Naive Bayesian Learning in Social Networks

JERRY ANUNROJWONG, Harvard University, USA

NAT SOTHANAPHAN, Massachusetts Institute of Technology, USA

The DeGroot model of naive social learning assumes that agents only communicate scalar opinions. In practice, agents communicate not only their opinions, but their confidence in such opinions. We propose a model that captures this aspect of communication by incorporating signal informativeness into the naive social learning scenario. Our proposed model captures aspects of both Bayesian and naive learning. Agents in our model combine their neighbors' beliefs using Bayes' rule, but the agents naively assume that their neighbors' beliefs are independent. Depending on the initial beliefs, agents in our model may not reach a consensus, but we show that the agents will reach a consensus under mild continuity and boundedness assumptions on initial beliefs. This eventual consensus can be explicitly computed in terms of each agent's centrality and signal informativeness, allowing joint effects to be precisely understood. We apply our theory to adoption of new technology. In contrast to Banerjee et al. [5], we show that information about a new technology can be seeded initially in a tightly clustered group without information loss, but only if agents can expressively communicate their beliefs.

CCS Concepts: • **Theory of computation → Social networks**; **Convergence and learning in games**; • **Networks** → *Network economics*;

1 INTRODUCTION

This paper introduces an intuitive naive social learning model that extends the DeGroot model when agents are differently informed. In standard naive learning, each agent's belief is a scalar, and each agent updates her beliefs by taking the weighted averages of her neighbors' beliefs with fixed weights. This implicitly assumes that every agent's signal quality is the same, which is often not true in practice. We propose a solution to this problem by representing each agent's belief as a probability distribution over possible states of the world. Signal quality can then be naturally communicated by using Bayes' rule to combine one's belief with the beliefs of one's neighbors. For example, if an agent has a scalar opinion μ and a scalar confidence τ indicating how sure she is of that scalar opinion, her belief can be represented by the Gaussian $\mathcal{N}(\mu, 1/\tau)$. Our model can also accommodate other forms of information. For example, an agent may believe that either one of the two values is very likely, while other values are less likely. This belief can be represented as a double-peaked distribution, but it cannot be represented in standard DeGroot models.

It is known separately in Bayesian learning and naive learning literatures that higher quality signals and more centrally located agents, respectively, have more influence in social learning. The innovation in our model is to introduce signal quality into the naive learning setting, allowing us to precisely investigate how signal quality and agent centrality interact. The two notions are

intertwined in the *weighted likelihood function*, where the formal expression is at equation (1) in Section 1.1. Informally, the value of this function at θ is the product of the votes of each agent, where an agent's vote is how likely the state value θ is after the agent received the signal compared to before (the signal quality part), and the number of votes is the agent's eigenvector centrality (the centrality part).

Our main theorem is that, under some conditions, long-run beliefs are concentrated near the maximizers of the weighted likelihood function. If we know the probability densities of initial beliefs, we can explicitly calculate the consensus belief. This allows us to analyze the interaction; for example, we can calculate how much more informative a signal has to be to offset the centrality disadvantage. This formula can have practical applications in the analysis of information spread in decentralized networks, such as propagation of rumors or adoption of new technologies. Practitioners interested in empirical results can also use our formula without having to fully grasp the theory in the paper. Some examples of applications are discussed in Section 7.

To the best of our knowledge, before this work, only Banerjee et al. [5] have directly tried to solve the issue of differently informed agents in DeGroot learning; they assume that each agent can either hold a scalar belief or be uninformed. While their model can capture the spread of awareness of new things, their model still has no notion of signal informativeness that our model provides.

Our model also gives a different prediction than that of Banerjee et al. [5] on how to seed a new technology for widespread adoption. Their model has a *clustered seeding* problem: there is substantial information loss if the initially informed agents ("seeds") are clustered in a small group of communities. This is a problem because a social planner who wants to maximize the impact of a newly introduced technology should seed it to centrally located opinion leaders, but these leaders are often closely connected to each other. Our model does not have this problem; the influence of each agent depends only on the agent's signal informativeness and centrality. The key modeling difference is that agents in our model can communicate confidence of their beliefs that their neighbors can take into account. Therefore, to solve the clustered seeding problem, the social planner should encourage open and expressive communication between agents.

1.1 Contributions

Our agents are in a social network, represented as a strongly connected directed graph. Assume that there is a true state of the world θ^*. Each agent gets a signal drawn from a distribution that depends on θ^*. She then forms a *belief*, the posterior probability distribution of θ^* given the signal. In each round, each agent has access to beliefs of her neighbors and naively assumes that these all come from *independent* signals. She then forms the Bayesian posterior of θ^* given the signals of her own and her neighbors as her new belief. This gives rise to an update rule that combines her belief with her neighbors' beliefs that she uses in every round.

We prove that our learning process can be viewed as follows. At every time step, each agent passes copies of signals she has to her neighbors, so copies of signals "flow" through the network. Every copy is taken as independent and no copy is discarded, so the number of copies of agent j's signal that agent i has at round t is the number of paths from j to i of length t. Agent i's belief at time t is then the posterior belief conditional on these repeated signals (Prop. 4.6).

We then tackle the important question of what happens to agents' beliefs in the long run. The answer depends on the following *weighted likelihood function L* (Def. 5.1). Let $f_i^{(t)}(\theta)$ be agent i's belief of θ^* in round t and $f_*(\theta)$ be the common prior of θ^*. Then

$$L(\theta) = \prod_{i=1}^{N} \left(\frac{f_i^{(0)}(\theta)}{f_*(\theta)} \right)^{v_i}, \tag{1}$$

where v_i is the eigenvector centrality of agent i. We show that if $L(\theta_1) < L(\theta_2)$, then as $t \to \infty$,

$$\frac{f_i^{(t)}(\theta_1)}{f_i^{(t)}(\theta_2)} \to 0 \tag{2}$$

(Prop. 5.3). In other words, agents' beliefs that the true parameter is θ_2 are much stronger than the corresponding beliefs for θ_1 as time goes on. Thus, in the long run, agents' beliefs should concentrate at values θ that maximize or nearly maximize L.

The form of the weighted likelihood function highlights two factors that determine the quality of learning. An agent who is more *centrally located* (high v_i) and has a more *informative signal* (high or low $f_i^{(0)}(\theta)/f_*(\theta)$) has more influence over the consensus belief. Note that our consensus belief is fundamentally different from that of DeGroot. In our model, the limit belief is the weighted likelihood maximizer, while in DeGroot, it is a weighted average of beliefs.

The fact that beliefs tend to converge to a point distribution agrees with the following intuition: agents take a growing number of repeated signals as independent, so they become increasingly confident in their beliefs. The fact that the limit belief is tractable and interpretable comes as a surprise, and we believe this to be an attractive feature of our model.

In spite of this intuition and much to our astonishment, the concentration previously discussed turns out not to hold in general if Θ, the state space, is infinite. We discover that even with a single $\theta_{\max}$ uniquely maximizing L, agents' beliefs may not converge to the point distribution at $\theta_{\max}$. We resolve this difficulty by introducing an extra condition: if $\theta_{\max}$ still maximizes a "perturbed" version of L, then the concentration result holds. We can think of this as $\theta_{\max}$ being a sufficiently robust maximizer of L. This result requires careful analysis of convergences of infinite sequences, and the generalization of this idea for an arbitrary state space Θ can be found in Proposition 5.5.

When the state space $\Theta = \mathbb{R}$ is a continuum, for example in the case of Gaussian beliefs, this perturbed L condition is still too strong: perturbing L will almost always shift the maximizer due to the continuity of the space. The correct condition needs to allow the maximizer to shift, but not too much. Furthermore, it is insufficient to consider the density at a single point, since this tells us nothing about densities at nearby points and so cannot establish concentration. Instead, we need to show that for any neighborhood of $\theta_{\max}$, agents' beliefs that the true state lies in that neighborhood tend to 1. Our Master Theorem 5.8 pulls together all of these ideas.

