nikodym derivative of the measures $\mathbb{F}_0$ and $\mathbb{F}_1$ (recall that these are absolutely continuous with respect to each other). If $\mathbb{F}_0$ and $\mathbb{F}_1$ have densities, then for each $j \in \{0, 1\}$, $d\mathbb{F}_j$ can be replaced by the density of $\mathbb{F}_j$. If both measures have atoms at some $s \in S$, then $d\mathbb{F}_0/d\mathbb{F}_1(s) = \mathbb{F}_0(s)/\mathbb{F}_1(s)$.

Definition 4 *The signal structure has **bounded private beliefs** if there exists some $0 < m,\ M < \infty$ such that the Radon-Nikodym derivative $d\mathbb{F}_0/d\mathbb{F}_1$ satisfies*

$$m < \frac{d\mathbb{F}_0}{d\mathbb{F}_1}(s) < M,$$

*for almost all $s \in S$ under measure $(\mathbb{F}_0 + \mathbb{F}_1)/2$. The signal structure has **unbounded private beliefs** if for any S' contained in S with probability 1 under measure $(\mathbb{F}_0+\mathbb{F}_1)/2$, we have*

$$\inf_{s\in S'} \frac{d\mathbb{F}_0}{d\mathbb{F}_1}(s) = 0, \quad and \quad \sup_{s\in S'} \frac{d\mathbb{F}_0}{d\mathbb{F}_1}(s) = \infty.$$

Bounded private beliefs imply that there is a maximum amount of information that an individual can derive from his private signal. Conversely, unbounded private beliefs correspond to a situation where an agent can receive an arbitrarily strong signal about the underlying state (see Section 5.2 for a more detailed discussion of this property). Smith and Sorensen (2000) show that, in the special case of full observation network topology, there will be asymptotic learning if and only if private beliefs are unbounded.

The following theorem shows that for general network topologies, unbounded private beliefs play a similar role. In particular, unbounded private beliefs and expanding observations are sufficient for asymptotic learning in all equilibria.

Theorem 2 *Assume that the signal structure $(\mathbb{F}_0, \mathbb{F}_1)$ has unbounded private beliefs and the network topology $\{\mathbb{Q}_n\}_{n\in\mathbb{N}}$ has expanding observations. Then, asymptotic learning occurs in every equilibrium $\sigma \in \Sigma^*$.*

The proof of this theorem, which is provided in Section 6, takes up a large part of the remainder of this paper. However, many of its implications can be discussed before presenting a detailed proof.

Theorem 2 is quite a striking result. It implies that unbounded private beliefs are sufficient for asymptotic learning for most (but not all) network topologies. In particular, the condition that the network topology has expanding observations is fairly mild and only requires a minimum amount of arrival of recent information to the network. Social networks in which each individual observes all past actions, those in which each observes just his neighbor, and those in which each individual observes $M \geq 1$ agents independently and uniformly drawn from his predecessors are all examples of network topologies with expanding observations. Theorem 2 therefore implies that unbounded private beliefs are sufficient to guarantee asymptotic learning in social networks with these properties and many others.

Nevertheless, there are interesting network topologies where asymptotic learning does not occur even with unbounded private signals. The following corollary to Theorems 1 and 2 shows that for an interesting class of stochastic network topologies, there is a critical topology at which there is a *phase transition*—that is, for all network topologies with greater expansion of observations than this critical topology, there will be asymptotic learning and for all topologies with less expansion, asymptotic learning will fail. The proof of this corollary is also provided in Section 6.

Corollary 1 *Assume that the signal structure $(\mathbb{F}_0, \mathbb{F}_1)$ has unbounded private beliefs. Assume also that the network topology is given by $\{\mathbb{Q}_n\}_{n \in \mathbb{N}}$ such that*

$$\mathbb{Q}_n(m \in B(n)) = \frac{A}{(n-1)^C} \quad \text{for all } n \text{ and all } m < n,$$

where, given n, the draws for $m, m' < n$ are independent and A and C are positive constants. If $C < 1$ then asymptotic learning occurs in all equilibria. If $C \geq 1$, then asymptotic learning does not occur in any equilibrium.

Given the class of network topologies in this corollary, $C < 1$ implies that as the network becomes large, there will be sufficient expansion of observations. In contrast, for $C \geq 1$, stochastic process $\mathbb{Q}_n$ does not place enough probability on observing recent actions and the network topology is nonexpanding. Consequently, Theorem 1 applies and there is no asymptotic learning.

To highlight the implications of Theorems 1 and 2 for deterministic network topologies, let us introduce the following definition.

Definition 5 *Assume that the network topology is deterministic. Then, we say a finite sequence of agents π is an **information path** of agent n if for each i, $\pi_i \in B(\pi_{i+1})$ and the last element of π is n. Let $\bar{\pi}(n)$ be an information path of agent n that has maximal length. Then, we let $L(n)$ denote the number of elements in $\bar{\pi}(n)$ and call it agent n's **information depth**.*

Intuitively, the concepts of information path and information depth capture the intuitive notion of how long the "trail" of the information in the neighborhood of an

individual is. For example, if each individual observes only his immediate neighbor (i.e., $B(n) = \{n - 1\}$ with probability one), each will have a small neighborhood, but the information depth of a high-indexed individual will be high (or the "trail" will be long), because the immediate neighbor's action will contain information about the signals of all previous individuals. The next corollary shows that with deterministic network topologies, asymptotic learning will occur if only if the information depth (or the trail of the information) increases without bound as the network becomes larger.

Corollary 2 *Assume that the signal structure $(\mathbb{F}_0, \mathbb{F}_1)$ has unbounded private beliefs. Assume that the network topology is deterministic. Then, asymptotic learning occurs for all equilibria if the sequence of information depths $\{L(n)\}_{n\in\mathbb{N}}$ goes to infinity. If the sequence $\{L(n)\}_{n\in\mathbb{N}}$ does not go to infinity, then asymptotic learning does not occur in any equilibrium.*

In the full observation network topology, bounded beliefs imply lack of asymptotic learning. One might thus expect a converse to Theorem 2, whereby asymptotic learning fails whenever signals are bounded. Under general network topologies, learning dynamics turn out to be more interesting and richer. The next theorem provides a partial converse to Theorem 2 and shows that for a wide range of deterministic and stochastic network topologies, bounded beliefs imply no asymptotic learning. However, somewhat surprisingly, Theorem 4 will show that the same is not true with more general stochastic network topologies.

Theorem 3 *Assume that the signal structure $(\mathbb{F}_0, \mathbb{F}_1)$ has bounded private beliefs. If the network topology $\{\mathbb{Q}_n\}_{n\in\mathbb{N}}$ satisfies one of the following conditions,*

(a) $B(n) = \{1, \ldots, n - 1\}$ for all n,

(b) $|B(n)| \leq 1$ for all n, or

(c) there exists some constant M such that $|B(n)| \leq M$ for all n and

$$\lim_{n\to\infty} \max_{b\in B(n)} b = \infty \qquad \text{with probability } 1,$$

then, asymptotic learning does not occur in any equilibrium $\sigma \in \Sigma^*$.

This theorem implies that in most common deterministic and stochastic network topologies, bounded private beliefs imply lack of asymptotic learning. The proof of Theorem 3 is presented in Section 7. Although Part (a) of this theorem is already proved by Smith and Sorensen (2000), we provide an alternative proof that highlights the importance of the concepts emphasized here.

The following corollary, which is also proved in Section 7, illustrates the implications of Theorem 3. It shows that, when private beliefs are bounded, there will be no asymptotic learning (in any equilibrium) in stochastic networks with random sampling.

Corollary 3 *Assume that the signal structure $(\mathbb{F}_0, \mathbb{F}_1)$ has bounded private beliefs. Assume that each agent n samples M agents uniformly and independently among $\{1, ..., n - 1\}$, for some $M \geq 1$. Then, asymptotic learning does not occur in any equilibrium $\sigma \in \Sigma^*$.*

We next show that with general stochastic topologies asymptotic learning is possible. Let us first define the notion of a *nonpersuasive neighborhood*.

Definition 6 *A finite set $B \subset \mathbb{N}$ is a* **nonpersuasive neighborhood** *in equilibrium $\sigma \in \Sigma^*$ if*

$$\mathbb{P}_\sigma \left(\theta = 1 | x_k = y_k \text{ for all } k \in B \right) \in \left(\underline{\beta}, \overline{\beta} \right)$$

for any set of values $y_k \in \{0,1\}$ for each k. We denote the set of all nonpersuasive neighborhoods by $\mathcal{U}_\sigma$.

A neighborhood B is nonpersuasive in equilibrium $\sigma \in \Sigma^*$ if for any set of decisions that agent n observes, his behavior may still depend on his private signal. A nonpersuasive neighborhood is defined with respect to a particular equilibrium. However, it is straightforward to see that $B = \emptyset$, i.e., the empty neighborhood, is nonpersuasive in any equilibrium. Moreover the set $B = \{1\}$ is nonpersuasive as long as $\mathbb{P}_\sigma(\theta = 1 | x_1 = 1) < \overline{\beta}$ and $\mathbb{P}_\sigma(\theta = 1 | x_1 = 0) > \underline{\beta}$. It can be verified that this condition is equivalent to

$$\mathbb{G}_1(1/2) < \min \left\{ \left(\frac{\overline{\beta}}{1 - \overline{\beta}} \right) \mathbb{G}_0(1/2), \left(\frac{\underline{\beta}}{1 - \underline{\beta}} \right) \mathbb{G}_0(1/2) + \frac{1 - 2\underline{\beta}}{1 - \underline{\beta}} \right\}.$$

Our main theorem for learning with bounded beliefs, which we state next, provides a class of stochastic social networks where asymptotic learning takes place for any signal structure. The proof of this theorem is presented in Section 6.

Theorem 4 *Let $(\mathbb{F}_0, \mathbb{F}_1)$ be an arbitrary signal structure. Let M be a positive integer and let $C_1, ..., C_M$ be sets such that $C_i \in \mathcal{U}_\sigma$ for all $i = 1, \ldots, M$ for some equilibrium $\sigma \in \Sigma^*$. For each $i = 1, \ldots, M$, let $\{r_i(n)\}$ be a sequence of non-negative numbers such that*

$$\lim_{n \to \infty} \sum_{i=1}^{M} r_i(n) = 0 \quad \text{and} \quad \sum_{n=1}^{\infty} \sum_{i=1}^{M} r_i(n) = \infty, \tag{3}$$

with $\sum_{i=1}^{M} r_i(n) \leq 1$ for all n and $r_i(n) = 0$ for all $n \leq \max_{b \in C_i} b$. Assume the network topology satisfies

$$B(n) = \begin{cases} C_i, & \text{with probability } r_i(n) \text{ for each } i \text{ from 1 to } M, \\ \{1, 2, ..., n-1\}, & \text{with probability } 1 - \sum_{i=1}^{M} r_i(n). \end{cases}$$

Then, asymptotic learning occurs in equilibrium σ.

Clearly, this theorem could have been stated with $C_i = \emptyset$ for all $i = 1, \ldots, m$, which would correspond to agent n making a decision without observing anybody else's action with some probability $r(n) = \sum_{i=1}^{M} r_i(n) \leq 1$.

This is a rather surprising result, particularly in view of existing results in the literature, which generate herds and information cascades (and no learning) with bounded beliefs. This theorem indicates that learning dynamics become significantly richer when we consider general social networks. In particular, certain stochastic network topologies enable a significant amount of new information to arrive into the network, because some

agents make decisions with limited information (nonpersuasive neighborhoods). As a result, the relevant information can be aggregated in equilibrium, leading to individuals' decisions eventually converging to the right action (in probability).

It is important to emphasize the difference between this result and that in Sgroi (2002), which shows that a social planner can ensure some degree of information aggregation by forcing a subsequence of agents to make decisions without observing past actions. With the same reasoning, one might conjecture that asymptotic learning may occur if a particular subsequence of agents, such as that indexed by prime numbers, has empty neighborhoods. However, there will not be asymptotic learning in this deterministic topology since $\liminf_{n\to\infty} \mathbb{P}_\sigma(x_n = \theta) < 1$. For the result that there is asymptotic learning (i.e., $\liminf_{n\to\infty} \mathbb{P}_\sigma(x_n = \theta) = 1$) in Theorem 4, the feature that the network topology is stochastic is essential.

5   Equilibrium Strategies

In this section, we provide a characterization of equilibrium strategies. We show that equilibrium decision rules of individuals can be decomposed into two parts, one that only depends on an individual's private signal, and the other that is a function of the observations of past actions. We also show why a full characterization of individual decisions is nontrivial and motivate an alternative proof technique, relying on developing bounds on improvements in the probability of the correct decisions, that will be used in the rest of our analysis.

5.1   Characterization of Individual Decisions

Our first lemma shows that individual decisions can be characterized as a function of the sum of two posteriors. These posteriors play an important role in our analysis. We will refer to these posteriors as the individual's *private belief* and the *social belief*.

Lemma 1 *Let $\sigma \in \Sigma^*$ be an equilibrium of the game. Let $I_n \in \mathcal{I}_n$ be an information set of agent n. Then, the decision of agent n, $x_n = \sigma(I_n)$, satisfies*

$$x_n = \begin{cases} 1, & if \;\; \mathbb{P}_\sigma(\theta = 1 \mid s_n) + \mathbb{P}_\sigma\big(\theta = 1 \mid B(n), x_k, \;\; k \in B(n)\big) > 1, \\ 0, & if \;\; \mathbb{P}_\sigma(\theta = 1 \mid s_n) + \mathbb{P}_\sigma\big(\theta = 1 \mid B(n), x_k, \;\; k \in B(n)\big) < 1, \end{cases}$$

and $x_n \in \{0, 1\}$ otherwise.

Proof. See Appendix B. ∎

The lemma above establishes an additive decomposition in the equilibrium decision rule between the information obtained from the private signal of the individual and from the observations of others' actions (in his neighborhood). The next definition formally distinguishes between the two components of an individual's information.

Definition 7 *We refer to the probability $\mathbb{P}_\sigma(\theta = 1 \mid s_n)$ as the **private belief** of agent n, and the probability*

$$\mathbb{P}_\sigma\big(\theta = 1 \mid B(n), x_k \text{ for all } k \in B(n)\big),$$

*as the **social belief** of agent n.*

Notice that the social belief depends on n since it is a function of the (realized) neighborhood of agent n.

Lemma 1 and Definition 7 imply that the equilibrium decision rule for agent $n \in \mathbb{N}$ is equivalent to choosing $x_n = 1$ when the sum of his private and social beliefs is greater than 1. Consequently, the properties of private and social beliefs will shape equilibrium learning behavior. In the next subsection, we provide a characterization for the dynamic behavior of private beliefs, which will be used in the analysis of the evolution of decision rules.

5.2   Private Beliefs

In this subsection, we study properties of private beliefs. Note that the private belief is a function of the private signal $s \in S$ and is not a function of the strategy profile σ since it does not depend on the decisions of other agents. We represent probabilities that do not depend on the strategy profile by $\mathbb{P}$. We use the notation p_n to represent the private belief of agent n, i.e.,

$$p_n = \mathbb{P}(\theta = 1 \mid s_n).$$

The next lemma follows from a simple application of Bayes' Rule.

Lemma 2 *For any n and any signal $s_n \in S$, the private belief p_n of agent n is given by*

$$p_n = \left(1 + \frac{d\mathbb{F}_0}{d\mathbb{F}_1}(s_n)\right)^{-1}. \tag{4}$$

We next define the support of a private belief. In our subsequent analysis, we will see that properties of the support of private beliefs play a key role in asymptotic learning behavior. Since the p_n are identically distributed for all n (which follows by the assumption that the private signals s_n are identically distributed), in the following, we will use agent 1's private belief p_1 to define the support and the conditional distributions of private beliefs.

Definition 8 *The **support of the private beliefs** is the interval $[\underline{\beta}, \overline{\beta}]$, where the end points of the interval are given by*

$$\underline{\beta} = \inf \{r \in [0,1] \mid \mathbb{P}(p_1 \le r) > 0\}, \quad and \quad \overline{\beta} = \sup \{r \in [0,1] \mid \mathbb{P}(p_1 \le r) < 1\}.$$

Combining Lemma 2 with Definition 4, we see that beliefs are unbounded if and only if $\underline{\beta} = 1 - \overline{\beta} = 0$. When the private beliefs are bounded, there is a maximum informativeness to any signal. When they are unbounded, agents may receive arbitrarily strong signals favoring either state (this follows from the assumption that $(\mathbb{F}_0, \mathbb{F}_1)$ are absolutely continuous with respect to each other). When both $\underline{\beta} > 0$ and $\overline{\beta} < 1$, private beliefs are bounded.

We represent the conditional distribution of a private belief given the underlying state by $\mathbb{G}_j$ for each $j \in \{0, 1\}$, i.e.,

$$\mathbb{G}_j(r) = \mathbb{P}(p_1 \leq r \mid \theta = j). \tag{5}$$

We say that a pair of distributions $(\mathbb{G}_0, \mathbb{G}_1)$ are *private belief distributions* if there exist some signal space S and conditional private signal distributions $(\mathbb{F}_0, \mathbb{F}_1)$ such that the conditional distributions of the private beliefs are given by $(\mathbb{G}_0, \mathbb{G}_1)$. The next lemma presents key relations for private belief distributions.

Lemma 3 *For any private belief distributions $(\mathbb{G}_0, \mathbb{G}_1)$, the following relations hold.*

(a) *For all $r \in (0, 1)$, we have*
$$\frac{d\mathbb{G}_0}{d\mathbb{G}_1}(r) = \frac{1 - r}{r}.$$

(b) *We have*

$$\mathbb{G}_0(r) \geq \left(\frac{1 - r}{r}\right) \mathbb{G}_1(r) + \frac{r - z}{2}\mathbb{G}_1(z) \ \text{ for all } 0 < z < r < 1,$$

$$1 - \mathbb{G}_1(r) \geq (1 - \mathbb{G}_0(r)) \left(\frac{r}{1 - r}\right) + \frac{w - r}{2}(1 - \mathbb{G}_1(w)) \ \text{ for all } 0 < r < w < 1.$$

(c) *The ratio $\mathbb{G}_0(r)/\mathbb{G}_1(r)$ is nonincreasing in r and $\mathbb{G}_0(r)/\mathbb{G}_1(r) > 1$ for all $r \in (\underline{\beta}, \overline{\beta})$.*

The proof of this lemma is provided in Appendix B. Part (a) establishes a basic relation for private belief distributions, which is used in some of the proofs below. The inequalities presented in part (b) of this lemma play an important role in quantifying how much information an individual obtains from his private signal. Part (c) will be used in our analysis of learning with bounded beliefs.

5.3 Social Beliefs

In this subsection, we illustrate the difficulties involved in determining equilibrium learning in general social networks. In particular, we show that social beliefs, as defined in Definition 4, may be nonmonotone, in the sense that additional observations of $x_n = 1$ in the neighborhood of an individual may reduce the social belief (i.e., the posterior derived from past observations that $x_n = 1$ is the correct action).

The following example establishes this point. Suppose the private signals are such that $\mathbb{G}_0(r) = 2r - r^2$ and $\mathbb{G}_1(r) = r^2$, which is a pair of private belief distributions $(\mathbb{G}_0, \mathbb{G}_1)$. Suppose the network topology is deterministic and for the first eight agents, it has the following structure: $B(1) = \emptyset, B(2) = ... = B(7) = \{1\}$ and $B(8) = \{1, ..., 7\}$ (see Figure 2).

For this social network, agent 1 has 3/4 probability of making a correct decision in either state of the world. If agent 1 chooses the action that yields a higher payoff (i.e., the correct decision), then agents 2 to 7 each have 15/16 probability of choosing the
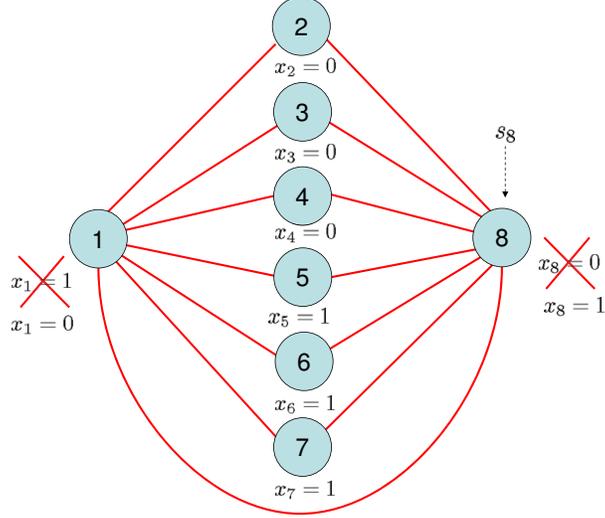
Figure 2: The figure illustrates a deterministic topology in which the social beliefs are nonmonotone.

correct decision. However, if agent 1 fails to choose the correct decision, then agents 2 to 7 have a 7/16 probability of choosing the correct decision. Now suppose agents 1 to 4 choose action $x_n = 0$, while agents 5 to 7 choose $x_n = 1$. The probability of this event happening in each state of the world is:

$$\mathbb{P}_\sigma(x_1 = ... = x_4 = 0, x_5 = x_6 = x_7 = 1 | \theta = 0) = \frac{3}{4}\left(\frac{15}{16}\right)^3\left(\frac{1}{16}\right)^3 = \frac{10125}{2^{26}},$$

$$\mathbb{P}_\sigma(x_1 = ... = x_4 = 0, x_5 = x_6 = x_7 = 1 | \theta = 1) = \frac{1}{4}\left(\frac{9}{16}\right)^3\left(\frac{7}{16}\right)^3 = \frac{250047}{2^{26}}.$$

Using Bayes' Rule, the social belief of agent 8 is given by

$$\left[1 + \frac{10125}{250047}\right]^{-1} \simeq 0.961.$$

Now, consider a change in x_1 from 0 to 1, while keeping all decisions as they are. Then,

$$\mathbb{P}_\sigma(x_1 = 1, x_2 = x_3 = x_4 = 0, x_5 = x_6 = x_7 = 1 | \theta = 0) = \frac{1}{4}\left(\frac{7}{16}\right)^3\left(\frac{9}{16}\right)^3 = \frac{250047}{2^{26}},$$

$$\mathbb{P}_\sigma(x_1 = 1, x_2 = x_3 = x_4 = 0, x_5 = x_6 = x_7 = 1 | \theta = 1) = \frac{1}{4}\left(\frac{1}{16}\right)^3\left(\frac{16}{16}\right)^3 = \frac{10125}{2^{26}}.$$

This leads to a social belief of agent 8 given by

$$\left[1 + \frac{250047}{10125}\right]^{-1} \simeq 0.039.$$

18

Therefore, this example has established that when x_1 changes from 0 to 1, agent 8's social belief declines from 0.961 to 0.039. That is, while the agent strongly believes the state is 1 when $x_1 = 0$, he equally strongly believes the state is 0 when $x_1 = 1$. This happens because when half of the agents in $\{2, \ldots, 7\}$ choose action 0 and the other half choose action 1, agent n places a high probability to the event that $x_1 \neq \theta$. This leads to a nonmonotonicity in social beliefs.

Since such nonmonotonicities cannot be ruled out in general, standard approaches to characterizing equilibrium behavior cannot be used. Instead, in the next section, we use an alternative approach, which develops a lower bound to the probability that an individual will make the correct decision relative to agents in his neighborhood.

6   Learning with Unbounded Private Beliefs and Expanding Observations

This section presents a proof of our main result, Theorem 2. The proof follows by combining several lemmas and propositions provided in this section. In the next subsection, we show that the expected utility of an individual is no less than the expected utility of any agent in his realized neighborhood. Though useful, this is a relatively weak result and is not sufficient to establish that asymptotic learning will take place in equilibrium. Subsection 6.2 provides the key result for the proof of Theorem 2. It focuses on the case in which each individual observes the action of a single agent and private beliefs are unbounded. Under these conditions, it establishes (a special case of) the *strong improvement principle*, which shows that the increase in expected utility is bounded away from zero (as long as social beliefs have not converged to the true state). Subsection 6.3 generalizes the strong improvement principle to the case in which each individual has a stochastically-generated neighborhood, potentially consisting of multiple (or no) agents. Subsection 6.4 then presents the proof of Theorem 2, which follows by combining these results with the fact that the network topology has expanding observations, so that the sequence of improvements will ultimately lead to asymptotic learning. Finally, subsection 6.5 provides proofs of Corollaries 1 and 2, which were presented in Section 4.

6.1   Information Monotonicity

As a first step, we show that the ex-ante probability of an agent making the correct decision (and thus his expected payoff) is no less than the probability of any of the agents in his realized neighborhood making the correct decision.

Proposition 2 (Information Monotonicity) *Let $\sigma \in \Sigma^*$ be an equilibrium. For any agent n and neighborhood $\mathfrak{B}$, we have*

$$\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \mathfrak{B}) \geq \max_{b \in \mathfrak{B}} \mathbb{P}_\sigma(x_b = \theta).$$

Proof. See Appendix B. ∎

Information monotonicity is similar to the *(expected) welfare improvement principle* in Banerjee and Fudenberg (2004) and in Smith and Sorensen (1998), and the *imitation principle* in Gale and Kariv (2003) and is very intuitive. However, it is not sufficiently strong to establish asymptotic learning. To ensure that, as the network becomes large, decisions converge (in probability) to the correct action, we need *strict* improvements. This will be established in the next two subsections.

6.2 Observing a Single Agent

In this subsection, we focus on a specific network topology where each agent observes the decision of a single agent. For this case, we provide an explicit characterization of the equilibrium, and under the assumption that private beliefs are unbounded, we establish a preliminary version of the *strong improvement principle*, which provides a lower bound on the increase in the ex-ante probability that an individual will make a correct decision over his neighbor's probability (recall that for now there is a single agent in each individual's neighborhood, thus each individual has a single "neighbor"). This result will be generalized to arbitrary networks in the next subsection.

For each n and strategy profile σ, let us define Y_n^σ and N_n^σ as the probabilities of agent n making the correct decision conditional on state θ. More formally, these are defined as

$$Y_n^\sigma = \mathbb{P}_\sigma(x_n = 1 \mid \theta = 1), \qquad N_n^\sigma = \mathbb{P}_\sigma(x_n = 0 \mid \theta = 0).$$

The unconditional probability of a correct decision is then

$$\frac{1}{2}(Y_n^\sigma + N_n^\sigma) = \mathbb{P}_\sigma(x_n = \theta). \tag{6}$$

We also define the *thresholds* L_n^σ and U_n^σ in terms of these probabilities:

$$L_n^\sigma = \frac{1 - N_n^\sigma}{1 - N_n^\sigma + Y_n^\sigma}, \qquad U_n^\sigma = \frac{N_n^\sigma}{N_n^\sigma + 1 - Y_n^\sigma}. \tag{7}$$

The next proposition shows that the equilibrium decisions are fully characterized in terms of these thresholds.

Proposition 3 *Let $B(n) = \{b\}$ for some agent n. Let $\sigma \in \Sigma^*$ be an equilibrium, and let L_b^σ and U_b^σ be given by Eq. (7). Then, agent n's decision x_n in equilibrium σ satisfies*

$$x_n = \begin{cases} 0, & \text{if } p_n < L_b^\sigma \\ x_b, & \text{if } p_n \in (L_b^\sigma, U_b^\sigma) \\ 1, & \text{if } p_n > U_b^\sigma. \end{cases}$$

The proof is omitted since it is an immediate application of Lemma 1 [use Bayes' Rule to determine $\mathbb{P}_\sigma(\theta = 1|x_b = j)$ for each $j \in \{0, 1\}$].

Note that the sequence $\{(U_n, L_n)\}$ only depends on $\{(Y_n, N_n)\}$, and is thus deterministic. This reflects the fact that each individual recognizes the amount of information that will be contained in the action of the previous agent, which determines his own
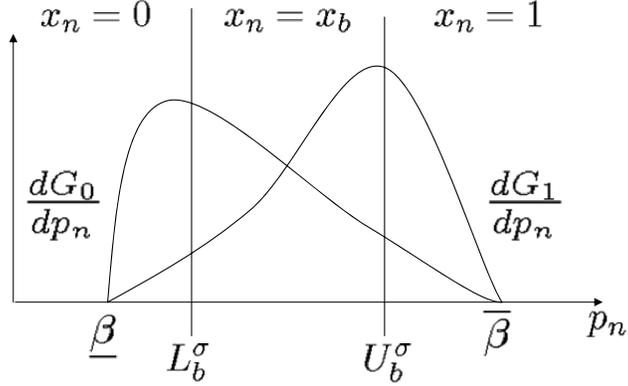
Figure 3: The equilibrium decision rule when observing a single agent, illustrated on the private belief space.

decision thresholds. Individual actions are still stochastic since they are determined by whether the individual's private belief is below L_b, above U_b, or in between (see Figure 3).

Using the structure of the equilibrium decision rule, the next lemma provides an expression for the probability of agent n making the correct decision conditional on his observing agent $b < n$, in terms of the private belief distributions and the thresholds L_b^σ and U_b^σ.

Lemma 4 *Let $B(n) = \{b\}$ for some agent n. Let $\sigma \in \Sigma^*$ be an equilibrium, and let L_b^σ and U_b^σ be given by Eq. (7). Then,*

$$\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\})$$
$$= \frac{1}{2}\Big[\mathbb{G}_0(L_b^\sigma) + \big(\mathbb{G}_0(U_b^\sigma) - \mathbb{G}_0(L_b^\sigma)\big)N_b^\sigma + (1 - \mathbb{G}_1(U_b^\sigma)) + \big(\mathbb{G}_1(U_b^\sigma) - \mathbb{G}_1(L_b^\sigma)\big)Y_b^\sigma\Big].$$

Proof. By definition, agent n receives the same expected utility from all his possible equilibrium choices. We can thus compute the expected utility by supposing that the agent will choose $x_n = 0$ when indifferent. Then, the expected utility of agent n (the probability of the correct decision) can be written as

$$\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\})$$
$$= \mathbb{P}_\sigma(p_n \le L_b^\sigma \mid \theta = 0)\mathbb{P}(\theta = 0) + \mathbb{P}_\sigma(p_n \in (L_b^\sigma, U_b^\sigma], x_b = 0 \mid \theta = 0)\mathbb{P}(\theta = 0)$$
$$+ \mathbb{P}_\sigma(p_n > U_b^\sigma \mid \theta = 1)\mathbb{P}(\theta = 1) + \mathbb{P}_\sigma(p_n \in (L_b^\sigma, U_b^\sigma], x_b = 1 \mid \theta = 1)\mathbb{P}(\theta = 1).$$

The result then follows using the fact that p_n and x_b are conditionally independent given θ and the notation for the private belief distributions [cf. Eq. (5)]. ∎

Using the previous lemma, we next strengthen Proposition 2 and provide a lower bound on the amount of improvement in the ex-ante probability of making the correct decision between an agent and his neighbor.

Lemma 5 *Let $B(n) = \{b\}$ for some agent n. Let $\sigma \in \Sigma^*$ be an equilibrium, and let L_b^σ and U_b^σ be given by Eq. (7). Then,*

$$
\begin{aligned}
\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\}) \geq{} & \mathbb{P}_\sigma(x_b = \theta) + \frac{(1 - N_b^\sigma)L_b^\sigma}{8} \mathbb{G}_1\left(\frac{L_b^\sigma}{2}\right) \\
& + \frac{(1 - Y_b^\sigma)(1 - U_b^\sigma)}{8}\left[1 - \mathbb{G}_0\left(\frac{1 + U_b^\sigma}{2}\right)\right].
\end{aligned}
$$

Proof. In Lemma 3(b), let $r = L_b^\sigma$, $z = L_b^\sigma/2$, so that we obtain

$$
(1 - N_b^\sigma)\mathbb{G}_0(L_b^\sigma) \geq Y_b^\sigma \mathbb{G}_1(L_b^\sigma) + \frac{(1 - N_b^\sigma)L_b^\sigma}{4}\mathbb{G}_1\left(\frac{L_b^\sigma}{2}\right).
$$

Next, again using Lemma 3(b) and letting $r = U_b^\sigma$ and $w = (1 + U_b^\sigma)/2$, we have

$$
(1 - Y_b^\sigma)[1 - \mathbb{G}_1(U_b^\sigma)] \geq N_b^\sigma[1 - \mathbb{G}_0(_b^\sigma)] + \frac{(1 - Y_b^\sigma)(1 - U_b^\sigma)}{4}\left[1 - \mathbb{G}_0\left(\frac{1 + U_b^\sigma}{2}\right)\right].
$$

Combining the preceding two relations with Lemma 4 and using the fact that $Y_b^\sigma + N_b^\sigma = 2\,\mathbb{P}_\sigma(x_b = \theta)$ [cf. Eq. (6)], the desired result follows. ∎

The next lemma establishes that the lower bound on the amount of improvement in the ex-ante probability is uniformly bounded away from zero for unbounded private beliefs and when $\mathbb{P}_\sigma(x_b = \theta) < 1$, i.e., when asymptotic learning is not achieved.

Lemma 6 *Let $B(n) = \{b\}$ for some n. Let $\sigma \in \Sigma^*$ be an equilibrium, and denote $\alpha = \mathbb{P}_\sigma(x_b = \theta)$. Then,*

$$
\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\}) \geq \alpha + \frac{(1 - \alpha)^2}{8} \min\left\{\mathbb{G}_1\left(\frac{1 - \alpha}{2}\right), 1 - \mathbb{G}_0\left(\frac{1 + \alpha}{2}\right)\right\}.
$$

Proof. We consider two cases separately.