With the Master Theorem 5.8, we derive a convergence theorem when the state space is a continuum, provided that the initial beliefs and priors are sufficiently continuous (Def. 6.4). Our main result in this most difficult case is Theorem 6.5. The following is a simplified version of this theorem, simplified from Corollary 6.6.

THEOREM 1.1. *Assume that the state space* $\Theta = \mathbb{R}^k$. *Suppose that the normalized beliefs* $f_i^{(0)}(\theta)/f_*(\theta)$ *are bounded continuous functions; the weighted likelihood function* L *attains the maximum* $L_{\max}$ *uniquely at the point* $\theta_{\max}$; *and* L *decays in the sense that there is a bounded set* S *such that* $\sup_{\theta \notin S} L(\theta) < L_{\max}$. *Then, as* $t \to \infty$, *beliefs* $f_i^{(t)}$ *converge to the point distribution at* $\theta_{\max}$.

This result states that whenever the initial beliefs are continuous functions that are bounded with respect to the prior and they decay sufficiently fast, agents' beliefs converge to the maximizer of the weighted likelihood function if this maximizer is unique. This result encompasses a wide range of probability distributions in practice. For example, the theorem applies when beliefs are multivariate Gaussian. Therefore, Theorem 1.1 is widely applicable.

The rest of the paper is organized as follows. We discuss related literature in the remaining of this section. Section 2 describes our model. Section 3 discusses mathematical preliminaries. Section 4 gives fundamental definitions, interprets the updated beliefs, and analyzes feasibility of initial

conditions. Section 5 proves the "Master Theorem" that establishes concentration of beliefs near the weighted likelihood maximizers. Section 6 uses the Master Theorem to prove convergence of beliefs in the finite, infinite discrete, and $\mathbb{R}^k$ cases. Section 7 gives applications in the cases of binary, Poisson, and Gaussian beliefs.

1.2 Related Literature

Our paper is closely related to a growing literature on learning and belief formation in social networks [3, 4, 6, 11, 21, 22]. Much of this literature focuses on Bayesian learning [1, 7, 15, 18] and DeGroot learning [8, 12]. Comparable to Golub and Jackson [12], our model shows that centrally located agents, as measured by eigenvector centralities, have more influence on the consensus belief. However, in our model, eigenvector centralities appear as duplicate agents rather than linear weights in the limit belief.

Specifically, our paper belongs to the quasi-Bayesian social learning literature. Eyster and Rabin [9, 10] and Rahimian et al. [19] study long-run aggregation of information when agents ignore repetition. Li and Tan [14] study learning where each agent updates both her belief on the state of the world and her belief on her neighbors' beliefs. Jadbabaie et al. [13] also combine DeGroot learning with Bayesian updates. Lastly, likelihood ratios play a central role in our model; they appear in Rahimian et al. [20] and Molavi et al. [17] in a different context.

Finally, our work is related to Banerjee et al. [5]. In their model, the informedness of agents is a binary state, informed individuals hold scalar beliefs, and the learning rule is to take the average belief of informed neighbors. This contrasts with our model where informedness is a continuous attribute. Their work identifies the clustered seeding problem discussed in the introduction, which does not occur in our model. See further discussion of this problem in Section 7.3.

2 MODEL

We discuss the assumptions of our model in this section.

2.1 Social Network

Our agents are in a social network, modeled as a directed graph G with $N \geq 2$ nodes, labeled $1, 2, \ldots, N$ corresponding to N agents. An edge from i to j is denoted (i, j) and means that there is information flow from agent i to agent j. Assume that G is strongly connected, that is, for every i and j there is a path from i to j. Suppose that there is a self-loop (i, i) at every node, meaning that information flows from every agent to herself. If there is an edge (i, j), we say that i is an *in-neighbor* of j. Let $N(i)$ denote the set of in-neighbors of i.

2.2 Beliefs

Let Θ be the space of parameters θ, modeled as a measure space. We model beliefs as *probability density functions* on Θ: nonnegative measurable functions $f : \Theta \to \mathbb{R}_{\geq 0}$ such that the integral $\int_{\theta \in \Theta} f(\theta) = 1$. These reduce to probability mass functions in a discrete setting, i.e. when Θ is discrete with the counting measure. Note that we model beliefs as probability densities, which is a special case of general probability distributions. For example, a point distribution on $\mathbb{R}$ cannot be represented as a density, but most probability distributions in practice are densities. A *prior* on Θ is any nonnegative measurable function $f : \Theta \to \mathbb{R}_{\geq 0}$. If the integral $\int_{\theta \in \Theta} f(\theta)$ diverges, we call the prior f *improper*. Note that our improper priors are not necessarily limits of proper priors, but simply densities whose integrals diverge and so cannot be normalized. An example of an improper prior is the uninformative prior, $f(\theta) \equiv 1$, where Θ is an infinite discrete set or an unbounded interval of $\mathbb{R}$.

Suppose that each agent has a common (possibly improper) prior f_* for the true state of the world θ^* on Θ. Assume that $f_*(\theta) > 0$ for all $\theta \in \Theta$. This assumption is important in order to state the update rule for beliefs (e.g. in Definition 2.1).

2.3 Belief Updates

Agents update their beliefs based on the beliefs of their in-neighbors using the following rule, which will be justified shortly.

Definition 2.1. The *naive Bayesian updated belief* for agent i with respect to beliefs $f_1, \ldots, f_N$ is the belief

$$f(\theta) = \frac{f_*(\theta) \prod_{j \in N(i)} \left(f_j(\theta)/f_*(\theta) \right)}{\int_{\theta' \in \Theta} f_*(\theta') \prod_{j \in N(i)} \left(f_j(\theta')/f_*(\theta') \right)},$$

which is well-defined whenever the integral in the denominator is nonzero and finite.

It is easily checked that this definition defines a belief. We now provide a justification for this definition. In the justification, we will assume that f_* is a probability density function, but the formula in Definition 2.1 also makes sense for improper priors.

Consider the following scenario called an *underlying scenario*. Suppose that agent i knows a family of probability density functions $(p_{\theta,i})_{\theta \in \Theta}$, indexed by the parameter θ and the agent i, over a common sample space Ω_i. We make the following marginalizability assumption, which is important in order to reason consistently about marginal densities. For any integers $n_1, \ldots, n_N \geq 0$, let $X_{i,1}, \ldots, X_{i,n_i}$ be random variables drawn independently from $p_{\theta,i}$. We thus have a joint probability density for these random variables $p\left((X_{i,j})_{1 \leq i \leq n, 1 \leq j \leq n_i}, \theta\right) = p(\theta) \prod_{i=1}^{N} \prod_{j=1}^{n_i} p(X_{i,j}|\theta)$. Assume that this distribution can be marginalized over θ with positive density.

Agent i receives a signal X_i drawn from $p_{\theta^*,i}$, the distribution corresponding to the true parameter θ^*, without knowing θ^*. She then forms a belief $f_i(\theta)$ for θ^* conditional on the signal X_i. Suppose that agent i learns about beliefs $f_j(\theta)$ of all her in-neighbors $j \in N(i)$, formed in the same way, and wishes to perform a Bayesian update on her own belief. The result will be the belief $f(\theta)$ for θ^* conditional on signals X_j for every $j \in N(i)$. Even though the resulting density seems to depend on X_j, we will show in the next proposition that this update can be performed knowing only the beliefs f_j but not the private signals X_j, with the formula as in Definition 2.1.