Case 1: $N_b^\sigma \leq \alpha$. From the definition of L_b^σ and the fact that $Y_b^\sigma = 2\alpha - N_b^\sigma$ [cf. Eq. (6)], we have

$$
L_b^\sigma = \frac{1 - N_b^\sigma}{1 - 2N_b^\sigma + 2\alpha}.
$$

Since σ is an equilibrium, we have $\alpha \geq 1/2$, and thus the right hand-side of the preceding inequality is a nonincreasing function of N_b^σ. Since $N_b^\sigma \leq \alpha$, this relation therefore implies that $L_b^\sigma \geq 1 - \alpha$. Combining the relations $1 - N_b^\sigma \geq 1 - \alpha$ and $L_b^\sigma \geq 1 - \alpha$, we obtain

$$
\frac{(1 - N_b^\sigma)L_b^\sigma}{8} \mathbb{G}_1\left(\frac{L_b^\sigma}{2}\right) \geq \frac{(1 - \alpha)^2}{8} \mathbb{G}_1\left(\frac{1 - \alpha}{2}\right). \tag{8}
$$

22

Case 2: $N_b^\sigma \geq \alpha$. Since $Y_b^\sigma + N_b^\sigma = 2\alpha$, this implies that $Y_b^\sigma \leq \alpha$. Using the definition of U_b^σ and a similar argument as the one above, we obtain

$$\frac{(1 - Y_b^\sigma)(1 - U_b^\sigma)}{8}\left[1 - \mathbb{G}_0\left(\frac{1 + U_b^\sigma}{2}\right)\right] \geq \frac{(1 - \alpha)^2}{8}\left[1 - \mathbb{G}_0\left(\frac{1 + \alpha}{2}\right)\right]. \qquad (9)$$

Combining Eqs. (8) and (9), we obtain

$$\frac{(1 - N_b^\sigma)L_b^\sigma}{8}\mathbb{G}_1\left(\frac{L_b^\sigma}{2}\right) \quad + \quad \frac{(1 - Y_b^\sigma)(1 - U_b^\sigma)}{8}\left[1 - \mathbb{G}_0\left(\frac{1 + U_b^\sigma}{2}\right)\right]$$

$$\geq \frac{(1 - \alpha)^2}{8} \ \min\left\{\mathbb{G}_1\left(\frac{1 - \alpha}{2}\right), 1 - \mathbb{G}_0\left(\frac{1 + \alpha}{2}\right)\right\},$$

where we also used the fact that each term on the left hand-side of the preceding inequality is nonnegative. Substituting this into Lemma 5, the desired result follows. ∎

The preceding lemma characterizes the improvements in the probability of making the correct decision between an agent and his neighbor. To study the limiting behavior of these improvements, we introduce the function $\overline{\mathcal{Z}} : [1/2, 1] \to [1/2, 1]$ defined by

$$\overline{\mathcal{Z}}(\alpha) = \alpha + \frac{(1 - \alpha)^2}{8} \ \min\left\{\mathbb{G}_1\left(\frac{1 - \alpha}{2}\right), 1 - \mathbb{G}_0\left(\frac{1 + \alpha}{2}\right)\right\}. \qquad (10)$$

Lemma 6 establishes that for n, which has $B(n) = \{b\}$, we have

$$\mathbb{P}_\sigma(x_n = \theta | B(n) = \{b\}) \geq \overline{\mathcal{Z}}\left(\mathbb{P}_\sigma(x_b = \theta)\right), \qquad (11)$$

i.e., the function $\overline{\mathcal{Z}}$ acts as an improvement function for the evolution of the probability of making the correct decision. The function $\overline{\mathcal{Z}}(\cdot)$ has several important properties, which are formally stated in the next lemma.

Lemma 7 *The function* $\overline{\mathcal{Z}} : [1/2, 1] \to [1/2, 1]$ *given in (10) satisfy the following properties:*

(a) *The function* $\overline{\mathcal{Z}}$ *has no upwards jumps. That is, for any* $\alpha \in [1/2, 1]$,

$$\overline{\mathcal{Z}}(\alpha) = \lim_{r \uparrow \alpha} \overline{\mathcal{Z}}(r) \geq \lim_{r \downarrow \alpha} \overline{\mathcal{Z}}(r).$$

(b) *For any* $\alpha \in [1/2, 1]$, $\overline{\mathcal{Z}}(\alpha) \geq \alpha$.

(c) *If the private beliefs are unbounded, then for any* $\alpha \in [1/2, 1)$, $\overline{\mathcal{Z}}(\alpha) > \alpha$.

Proof. Since $\mathbb{G}_0$ and $\mathbb{G}_1$ are cumulative distribution functions, they cannot have downwards jumps, i.e., for each $j \in \{0, 1\}$, $\lim_{r \uparrow \alpha} \mathbb{G}_j(r) \leq \lim_{r \downarrow \alpha} \mathbb{G}_j(r)$ for any $\alpha \in [1/2, 1]$, establishing Part (a). Part (b) follows from the fact that cumulative distribution functions take values in $[0, 1]$. For Part (c), suppose that for some $\alpha \in [1/2, 1)$, $\mathcal{Z}(\alpha) = \alpha$. This implies that

$$\min\left\{\mathbb{G}_1\left(\frac{1 - \alpha}{2}\right), 1 - \mathbb{G}_0\left(\frac{1 + \alpha}{2}\right)\right\} = 0. \qquad (12)$$

However, from the assumption on the private beliefs, we have that for all $\alpha \in (0,1)$ and any $j \in \{0,1\}$, $\mathbb{G}_j(\alpha) \in (0,1)$, contradicting Eq. (12). $\blacksquare$

The properties of the $\overline{\mathcal{Z}}$ function will be used in the analysis of asymptotic learning in general networks in subsection 6.3. The analysis of asymptotic learning requires the relevant improvement function to be both continuous and monotone. However, $\overline{\mathcal{Z}}$ does not necessarily satisfy these properties. We next construct a related function $\mathcal{Z} : [1/2,1] \rightarrow [1/2,1]$ that satisfies these properties and can be used as the improvement function in the asymptotic analysis. Let $\mathcal{Z}$ be defined as:

$$\mathcal{Z}(\alpha) = \frac{1}{2}\left(\alpha + \sup_{r \in [1/2,\alpha]} \overline{\mathcal{Z}}(r)\right). \tag{13}$$

This function shares the same "improvement" properties as $\overline{\mathcal{Z}}$, but is also nondecreasing and continuous. The properties of the function $\mathcal{Z}(\cdot)$ stated in the following lemma.

Lemma 8 *The function $\mathcal{Z} : [1/2,1] \rightarrow [1/2,1]$ given in (13) satisfy the following properties:*

(a) *For any $\alpha \in [1/2,1]$, $\mathcal{Z}(\alpha) \geq \alpha$.*

(b) *If the private beliefs are unbounded, then for any $\alpha \in [1/2,1)$, $\mathcal{Z}(\alpha) > \alpha$.*

(c) *The function $\mathcal{Z}$ is increasing and continuous.*

Proof. Parts (a) and (b) follow immediately from Lemma 7, parts (b) and (c) respectively. The function $\sup_{r \in [1/2,\alpha]} \overline{\mathcal{Z}}(r)$ is nondecreasing and the function α is increasing, therefore the average of these two functions, which is $\mathcal{Z}$, is an increasing function, establishing the first part of part (c).

We finally show that $\mathcal{Z}$ is a continuous function. We first show $\mathcal{Z}(\alpha)$ is continuous for all $\alpha \in [1/2,1)$. To obtain a contradiction, assume that $\mathcal{Z}$ is discontinuous at some $\alpha^* \in [1/2,1)$. This implies that $\sup_{r \in [1/2,\alpha]} \overline{\mathcal{Z}}(r)$ is discontinuous at α^*. Since $\sup_{r \in [1/2,\alpha]} \overline{\mathcal{Z}}(r)$ is a nondecreasing function, we have

$$\lim_{\alpha \downarrow \alpha^*} \sup_{r \in [1/2,\alpha]} \overline{\mathcal{Z}}(r) > \sup_{r \in [1/2,\alpha^*]} \overline{\mathcal{Z}}(r),$$

from which it follows that there exists some $\epsilon > 0$ such that for any $\delta > 0$

$$\sup_{r \in [1/2,\alpha^*+\delta]} \overline{\mathcal{Z}}(r) > \overline{\mathcal{Z}}(\alpha) + \epsilon \qquad \text{for all } \alpha \in [1/2,\alpha^*).$$

This contradicts the fact that the function $\overline{\mathcal{Z}}$ does not have an upward jump [cf. Lemma 7 (a)], and establishes the continuity of $Z(\alpha)$ for all $\alpha \in [1/2,1)$. The continuity of the function $\mathcal{Z}(\alpha)$ at $\alpha = 1$ follows from part (a). $\blacksquare$

The next proposition shows that the function $\mathcal{Z}$ is also a (strong) improvement function for the evolution of the probability of making the correct decision.

Proposition 4 (Strong Improvement Principle) *Let $B(n) = \{b\}$ for some n. Let $\sigma \in \Sigma^*$ be an equilibrium. Then, we have*

$$\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\}) \geq \mathcal{Z}\left(\mathbb{P}_\sigma(x_b = \theta)\right). \tag{14}$$

Proof. Let α denote $\mathbb{P}_\sigma(x_b = \theta)$. If $\mathcal{Z}(\alpha) = \alpha$, then the result follows immediately from Proposition 2. Suppose next that $\mathcal{Z}(\alpha) > \alpha$. This implies that $\mathcal{Z}(\alpha) < \sup_{r \in [1/2,\alpha]} \overline{\mathcal{Z}}(r)$. Therefore, there exists some $\overline{\alpha} \in [1/2, \alpha]$ such that

$$\overline{\mathcal{Z}}(\overline{\alpha}) > \mathcal{Z}(\alpha). \tag{15}$$

We next show that $\mathbb{P}_\sigma(x_n = \theta | B(n) = b) \geq \overline{\mathcal{Z}}(\overline{\alpha})$. Agent n can always (privately) make the information from his observation of x_b coarser (i.e., not observe x_b according to some probability). Let the observation thus generated by agent n be denoted by $\tilde{x}_b$, and suppose that it is given by

$$\tilde{x}_b = \begin{cases} x_b, & \text{with probability } (2\overline{\alpha} - 1)/(2\alpha - 1) \\ 0, & \text{with probability } (\alpha - \overline{\alpha})/(2\alpha - 1) \\ 1, & \text{with probability } (\alpha - \overline{\alpha})/(2\alpha - 1), \end{cases}$$

where the realizations of $\tilde{x}_b$ are independent from agent n's information set. Next observe that $\mathbb{P}_\sigma(\tilde{x}_b = \theta) = \overline{\alpha}$. Then, Lemma 6 implies that $\mathbb{P}_\sigma(x_n = \theta | B(n) = b) \geq \overline{\mathcal{Z}}(\overline{\alpha})$. Since $\overline{\mathcal{Z}}(\overline{\alpha}) > \mathcal{Z}(\alpha)$ [cf. Eq. (15)], the desired result follows. ∎

6.3 Learning from Multiple Agents

In this subsection, we generalize the results of the previous subsection to an arbitrary network topology. We first present a stronger version of the information monotonicity relation (cf. Proposition 2), where the amount of improvement is given by the improvement function $\mathcal{Z}$ defined in Eq. (14). Even though a full characterization of equilibrium decisions in general network topologies is a nontractable problem (recall the discussion in subsection 5.3), it is possible to establish an analogue of Proposition 4, that is, a *generalized strong improvement principle*, which provides a lower bound on the amount of increase in the probabilities of making the correct decision. The idea of the proof is to show that improvements can be no less than the case in which each individual's neighborhood consisted of a single agent.

Proposition 5 (Generalized Strong Improvement Principle) *For any $n \in \mathbb{N}$, any set $\mathfrak{B} \subseteq \{1, ..., n-1\}$ and any equilibrium $\sigma \in \Sigma^*$, we have*

$$\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \mathfrak{B}) \geq \mathcal{Z}\left(\max_{b \in \mathfrak{B}} \mathbb{P}_\sigma(x_b = \theta)\right).$$

Proof. Given an equilibrium $\sigma \in \Sigma^*$ and agent n, let h_σ be a function that maps any subset of $\{1, \ldots, n-1\}$ to an element of $\{1, \ldots, n-1\}$ such that for any $\mathfrak{B} \subset \{1, \ldots, n-1\}$, we have

$$h_\sigma(\mathfrak{B}) \in \arg\max_{b \in \mathfrak{B}} \mathbb{P}_\sigma(x_b = \theta). \tag{16}$$

We define w_n as the decision that maximizes the conditional probability of making the correct decision given the private signal s_n and the decision of the agent $h_\sigma(B(n))$, i.e.,

$$w_n \in \arg\max_{y \in \{0,1\}} \mathbb{P}_\sigma \left(y = \theta \mid s_n, x_{h_\sigma(B(n))} \right).$$

The equilibrium decision x_n of agent n satisfies

$$\mathbb{P}_\sigma(x_n = \theta \mid s_n, B(n), x_k, \ k \in B(n)) \geq \mathbb{P}_\sigma(w_n = \theta \mid s_n, B(n), x_k, \ k \in B(n)),$$

[cf. the characterization of the equilibrium decision rule in Eq. (2)]. Integrating over all possible private signals and decisions of neighbors, we obtain for any $\mathcal{B} \subset \{1, \ldots, n-1\}$,

$$\mathbb{P}_\sigma \left(x_n = \theta \mid B(n) = \mathcal{B} \right) \geq \mathbb{P}_\sigma \left(w_n = \theta \mid B(n) = \mathcal{B} \right). \tag{17}$$

Because w_n is an optimal choice given a single observation, Eq. (14) holds and yields

$$\mathbb{P}_\sigma(w_n = \theta \mid B(n) = \mathcal{B}) \geq \mathcal{Z} \left(\mathbb{P}_\sigma(x_{h_\sigma(\mathcal{B})} = \theta) \right). \tag{18}$$

Combining Eqs. (16), (17) and (18) we obtain the desired result. ∎

This proposition is a key result, since it shows that, under unbounded private beliefs, there are improvements in payoffs (probabilities of making correct decisions) that are bounded away from zero. We will next use this generalized strong improvement principle to prove Theorem 2. The proof involves showing that under the expanding observations and the unbounded private beliefs assumptions, the amount of improvement in the probabilities of making the correct decision given by $\mathcal{Z}$ accumulates until asymptotic learning is reached.

6.4 Proof of Theorem 2

The proof consists of two parts. In the first part of the proof, we construct two sequences $\{\alpha_k\}$ and $\{\phi_k\}$ such that for all $k \geq 0$, there holds

$$\mathbb{P}_\sigma(x_n = \theta) \geq \phi_k \quad \text{for all } n \geq \alpha_k. \tag{19}$$

The second part of the proof shows that ϕ_k converges to 1, thus establishing the result.

Given some integer $K > 0$ and scalar $\epsilon > 0$, let $N(K, \epsilon) > 0$ be an integer such that for all $n \geq N(K, \epsilon)$,

$$\mathbb{Q}_n \left(\max_{b \in B(n)} b < K \right) < \epsilon,$$

(such an integer exists in view of the fact that, by hypothesis, the network topology features expanding observations). We let $\alpha_1 = 1$ and $\phi_1 = 1/2$ and define the sequences $\{\alpha_k\}$ and ϕ_k recursively by

$$\alpha_{k+1} = N \left(\alpha_k, \frac{1}{2} \left[1 - \frac{\phi_k}{\mathcal{Z}(\phi_k)} \right] \right), \qquad \phi_{k+1} = \frac{\phi_k + \mathcal{Z}(\phi_k)}{2}.$$

Using the fact that the range of the function $\mathcal{Z}$ is $[1/2, 1]$, it can be seen that $\phi_k \in [1/2, 1]$ for all k, therefore the preceding sequences are well-defined.

We use induction on the index k to prove relation (19). Since σ is an equilibrium, we have

$$\mathbb{P}_\sigma(x_n = \theta) \geq \frac{1}{2} \qquad \text{for all } n \geq 1,$$

which together with $\alpha_1 = 1$ and $\phi_1 = 1/2$ shows relation (19) for $k = 1$. Assume that the relation (19) holds for an arbitrary k, i.e.,

$$\mathbb{P}_\sigma(x_j = \theta) \geq \phi_k \text{ for all } j \geq \alpha_k. \tag{20}$$

Consider some agent n with $n \geq \alpha_{k+1}$. By integrating the relation from Lemma 5 over all possible neighborhoods $B(n)$, we obtain

$$\mathbb{P}_\sigma(x_n = \theta) \geq \mathbb{E}_{B(n)} \left[\mathcal{Z} \left(\max_{b \in B(n)} \mathbb{P}_\sigma(x_b = \theta) \right) \right],$$

where $\mathbb{E}_{B(n)}$ denotes the expectation with respect to the neighborhood $B(n)$ (i.e., the weighted sum over all possible neighborhoods $B(n)$). We can rewrite the preceding as

$$
\begin{aligned}
\mathbb{P}_\sigma(x_n = \theta) \geq{}& \mathbb{E}_{B(n)} \left[\mathcal{Z} \left(\max_{b \in B(n)} \mathbb{P}_\sigma(x_b = \theta) \right) \;\middle|\; \max_{b \in B(n)} b \geq \alpha_k \right] \mathbb{Q}_n \left(\max_{b \in B(n)} b \geq \alpha_k \right) \\
&+ \mathbb{E}_{B(n)} \left[\mathcal{Z} \left(\max_{b \in B(n)} \mathbb{P}_\sigma(x_b = \theta) \right) \;\middle|\; \max_{b \in B(n)} b < \alpha_k \right] \mathbb{Q}_n \left(\max_{b \in B(n)} b < \alpha_k \right).
\end{aligned}
$$

Since the terms on the right hand-side of the preceding relation are nonnegative, this implies that

$$\mathbb{P}_\sigma(x_n = \theta) \geq \mathbb{E}_{B(n)} \left[\mathcal{Z} \left(\max_{b \in B(n)} \mathbb{P}_\sigma(x_b = \theta) \right) \;\middle|\; \max_{b \in B(n)} b \geq \alpha_k \right] \mathbb{Q}_n \left(\max_{b \in B(n)} b \geq \alpha_k \right).$$

Since $\max_{b \in B(n)} b \geq \alpha_k$, Eq. (20) implies that

$$\max_{b \in B(n)} \mathbb{P}_\sigma(x_b = \theta) \geq \phi_k.$$

Since the function $\mathcal{Z}$ is nondecreasing [cf. Lemma 8(c)], combining the preceding two relations, we obtain

$$
\begin{aligned}
\mathbb{P}_\sigma(x_n = \theta) \geq{}& \mathbb{E}_{B(n)} \left[\mathcal{Z}(\phi_k) \;\middle|\; \max_{b \in B(n)} b \geq \alpha_k \right] \mathbb{Q}_n \left(\max_{b \in B(n)} b \geq \alpha_k \right) \\
={}& \mathcal{Z}(\phi_k) \mathbb{Q}_n \left(\max_{b \in B(n)} b \geq \alpha_k \right),
\end{aligned}
$$

where the equality follows since the sequence $\{\phi_k\}$ is deterministic. Using the definition of α_k, this implies that

$$\mathbb{P}_\sigma(x_n = \theta) \geq \mathcal{Z}(\phi_k) \frac{1}{2} \left[1 + \frac{\phi_k}{\mathcal{Z}(\phi_k)} \right] = \phi_{k+1},$$

thus completing the induction.

We finally prove that $\phi_k \to 1$ as $k \to \infty$. Since $\mathcal{Z}(\alpha) \geq \alpha$ for all $\alpha \in [1/2, 1]$ [cf. Lemma 8(a)], it follows from the definition of ϕ_k that $\{\phi_k\}_{k \in \mathbb{N}}$ is a nondecreasing sequence. It is also bounded and therefore it converges to some ϕ^*. Taking the limit in the definition of ϕ_k, we obtain

$$2\phi^* = 2 \lim_{k \to \infty} \phi_k = \lim_{k \to \infty} \left[\phi_k + \mathcal{Z}(\phi_k) \right] = \phi^* + \mathcal{Z}(\phi^*),$$

where the third equality follows since $\mathcal{Z}$ is a continuous function [cf. Lemma 8(c)]. This shows that $\phi^* = \mathcal{Z}(\phi^*)$, i.e., ϕ^* is a fixed point of $\mathcal{Z}$. Since the private beliefs are unbounded, the unique fixed point of $\mathcal{Z}$ is 1, showing that $\phi_k \to 1$ as $k \to \infty$ and completing the proof. ∎

6.5   Proofs of Corollaries 1 and 2

Proof of Corollary 1. We first show that if $C \geq 1$, then the network topology has nonexpanding observations. To show this, we set $K = 1$ in Definition 3 and show that the probability of infinitely many agents having empty neighborhoods is uniformly bounded away from 0. We first consider the case $C > 1$. Then, the probability that the neighborhood of agent $n + 1$ is the empty set is given by

$$\mathbb{Q}_{n+1}(B(n+1) = \emptyset) = \left(1 - \frac{A}{n^C} \right)^n,$$

which converges to 1 as n goes to infinity. If $C = 1$, then

$$\lim_{n \to \infty} \mathbb{Q}_{n+1}(B(n+1) = \emptyset) = \lim_{n \to \infty} \left(1 - \frac{A}{n} \right)^n = e^{-A}.$$

Therefore, for infinitely many agents, $\mathbb{Q}_{n+1}(B(n+1) = \emptyset) \geq e^{-A}/2$. The preceding show that the network topology has nonexpanding observations for $C \geq 1$, hence the result follows from Theorem 1.

We next assume that $C < 1$. For any K and all $n \geq K$, we have

$$\mathbb{Q}_{n+1}\left(\max_{b \in B(n+1)} b \leq K \right) = \left(1 - \frac{A}{n^C} \right)^{n-K},$$

which converges to 0 as n goes to infinity. Hence the network topology is expanding in observations and the result follows from Theorem 2. ∎

Proof of Corollary 2. We show that $\{L(n)\}_{n \in \mathbb{N}}$ goes to infinity if and only if the deterministic sequence $\{\max_{b \in B(n)} b\}_{n \in \mathbb{N}}$ goes to infinity. Suppose first $\{L(n)\}_{n \in \mathbb{N}}$ diverges to infinity. Then, for every K there exists N such that for all $n \geq N$, $L(n) \geq K$. Note that

$$L(n) \leq 1 + \max_{b \in B(n)} b$$

because the longest information path must be a subset of the sequence $(1, 2, ..., \max_{b \in B(n)} b, n)$. So, for $n \geq N$, if $L(n) \geq K$, then $\max_{b \in B(n)} b > K$, thus proving the first part of the

lemma. Suppose next that $\{\max_{b \in B(n)} b\}_{n \in \mathbb{N}}$ goes to infinity as n goes to infinity. We show by induction that for each $d \in \mathbb{N}$, there exists some integer C_d such that $L(n) \geq d$ for all $n \geq C_d$. Since $L(n) \geq 1$ for all n, then $C_1 = 1$. Assume such C_d exists for some d. Then, we show that such a C_{d+1} also exists. Since $\{\max_{b \in B(n)} b\}_{n \in \mathbb{N}}$ goes to infinity, there exists some N_d such that for all $n \geq N_d$,

$$\max_{b \in B(n)} b \geq C_d.$$

Now, for any $n \geq N_d$, there exists a path with size d up to some $k \geq C_d$ and then another observation from k to n, therefore $L(n) \geq d + 1$. Hence, $C_{d+1} = N_d$. ∎

7 No Learning with Bounded Private Beliefs

In this section, we study asymptotic learning when the private beliefs are bounded. In particular, we consider the three cases of Theorem 3 separately. For parts (a) and (b), we sketch the proofs highlighting the intuition for why asymptotic learning does not occur (detailed proofs are contained in Appendix B). The proofs of part (c) and Corollary 3 are presented here.

The next proposition states the result corresponding to part (a) of Theorem 3. This result is proved in Smith and Sorensen (2000). In Appendix B, we provide an alternative proof that illustrates the parallel with the remainder of Theorem 3.

Proposition 6 *Assume that the signal structure* $(\mathbb{F}_0, \mathbb{F}_1)$ *has bounded private beliefs and* $B(n) = \{1, \ldots, n-1\}$ *for all* n. *Then, asymptotic learning does not occur in any equilibrium.*

Briefly, this result follows by showing that under bounded private beliefs there exist $0 < \underline{\Delta} < \overline{\Delta} < 1$ such that the social belief of each agent belongs to the interval $[\underline{\Delta}, \overline{\Delta}]$. This establishes that either individuals always make use of their own signals in taking their actions, leading to a positive probability of a mistake, or individuals follow a potentially incorrect social belief.

The next proposition shows that asymptotic learning fails when each individual observes the action of at most one agent from the past.

Proposition 7 *Assume that the signal structure* $(\mathbb{F}_0, \mathbb{F}_1)$ *has bounded private beliefs and* $|B(n)| \leq 1$ *for all* n. *Then, asymptotic learning does not occur in any equilibrium.*

The proof of this result follows by establishing an upper bound on the amount of improvement in the ex-ante probability of the correct action, hence providing a converse to the Strong Improvement Principle (cf. Propositions 4 and 5). Under bounded private beliefs, this upper bound is uniformly bounded away from 1, establishing no learning.

Finally, the following proposition establishes Part (c) of Theorem 3.

Proposition 8 *Assume that the signal structure* $(\mathbb{F}_0, \mathbb{F}_1)$ *has bounded private beliefs. Assume that there exists some constant M such that $|B(n)| \leq M$ for all n and*

$$\lim_{n \to \infty} \max_{b \in B(n)} b = \infty \qquad \text{with probability } 1.$$

Then, asymptotic learning does not occur in any equilibrium.

Proof. We start with the following lemma, which will be used subsequently in the proof.

Lemma 9 *Assume that asymptotic learning occurs in some equilibrium σ, i.e., we have $\lim_{n \to \infty} \mathbb{P}_\sigma(x_n = \theta) = 1$. For some constant K, let $\mathcal{D}$ be the set of all subsets of $\{1, ..., K\}$. Then,*

$$\lim_{n \to \infty} \min_{D \in \mathcal{D}} \mathbb{P}_\sigma(x_n = \theta \mid x_k = 1, \ k \in D) = 1.$$

Proof. First note that since the event $x_k = 1$ for all $k \leq K$ is the intersection of events $x_k = 1$ for each $k \leq K$,

$$\min_{D \in \mathcal{D}} \mathbb{P}_\sigma(x_k = 1, \ k \in D) = \mathbb{P}_\sigma(x_k = 1, \ k \leq K).$$

Let $\Delta = \mathbb{P}_\sigma(x_k = 1, \ k \leq K)$. Fix some $\tilde{D} \in \mathcal{D}$. Then,

$$\mathbb{P}_\sigma(x_k = 1, \ k \in \tilde{D}) \geq \Delta > 0,$$

where the second inequality follows from the fact that there is a positive probability of the first K agents choosing $x_n = 1$. Let $\mathcal{A} = \{0, 1\}^{|\tilde{D}|}$, i.e., $\mathcal{A}$ is the set of all possible actions for the set of agents $\tilde{D}$. Then,

$$\mathbb{P}_\sigma(x_n = \theta) = \sum_{a_k \in \mathcal{A}} \mathbb{P}_\sigma(x_n = \theta \mid x_k = a_k, \ k \in \tilde{D}) \mathbb{P}_\sigma(x_k = a_k, \ k \in \tilde{D}).$$

Since $\mathbb{P}_\sigma(x_n = \theta)$ converges to 1 and all elements in the sequence $\mathbb{P}_\sigma(x_k = 1, \ k \in \tilde{D})$ are greater than or equal to $\Delta > 0$, it follows that the sequence $\mathbb{P}_\sigma(x_n = \theta \mid x_k = 1, \ k \in \tilde{D})$ also converges to 1. Hence, for each $\epsilon > 0$, there exists some $N_\epsilon(\tilde{D})$ such that for all $n \geq N_\epsilon(\tilde{D})$,

$$\mathbb{P}_\sigma(x_n = \theta \mid x_k = 1, \ k \in \tilde{D}) \geq 1 - \epsilon.$$

Therefore, for any $\epsilon > 0$,

$$\min_{D \in \mathcal{D}} \mathbb{P}_\sigma(x_n = \theta \mid x_k = 1, \ k \in D) \geq 1 - \epsilon \quad \text{for all} \quad n \geq \max_{D \in \mathcal{D}} N_\epsilon(D),$$

thus completing the proof. ∎

Proof of Proposition 8. To arrive at a contradiction, we assume that in some equilibrium $\sigma \in \Sigma^*$, $\lim_{n \to \infty} \mathbb{P}_\sigma(x_n = \theta) = 1$. The key part of the proof is to show that this implies

$$\lim_{n \to \infty} \mathbb{P}_\sigma(\theta = 1 \mid x_k = 1, \ k \in B(n)) = 1. \tag{21}$$

To prove this claim, we show that for any $\epsilon > 0$, there exists some $\tilde{K}(\epsilon)$ such that for any neighborhood $\mathfrak{B}$ with $|\mathfrak{B}| \leq M$ and $\max_{b \in \mathfrak{B}} b \geq \tilde{K}(\epsilon)$ we have

$$\mathbb{P}_\sigma(\theta = 1 \mid x_k = 1, \ k \in \mathfrak{B}) \geq 1 - \epsilon. \tag{22}$$

In view of the assumption that $\max_{b \in B(n)} b$ converges to infinity with probability 1, this implies the desired claim (21).

For a fixed $\epsilon > 0$, we define $\tilde{K}(\epsilon)$ as follows: We recursively construct M thresholds $K_0 < ... < K_{M-1}$ and let $\tilde{K}(\epsilon) = K_{M-1}$. We consider an arbitrary neighborhood $\mathfrak{B}$ with $|\mathfrak{B}| \leq M$ and $\max_{b \in \mathfrak{B}} b \geq K_{M-1}$, and for each $d \in \{0, ..., M-1\}$, define the sets

$$\mathfrak{B}_d = \{b \in \mathfrak{B} : b \geq K_d\} \text{ and } \mathfrak{C}_d = \{b \in \mathfrak{B} : b < K_{d-1}\},$$

where $\mathfrak{C}_0 = \emptyset$. With this construction, it follows that there exists at least one $d \in \{0, ..., M-1\}$ such that $\mathfrak{B} = \mathfrak{B}_d \cup \mathfrak{C}_d$, in which case we say $\mathfrak{B}$ is of type d. We show below that for any $\mathfrak{B}$ of type d, we have

$$\mathbb{P}_\sigma(\theta = 1 \mid x_k = 1, \ k \in \mathfrak{B}_d \cup \mathfrak{C}_d) \geq 1 - \epsilon, \tag{23}$$

which implies the relation in (22).

We first define K_0 and show that for any $\mathfrak{B}$ of type 0, relation (23) holds. Since $\lim_{n \to \infty} \mathbb{P}_\sigma(x_n = \theta) = 1$ by assumption, there exists some N_0 such that for all $n \geq N_0$,

$$\mathbb{P}_\sigma(x_n = \theta) \geq 1 - \frac{\epsilon}{2M}.$$

Let $K_0 = N_0$. Let $\mathfrak{B}$ be a neighborhood of type 0, implying that $\mathfrak{B} = \mathfrak{B}_0$ and all elements $b \in \mathfrak{B}_0$ satisfy $b \geq K_0$. By using a union bound, the preceding inequality implies

$$\mathbb{P}_\sigma(x_k = \theta, \ k \in \mathfrak{B}_0) \geq 1 - \sum_{k \in \mathfrak{B}_0} \mathbb{P}_\sigma(x_k \neq \theta) \geq 1 - \frac{\epsilon}{2}.$$

Hence, we have

$$\mathbb{P}_\sigma(x_k = \theta, \ k \in \mathfrak{B}_0 \mid \theta = 1)\frac{1}{2} + \mathbb{P}_\sigma(x_k = \theta, \ k \in \mathfrak{B}_0 \mid \theta = 0)\frac{1}{2} \geq 1 - \frac{\epsilon}{2},$$

and for any $j \in \{0, 1\}$,

$$\mathbb{P}_\sigma(x_k = \theta, \ k \in \mathfrak{B}_0 \mid \theta = j) \geq 1 - \epsilon. \tag{24}$$

Therefore, for any such $\mathfrak{B}_0$,

$$\mathbb{P}_\sigma(\theta = 1 \mid x_k = 1, \ k \in \mathfrak{B}_0) = \left[1 + \frac{\mathbb{P}_\sigma(x_k = 1, \ k \in \mathfrak{B}_0 \mid \theta = 0)\mathbb{P}_\sigma(\theta = 0)}{\mathbb{P}_\sigma(x_k = 1, \ k \in \mathfrak{B}_0 \mid \theta = 1)\mathbb{P}_\sigma(\theta = 1)}\right]^{-1}$$

$$\geq \left[1 + \frac{\epsilon}{1 - \epsilon}\right]^{-1} = 1 - \epsilon,$$

showing that relation (23) holds for any $\mathfrak{B}$ of type 0.