PROPOSITION 2.2. *In an underlying scenario described above, the updated belief for agent i is the naive Bayesian updated belief with respect to beliefs $f_1, \ldots, f_N$ as in Definition 2.1.*

PROOF SKETCH. Use Bayes' rule to compute the updated belief as

$$f(\theta) = p\left(\theta|(X_j)_{j \in N(i)}\right) = \frac{p(\theta) \prod_{j \in N(i)} p\left(X_j|\theta\right)}{\int_{\theta' \in \Theta} p(\theta') \prod_{j \in N(i)} p\left(X_j|\theta'\right)}.$$

Another application of Bayes' rule on $p(X_j|\theta)$ finishes the proof.   □

2.4 Belief Dynamics

Here we describe the evolution of beliefs. Agents share a common prior f_* of θ^*, which may be improper, and agent i starts off with *initial belief* $f_i^{(0)}$ of θ^*. The functions $f_*, f_1^{(0)}, \ldots, f_N^{(0)}$ are the *initial condition* of the model. The subtlety is that not all initial conditions lead to well-defined updates. Nevertheless, in Section 4.3, we will provide simple checks that guarantee the existence of updated beliefs. For $t = 0, 1, 2, \ldots$, let $f_i^{(t)}$ be agent i's belief of θ^* at round t. Agents update beliefs as follows. For $t \geq 0$, $f_i^{(t+1)}$ is the naive Bayesian updated belief (Def. 2.1) for agent i with respect to beliefs $f_1^{(t)}, \ldots, f_N^{(t)}$.

We call an initial condition *feasible* if the updated beliefs are well-defined in every round, that is, the integral in Definition 2.1 is nonzero and finite. Moreover, an initial condition *has an underlying scenario* if it arises from an underlying scenario in the above justification of Definition 2.1. Proposition 4.6 shows that an initial condition with an underlying scenario is always feasible.

For an initial condition with an underlying scenario, note that although the update rule is rationally justified, it is only optimal for the first round. In later rounds, an optimal update needs to take into account that the agents' beliefs are already updated at least once and so no longer independent. Still, our naive agents use the same update rule for every round. This modeling choice formalizes what we said earlier that people use Bayes' rule to learn, but they ignore repeated information.

We name our model *naive Bayesian* for two reasons. First, our model combines features from naive learning and Bayesian learning. Second, our agents assume that distributions of neighbors are independent conditional on the state, which is also the assumption made in Naive Bayes classifiers in machine learning.

3 MATHEMATICAL PRELIMINARIES

Here, we review Perron-Frobenius theory, which helps us determine the behavior of large powers of matrices and establish convergence of beliefs. In the full version of this paper, we also review the theory of L^p spaces. That material helps us bound functions and establish feasibility of initial conditions, but it is not used after we assume feasibility, so it is omitted for lack of space.

Let G be the graph of our model, and let A be the adjacency matrix of G: $A_{ij} = 1$ if there is an edge (i, j) and 0 otherwise. The Perron-Frobenius theory provides us with properties of eigenvalues and eigenvectors of A.

Definition 3.1. A square matrix M is *primitive* if its entries are nonnegative and the entries of M^k are positive for some $k > 0$.

Lemma 3.2. *A is a primitive matrix.*

The following theorem from Meyer [16, Chapter 8], a version of Perron-Frobenius for primitive matrices, characterizes the largest eigenvalue of a primitive matrix and the growth rate of the matrix's powers.

Theorem 3.3 (Perron-Frobenius for Primitive Matrices). *Let M be a primitive matrix. Let $r = \rho(M)$ be the spectral radius of M, that is, the maximum absolute value of eigenvalues of M. Then r is an eigenvalue of M, any other eigenvalue of M has absolute value less than r, and there are unique vectors v and w with positive entries such that*

$$Mv = rv, \quad wM = rw, \quad \|v\|_1 = \|w\|_1 = 1.$$

Note that $\|\cdot\|_1$ denotes the sum of entries. Call v the Perron vector *of M. Then w is the Perron vector of $M^\top$.*

Finally, the limit $\lim_{k \to \infty} M^k / r^k$ exists and is given by

$$P := \lim_{k \to \infty} \frac{M^k}{r^k} = \frac{vw^\top}{w^\top v}.$$

This limit P is called the Perron projection *of M.*

The following lemma will be useful when we reason that the entries of A^k tend to infinity.

Lemma 3.4. *The spectral radius of A is greater than 1, that is, $\rho(A) > 1$.*

We now define the notion of centrality alluded to in the introduction.

Definition 3.5. If v is the Perron vector of A, then call v_i the *eigenvector centrality* of agent i.

4 UNDERSTANDING THE MODEL

In Section 4.1, we introduce some preliminary definitions that will be useful in understanding the model in Sections 4.2 and 4.3. Then we interpret the updated beliefs in Section 4.2. Finally, we find the conditions that guarantee that an initial condition is feasible in Section 4.3.

4.1 Preliminary Definitions

We introduce the notion of *normalized beliefs*.

Definition 4.1. The *normalized belief* of agent i at round t is the function

$$g_i^{(t)}(\theta) := \frac{f_i^{(t)}(\theta)}{f_*(\theta)}.$$

We can interpret this as how large a a belief is compared to the prior. This quantity will appear in the weighted likelihood function in Definition 5.1 and play a central role in our analysis.

With normalized beliefs, the update rule in Definition 2.1 assumes the simpler form

$$g_i^{(t+1)}(\theta) = \frac{\prod_{j \in N(i)} g_j^{(t)}(\theta)}{\int_{\theta' \in \Theta} f_*(\theta') \prod_{j \in N(i)} g_j^{(t)}(\theta')}. \tag{3}$$

Definition 4.2. For nodes i and j of G and $t \geq 0$, let $P_{ij}^{(t)}$ be the number of paths from i to j of length t. By convention, there is one path of length 0 from a node to itself and no path of length 0 from a node to a different node.

LEMMA 4.3. *The numbers of paths satisfy the recurrence $P_{ij}^{(t+1)} = \sum_{k \in N(j)} P_{ik}^{(t)}$. Moreover, $P_{ij}^{(t)}$ is the (i, j) entry of the matrix A^t.*

Finally, we formally define the notion of *k-feasibility* and *feasibility*.

Definition 4.4. For any $k \geq 1$, an initial condition $f_*, f_1, \ldots, f_N$ is *k-feasible* if the updated beliefs are well-defined up to at least round k. An initial condition is *feasible* if it is *k*-feasible for every k.

4.2 Interpretation of Updated Beliefs

We interpret what agents are doing in our model. Our goal is to show that each agent aggregates information where another agent's initial belief is weighted by the number of paths from that agent to the aggregating agent. This means that an agent who has many paths to other agents, i.e. an agent with high centrality, will be influential in our model. This result will lead to conditions for feasible initial conditions in Section 4.3 and show concentration of beliefs in Section 5.

PROPOSITION 4.5. *An initial condition $f_*, f_1^{(0)}, \ldots, f_N^{(0)}$ is k-feasible if and only if the expression*

$$\frac{\prod_{j=1}^N g_j^{(0)}(\theta)^{P_{ji}^{(t)}}}{\int_{\theta' \in \Theta} f_*(\theta') \prod_{j=1}^N g_j^{(0)}(\theta')^{P_{ji}^{(t)}}}$$

is well-defined for all i and $0 \leq t \leq k$, and that expression is the formula for $g_i^{(t)}(\theta)$.

PROOF SKETCH. Induct on k and use Lemma 4.3 to get exponents into the form $P_{ji}^{(t)}$. □

Now we claim that an initial condition with an underlying scenario is feasible, as well as give an explicit formula for the updated beliefs. This is important as it gives us insight into the update process.

PROPOSITION 4.6. *An initial condition* $f_*, f_1^{(0)}, \ldots, f_N^{(0)}$ *with an underlying scenario is feasible. If agent i receives signal* X_i *in the underlying scenario and* $I_i^{(t)}$ *is the ordered tuple consisting of* $P_{ji}^{(t)}$ *independent copies of* X_j *for each j, then*

$$f_i^{(t)}(\theta) = p(\theta | I_i^{(t)}).$$

In other words, the belief of agent i in round t for each parameter θ *is the belief for that parameter conditional on the information* $I_i^{(t)}$.

Proposition 4.6 gives us an intuition of the process. At every time step, each agent passes copies of signals she has to her neighbors, so copies of signals "flow" through the network. Every copy is taken as independent and no copy is discarded, so the number of copies of agent j's signal that agent i has at round t is the number of paths from j to i of length t. Moreover, we can interpret the normalized belief $g_i^{(t)}(\theta)$ as the amount of support that $I_i^{(t)}$ provides for θ. Since $g_i^{(t)}(\theta) = p(\theta|I_i^{(t)})/p(\theta)$ by Proposition 4.6, this quantity is larger or smaller than 1 if learning $I_i^{(t)}$ increases or decreases the probability density at θ, respectively.

From Proposition 4.6, we can qualitatively predict how our model behaves in the long term. Because repeated signals in $I_i^{(t)}$ are taken as independent, agents should become overconfident over time. Furthermore, because the numbers of repeated signals are the numbers of paths, agents with more paths to other agents (those with higher eigenvector centralities) should be more influential. Our results on concentration and convergence of beliefs in Sections 5 and 6 confirm these predictions.

4.3 Initial Conditions

In the full version of this paper, we use the theory of L^p spaces to investigate containment relationships between classes of initial conditions and present simple sufficient conditions for feasibility. The only result that is used in this paper is the following sufficient condition (Corollary 4.14 in the full paper) that is easy to check in practice.

COROLLARY 4.7. *An initial condition* $f_*, f_1^{(0)}, \ldots, f_N^{(0)}$ *is feasible if the product of initial beliefs* $f_1^{(0)} \ldots f_N^{(0)}$ *does not vanish almost everywhere and, for every i,* $g_i^{(0)}$ *is bounded almost everywhere.*

5 CONCENTRATION OF BELIEFS

With the tools we have built up, we can now start seriously analyzing the model. Assume from this point onwards that the initial condition is feasible. Our goal in this section is to establish the following remarkable result about concentration of beliefs. In the long term, beliefs of agents in our model tend to concentrate at points of Θ which maximize the *weighted likelihood function L* (Def. 5.1). Specifically, for a set $A \subseteq \Theta$ such that the values of L on this set are dominated in some sense by $L(\theta')$ for another $\theta' \in \Theta$,

$$\int_{\theta \in A} f_i^{(t)}(\theta) \to 0$$

as $t \to \infty$. We can think of this as saying that the agents' beliefs are being concentrated more and more near the maximizer of L, and so the beliefs on any set that stays clear of the maximizer tend to zero. The precise statement is in the Master Theorem 5.8. Once we have this result, it will become relatively easy to show convergence of the agents' beliefs towards maximizers of L in various settings; this will be done in Section 6.

5.1 Weighted Likelihood

We first define the weighted likelihood function mentioned in the introduction.

Definition 5.1. Assume that the initial condition is feasible. The *weighted likelihood function* is defined for each $\theta \in \Theta$ by

$$L(\theta) = \prod_{i=1}^{N} (g_i^{(0)}(\theta))^{v_i}.$$

Recall that $g_i^{(0)}$ is the normalized beliefs defined in Definition 4.1 and that the eigenvector centrality v_i can be thought of as the influence of agent i. So this function aggregates the normalized beliefs of agents while being biased by the agents' influences. In the full paper, we discuss how the weighted likelihood function differs from the *unbiased* likelihood that results from optimal information aggregation. To see how this function is being "biased," the *unbiased* likelihood of each parameter θ given the initial beliefs of agents formed according to an underlying scenario (Sec. 2.3) is given by

$$L_{\text{unbiased}}(\theta) = c f_*(\theta) \prod_{i=1}^{N} g_i^{(0)}(\theta),$$

where c is a normalizing constant. Under optimal information aggregation, the posterior belief is proportional to L_{unbiased}. To derive this, simply do the computation analogous to the one in Proposition 2.2. Notice that L and L_{unbiased} are different in two ways. First, each $g_i^{(0)}$ factor in L is to the power of v_i, suggesting that L is biased by the centrality of each agent as already pointed out. Second, there is an extra factor of f_*, the prior, in L_{unbiased}. This is because agents in our model become so confident in their opinions that the prior is disregarded entirely, so this factor does not appear in L.

We now show that the weighted likelihood function guides the convergence of the ratio of beliefs at two different points (Prop. 5.3). We can think of this as a "pointwise" concentration result. It is much weaker than the actual concentration result we seek (except when Θ is finite; see Theorem 6.2). Nonetheless, it is the basis of the proof of the Master Theorem 5.8.

The next proposition gives an explicit formula for the ratio of beliefs at two different points, highlighting the terms that become small in the limit. It is a crucial component that is used in Section 5.2 to prove the Master Theorem 5.8.

PROPOSITION 5.2. *Let $r = \rho(A)$ be the spectral radius of A, and v and w be the Perron vectors of A and $A^\top$, respectively. Then there are $\varepsilon_{i,j}^{(t)} \to 0$ as $t \to \infty$ such that for any $\theta_1, \theta_2 \in \Theta$ with $L(\theta_2) > 0$,*

$$\frac{f_i^{(t)}(\theta_1)}{f_i^{(t)}(\theta_2)} = \frac{f_*(\theta_1)}{f_*(\theta_2)} \left[\prod_{j=1}^{N} \left(\frac{g_j^{(0)}(\theta_1)}{g_j^{(0)}(\theta_2)} \right)^{v_j + \varepsilon_{i,j}^{(t)}} \right]^{w_i r^t / w^\top v}.$$

From this proposition, we can begin to see the weighted likelihood function take shape: in the formula above, if the prior factors in front are ignored and the small terms $\varepsilon_{i,j}^{(t)}$ disregarded, what is left is the ratio $L(\theta_1)/L(\theta_2)$ raised to some power.

PROOF SKETCH. The proof is essentially by direct computation. Use Proposition 4.5 to get the expression for the ratio $g_i^{(t)}(\theta_1)/g_i^{(t)}(\theta_2)$. Then use Lemma 4.3 and Theorem 3.3 to transform it into the desired formula.   □

We can now show the "pointwise" concentration of beliefs at maximizers of L. This hints at the actual concentration result in the next subsection.

PROPOSITION 5.3. *If $\theta_1, \theta_2 \in \Theta$ are such that $L(\theta_1) < L(\theta_2)$, then as $t \to \infty$,*

$$\frac{f_i^{(t)}(\theta_1)}{f_i^{(t)}(\theta_2)} \to 0.$$

5.2 Master Theorem

This is a technical section; its aim is to prove the Master Theorem 5.8 which establishes concentration of beliefs in very general settings. In order to do this, subtle issues of convergence arise which can be best dealt with using ideas from measure theory (e.g. Lebesgue's dominated convergence theorem). Even in the relatively benign setting $\Theta = \mathbb{N}$, these issues are not so easy to circumnavigate. We first present the general idea of the Master Theorem. Suppose $A \subseteq \Theta$ and $\theta' \in \Theta$ are such that $L(\theta) < L(\theta')$ for all $\theta \in A$. Thus, θ' "dominates" all of A in the sense of the weighted likelihood function. We hope that $\int_{\theta \in A} f_i^{(t)}(\theta) \to 0$ as $t \to \infty$. Proposition 5.3 implies the pointwise result $f_i^{(t)}(\theta)/f_i^{(t)}(\theta') \to 0$ as $t \to \infty$. We would like to aggregate this pointwise result into

$$\frac{\int_{\theta \in A} f_i^{(t)}(\theta)}{f_i^{(t)}(\theta')} \to 0 \tag{4}$$

(Prop. 5.5), so that we may conclude $\int_{\theta \in A} f_i^{(t)}(\theta) \to 0$. To do so, we need to control the convergence in Proposition 5.3; the convergence needs to be "uniform" in some way. It turns out that the correct hypothesis is that the inequality $L(\theta) < L(\theta')$ holds when the exponents in L are slightly perturbed: $L(\theta)g_i^{(0)}(\theta)^{\delta_i} < L(\theta')g_i^{(0)}(\theta')^{\delta_i}$ for small $\delta_i > 0$. As explained in Section 1.1, we can think of this as θ' dominating A *robustly*, i.e. the domination does not fail even if the centralities of agents are being slightly perturbed. In the full version, we offer an example where this robustness condition fails and concentration also fails, which shows that this hypothesis cannot be omitted.