31

We proceed recursively, i.e., given K_{d-1} we define K_d and show that relation (23) holds for any neighborhood $\mathfrak{B}$ of type d. Lemma 9 implies that

$$\lim_{n \to \infty} \min_{D \subseteq \{1,\ldots,K_{d-1}-1\}} \mathbb{P}_\sigma(x_n = \theta \mid x_k = 1, \ k \in D) = 1.$$

Therefore, for any $\delta > 0$, there exists some K_d such that for all $n \geq K_d$,

$$\min_{D \subseteq \{1,\ldots,K_{d-1}-1\}} \mathbb{P}_\sigma(x_n = \theta \mid x_k = 1, \ k \in D) \geq 1 - \delta\epsilon.$$

From the equation above and definition of $\mathfrak{C}_d$ it follows that for any $\mathfrak{C}_d$,

$$\mathbb{P}_\sigma(x_n = \theta \mid x_k = 1, \ k \in \mathfrak{C}_d) \geq 1 - \delta\epsilon.$$

By a union bound,

$$\begin{aligned}
\mathbb{P}_\sigma(x_k = \theta, \ k \in \mathfrak{B}_d \mid x_k = 1, \ k \in \mathfrak{C}_d) &\geq& 1 - \sum_{k \in \mathfrak{B}_d} \mathbb{P}_\sigma(x_k \neq \theta \mid x_k = 1, \ k \in \mathfrak{C}_d) \\
&\geq& 1 - (M - d)\delta\epsilon.
\end{aligned}$$

Repeating the argument from Eq. (24), for any $j \in \{0, 1\}$,

$$\mathbb{P}_\sigma(x_k = \theta, \ k \in \mathfrak{B}_d \mid \theta = j, x_k = 1, \ k \in \mathfrak{C}_d) \geq 1 - \frac{(M - d)\delta\epsilon}{\mathbb{P}_\sigma(\theta = j \mid x_k = 1, \ k \in \mathfrak{C}_d)}.$$

Hence, for any such $\mathfrak{B}_d$,

$$\begin{aligned}
&\mathbb{P}_\sigma(\theta = 1 \mid x_k = 1, \ k \in \mathfrak{B}_d \cup \mathfrak{C}_d) \\
&= \left[1 + \frac{\mathbb{P}_\sigma(x_k = 1, \ k \in \mathfrak{B}_d \mid \theta = 0, x_k = 1, \ k \in \mathfrak{C}_d)\mathbb{P}_\sigma(\theta = 0, x_k = 1, \ k \in \mathfrak{C}_d)}{\mathbb{P}_\sigma(x_k = 1, \ k \in \mathfrak{B}_d \mid \theta = 1, x_k = 1, \ k \in \mathfrak{C}_d)\mathbb{P}_\sigma(\theta = 1, x_k = 1, \ k \in \mathfrak{C}_d)}\right]^{-1} \\
&\geq \left[1 + \frac{\frac{(M-d)\delta\epsilon}{\mathbb{P}_\sigma(\theta=0 \mid x_k=1, \ k \in \mathfrak{C}_d)}\mathbb{P}_\sigma(\theta = 0, x_k = 1, \ k \in \mathfrak{C}_d)}{\left(1 - \frac{(M-d)\delta\epsilon}{\mathbb{P}_\sigma(\theta=1 \mid x_k=1, \ k \in \mathfrak{C}_d)}\right)\mathbb{P}_\sigma(\theta = 1, x_k = 1, \ k \in \mathfrak{C}_d)}\right]^{-1} \\
&= 1 - \frac{(M - d)\delta\epsilon}{\mathbb{P}_\sigma(\theta = 1 \mid x_k = 1, \ k \in \mathfrak{C}_d)}.
\end{aligned}$$

Choosing

$$\delta = \left(\frac{1}{M - d}\right) \min_{D \subseteq \{1,\ldots,K_{d-1}-1\}} \mathbb{P}_\sigma(\theta = 1 \mid x_k = 1, \ k \in D),$$

we obtain that for any neighborhood $\mathfrak{B}$ of type d

$$\mathbb{P}_\sigma(\theta = 1 \mid x_k = 1, \ k \in \mathfrak{B}_d \cup \mathfrak{C}_d) \geq 1 - \epsilon.$$

This proves that Eq. (23) holds for any neighborhood $\mathfrak{B}$ of type d, and completing the proof of Eq. (22) and therefore of Eq. (21).

Since the private beliefs are bounded, we have $\underline{\beta} > 0$. By Eq. (21), there exists some $\overline{N}$ such that

$$\mathbb{P}_\sigma(\theta = 1 | x_k = 1, \ k \in B(n)) \geq 1 - \frac{\underline{\beta}}{2} \qquad \text{for all } n \geq \overline{N}.$$

Suppose the first $\overline{N}$ agents choose 1, i.e., $x_k = 1$ for all $k \leq \overline{N}$, which is an event with positive probability for any state of the world θ. We now prove inductively that this event implies that $x_n = 1$ for all $n \in \mathbb{N}$. Suppose it holds for some $n \geq \overline{N}$. Then, by Eq. (22),

$$\mathbb{P}_\sigma(\theta = 1 \mid s_{n+1}) + \mathbb{P}_\sigma(\theta = 1 \mid x_k = 1, \ \in B(n+1)) \geq \underline{\beta} + 1 - \frac{\underline{\beta}}{2} > 1.$$

By Lemma 1, this implies that $x_{n+1} = 1$. Hence, we conclude there is a positive probability $x_n = 1$ for all $n \in \mathbb{N}$ in any state of the world, contradicting $\lim_{n \to \infty} \mathbb{P}_\sigma(x_n = \theta) = 1$, and completing the proof. ∎

Proof of Corollary 3. For $M = 1$, the result follows from Theorem 3 part (b). For $M \geq 2$, we show that, under the assumption on the network topology, $\max_{b \in B(n)} b$ goes to infinity with probability one. To arrive at a contradiction, suppose this is not true. Then, there exists some $K \in \mathbb{N}$ and scalar $\epsilon > 0$ such that

$$\mathbb{Q}\left(\max_{b \in B(n)} b \leq K \text{ for infinitely many } n\right) \geq \epsilon.$$

By the Borel-Cantelli Lemma (see, e.g., Breiman, Lemma 3.14, p. 41), this implies that

$$\sum_{n=1}^{\infty} \mathbb{Q}_n\left(\max_{b \in B(n)} b \leq K\right) = \infty.$$

Since the samples are all uniformly drawn and independent, for all $n \geq 2$,

$$\mathbb{Q}_n\left(\max_{b \in B(n)} b \leq K\right) = \left(\frac{\min\{K, n-1\}}{n-1}\right)^M.$$

Therefore,

$$\sum_{n=1}^{\infty} \mathbb{Q}_n\left(\max_{b \in B(n)} b \leq K\right) = 1 + \sum_{n=1}^{\infty} \left(\frac{\min\{K, n-1\}}{n}\right)^M \leq 1 + \sum_{n=1}^{\infty} \left(\frac{K}{n}\right)^M < \infty,$$

where the last inequality holds since $M \geq 2$. Hence, we obtain a contradiction. The result follows by using Theorem 3 part (c). ∎

8 Proof of Theorem 4

To simplify the exposition of the proof, we assume that the corresponding private belief distributions $(\mathbb{G}_0, \mathbb{G}_1)$ are continuous, which implies that the equilibrium is unique.

For each n, let $x^n = (x_1, \ldots, x_n)$ represent the sequence of decisions up to and including x_n. Let $q^*(x^n)$ denote the "social belief" when x^n is observed under equilibrium σ, i.e.,

$$q^*(x^n) = \mathbb{P}_\sigma(\theta = 1 \mid x^n).$$

The social belief $q^*(x^n)$ is a martingale and, by the martingale convergence theorem, converges with probability 1 to some random variable $\hat{q}$. Conditional on $\theta = 1$, the likelihood ratio

$$\frac{1 - q^*(x^n)}{q^*(x^n)} = \frac{\mathbb{P}_\sigma(\theta = 0 \mid x^n)}{\mathbb{P}_\sigma(\theta = 1 \mid x^n)}$$

is also a martingale [see Doob, 1953, Eq. (7.12)]. Therefore, conditional on $\theta = 1$, the ratio $(1 - q^*(x^n))/q^*(x^n)$ converges with probability 1 to some random variable $(1 - \hat{q}_1)/\hat{q}_1$. In particular, we have

$$\mathbb{E}_\sigma \left[\frac{1 - \hat{q}_1}{\hat{q}_1} \right] < \infty,$$

[see Breiman, Theorem 5.14], and therefore $\hat{q}_1 > 0$ with probability 1. Similarly, $q^*(x^n)/(1 - q^*(x^n))$ is a martingale conditional on $\theta = 0$ and converges with probability 1 to some random variable $\hat{q}_0/(1 - \hat{q}_0)$, where $\hat{q}_0 < 1$ with probability 1. Therefore,

$$\mathbb{P}_\sigma(\hat{q} > 0 \mid \theta = 1) = 1 \text{ and } \mathbb{P}_\sigma(\hat{q} < 1 \mid \theta = 0) = 1. \tag{25}$$

The key part of the proof is to show that the support of $\hat{q}$ is contained in the set $\{0, 1\}$. This fact combined with Eq. (25) guarantees that $\hat{q} = \theta$ (i.e., the agents that observe the entire history eventually know what the state of the world θ is).

To show that the support of $\hat{q}$ is contained in $\{0, 1\}$, we study the evolution dynamics of $q^*(x^n)$. Suppose $x_{n+1} = 0$. Using Bayes' Rule twice, we have

$$
\begin{aligned}
q^*((x^n, 0)) &= \frac{\mathbb{P}_\sigma(x_{n+1} = 0, x^n \mid \theta = 1)}{\sum_{j=0}^1 \mathbb{P}_\sigma(x_{n+1} = 0, x^n \mid \theta = j)} = \left[1 + \frac{\mathbb{P}_\sigma(x_{n+1} = 0, x^n \mid \theta = 0)}{\mathbb{P}_\sigma(x_{n+1} = 0, x^n \mid \theta = 1)} \right]^{-1} \\
&= \left[1 + \left(\frac{1}{q^*(x^n)} - 1 \right) \frac{\mathbb{P}_\sigma(x_{n+1} = 0 \mid \theta = 0, x^n)}{\mathbb{P}_\sigma(x_{n+1} = 0 \mid \theta = 1, x^n)} \right]^{-1}.
\end{aligned}
$$

To simplify notation, let

$$f_n(x^n) = \frac{\mathbb{P}_\sigma(x_{n+1} = 0 \mid \theta = 0, x^n)}{\mathbb{P}_\sigma(x_{n+1} = 0 \mid \theta = 1, x^n)}, \tag{26}$$

so that

$$q^*((x^n, 0)) = \left[1 + \left(\frac{1}{q^*(x^n)} - 1 \right) f_n(x^n) \right]^{-1}. \tag{27}$$

We next show that if $q^*(x^n) \geq 1/2$ for some n, we have

$$f_n(x^n) \geq 1 + \delta \qquad \text{for some } \delta > 0, \tag{28}$$

34

which will allow us to establish a bound on the difference between $q^*(x^n)$ and $q^*((x^n, 0))$. By conditioning on the neighborhood $B(n+1)$, we have for any $j \in \{0, 1\}$,

$$\mathbb{P}_\sigma(x_{n+1} = 0 \mid \theta = j, x^n) = \sum_{i=1}^{M} r_i(n+1)\mathbb{P}_\sigma(x_{n+1} = 0 \mid \theta = j, x^n, B(n+1) = C_i)$$

$$+ \left(1 - \sum_{i=1}^{M} r_i(n+1)\right) \mathbb{P}_\sigma(x_{n+1} = 0 \mid \theta = j, x^n, B(n+1) = \{1, ..., n\}).$$

For each $i = 1, \ldots, M$, we define

$$q_i(x^n) = \mathbb{P}_\sigma(\theta = 1 | x_k, k \in C_i).$$

Using the assumption that $\mathbb{G}_0$ and $\mathbb{G}_1$ are continuous, Lemma 1 implies that

$$\mathbb{P}_\sigma(x_{n+1} = 0 \mid \theta = j, x^n, B(n+1) = C_i) = \mathbb{P}_\sigma(p_{n+1} \le 1 - q_i(x^n) \mid \theta = j, x^n, B(n+1) = C_i)$$
$$= \mathbb{G}_j\left(1 - q_i(x^n)\right).$$

Therefore,

$$\mathbb{P}_\sigma(x_{n+1} = 0 \mid \theta = j, x^n) = \sum_{i=1}^{M} r_i(n+1)\mathbb{G}_j(1 - q_i(x^n)) + \left(1 - \sum_{i=1}^{M} r_i(n+1)\right)\mathbb{G}_j(1 - q^*(x^n)).$$

Hence, Eq. (26) can be rewritten as

$$f_n(x^n) = \frac{\sum_{i=1}^{M} r_i(n+1)\mathbb{G}_0(1 - q_i(x^n)) + \left(1 - \sum_{i=1}^{M} r_i(n+1)\right)\mathbb{G}_0(1 - q^*(x^n))}{\sum_{i=1}^{M} r_i(n+1)\mathbb{G}_1(1 - q_i(x^n)) + \left(1 - \sum_{i=1}^{M} r_i(n+1)\right)\mathbb{G}_1(1 - q^*(x^n))}.$$

Let $\mathcal{N} \subseteq \mathbb{N}$ be the set of all n such that $\sum_{i=1}^{M} r_i(n) \in (0, 1)$. The set $\mathcal{N}$ has infinitely many elements in view of the assumptions that $\lim_{n\to\infty} \sum_{i=1}^{M} r_i(n) = 0$ and $\sum_{n=1}^{\infty} \sum_{i=1}^{M} r_i(n) = \infty$ [cf. Eq. (3)]. To simplify notation, let $\omega_i(n) = r_i(n)/(1 - \sum_{k=1}^{M} r_k(n))$ for all $n \in \mathcal{N}$. Note that for all $n \in \mathcal{N}$, there exists some i such that $\omega_i(n) > 0$. Then, for any $n \in \mathcal{N}$,

$$f_n(x^n) = \frac{\sum_{i=1}^{M} \omega_i(n+1)\mathbb{G}_0(1 - q_i(x^n)) + \mathbb{G}_0(1 - q^*(x^n))}{\sum_{i=1}^{M} \omega_i(n+1)\mathbb{G}_1(1 - q_i(x^n)) + \mathbb{G}_1(1 - q^*(x^n))}. \tag{29}$$

If $q^*(x^n) \ge 1/2$, it follows from Lemma 3(c) that

$$\mathbb{G}_0(1 - q^*(x^n)) \ge \left(\frac{\mathbb{G}_0(1/2)}{\mathbb{G}_1(1/2)}\right)\mathbb{G}_1(1 - q^*(x^n)). \tag{30}$$

Furthermore, since C_i is a nonpersuasive neighborhood, we have

$$q_i(x^n) \in (\underline{\beta}, \overline{\beta}) \text{ for all } x^n,$$

which implies the existence of some c_i' with $\inf_{x^n} q_i(x^n) = c_i' > \beta$, where the strict inequality follows since the infimum is over a finite set. Lemma 3(c) then implies that

$$\mathbb{G}_0(1 - q_i(x^n)) \geq \left(\frac{\mathbb{G}_0(1 - c_i')}{\mathbb{G}_1(1 - c_i')} \right) \mathbb{G}_1(1 - q_i(x^n)). \tag{31}$$

Combining Eqs. (29), (30) and (31), we see that for all $n \in \mathcal{N}$ with $q^*(x^n) \geq 1/2$,

$$f_n(x^n) \geq \frac{\sum_{i=1}^{M} \omega_i(n+1) \left(\frac{\mathbb{G}_0(1-c_i')}{\mathbb{G}_1(1-c_i')} \right) \mathbb{G}_1(1 - q_i(x^n)) + \left(\frac{\mathbb{G}_0(1/2)}{\mathbb{G}_1(1/2)} \right) \mathbb{G}_1(1 - q^*(x^n))}{\sum_{i=1}^{M} \omega_i(n+1)\mathbb{G}_1(1 - q_i(x^n)) + \mathbb{G}_1(1 - q^*(x^n))}.$$

Notice the right-hand side of the equation above is merely a weighted average. Therefore, if $n \in \mathcal{N}$ and $q^*(x^n) \geq 1/2$, then

$$f_n(x^n) \geq \min \left\{ \frac{\mathbb{G}_0(1/2)}{\mathbb{G}_1(1/2)}, \min_{i \in \{1,\dots,M\}} \frac{\mathbb{G}_0(1 - c_i')}{\mathbb{G}_1(1 - c_i')} \right\} = 1 + \delta,$$

for some $\delta > 0$ using Lemma 3(c), proving Eq. (28).

Combining Eq. (28) with Eq. (27) yields

$$q^* \left((x^n, 0) \right) \leq \left[1 + \left(\frac{1}{q^*(x^n)} - 1 \right)(1 + \delta) \right]^{-1} \qquad \text{for all } n \in \mathcal{N}, \ q^*(x^n) \geq 1/2. \tag{32}$$

Suppose now $n \in \mathcal{N}$ and $q^*(x^n) \in [1/2, 1 - \epsilon]$ for some $\epsilon > 0$. We show that there exists some constant $K(\delta, \epsilon) > 0$ such that

$$q^*(x^n) - q^* \left((x^n, 0) \right) \geq K(\delta, \epsilon). \tag{33}$$

Define $g : [1/2, 1 - \epsilon] \rightarrow [0, 1]$ as

$$g(q) = q - \left[1 + \left(\frac{1}{q} - 1 \right)(1 + \delta) \right]^{-1}.$$

It can be seen that $g(q)$ is a concave function over $q \in [1/2, 1 - \epsilon]$. Let $K(\delta, \epsilon)$ be

$$K(\delta, \epsilon) = \inf_{q \in [1/2, 1-\epsilon]} g(q) = \min\{g(1/2), g(1 - \epsilon)\} = \min \left\{ \frac{\delta}{2(2 + \delta)}, \frac{\epsilon\delta(1 - \epsilon)}{1 + \epsilon\delta} \right\} > 0.$$

From Eq. (32), it follows that

$$q^*(x^n) - q^* \left((x^n, 0) \right) \geq q^*(x^n) - \left[1 + \left(\frac{1}{q^*(x^n)} - 1 \right)(1 + \delta) \right]^{-1} \geq g \left(q^*(x^n) \right) \geq K(\delta, \epsilon),$$

thus proving Eq. (33).

Recall that $q^*(x^n)$ converges to $\hat{q}$ with probability 1. We show that for any $\epsilon > 0$, the support of $\hat{q}$ does not contain $(1/2, 1 - \epsilon)$. Assume, to arrive at a contradiction, that it does. Consider a sample path that converges to a value in the interval $(1/2, 1 - \epsilon)$. For this sample path, there exists some N such that for all $n \geq N$, $q^*(x^n) \in [1/2, 1 -$

ϵ]. By the Borel-Cantelli Lemma, there are infinitely many agents n within $\mathcal{N}$ that observe a neighborhood C_i for some i because the neighborhoods are independent and $\sum_{n=1}^{\infty} \sum_{i=1}^{M} r_i(n) = \infty$. Since these are nonpersuasive neighborhoods, an infinite subset will choose action 0. Therefore, for infinitely many n that satisfy $n \geq N$ and $n \in \mathcal{N}$, by Eq. (33),

$$q^*(x^{n+1}) \leq q^*(x^n) - K(\delta, \epsilon).$$

But this implies the sequence $q^*(x^n)$ is not Cauchy and, therefore, contradicts the fact that $q^*(x^n)$ is a convergent sequence. Hence, we conclude that the support of $\hat{q}$ does not contain $(1/2, 1 - \epsilon)$. Since this argument holds for any $\epsilon > 0$, the support of $\hat{q}$ cannot contain $(1/2, 1)$. A similar argument leads to the conclusion that the support of $\hat{q}$ does not include $(0, 1/2]$. Therefore, the support of $\hat{q}$ is a subset of the set $\{0, 1\}$. By Eq. (25), this implies that $\hat{q} = \theta$ with probability 1.