After obtaining the bound (4), if Θ is discrete, i.e. $\Theta = \mathbb{N}$, then the obvious bound $f_i^{(t)}(\theta') \le 1$ finishes the proof. But this is not so easy if Θ is continuous, i.e. $\Theta = \mathbb{R}^k$. In fact, if concentration is to be true, it is likely that $f_i^{(t)}(\theta') \to \infty$ as $t \to \infty$ if θ' maximizes L, and so we cannot get from (4) to the desired result directly. The second idea is to pick θ' that is *not* a maximizer of L and show that $f_i^{(t)}(\theta') \to 0$ (which is likely to happen if concentration near the maximizer is to occur), which will then allow us to get the desired result. To do this, we find a set B such that $L(\theta) > L(\theta')$ for all $\theta \in B$, i.e. now the set B dominates θ' instead. Then if the convergence is similarly controlled, we should get

$$\frac{\int_{\theta \in B} f_i^{(t)}(\theta)}{f_i^{(t)}(\theta')} \to \infty$$

(Prop. 5.7), which implies $f_i^{(t)}(\theta') \to 0$. Putting all of this together completes the proof.

To summarize, the mechanics of the Master Theorem are a two-step domination process. Suppose that we have sets A and B and a point θ', such that θ' dominates A and B dominates θ', and both dominations are robust. Then the latter domination forces the density at θ' to go to zero, which, by the former domination, in turn forces the density on A to go to zero.

In what follows, we implement the program outlined above. First we formalize the notion of this "robust domination" which allows us to control convergence. The following lemma is a purely analytical statement that contains the essential ingredients to carrying out the "θ' dominates A" part; then Proposition 5.5 will actually carry it out. The proof of the lemma is given in the full version.

LEMMA 5.4. *Let S be a measure space. For $t \in \mathbb{Z}_{\geq 0}$ and $1 \leq i \leq N$, let $a_i, c : S \to \mathbb{R}_{\geq 0}$ be measurable functions, $r_t > 0$ and $\varepsilon_i^{(t)} \geq -1$. As $t \to \infty$, $r_t \to \infty$ and, for every i, $\varepsilon_i^{(t)} \to 0$. The functions a_i has product $A(x) := \prod_{i=1}^{N} a_i(x) \leq 1$ for almost every $x \in S$. Moreover, for every i, there is $\delta_i > 0$ such that $A(x)a_i(x)^{\delta_i} \leq 1$ for almost every $x \in S$. Suppose that for every t, the integral*

$$I_t := \int_{x \in S} c(x) \left(\prod_{i=1}^{N} a_i(x)^{1+\varepsilon_i^{(t)}} \right)^{r_t} < \infty.$$

Then, as $t \to \infty$,

$$I_t \to \int_{A(x)=1} c(x) < \infty.$$

In particular, if $A(x) < 1$ almost everywhere, then $I_t \to 0$ as $t \to \infty$.

Note that the form of I_t is reminiscent of the expression in Proposition 5.2 and indeed this is what the lemma will be used for; but this abstraction in terms of these various variables is necessary in order for the proof of the lemma to not become too large and convoluted.

Once we have Lemma 5.4, Proposition 5.5 follows immediately, although it may be a little tedious to check that all conditions are satisfied.

PROPOSITION 5.5. *Let $A \subseteq \Theta$ be a measurable set. Let $\theta' \in \Theta$ be such that $L(\theta') > 0$ and $L(\theta) \leq L(\theta')$ for almost every $\theta \in A$. Moreover, for each i, there is $\delta_i > 0$ such that $L(\theta)g_i^{(0)}(\theta)^{\delta_i} \leq L(\theta')g_i^{(0)}(\theta')^{\delta_i}$ for almost every $\theta \in A$. Then there is $0 \leq M < \infty$ such that, as $t \to \infty$,*

$$\frac{\int_{\theta \in A} f_i^{(t)}(\theta)}{f_i^{(t)}(\theta')} \to M.$$

If the set $\{\theta \in A : L(\theta) = L(\theta')\}$ has measure zero, then $M = 0$.

The previous proposition completes the "θ' dominates A" part of the program. To carry out the "B dominates θ'" part, we prove the following Lemma 5.6 and Proposition 5.7, which can be thought of as analogous versions of Lemma 5.4 and Proposition 5.5 but with reversed inequalities and slight changes in details. In particular, the part that changes is the final step of the lemma where here we use Fatou's lemma to conclude instead.

LEMMA 5.6. *Let $S, a_i, c, r_t, \varepsilon_i^{(t)}, A$ and I_t be as in Lemma 5.4 with the following changes. The inequalities involving A are reversed: $A(x) \geq 1$ for almost every $x \in S$, and for every i, there is $\delta_i > 0$ such that $A(x)a_i(x)^{\delta_i} \geq 1$ for almost every $x \in S$. Moreover, the integral I_t may be infinite. If $A(x) = 1$ almost everywhere, then as $t \to \infty$,*

$$I_t \to \int_{x \in S} c(x),$$

where the limit can be infinite. Otherwise, $I_t \to \infty$ as $t \to \infty$.

One can prove the following proposition by applying Lemma 5.6 in the same way that one applies Lemma 5.4 to prove Proposition 5.5. So we omit the proof of the following proposition.

PROPOSITION 5.7. *Let $B \subseteq \Theta$ be a measurable set. Let $\theta' \in \Theta$ be such that $L(\theta') > 0$ and $L(\theta) \geq L(\theta')$ for almost every $\theta \in B$. Moreover, for each i, there is $\delta_i > 0$ such that $L(\theta)g_i^{(0)}(\theta)^{\delta_i} \geq L(\theta')g_i^{(0)}(\theta')^{\delta_i}$ for almost every $\theta \in B$. Then there is $0 \leq M \leq \infty$ such that, as $t \to \infty$,*

$$\frac{\int_{\theta \in B} f_i^{(t)}(\theta)}{f_i^{(t)}(\theta')} \to M.$$

If B has positive measure, then $M > 0$. If $\{\theta \in B : L(\theta) > L(\theta')\}$ has positive measure, then $M = \infty$.

From here it is easy to combine Propositions 5.5 and 5.7 into the Master Theorem. Recall that the various technical hypotheses of the theorem should be intuitively read as "θ' dominates A, and B dominates θ' robustly in the sense of the weighted likelihood function," as explained earlier in this section.

THEOREM 5.8 (MASTER THEOREM). *Let $A, B \subseteq \Theta$ be measurable sets, where B has positive measure. Let $\theta' \in \Theta$ be such that $L(\theta') > 0$, $L(\theta) \leq L(\theta')$ for almost every $\theta \in A$, and $L(\theta) \geq L(\theta')$ for almost every $\theta \in B$. Moreover, for each i, there are $\delta_i, \delta_i' > 0$ such that $L(\theta)g_i^{(0)}(\theta)^{\delta_i} \leq L(\theta')g_i^{(0)}(\theta')^{\delta_i}$ for almost every $\theta \in A$ and $L(\theta)g_i^{(0)}(\theta)^{\delta_i'} \geq L(\theta')g_i^{(0)}(\theta')^{\delta_i'}$ for almost every $\theta \in B$. Suppose that either $\{\theta \in A : L(\theta) = L(\theta')\}$ has measure zero or $\{\theta \in B : L(\theta) > L(\theta')\}$ has positive measure. Then*

$$\int_{\theta \in A} f_i^{(t)}(\theta) \to 0.$$

PROOF. By Propositions 5.5 and 5.7,

$$\frac{\int_{\theta \in A} f_i^{(t)}(\theta)}{\int_{\theta \in B} f_i^{(t)}(\theta)} \to \frac{M_1}{M_2},$$

where M_1 and M_2 are limits in the two respective propositions. Because B has positive measure, $M_2 > 0$. Our final hypothesis implies that either $M_1 = 0$ or $M_2 = \infty$, so $M_1/M_2 = 0$. Then our result follows from $\int_{\theta \in B} f_i^{(t)}(\theta) \leq 1$. □

6 CONVERGENCE OF BELIEFS

We apply the Master Theorem 5.8 to show convergence of beliefs of agents toward a point distribution at the maximizer or near-maximizer of L. We establish separate results for the finite case (Sec. 6.2), the infinite discrete case (Sec. 6.3) and the $\mathbb{R}^k$ case (Sec. 6.4). Our results in the $\mathbb{R}^k$ case are the basis of Theorem 1.1. Results in this section will be used in applications in Section 7.

We first explain the general idea of this section. Recall that to use the Master Theorem, for a given set A, we need to find a point θ' that dominates A and a set B that dominates θ'. What this section aims to do is the following. First, for a given set A, the choice of θ' and B will be automated and abstracted away from the user. Second, even the choice of A will be abstracted away, leaving only the conclusion that "beliefs converge to a point distribution," at least in the case of a unique maximizer. The general idea is not hard, but the details can be quite formidable.

6.1 Preliminaries

First we need some preliminaries. This subsection presents the technical details in carrying out the "automatically choosing θ' and B" part. Let $L_{\text{ess sup}} = \text{ess sup}_{\theta \in \Theta} L(\theta) \leq \infty$, that is, $L_{\text{ess sup}}$ is the smallest M for which $\{\theta \in \Theta : L(\theta) > M\}$ has measure zero. Note that we can replace "ess sup" by "sup" if Θ is infinite discrete or "max" if Θ is finite. The following proposition gives sufficient conditions for beliefs to concentrate in places where L is near $L_{\text{ess sup}}$.

PROPOSITION 6.1. *Let $0 \leq M < L_{\text{ess sup}}$ and let $S = \{\theta \in \Theta : L(\theta) > M\}$. Assume that for every i, $g_i^{(0)}$ is bounded a.e. on $\Theta \setminus S$. Suppose that one of the following holds:*

(1) L is not constant a.e. on S and for every i, $g_i^{(0)}$ is bounded below by a positive number a.e. on S;

(2) there is a subset $T \subseteq S$ of positive measure such that, for every i, $g_i^{(0)}$ is constant on T.

Then, as $t \to \infty$,

$$\int_{\theta \in S} f_i^{(t)}(\theta) \to 1.$$

To better understand the above proposition, it will be useful to understand what roles the various hypotheses play. The "bounded" and "bounded below by a positive number" hypotheses ensure that the gaps between A and θ' and θ' and B are robust as needed in the Master Theorem 5.8. Now we note that S dominates $\Theta \setminus S$; however, we still need to pick θ' and B out of S. There are two ways to do this depending on the situation. First, think of $\Theta = \mathbb{R}$. In this case, if L is constant on S, it will be hard to pick B that dominates θ' robustly; but we can do this otherwise. This is condition (1) in the proposition. Second, think of $\Theta = \mathbb{N}$. In this case, even if S is a single point (and so L is constant on it), it will be easy to pick θ' and B; but this case does not satisfy (1). This is where condition (2) comes in. Putting all these details together gives us the proposition.

6.2 Finite Case

We now apply Proposition 6.1 to show convergence in the case where Θ is finite. The theorem in this case is quite easy and we can appeal to either case (2) of Proposition 6.1 or just Proposition 5.3.

THEOREM 6.2. *Assume that Θ is finite and the initial condition is feasible. Let $\theta_{\max}$ be the set of points that maximize L. Then as $t \to \infty$,*

$$\sum_{\theta \in \theta_{\max}} f_i^{(t)}(\theta) \to 1.$$

In particular, if $\theta_{\max}$ consists of a single point, then $f_i^{(t)}$ converges to the point distribution at that point.

One may wonder what more can be said when $\theta_{\max}$ consists of several points. The example given in the full version shows that the behavior can greatly vary.

6.3 Infinite Discrete Case

We now move on to the theorem for the infinite discrete case, which follows from the second case of Proposition 6.1. This is the first nontrivial case and contains hypotheses that the the initial conditions have to verify in order for beliefs to converge.

THEOREM 6.3. *Assume that Θ is infinite discrete and the initial condition is feasible.*

(1) Let L attain the maximum $L_{\max}$ at the set of points $\theta_{\max} \neq \varnothing$. Suppose that $\sup_{\theta \notin \theta_{\max}} L(\theta) < L_{\max}$ and $g_i^{(0)}$ is bounded on $\Theta \setminus \theta_{\max}$. Then as $t \to \infty$,

$$\sum_{\theta \in \theta_{\max}} f_i^{(t)}(\theta) \to 1.$$

In particular, if $\theta_{\max}$ consists of a single point, then $f_i^{(t)}$ converges to the point distribution at that point.

(2) If L does not attain a maximum, let $L_{\sup} = \sup_{\theta \in \Theta} L(\theta)$. For any $M < L_{\sup}$, if $g_i^{(0)}$ is bounded on $\{\theta \in \Theta : L(\theta) \leq M\}$, then as $t \to \infty$,

$$\sum_{L(\theta) > M} f_i^{(t)}(\theta) \to 1.$$

The theorem is essentially saying the following. Note first that there are two separate cases. If maximizers of L exist, then, assuming a "positive gap" condition and a "boundedness" condition, beliefs converge to those maximizers. But if maximizers of L do not exist (for example, if $L(\theta)$ increases to but is never equal to a positive number), then all that we can say is that the beliefs concentrate at points where L are near the supremum, with a boundedness condition similar to the first case. We give examples in the full version to show that both "positive gap" and "boundedness" conditions are necessary and not just the artifact of the proof.

6.4 $\mathbb{R}^k$ Case

Finally, we deal with convergence in the most difficult case, $\Theta = \mathbb{R}^k$. Convergence here depends on the topology of $\mathbb{R}^k$, so we need to work with functions that are sufficiently continuous. Concepts from real analysis will be freely used. Let $\overline{A}$ denote the closure of A and $B_r(a)$ the open ball of radius r around a.

First we introduce the notion of *piecewise continuity*. There is no universally agreed-upon definition of piecewise continuity, so we define our own in Definition 6.4. This definition is intended to rule out pathological behaviors, and should cover most densities in practice.

Definition 6.4. Let $\Theta \subseteq \mathbb{R}^k$. A function $f : \Theta \to \mathbb{R}_{\geq 0}$ is *piecewise continuous* if there is a countable collection of disjoint open sets $U_i \subseteq \Theta$ such that $\Theta \setminus \cup_i U_i$ has measure zero; only finitely many U_i's intersect any given bounded set $V \subseteq \mathbb{R}^k$; and for each i, the restriction $f : U_i \to \mathbb{R}_{\geq 0}$ can be extended to a continuous function $\widetilde{f_i} : \overline{U_i} \to [0, \infty]$.

We now show convergence of beliefs when $\Theta \subseteq \mathbb{R}^k$ and $g_i^{(0)}$ is piecewise continuous. The following theorem can be regarded as the most significant result in this paper.