We finally show that $\hat{q} = \theta$ with probability 1 implies the convergence of actions x_n to θ in probability. Suppose first $\theta = 1$. Then, $q^*(x^n)$ converges to 1 with probability 1. Therefore, $\mathbb{P}_\sigma(p_{n+1} + q^*(x^n) \geq 1 \mid \theta = 1)$ converges to 1. Therefore, for any n,

$$\mathbb{P}_\sigma(x_n = \theta \mid \theta = 1) \geq \mathbb{P}_\sigma(x_n = \theta \mid \theta = 1, B(n) = \{1, ..., n-1\}) \mathbb{Q}(B(n) = \{1, ..., n-1\})$$

$$= \mathbb{P}_\sigma(p_{n+1} + q^*(x^n) \geq 1 | \theta = 1) \left(1 - \sum_{i=1}^{M} r_i(n)\right).$$

Since $\sum_{i=1}^{M} r_i(n)$ converges to 0, the preceding relation implies that $\mathbb{P}_\sigma(x_n = \theta \mid \theta = 1)$ converges to 1. By the same argument, $\mathbb{P}_\sigma(x_n = \theta \mid \theta = 0)$ also converges to 1, thus proving asymptotic learning occurs in equilibrium σ. ∎

9 Conclusion

In this paper, we studied the problem of Bayesian (equilibrium) learning over a general social network. A large social learning literature, pioneered by Bikhchandani, Hirshleifer and Welch (1992), Banerjee (1992) and Smith and Sorensen (2000), has studied equilibria of sequential-move games, where each individual observes all past actions. The focus has been on whether equilibria lead to aggregation of information (and thus to asymptotic learning).

In many relevant situations, individuals obtain their information not by observing all past actions, but from their "social network". This raises the question of how the structures of social networks in which individuals are situated affects learning behavior. To address these questions, we formulated a sequential-move equilibrium learning model over a general social network.

In our model, each individual receives a signal about the underlying state of the world and observes the past actions of a stochastically-generated neighborhood of individuals. The stochastic process generating the neighborhoods defines the network topology. The signal structure determines the conditional distributions of the signals received by each individual as a function of the underlying state. The social network

consists of the network topology and the signal structure. Each individual then chooses one of two possible actions depending on his posterior beliefs given his signal and the realized neighborhood. We characterized pure-strategy (perfect Bayesian) equilibria for arbitrary stochastic and deterministic social networks, and characterized the conditions under which there is asymptotic learning. Asymptotic learning corresponds to individual decisions converging (in probability) to the right action as the social network becomes large.

Two concepts turn out to be crucial in determining whether there will be asymptotic learning. The first is common with the previous literature. Following Smith and Sorensen (2000), we say that private beliefs are bounded if the likelihood ratio implied by individual signals is bounded and there is a maximum amount of information that can be generated from these signals. Conversely, private beliefs are unbounded if the corresponding likelihood ratio is unbounded. The second important concept is that of expanding or nonexpanding observations. A network topology has nonexpanding observations if there exists infinitely many agents observing the actions of only a finite subset of (excessively influential) agents. Most network topologies feature expanding observations.

Nonexpanding observations do not allow asymptotic learning, since there exists infinitely many agents who do not receive sufficiently many observations to be able to aggregate information.

Our main theorem shows that expanding observations and unbounded private signals are sufficient to ensure asymptotic learning. Since expanding observations is a relatively mild restriction, to the extent that unbounded private beliefs constitute a good approximation to the informativeness of individual signals, this result implies that all equilibria feature asymptotic learning applies in a wide variety of settings. Another implication is that asymptotic learning is possible even when there are "influential agents" or "information leaders", that is, individuals who are observed by many, most or even all agents (while others may be observed not at all or much less frequently). It is only when individuals are excessively influential—loosely speaking when they act as the sole source of information for infinitely many agents—that asymptotic learning ceases to apply.

We also provide a partial converse to this result, showing that under the most common deterministic or stochastic network topologies, bounded beliefs imply no asymptotic learning. However, we show that asymptotic learning is possible even with bounded beliefs for a certain class of stochastic network topologies.

Beyond the specific results presented in this paper, we believe that the framework developed here opens the way for a more general analysis of the structure of social networks on learning dynamics. Among the questions that can be studied using this framework are the following: (1) the effect of network structure on the speed (rate of convergence) of Bayesian learning; (2) equilibrium learning when there are heterogeneous preferences; (3) equilibrium learning when the underlying state is changing dynamically; (4) the influence of a subset of a social network (for example, the media or interested parties) in influencing the views of the rest as a function of the network structure. We intend to investigate these questions in future work.

Appendix A

9.1 Proof of Theorem 1

Suppose that the network has nonexpanding observations. This implies that there exists some $K \in \mathbb{N}$, $\epsilon > 0$, and a subsequence of agents $\mathcal{N}$ such that for all $n \in \mathcal{N}$,

$$\mathbb{Q}_n \left(\max_{b \in B(n)} b < K \right) \geq \epsilon. \tag{34}$$

For any such agent $n \in \mathcal{N}$, we have

$$
\begin{aligned}
\mathbb{P}_\sigma(x_n = \theta) \;=\; & \mathbb{P}_\sigma \left(x_n = \theta \;\Big|\; \max_{b \in B(n)} b < K \right) \mathbb{Q}_n \left(\max_{b \in B(n)} b < K \right) \\
& + \mathbb{P}_\sigma \left(x_n = \theta \;\Big|\; \max_{b \in B(n)} b \geq K \right) \mathbb{Q}_n \left(\max_{b \in B(n)} b \geq K \right) \\
\leq \;& \mathbb{P}_\sigma \left(x_n = \theta \;\Big|\; \max_{b \in B(n)} b < K \right) \mathbb{Q}_n \left(\max_{b \in B(n)} b < K \right) + \mathbb{Q}_n \left(\max_{b \in B(n)} b \geq K \right) \\
\leq \;& 1 - \epsilon + \epsilon \mathbb{P}_\sigma \left(x_n = \theta \;\Big|\; \max_{b \in B(n)} b < K \right),
\end{aligned}
\tag{35}
$$

where the second inequality follows from Eq. (34).

Given some equilibrium $\sigma \in \Sigma^*$ and agent n, we define z_n as the decision that maximizes the conditional probability of making a correct decision given the private signals and neighborhoods of the first $K - 1$ agents and agent n, i.e.,

$$z_n = \arg\max_{y \in \{0,1\}} \mathbb{P}_{y, \sigma_{-n}}(y = \theta \mid s_i, B(i), \text{ for } i = 1, \dots, K-1, n). \tag{36}$$

We denote a particular realization of private signal s_i by $\mathfrak{s}_i$ and a realization of neighborhood $B(i)$ by $\mathfrak{B}(i)$ for all i. Given the equilibrium σ and the realization $\mathfrak{s}_1, ..., \mathfrak{s}_{K-1}$ and $\mathfrak{B}(1), ..., \mathfrak{B}(K-1)$, all decisions $x_1, \dots, x_{K-1}$ are recursively defined [i.e., they are non-stochastic; see the definition of the information set in Eq. (1)]. Therefore, for any $\mathfrak{B}(n)$ that satisfies $\max_{b \in \mathfrak{B}(n)} b < K$, the decision x_n is also defined. By the definition of z_n [cf. Eq. (36)], this implies that

$$
\begin{aligned}
\mathbb{P}_\sigma(x_n = \theta \;\mid\;& s_i = \mathfrak{s}_i, \; B(i) = \mathfrak{B}(i), \text{ for } i = 1, \dots, K-1, n) \\
& \leq \mathbb{P}_\sigma(z_n = \theta \mid s_i = \mathfrak{s}_i, \; B(i) = \mathfrak{B}(i), \text{ for } i = 1, \dots, K-1, n).
\end{aligned}
$$

By integrating over all possible $\mathfrak{s}_1, ..., \mathfrak{s}_{K-1}, \mathfrak{s}_n, \mathfrak{B}(1), ..., \mathfrak{B}(K-1)$, this yields

$$\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \mathfrak{B}(n)) \leq \mathbb{P}_\sigma(z_n = \theta \mid B(n) = \mathfrak{B}(n)),$$

for any $\mathfrak{B}(n)$ that satisfies $\max_{b \in \mathfrak{B}(n)} b < K$. By integrating over all $\mathfrak{B}(n)$ that satisfy this condition, we obtain

$$\mathbb{P}_\sigma \left(x_n = \theta \;\Big|\; \max_{b \in B(n)} b < K \right) \leq \mathbb{P}_\sigma \left(z_n = \theta \;\Big|\; \max_{b \in B(n)} b < K \right). \tag{37}$$

Moreover, since the sequence of neighborhoods $\{B(i)\}_{i \in \mathbb{N}}$ is independent of θ and the sequence of private signals $\{s_i\}_{i \in \mathbb{N}}$, it follows from Eq. (36) that the decision z_n is given by

$$z_n = \arg\max_{y \in \{0,1\}} \mathbb{P}_\sigma(y = \theta \mid s_1, ..., s_{K-1}, s_n). \tag{38}$$

Therefore, z_n is also independent of the sequence of neighborhoods $\{B(i)\}_{i \in \mathbb{N}}$ and we have

$$\mathbb{P}_\sigma\left(z_n = \theta \mid \max_{b \in B(n)} b < K\right) = \mathbb{P}_\sigma\left(z_n = \theta\right).$$

Since the private signals have the same distribution, it follows from Eq. (38) that for any $n, m \geq K$, the random variables z_n and z_m have identical probability distributions. Hence, for any $n \geq K$, Eq. (38) implies that

$$\mathbb{P}_\sigma\left(z_n = \theta\right) = \mathbb{P}_\sigma\left(z_K = \theta\right).$$

Combining the preceding two relations with Eq. (37), we have for any $n \geq K$,

$$\mathbb{P}_\sigma\left(x_n = \theta \mid \max_{b \in B(n)} b < K\right) \leq \mathbb{P}_\sigma\left(z_n = \theta \mid \max_{b \in B(n)} b < K\right) = \mathbb{P}_\sigma\left(z_n = \theta\right) = \mathbb{P}_\sigma\left(z_K = \theta\right).$$

Substituting this relation in Eq. (35), we obtain for any $n \in \mathcal{N}, n \geq K$,

$$\mathbb{P}_\sigma(x_n = \theta) \leq 1 - \epsilon + \epsilon \mathbb{P}_\sigma\left(z_K = \theta\right).$$

Therefore,

$$\liminf_{n \to \infty} \mathbb{P}_\sigma(x_n = \theta) \leq 1 - \epsilon + \epsilon \mathbb{P}_\sigma(z_K = \theta). \tag{39}$$

We finally show that in view of the assumption that $\mathbb{F}_0$ and $\mathbb{F}_1$ are absolutely continuous with respect to each other [which implies $\mathbb{P}_\sigma(x_1 = \theta) < 1$], we have $\mathbb{P}_\sigma(z_K = \theta) < 1$ for any given K. If $\mathbb{P}_\sigma(x_1 = \theta) < 1$ holds, then we have either $\mathbb{P}_\sigma(x_1 = \theta \mid \theta = 1) < 1$ or $\mathbb{P}_\sigma(x_1 = \theta \mid \theta = 0) < 1$. Assume without loss of generality that we have

$$\mathbb{P}_\sigma(x_1 = \theta \mid \theta = 1) < 1. \tag{40}$$

Let $\overline{S}$ denote the set of all private signals such that if $s_1 \in \overline{S}_\sigma$, then $x_1 = 0$ in equilibrium σ. Since the first agent's decision is a function of s_1, then Eq. (40) is equivalent to

$$\mathbb{P}_\sigma(s_1 \in \overline{S}_\sigma \mid \theta = 1) > 0.$$

Since the private signals are conditionally independent given θ, this implies that

$$\mathbb{P}_\sigma(s_i \in \overline{S}_\sigma \text{ for all } i \leq K \mid \theta = 1) = \mathbb{P}_\sigma(s_1 \in \overline{S}_\sigma \mid \theta = 1)^K > 0. \tag{41}$$

We next show that if $s_i \in \overline{S}_\sigma$ for all $i \leq K$, then $z_K = 0$. Using Bayes' Rule, we have

$$\begin{aligned}
\mathbb{P}_\sigma(\theta = 0 \mid s_i \in \overline{S}_\sigma \text{ for all } i \leq K) &= \left[1 + \frac{\mathbb{P}_\sigma(s_i \in \overline{S}_\sigma \text{ for all } i \leq K \mid \theta = 1)}{\mathbb{P}_\sigma(s_i \in \overline{S}_\sigma \text{ for all } i \leq K \mid \theta = 0)}\right]^{-1} \\
&= \left[1 + \frac{\prod_{i=1}^{K} \mathbb{P}_\sigma(s_i \in \overline{S}_\sigma \mid \theta = 1)}{\prod_{i=1}^{K} \mathbb{P}_\sigma(s_i \in \overline{S}_\sigma \mid \theta = 0)}\right]^{-1} \\
&= \left[1 + \left(\frac{\mathbb{P}_\sigma(s_1 \in \overline{S}_\sigma \mid \theta = 1)}{\mathbb{P}_\sigma(s_1 \in \overline{S}_\sigma \mid \theta = 0)}\right)^K\right]^{-1}, \tag{42}
\end{aligned}$$

40

where the second equality follows from the conditional independence of the private signals and the third equality holds since private signals are identically distributed. Applying Bayes' Rule on the second term in parentheses in Eq. (42), this implies that

$$\mathbb{P}_\sigma(\theta = 0 \mid s_i \in \overline{S}_\sigma \text{ for all } i \leq K) \;=\; \left[1 + \left(\frac{1}{\mathbb{P}_\sigma(\theta = 0 \mid s_1 \in \overline{S}_\sigma)} - 1 \right)^K \right]^{-1}. \quad (43)$$

Since $s_1 \in \overline{S}_\sigma$ induces $x_1 = 0$, we have $\mathbb{P}_\sigma(\theta = 0 \mid s_1 \in \overline{S}_\sigma) \geq 1/2$. Because the function on the right-handside of Eq. (43) is nondecreasing in $\mathbb{P}_\sigma(\theta = 0 \mid s_1 \in \overline{S}_\sigma)$ for any value in $[1/2,1]$, we obtain

$$\mathbb{P}_\sigma(\theta = 0 \mid s_i \in \overline{S}_\sigma \text{ for all } i \leq K) \geq \frac{1}{2}.$$

By the definition of z_K, this implies that if $s_i \in \overline{S}_\sigma$ for all $i \leq K$, then $z_K = 0$ (we can let z_n be equal to 0 whenever both states are equally likely given the private signals). Combined with the fact that the event $\{s_i \in \overline{S}_\sigma \text{ for all } i \leq K\}$ has positive conditional probability given $\theta = 1$ under measure $\mathbb{P}_\sigma$ [cf. Eq. (41)], this implies that $\mathbb{P}_\sigma(z_K = \theta) < 1$. Substituting this relation in Eq. (39), we have

$$\liminf_{n \to \infty} \mathbb{P}_\sigma(x_n = \theta) < 1,$$

thus showing that asymptotic learning does not occur. ∎

References

[1] Austen-Smith D. and Banks J., "Information Aggregation, Rationality, and the Condorcet Jury Theorem," *The American Political Science Review*, vol. 90, no. 1, pp. 34-45, 1996.

[2] Bala V. and Goyal S., "Learning from Neighbours," *The Review of Economic Studies*, vol. 65, no. 3, pp. 595-621, 1998.

[3] Bala V. and Goyal S., "Conformism and Diversity under Social Learning," *Economic Theory*, vol. 17, pp. 101-120, 2001.

[4] Banerjee A., "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics,* vol. 107, pp. 797-817, 1992.

[5] Banerjee A. and Fudenberg D., "Word-of-mouth Learning," *Games and Economic Behavior*, vol. 46, pp. 1-22, 2004.

[6] Besley T. and Case A., "Diffusion as a Learning Process: Evidence from HYV Cotton," Working Papers 228, Princeton University, Woodrow Wilson School of Public and International Affairs, Research Program in Development Studies, 1994.

[7] Bikhchandani S., Hirshleifer D., and Welch I., "A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades," *The Journal of Political Economy*, vol. 100, pp. 992-1026, 1992.

[8] Bikhchandani S., Hirshleifer D., and Welch I., "Learning from the Behavior of Others: Conformity, Fats, and Informational Cascades," *The Journal of Economic Perspectives*, vol. 12, pp. 151-170, 1998.

[9] Breiman L., "Probability," Addison-Wellesley, 1968.