THEOREM 6.5. *Assume that $\Theta \subseteq \mathbb{R}^k$ and the initial condition is feasible. Let $g_i^{(0)}$ be piecewise continuous. Let $L_{\mathrm{ess\,sup}} = \mathrm{ess\,sup}_{\theta \in \Theta} L(\theta) \leq \infty$, and for any M, let $S_M = \{\theta \in \Theta : L(\theta) > M\}$. Let $\theta_{\mathrm{ess\,sup}}$ be the set of all $\theta \in \overline{\Theta}$ such that for every $M < L_{\mathrm{ess\,sup}}$ and $\delta > 0$, $B_\delta(\theta) \cap S_M$ has positive measure. Suppose that there is $r > 0$ such that $\mathrm{ess\,sup}_{|\theta| > r} L(\theta) < L_{\mathrm{ess\,sup}}$. Moreover, there is $M' < L_{\mathrm{ess\,sup}}$ such that for every $M \in (M', L_{\mathrm{ess\,sup}})$, L is not constant a.e. on S_M; $g_i^{(0)}$ is bounded a.e. on $\Theta \setminus S_M$; and $g_i^{(0)}$ is bounded below by a positive number a.e. on S_M. Then $\theta_{\mathrm{ess\,sup}}$ is a nonempty compact set, and for any open set U containing $\theta_{\mathrm{ess\,sup}}$, as $t \to \infty$,*

$$\int_{\theta \in U \cap \Theta} f_i^{(t)}(\theta) \to 1.$$

In particular, if $\theta_{\mathrm{ess\,sup}}$ consists of a single point, then $f_i^{(t)}$, viewed as a density on $\mathbb{R}^k$ that vanishes outside of Θ, converges to the point distribution at that point.

The various hypotheses of the theorem fit together in the following way. First, $\theta_{\mathrm{ess\,sup}}$ is supposed to be the "maximizers of L," except that there are two complications. One, we must not be distracted by sets of measure zero. Two, these maximizers may lie outside of Θ; think $1/x^2$ with $\Theta = (0, \infty)$. The definition in the theorem that captures both of these aspects is the following: the "maximizers" are θ such that in small balls around θ, the set where L is near its essential supremum has positive measure. Then to show that beliefs converge to a point distribution at $\theta_{\mathrm{ess\,sup}}$ (if it consists of a single point), we require L to decay away at infinity, in the sense that the set where L is near its essential supremum is bounded; this allows us to use compactness arguments. Then several hypotheses on nonconstancy and boundedness from above and below are required in order to use case (1) of Proposition 6.1. The conclusion is that beliefs concentrate in any open set containing $\theta_{\mathrm{ess\,sup}}$, and this shows convergence in the case where $\theta_{\mathrm{ess\,sup}}$ consists of a single point.

PROOF SKETCH. The proof is unfortunately rather technical. We will only give sketches of key steps here and refer the reader to the full version for details. The first step is to show that $\theta_{\mathrm{ess\,sup}}$ is compact by showing that it is bounded (using the fact that L decays away at infinity) and closed (by combining the U_i's from the piecewise continuity definition judiciously). In particular, it will be important to characterize $\theta_{\mathrm{ess\,sup}}$ as the set on which the continuous extensions of L (in the sense of Definition 6.4) achieve $L_{\mathrm{ess\,sup}}$. Next, we show that whenever U is an open set containing $\theta_{\mathrm{ess\,sup}}$, beliefs on $U \cap \Theta$ tend to 1. Our strategy is the following. Proposition 6.1 implies that for any

$M' < M < L_{\text{ess sup}}$, as $t \to \infty$, $\int_{\theta \in S_M} f_i^{(t)}(\theta) \to 1$. It therefore suffices to show that there is such an M such that $S_M \setminus U$ has measure zero. Suppose the contrary. By restricting attention to a bounded set and a single U_i and letting $M \to L_{\text{ess sup}}$, it will be possible to construct a sequence $\theta_j \in S_{M_j} \setminus U$ that tends to some $\theta \in \theta_{\text{ess sup}}$. This is a contradiction because U is an open set containing θ.   □

Lastly, we present a simplified version of Theorem 6.5 when $\Theta = \mathbb{R}^k$ and $g_i^{(0)}$ is bounded and continuous. This version can deal with densities that are continuous on the whole of $\mathbb{R}^k$ with some mild decay conditions, and it is applicable to e.g. Gaussians and many similar densities. To prove this corollary, we directly apply Theorem 6.5, where Corollary 4.7 is used to prove that the initial condition is feasible.

COROLLARY 6.6. *Assume that* $\Theta = \mathbb{R}^k$. *Let* $g_i^{(0)}$ *be bounded continuous functions. Suppose that there is a bounded set S such that* $\sup_{\theta \notin S} L(\theta) < \sup_{\theta \in \mathbb{R}^k} L(\theta)$. *Then the initial condition is feasible, L attains a maximum at the set of points $\theta_{\max} \neq \varnothing$, and for any open set U containing $\theta_{\max}$, as $t \to \infty$,*

$$\int_{\theta \in U} f_i^{(t)}(\theta) \to 1.$$

In particular, if $\theta_{\max}$ consists of a single point, then $f_i^{(t)}$ converges to the point distribution at that point.

Another (further simplified) version of Corollary 6.6 that should be useful in practice is Theorem 1.1 in the introduction.

7 APPLICATIONS

We now seek to apply our theoretical results to model real-world situations. One example is given for each of the cases where Θ is finite, infinite discrete, and continuous to demonstrate the use of our theorems.

7.1 Binary Beliefs

Suppose that the underlying state of the world is binary, $\Theta = \{0, 1\}$. For example, agents are trying to determine the truth value of a statement. At first, agent i believes that the statement holds with probability x_i. By Theorem 6.2, we can conclude the following.

PROPOSITION 7.1. *Let the initial belief of agent i be Bernoulli distributed as* $\text{Bern}(x_i)$, $0 < x_i < 1$, *and the common prior be* $\text{Bern}(1/2)$. *Then the initial condition is feasible. If* $\prod_{i=1}^n x_i^{v_i} > \prod_{i=1}^n (1 - x_i)^{v_i}$, *then $f_i^{(t)}$ converges to the point distribution at 1; if* $\prod_{i=1}^n x_i^{v_i} < \prod_{i=1}^n (1 - x_i)^{v_i}$, *then $f_i^{(t)}$ converges to the point distribution at 0.*

Note that, except in the borderline case $\prod_{i=1}^n x_i^{v_i} = \prod_{i=1}^n (1 - x_i)^{v_i}$, agents will reach a consensus on whether the statement is true or false. This process can be viewed as if agents cast votes on what everyone should believe: everyone will unanimously agree that the statement is true if

$$\sum_{i=1}^n v_i \log \left(\frac{x_i}{1 - x_i} \right) > 0,$$

and that the statement is false if this expression is negative. Thus it is as if agent i casts votes of $\log(x_i/(1 - x_i))$ in support of the statement and gets v_i votes. The expression $\log(x_i/(1 - x_i))$, called the *log-odds*, transforms probability in $(0, 1)$ to the real line. Moreover, we can see that the strength of an agent's vote indeed comes from her confidence and her centrality.

A similar analysis holds when agents are trying to decide between a finite number of alternatives. In this case, if agent i believes in alternative j with probability $x_{i,j}$, then it is as if she casts votes of $\log x_{i,j}$ in support of alternative j. The alternative with the most votes becomes the consensus.

7.2 Poisson Beliefs

Now let us consider the case where there are an infinite number of alternatives, e.g. when the underlying state of the world is the number of events in a time period. The initial beliefs can be modeled by Poisson distributions.