[10] Callander S. and Horner J., "The Wisdom of the Minority," Northwestern mimeo, 2006.

[11] Celen B. and Kariv S., "Observational Learning Under Imperfect Information," *Games and Economic Behavior*, vol. 47, no. 1, pp. 72-86, 2004.

[12] Celen B. and Kariv S., "Distinguishing Informational Cascades from Herd Behavior in the Laboratory," *The American Economic Review*, vol. 94, no. 3, pp. 484-498, 2004.

[13] Chamley C. and Gale D., "Information Revelation and Strategic Delay in a Model of Investment," *Econometrica*, vol. 62, pp. 1065-1086, 1994.

[14] Choi S., Celen B., and Kariv S., "Learning in Networks: An Experimental Study," *UCLA Technical Report*, Dec. 2005.

[15] Condorcet N.C. , "Essai sur l'Application de l'Analyze à la Probabilité des Décisions Rendues à la Pluralité des Voix," Imprimérie Royale, Paris, France, 1785.

[16] Cover T., "Hypothesis Testing with Finite Statistics," *The Annals of Mathematical Statistics*, vol. 40, no. 3, pp. 828-835, 1969.

[17] DeMarzo P.M., Vayanos D., and Zwiebel J., "Persuasion Bias, Social Influence, and Unidimensional Opinions," *The Quarterly Journal of Economics*, vol. 118, no. 3, pp. 909-968, 2003.

[18] Doob J.L., "Stochastic Processes," John Wiley and Sons Inc., 1955.

[19] Ellison G. and Fudenberg D., "Rules of Thumb for Social Learning," *The Journal of Political Economy*, vol. 101, no. 4, pp. 612-643, 1993.

[20] Ellison G. and Fudenberg D., "Word-of-mouth communication and social learning," *The Quarterly Journal of Economics*, vol. 110, pp. 93-126, 1995.

[21] Feddersen T. and Pesendorfer W., "The Swing Voter's Curse," *The American Economic Review*, vol. 86, no. 3, pp. 408-424, 1996.

[22] Feddersen T. and Pesendorfer W., "Voting Behavior and Information Aggregation in Elections With Private Information," *Econometrica*, vol. 65, no. 5, pp. 1029-1058, 1997.

[23] Foster A. and Rosenzweig M., "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *The Journal of Political Economy*, vol. 103, no. 6, pp. 1176-1209, 1995.

[24] Gale D. and Kariv S., "Bayesian Learning in Social Networks," *Games and Economic Behavior*, vol. 45, no. 2, pp. 329-346, 2003.

[25] Golub B. and Jackson M.O., "Naive Learning in Social Networks: Convergence, Influence, and the Wisdom of Crowds," unpublished manuscript, 2007.

[26] Granovetter M., "The Strength of Weak Ties," *The American Journal of Sociology*, vol. 78, no. 6, pp. 1360-1380, 1973.

[27] Ioannides M. and Loury L., "Job Information Networks, Neighborhood Effects, and Inequality," *The Journal of Economic Literature*, vol. 42, no. 2, pp. 1056-1093, 2004.

[28] Jackson M.O., "The Economics of Social Networks," chapter in "Volume I of Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society", Richard Blundell, Whitney Newey, and Torsten Persson, Cambridge University Press, 2006.

[29] Jackson M.O., "The Study of Social Networks In Economics," chapter in "The Missing Links: Formation and Decay of Economic Networks", edited by James E. Rauch; Russell Sage Foundation, 2007.

[30] Jackson M.O. and Wolinsky A., "A Strategic Model of Social and Economic Networks," *The Journal of Economic Theory*, vol. 71, pp. 44-74, 1996.

[31] Jackson M.O. and Watts A., "The Evolution of Social and Economic Networks," *The Journal of Economic Theory*, vol. 106, pp. 265-295, 2002.

[32] Lee I., "On the Convergence of Informational Cascades," *The Journal of Economic Theory*, vol. 61, pp. 395-411, 1993.

[33] Lorenz J., Marciniszyn M., and Steger A., "Observational Learning in Random Networks," chapter in "Learning Theory", Springer Berlin / Heidelberg, 2007.

[34] Montgomery J.D., "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis," *The American Economic Review*, vol. 81 no. 5, pp. 1408-1418, 1991.

[35] Munshi K., "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *The Quarterly Journal of Economics*, vol. 118 no. 2, pp. 549-597, 2003.

[36] Munshi K., "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution," *The Journal of Development Economics*, vol. 73 no. 1, pp. 185-215, 2004.

[37] Myerson R.B., "Extended Poisson Games and the Condorcet Jury Theorem," *Games and Economic Behavior*, vol. 25, no. 1, pp. 111-131, 1998.

[38] Myerson R.B., "Large Poisson Games," *The Journal of Economic Theory*, vol. 94 no. 1, pp. 7-45, 2000.

[39] Papastavrou J.D. and Athans M., "Distributed Detection by a Large Team of Sensors in Tandem," *Proceedings of IEEE CDC*, 1990.

[40] Sgroi D., "Optimizing Information in the Herd: Guinea Pigs, Profits, and Welfare," *Games and Economic Behavior*, vol. 39, pp. 137-166, 2002.

[41] Smith L. and Sorensen P., "Rational Social Learning with Random Sampling," unpublished manuscript, 1998.

[42] Smith L. and Sorensen P., "Pathological Outcomes of Observational Learning," *Econometrica*, vol. 68, no. 2, pp. 371-398, 2000.

[43] Tay W.P., Tsitsiklis J.N., and Win M.Z., "On the Sub-Exponential Decay of Detection Error Probabilities in Long Tandems," *Proceedings of ICASSP*, 2007.

[44] Udry C. and Conley T., "Social Learning through Networks: the Adoption of New Agricultural Technologies in Ghana," *American Journal of Agricultural Economics*, vol. 83, pp. 668-673, 2001.

[45] Vives X., "Learning from Others: A Welfare Analysis," *Games and Economic Behavior*, vol. 20, pp. 177-200, 1997.

[46] Welch I., "Sequential Sales, Learning and Cascades," *The Journal of Finance*, vol. 47, pp. 695-732, 1992.

[47] Young H.P., "Condorcet's Theory of Voting," *The American Political Science Review*, vol. 82, pp. 12311244, 1988.

Appendix B: Omitted Proofs (Not for Publication)

Proof of Lemma 1. We prove that if

$$\mathbb{P}_\sigma(\theta = 1 \mid s_n) + \mathbb{P}_\sigma\big(\theta = 1 \mid B(n), x_k \text{ for all } k \in B(n)\big) > 1, \tag{44}$$

then $x_n = 1$. The proofs of the remaining statements follow the same line of argument. We first show that Eq. (44) holds if and only if

$$\mathbb{P}_\sigma(\theta = 1 \mid I_n) > 1/2, \tag{45}$$

therefore implying that $x_n = 1$ by the equilibrium condition [cf. Eq. (2)]. By Bayes' Rule, Eq. (45) is equivalent to

$$\mathbb{P}_\sigma(\theta = 1 \mid I_n) = \frac{d\mathbb{P}_\sigma(I_n \mid \theta = 1)\mathbb{P}_\sigma(\theta = 1)}{\sum_{j=0}^1 d\mathbb{P}_\sigma(I_n \mid \theta = j)\mathbb{P}_\sigma(\theta = j)} = \frac{d\mathbb{P}_\sigma(I_n \mid \theta = 1)}{\sum_{j=0}^1 d\mathbb{P}_\sigma(I_n \mid \theta = j)} > 1/2, \tag{46}$$

where the second equality follows from the assumption that states 0 and 1 are equally likely. Hence, Eq. (45) holds if and only if

$$d\mathbb{P}_\sigma(I_n \mid \theta = 1) > d\mathbb{P}_\sigma(I_n \mid \theta = 0). \tag{47}$$

Conditional on state θ, the private signals and the observed decisions are independent, i.e.,

$$d\mathbb{P}_\sigma(I_n \mid \theta = j) = d\mathbb{P}_\sigma(s_n \mid \theta = j)\mathbb{P}_\sigma(B(n), x_k, \ k \in B(n) \mid \theta = j).$$

Combining the preceding two relations, it follows that Eq. (47) is equivalent to

$$\frac{\mathbb{P}_\sigma(B(n), x_k, \ k \in B(n) \mid \theta = 1)}{\sum_{j=0}^1 \mathbb{P}_\sigma(B(n), x_k, \ k \in B(n) \mid \theta = j)} > \frac{d\mathbb{P}_\sigma(s_n \mid \theta = 0)}{\sum_{j=0}^1 d\mathbb{P}_\sigma(s_n \mid \theta = j)}.$$

Since both states are equally likely, this can be rewritten as

$$\frac{\mathbb{P}_\sigma(B(n), x_k, \ k \in B(n) \mid \theta = 1)P_\sigma(\theta = 1)}{\sum_{j=0}^1 \mathbb{P}_\sigma(B(n), x_k, \ k \in B(n) \mid \theta = j)P_\sigma(\theta = j)} > \frac{d\mathbb{P}_\sigma(s_n \mid \theta = 0)P_\sigma(\theta = 0)}{\sum_{j=0}^1 d\mathbb{P}_\sigma(s_n \mid \theta = j)P_\sigma(\theta = j)}.$$

Applying Bayes' Rule on both sides of Eq. (48), we see that the preceding relation is identical to

$$\mathbb{P}_\sigma(\theta = 1 \mid B(n), x_k, \ k \in B(n)) > \mathbb{P}_\sigma(\theta = 0 \mid s_n) = 1 - \mathbb{P}_\sigma(\theta = 1 \mid s_n),$$

completing the proof. ∎

Proof of Lemma 3.
(a) By the definition of a private belief, we have for any $p_n \in (0, 1)$,

$$\mathbb{P}(\theta = 1|s_n) = \mathbb{P}(\theta = 1|p_n).$$

Using Bayes' Rule, it follows that

$$p_n = \mathbb{P}_\sigma(\theta = 1 | p_n) = \frac{d\mathbb{P}(p_n | \theta = 1)\mathbb{P}(\theta = 1)}{\sum_{j=0}^1 d\mathbb{P}(p_n | \theta = j)\mathbb{P}(\theta = j)} = \frac{d\mathbb{P}(p_n | \theta = 1)}{\sum_{j=0}^1 d\mathbb{P}(p_n | \theta = j)} = \frac{d\mathbb{G}_1(p_n)}{\sum_{j=0}^1 d\mathbb{G}_j(p_n)}.$$

Because of the assumption that no signal is completely informative, i.e., $p_n \notin \{0, 1\}$, we can rewrite this equation as

$$\frac{d\mathbb{G}_0}{d\mathbb{G}_1}(p_n) = \frac{1 - p_n}{p_n},$$

completing the proof.

(b) For any $p \in (0, 1)$,

$$\mathbb{G}_0(p) = \int_{r=0}^p d\mathbb{G}_0(r) = \int_{r=0}^p \frac{1 - r}{r} d\mathbb{G}_1(r) = \left(\frac{1 - p}{p}\right)\mathbb{G}_1(p) + \int_{r=0}^p \left(\frac{1}{r} - \frac{1}{p}\right) d\mathbb{G}_1(r),$$

where the second equality follows from part (a) of this lemma. We can provide a lower bound on the last integral as

$$\int_{r=0}^p \left(\frac{1}{r} - \frac{1}{p}\right) d\mathbb{G}_1(r) \geq \int_{r=0}^z \left(\frac{1}{r} - \frac{1}{p}\right) d\mathbb{G}_1(r) \geq \int_{r=0}^z \left(\frac{1}{z} - \frac{2}{z + p}\right) d\mathbb{G}_1(r) \geq \frac{p - z}{2}\mathbb{G}_1(z),$$

for any $z \in (0, p)$. Equivalently, the second relation is obtained by

$$
\begin{aligned}
1 - \mathbb{G}_1(p) = \int_{r=p}^1 d\mathbb{G}_1(r) &= \int_{r=p}^1 \frac{r}{1 - r} d\mathbb{G}_0(r) \\
&= (1 - \mathbb{G}_0(p))\left(\frac{p}{1 - p}\right) + \int_{r=p}^1 \left(\frac{r}{1 - r} - \frac{p}{1 - p}\right) d\mathbb{G}_0(r),
\end{aligned}
$$

where the following bound is valid for any $p < w < 1$,

$$
\begin{aligned}
\int_{r=p}^1 \left(\frac{r}{1 - r} - \frac{p}{1 - p}\right) d\mathbb{G}_0(r) &\geq \int_{r=w}^1 \left(\frac{r}{1 - r} - \frac{p}{1 - p}\right) d\mathbb{G}_0(r) \\
&\geq \int_{r=w}^1 \left(\frac{w}{1 - w} - \frac{p + w}{2 - p - w}\right) d\mathbb{G}_0(r) \geq \frac{w - p}{2}(1 - \mathbb{G}_0(w)).
\end{aligned}
$$

(c) From part (a), we have for any $r \in (0, 1)$,

$$\mathbb{G}_0(r) = \int_{x=0}^r d\mathbb{G}_0(x) = \int_{x=0}^r \left(\frac{1 - x}{x}\right) d\mathbb{G}_1(x) \geq \int_{x=0}^r \left(\frac{1 - r}{r}\right) d\mathbb{G}_1(x) = \left(\frac{1 - r}{r}\right)\mathbb{G}_1(r). \tag{48}$$

Using part (a) again,

$$
\begin{aligned}
d\left(\frac{\mathbb{G}_0(r)}{\mathbb{G}_1(r)}\right) &= \frac{d\mathbb{G}_0(r)\mathbb{G}_1(r) - \mathbb{G}_0(r)d\mathbb{G}_1(r)}{(\mathbb{G}_1(r))^2} \\
&= \frac{d\mathbb{G}_1(r)}{(\mathbb{G}_1(r))^2}\left[\left(\frac{1 - r}{r}\right)\mathbb{G}_1(r) - \mathbb{G}_0(r)\right].
\end{aligned}
$$

47

Since $\mathbb{G}_1(r) > 0$ for $r > \underline{\beta}$, $d\mathbb{G}_1(r) \geq 0$ and the term in brackets above is non-positive by Eq. (48), we have

$$d\left(\frac{\mathbb{G}_0(r)}{\mathbb{G}_1(r)}\right) \leq 0,$$

thus proving the ratio $\mathbb{G}_0(r)/\mathbb{G}_1(r)$ is non-increasing.

We now show that

$$\mathbb{G}_0(r) \geq \mathbb{G}_1(r) \text{ for all } r \in [0,1]. \tag{49}$$

From Eq. (48), we obtain that Eq. (49) is true for $r \leq 1/2$. For $r > 1/2$,

$$1 - \mathbb{G}_0(r) = \int_{x=r}^1 d\mathbb{G}_0(x) = \int_{x=r}^1 \left(\frac{1-x}{x}\right) d\mathbb{G}_1(x) \leq \int_{x=r}^1 d\mathbb{G}_1(x) = 1 - \mathbb{G}_1(r),$$

thus proving Eq. (49).

We proceed to prove the second part of the lemma. Suppose $\mathbb{G}_0(r)/\mathbb{G}_1(r) = 1$ for some $r < \overline{\beta}$. Suppose first $r \in (1/2, \overline{\beta})$. Then,

$$
\begin{aligned}
\mathbb{G}_0(1) &= \mathbb{G}_0(r) + \int_{x=r}^1 d\mathbb{G}_0(x) \\
&= \mathbb{G}_1(r) + \int_{x=r}^1 d\mathbb{G}_0(x) \\
&= \mathbb{G}_1(r) + \int_{x=r}^1 \left(\frac{1-x}{x}\right) d\mathbb{G}_1(x) \\
&\geq \mathbb{G}_1(r) + \left(\frac{1-r}{r}\right) \int_{x=r}^1 d\mathbb{G}_1(x) \\
&\geq \mathbb{G}_1(r) + \left(\frac{1-r}{r}\right) [1 - \mathbb{G}_1(r)],
\end{aligned}
$$

which yields a contradiction unless $\mathbb{G}_1(r) = 1$. However, $\mathbb{G}_1(r) = 1$ implies $r \geq \overline{\beta}$ – also a contradiction. Now, suppose $r \in (\underline{\beta}, 1/2]$. Since the ratio $\mathbb{G}_0(r)/\mathbb{G}_1(r)$ is non-increasing, this implies that for all $x \in (r, 1]$, $\mathbb{G}_0(x)/\mathbb{G}_1(x) \leq 1$. Combined with Eq. (49), this yields $\mathbb{G}_0(x)/\mathbb{G}_1(x) = 1$ for all $x \in (r, 1]$, which yields a contradiction for $x \in (1/2, \overline{\beta})$. ∎

Proof of Proposition 2. Let $h : \{(n, B(n)) : n \in N, B(n) \subseteq \{1, 2, ..., n-1\}\} \to \mathbb{N}$ be an arbitrary function that maps an agent and a neighborhood of the agent into an element of the neighborhood, i.e., $h(n, B(n)) \in B(n)$. In view of the characterization of the equilibrium decision x_n [cf. Eq. (2)], it follows that for any private signal s_n, neighborhood $B(n) \subseteq \{1, 2, ..., n-1\}$, and decisions x_k, $k \in B(n)$, we have

$$\mathbb{P}_\sigma(x_n = \theta \mid s_n, B(n), x_k, \ k \in B(n)) \geq \mathbb{P}_\sigma(x_{h(n,B(n))} = \theta \mid s_n, B(n), x_k, \ k \in B(n)).$$

By integrating over all possible private signals and decisions of agents in the neighborhood, we obtain that for any n and any $B(n) = \mathfrak{B}$,

$$\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \mathfrak{B}) \geq \mathbb{P}_\sigma(x_{h(n,B(n))} = \theta \mid B(n) = \mathfrak{B}) = \mathbb{P}_\sigma(x_{h(n,\mathfrak{B})} = \theta),$$

where the equality follows by the assumption that each neighborhood is generated independently from all other neighborhoods. By taking the maximum over all functions h, we obtain

$$\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \mathfrak{B}) \geq \max_{b \in \mathfrak{B}} \mathbb{P}_\sigma(x_b = \theta),$$

showing the desired relation. ∎

Proof of Proposition 6. The proof consists of two steps. We first show that the lower and upper supports of the social belief $q_n = \mathbb{P}_\sigma(\theta = 1 | x_1, ..., x_{n-1})$ are bounded away from 0 and 1. We next show that this implies that x_n does not converge to θ in probability.

Let $x^{n-1} = (x_1, ..., x_{n-1})$ denote the sequence of decisions up to and including $n - 1$. Let $\varphi_{\sigma,x^{n-1}}(q_n, x_n)$ represent the social belief q_{n+1} given the social belief q_n and the decision x_n, for a given strategy σ and decisions x^{n-1}. We use Bayes' Rule to determine the dynamics of the social belief. For any x^{n-1} compatible with q_n, and $x_n = \overline{x}$ with $\overline{x} \in \{0, 1\}$, we have

$$
\begin{aligned}
\varphi_{\sigma,x^{n-1}}(q_n, \overline{x}) &= \mathbb{P}_\sigma(\theta = 1 \mid x_n = \overline{x}, q_n, x^{n-1}) \\
&= \left[1 + \frac{\mathbb{P}_\sigma(x_n = \overline{x}, q_n, x^{n-1}, \theta = 0)}{\mathbb{P}_\sigma(x_n = \overline{x}, q_n, x^{n-1}, \theta = 1)} \right]^{-1} \\
&= \left[1 + \frac{\mathbb{P}_\sigma(q_n, x^{n-1} \mid \theta = 0)}{\mathbb{P}_\sigma(q_n, x^{n-1} \mid \theta = 1)} \frac{\mathbb{P}_\sigma(x_n = \overline{x} \mid q_n, x^{n-1}, \theta = 0)}{\mathbb{P}_\sigma(x_n = \overline{x} \mid q_n, x^{n-1}, \theta = 1)} \right]^{-1} \\
&= \left[1 + \left(\frac{1}{q_n} - 1 \right) \frac{\mathbb{P}_\sigma(x_n = \overline{x} \mid q_n, x^{n-1}, \theta = 0)}{\mathbb{P}_\sigma(x_n = \overline{x} \mid q_n, x^{n-1}, \theta = 1)} \right]^{-1}. \quad (50)
\end{aligned}
$$

Let $\alpha_{\sigma,x^{n-1}}$ denote the probability that agent n chooses $x = 0$ in equilibrium σ when he observes history x^{n-1} and is indifferent between the two actions. Let

$$\mathbb{G}_j^-(r) = \lim_{s \uparrow r} \mathbb{G}_j(s),$$

for any $r \in [0, 1]$ and any $j \in \{0, 1\}$. Then, for any $j \in \{0, 1\}$,

$$
\begin{aligned}
\mathbb{P}_\sigma(x_n = 0 \mid q_n, x^{n-1}, \theta = j) &= \mathbb{P}_\sigma(p_n < 1 - q_n \mid q_n, \theta = j) \\
&\quad + \alpha_{\sigma,x^{n-1}} \mathbb{P}_\sigma(p_n = 1 - q_n \mid q_n, \theta = j) \\
&= \mathbb{G}_j^-(1 - q_n) + \alpha_{\sigma,x^{n-1}} \left[\mathbb{G}_j(1 - q_n) - \mathbb{G}_j^-(1 - q_n) \right].
\end{aligned}
$$

From Lemma 3(a), $d\mathbb{G}_0/d\mathbb{G}_1(r) = (1 - r)/r$ for all $r \in [0, 1]$. Therefore,

$$\frac{1 - r}{r} \leq \frac{\mathbb{G}_0(r)}{\mathbb{G}_1(r)}, \quad \text{and} \quad \frac{\mathbb{G}_0^-(r)}{\mathbb{G}_1^-(r)} \leq \frac{1 - \underline{\beta}}{\underline{\beta}}.$$

Hence, for any $\alpha_{\sigma,x^{n-1}}$,

$$
\begin{aligned}
\frac{\mathbb{P}_\sigma(x_n = 0 \mid q_n, x^{n-1}, \theta = 0)}{\mathbb{P}_\sigma(x_n = 0 \mid q_n, x^{n-1}, \theta = 1)} &= \frac{\mathbb{G}_0^-(1 - q_n) + \alpha_{\sigma,x^{n-1}} \left[\mathbb{G}_0(1 - q_n) - \mathbb{G}_0^-(1 - q_n) \right]}{\mathbb{G}_1^-(1 - q_n) + \alpha_{\sigma,x^{n-1}} \left[\mathbb{G}_1(1 - q_n) - \mathbb{G}_0^-(1 - q_n) \right]} \\
&\in \left[\frac{q_n}{1 - q_n}, \frac{1 - \underline{\beta}}{\underline{\beta}} \right].
\end{aligned}
$$

Combining this with Eq. (50), we obtain

$$\varphi_{\sigma,x^{n-1}}(q_n,0) \in \left[\left(1+\left(\frac{1}{q_n}-1\right)\left(\frac{1-\bar{\beta}}{\underline{\beta}}\right)\right)^{-1}, \left(1+\left(\frac{1}{q_n}-1\right)\left(\frac{q_n}{1-q_n}\right)\right)^{-1}\right]$$

$$= \left[\frac{\underline{\beta}q_n}{1-\underline{\beta}-q_n+2\underline{\beta}q_n}, \frac{1}{2}\right]$$

Note that $\frac{\underline{\beta}q_n}{1-\underline{\beta}-q_n+2\underline{\beta}q_n}$ is an increasing function of q_n and if $q_n \in [1-\bar{\beta}, 1-\underline{\beta}]$, then this function is minimized at $1-\bar{\beta}$. This implies that

$$\varphi_{\sigma,x^{n-1}}(q_n,0) \in \left[\frac{\underline{\beta}(1-\bar{\beta})}{-\underline{\beta}+\bar{\beta}+2\underline{\beta}(1-\bar{\beta})}, \frac{1}{2}\right] = \left[\underline{\Delta}, \frac{1}{2}\right],$$

where $\underline{\Delta}$ is a constant strictly greater than 0. An analogous argument for $x_n = 1$ establishes that there exists some $\overline{\Delta} < 1$ such that if $q_n \in [1-\bar{\beta}, 1-\underline{\beta}]$, then

$$\varphi_{\sigma,x^{n-1}}(q_n,1) \in \left[\frac{1}{2}, \overline{\Delta}\right].$$

We next show that $q_n \in [\underline{\Delta}, \overline{\Delta}]$ for all n. Suppose this is not true. Let N be the first agent such that

$$q_N \in [0, \underline{\Delta}) \cup (\overline{\Delta}, 1] \tag{51}$$

in some equilibrium and some realized history. Then, $q_{N-1} \in [0, 1-\bar{\beta}) \cup (1-\underline{\beta}, 1]$ because otherwise, the dynamics of q_n implies a violation of Eq. (51) for any x_{N-1}. But note that if $q_{N-1} < 1 - \bar{\beta}$, then by Lemma 1 agent $N-1$ chooses action $x_{N-1} = 0$ and, thus by Eq. (50),

$$q_N = \left[1+\left(\frac{1}{q_{N-1}}-1\right)\frac{\mathbb{P}_\sigma(x_{N-1}=0 \mid q_{N-1}, x^{N-2}, \theta=0)}{\mathbb{P}_\sigma(x_{N-1}=0 \mid q_{N-1}, x^{N-2}, \theta=1)}\right]^{-1}$$

$$= \left[1+\left(\frac{1}{q_{N-1}}-1\right)\frac{1}{1}\right]^{-1}$$

$$= q_{N-1}.$$

By the same argument, if $q_{N-1} > 1-\underline{\beta}$, we have that $q_N = q_{N-1}$. Therefore, $q_N = q_{N-1}$, which contradicts the fact that N is the first agent that satisfies Eq. (51).

We next show that $q_n \in [\underline{\Delta}, \overline{\Delta}]$ for all n implies that x_n does not converge in probability to θ. Let y_k denote a realization of x_k. Then, for any n and any sequence of y_k's, we have

$$\mathbb{P}_\sigma(\theta=1, x_k=y_k \text{ for all } k \leq n) \leq \overline{\Delta}\mathbb{P}_\sigma(x_k=y_k \text{ for all } k \leq n),$$

$$\mathbb{P}_\sigma(\theta=0, x_k=y_k \text{ for all } k \leq n) \leq (1-\underline{\Delta})\mathbb{P}_\sigma(x_k=y_k \text{ for all } k \leq n).$$

By summing the preceding relations over all y_k for $k < n$, we obtain

$$\mathbb{P}_\sigma(\theta=1, x_n=1) \leq \overline{\Delta}\mathbb{P}_\sigma(x_n=1) \text{ and } \mathbb{P}_\sigma(\theta=0, x_n=0) \leq (1-\underline{\Delta})\mathbb{P}_\sigma(x_n=0).$$

Therefore, for any n, we have

$$\mathbb{P}_\sigma(x_n = \theta) \leq \overline{\Delta}\mathbb{P}_\sigma(x_n = 1) + (1 - \underline{\Delta})\mathbb{P}_\sigma(x_n = 0) \leq \max\{\overline{\Delta}, 1 - \underline{\Delta}\} < 1,$$

which completes the proof. ∎

Proof of Proposition 7. The first step is the following lemma.

Lemma 10 *Let $B(n) = \{b\}$ for some n. We define*

$$f(\underline{\beta}, \overline{\beta}) = \max\left\{1 - \frac{\underline{\beta}}{2(1 - \underline{\beta})}, \frac{3}{2} - \frac{1}{2\overline{\beta}}\right\}, \tag{52}$$

where $\overline{\beta}$ and $\underline{\beta}$ are the lower and upper supports of the private beliefs (cf. Definition 8). Let σ be an equilibrium. Assume that $\mathbb{P}_\sigma(x_b = \theta) \leq f(\underline{\beta}, \overline{\beta})$. Then, we have

$$\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\}) \leq f(\underline{\beta}, \overline{\beta}).$$

Proof. We first assume that $\mathbb{P}_\sigma(x_b = \theta) = f(\underline{\beta}, \overline{\beta})$ and show that this implies

$$U_b^\sigma \geq \overline{\beta}, \qquad \text{and} \qquad L_b^\sigma \leq \underline{\beta}, \tag{53}$$

where U_b^σ and L_b^σ are defined in Eq. (7). We can rewrite U_b^σ as

$$U_b^\sigma = \frac{N_b^\sigma}{1 - 2\mathbb{P}_\sigma(x_b = \theta) + 2N_b^\sigma} = \frac{N_b^\sigma}{1 - 2f(\underline{\beta}, \overline{\beta}) + 2N_b^\sigma}.$$

This is a decreasing function of N_b^σ and, therefore,

$$U_b^\sigma \geq \frac{1}{1 - 2f(\underline{\beta}, \overline{\beta}) + 2}.$$

Using $f(\underline{\beta}, \overline{\beta}) \geq \frac{3}{2} - \frac{1}{2\overline{\beta}}$, the preceding relation implies $U_b^\sigma \geq \overline{\beta}$. An analogous argument shows that $L_b^\sigma \leq \underline{\beta}$.

Since the support of the private beliefs is $[\underline{\beta}, \overline{\beta}]$, using Lemma 3 and Eq. (53), there exists an equilibrium $\sigma' = (\sigma'_n, \sigma_{-n})$ such that $x_n = x_b$ with probability one (with respect to measure $P_{\sigma'}$). Since this gives an expected payoff $\mathbb{P}_{\sigma'}(x_n = \theta \mid B(n) = b) = \mathbb{P}_\sigma(x_b = \theta)$, it follows that, $\mathbb{P}_\sigma(x_n = \theta \mid B(n) = b) = \mathbb{P}_\sigma(x_b = \theta)$. This establishes the claim that

$$\text{if } \mathbb{P}_\sigma(x_b = \theta) = f(\underline{\beta}, \overline{\beta}), \text{ then } \mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\}) = f(\underline{\beta}, \overline{\beta}). \tag{54}$$

We next assume that $\mathbb{P}_\sigma(x_b = \theta) < f(\underline{\beta}, \overline{\beta})$. To arrive at a contradiction, suppose that

$$\mathbb{P}_\sigma(x_n = \theta \mid B(n) = \{b\}) > f(\underline{\beta}, \overline{\beta}). \tag{55}$$

Now consider a hypothetical situation where agent n observes a private signal generated with conditional probabilities $(\mathbb{F}_0, \mathbb{F}_1)$ and a coarser version of the observation x_b, i.e., the random variable $\tilde{x}_b$ distributed according to

$$\mathbb{P}(\tilde{x}_b = 1 \mid \theta = 1) = 1 - Y_b^\sigma \left[\frac{1 - f(\underline{\beta}, \overline{\beta})}{\mathbb{P}_\sigma(x_b = \theta)}\right] \quad \text{and} \quad \mathbb{P}(\tilde{x}_b = 0 \mid \theta = 0) = 1 - N_b^\sigma \left[\frac{1 - f(\underline{\beta}, \overline{\beta})}{\mathbb{P}_\sigma(x_b = \theta)}\right].$$

It follows from the preceding conditional probabilities that $\mathbb{P}(\tilde{x}_b = \theta) = f(\underline{\beta}, \overline{\beta})$. We assume that agent n uses the equilibrium strategy σ_n. Using a similar argument as in the proof of Eq. (54), this implies that

$$\mathbb{P}(x_n = \theta \mid B(n) = \{b\}) = f(\underline{\beta}, \overline{\beta}). \tag{56}$$

Let z be a binary random variable with values $\{0, 1\}$ and is generated independent of θ with probabilities

$$\mathbb{P}(z = 1) = 1 - \frac{2Y_b^\sigma}{\mathbb{P}_\sigma(x_b = \theta)} \quad \text{and} \quad \mathbb{P}(z = 0) = 1 - \frac{2N_b^\sigma}{\mathbb{P}_\sigma(x_b = \theta)}.$$

This implies that $\mathbb{P}(z = j \mid \theta = j) = \mathbb{P}(z = j)$ for $j \in \{0, 1\}$. Using $\tilde{x}_b$ with probability $\frac{1}{1+f(\underline{\beta},\overline{\beta})}\left[2 + \frac{(Y_b^\sigma - 1)\mathbb{P}_\sigma(x_b=\theta)}{Y_b^\sigma}\right]$ and z otherwise generates the original observation (random variable) x_b. Therefore, from Eq. (55), $\mathbb{P}(x_n = \theta|B(n) = \{b\}) > f(\underline{\beta}, \overline{\beta})$, which contradicts Eq. (56), and completes the proof. ∎

Let f be defined in Eq. (52). We show by induction that

$$\mathbb{P}_\sigma(x_n = \theta) \leq f(\underline{\beta}, \overline{\beta}) \qquad \text{for all } n. \tag{57}$$

Suppose that for all agents up to $n-1$ the preceding inequality holds. Since $|B(n)| \leq 1$, we have

$$
\begin{aligned}
\mathbb{P}_\sigma(x_n = \theta) &= \mathbb{P}_\sigma(x_n = \theta \mid B(n) = \emptyset)\mathbb{Q}_n(B(n) = \emptyset) \tag{58} \\
&\quad + \sum_{b=1}^{n-1}\mathbb{P}_\sigma(x_n = \theta \mid B(n) = b)\mathbb{Q}_n(B(n) = \{b\}) \\
&\leq \mathbb{P}_\sigma(x_n = \theta \mid B(n) = \emptyset)\mathbb{Q}_n(B(n) = \emptyset) + \sum_{b=1}^{n-1} f(\underline{\beta}, \overline{\beta})\mathbb{Q}_n(B(n) = \{b\}),
\end{aligned}
$$

where the inequality follows from the induction hypothesis and Lemma 10. Note that having $B(n) = \emptyset$ is equivalent to observing a decision b such that $\mathbb{P}_\sigma(x_b = \theta) = 1/2$. Since $1/2 \leq f(\underline{\beta}, \overline{\beta})$, Lemma 10 implies that $\mathbb{P}_\sigma(x_n = \theta|B(n) = \emptyset) \leq f(\underline{\beta}, \overline{\beta})$. Combined with Eq. (58), this completes the induction.

Since the private beliefs are bounded, i.e., $\underline{\beta} > 0$ and $\overline{\beta} < 1$, we have $f(\underline{\beta}, \overline{\beta}) < 1$ [cf. Eq. (52)]. Combined with Eq. (57), this establishes that $\liminf_{n\to\infty} \mathbb{P}_\sigma(x_n = \theta) < 1$, showing that asymptotic learning does not occur at any equilibrium σ. ∎