PROPOSITION 7.2. *Assume that the initial belief of agent i is* $\mathrm{Pois}(\lambda_i)$ *and the common prior is a flat prior, then the initial condition is feasible. Let* $\lambda^* = \prod_{i=1}^{N} \lambda_i^{v_i}$. *If λ^* is not an integer, then $f_i^{(t)}$ converges to the point distribution at* $\lfloor \lambda^* \rfloor$. *If λ^* is an integer, then $f_i^{(t)}$ lies in $\{\lambda^* - 1, \lambda^*\}$ with probability going to 1 as $t \to \infty$.*

PROOF SKETCH. Compute $L(\theta) \propto (\lambda^*)^\theta / \theta!$ and apply Theorem 6.3. □

The Poisson parameter λ is the mean of $\mathrm{Pois}(\lambda)$ and is related to the average event rate. Our result shows that agents will reach a consensus at λ^*, the *weighted geometric mean* of the rates corresponding to agents' beliefs, where again the weight of each agent is her centrality. Note that our model indicates that Poisson rates should be combined multiplicatively rather than additively. This accords with intuition: if two connected people believe that there will be 2 and 1000 floods on average next year, respectively, then Proposition 7.2 says that they will come to believe that there will be $\lfloor \sqrt{2000} \rfloor = 44$ floods. In contrast, combining rates additively as in DeGroot leads to the counterintuitive answer of 501 floods. Thus, even though each agent carries essentially one piece of information: the event rate, by modeling the beliefs according to the actual underlying parameter space, our model leads us to a natural consensus. A similar situation is the Gaussian belief case in the next section where beliefs converge to the weighted arithmetic mean and agent i also has weight v_i.

7.3 Gaussian Beliefs

Finally, we consider the interesting case of each agent carrying *two* pieces of information, her belief and her confidence. Suppose that the underlying state of the world is $\Theta = \mathbb{R}$ with the flat prior $f_* \equiv 1$. Each agent observes the true state of the world θ^* with Gaussian noise: agent i's signal μ_i is drawn from the normal distribution $\mathcal{N}(\theta^*, 1/\tau_i)$, where τ_i is the *precision* known to agent i. Note that this is an example of an underlying scenario explained in Section 2.3, except that the prior in this case is improper. Then agent i's initial belief for θ^* (conditional on the signal μ_i) can be computed to be $\mathcal{N}(\mu_i, 1/\tau_i)$.

PROPOSITION 7.3. *Assume that the initial belief of agent i is normally distributed as $\mathcal{N}(\mu_i, 1/\tau_i)$ and the common prior is a flat prior, then this initial condition is feasible and $f_i^{(t)}$ converges to the point distribution at*

$$\theta_{\max} = \sum_{i=1}^{N} c_i \mu_i, \quad c_i = \frac{v_i \tau_i}{\sum_{j=1}^{N} v_j \tau_j}.$$

PROOF SKETCH. The weighted likelihood function is proportional to a Gaussian with the maximum at $\theta_{\max}$. Apply Corollary 6.6. □

If we interpret the initial belief $\mathcal{N}(\mu_i, 1/\tau_i)$ as a scalar belief μ_i with confidence τ_i, then the consensus belief $\theta_{\max}$ is the weighted average of agents' signals μ_i with weights c_i proportional to her centrality v_i and her precision τ_i. Therefore, an agent who is more centrally located and has a

more informative signal has more influence over the consensus belief. This phenomenon cannot be captured by the standard DeGroot model which has no notion of quality of signals.

We can evaluate quality of learning by the variance of the consensus. This works as follows. The consensus $\theta_{\max}$ is a function of the signals μ_i, and the signals are random, so the consensus can be viewed as a random variable. Specifically, it is an unbiased estimator of the true state θ^*, and the lower its variance, the higher the quality of learning. Since $\mu_i \sim \mathcal{N}(\theta^*, 1/\tau_i)$ are independent, $\theta_{\max} = \sum_i c_i \mu_i \sim \mathcal{N}(\theta^*, S)$ remains a Gaussian, with variance

$$S = \frac{\sum_i v_i^2 \tau_i}{(\sum_i v_i \tau_i)^2}.$$

We now investigate the effect of increasing the precision τ_k on the variance S. We find that, all other variables being fixed, the result depends on the threshold $V_k = 2 \sum_{i \neq k} v_i^2 \tau_i / \sum_{i \neq k} v_i \tau_i$ of the centrality v_k. By direct computation, if $v_k \leq V_k$, then increasing τ_k always decreases S, while if $v_k > V_k$, then increasing τ_k decreases S only if $\tau_k \geq (v_k - V_k) \sum_{i \neq k} v_i \tau_i / v_k^2$ and *increases* S otherwise. Thus, increasing the precision τ_k usually decreases the variance and improves the quality of learning, *except* when the centrality v_k is high and the precision τ_k is low. This result has an implication for social planners who want to encourage good learning. The social planner cannot control the network structure v_k but she might be able to control the signal precision τ_k. If an agent is central but relatively uninformed, increasing the precision of that agent can reduce learning quality. An intuitive explanation might be that the agent now has some confidence to use her centrality to impose her less informed opinion on the consensus. Only when that central agent is sufficiently informed will additional precision improve learning quality. Therefore, if a social planner wants to seed centrally located agents (opinion leaders) with new technologies, she must invest enough resources to make those agents well informed; otherwise, the effort can backfire.

In the full paper, we discuss the case of Gaussian beliefs with correlated initial signals. As expected, higher correlations in initial signals reduce learning quality.

Another point of interest is that our consensus does not depend on the geometry of the network beyond the eigenvector centralities; this is true for all of our analysis from Section 5 onwards and not only for this particular Gaussian case. Let us contrast this with the work of Banerjee et al. [5], which also incorporates quality of signals into DeGroot learning. In their model, the quality of signals is binary (informed/uninformed), and each agent takes the average of signals of only her informed neighbors. They identify the *clustered seeding* problem: if the initially informed agents are closely connected, some agents will "block" other agents, reducing influence of the blocked agents.

To illustrate this point, consider a social network in the shape of an undirected cycle of length N and assume that only three adjacent agents $1, 2, 3$ are initially informed with the same precision. In our model, the three agents' signals have equal weights: $c_1 = c_2 = c_3 = 1/3$. But in their model, agent 2 is blocked by agents 1 and 3: $c_1 = c_3 \to 1/2$ and $c_2 \to 0$ as the number of agents $N \to \infty$. Note that all three agents do have the same centralities and the same precisions, yet the geometry of the network still allows them to block each other.

Why this occurs may be explained as follows: as the "informed segment" of the network expands, an agent who has just become newly informed will get the opinion of agent 1 or 3 first. Surely she will not be very confident because she has just heard of one opinion, but in the model of Banerjee et al. [5], her opinion counts as much as everyone else, and so the opinions of agents 1 and 3 get injected back into the informed segment and reinforce themselves to an inappropriate degree. In contrast, our model circumvents this problem by allowing agents to express their confidence. Thus, a newly informed agent is not very confident and her opinion counts less; and it turns out that this alone is enough to offset the blocking disadvantage and allows the opinion of agent 2 to come through in the consensus. We conclude that the clustered seeding problem occurs when agents can

communicate whether they are informed but cannot communicate precisions of their signals. When agents can communicate precisions of signals, even if beliefs are naively updated, the problem disappears. Therefore, it is of vital importance that agents be able to communicate opinions freely and expressively, and not just state their beliefs, in order for blocking to not occur.

ACKNOWLEDGEMENTS

REFERENCES

[1] Daron Acemoglu, Munther A. Dahleh, Ilan Lobel, and Asuman Ozdaglar. 2011. Bayesian Learning in Social Networks. *Review of Economic Studies* (2011).

[2] Jerry Anunrojwong and Nat Sothanaphan. 2018. Naive Bayesian Learning in Social Networks. (2018). https://arxiv.org/abs/1805.05878

[3] Venkatesh Bala and Sanjeev Goyal. 2000. A Noncooperative Model of Network Formation. *Econometrica* (2000).

[4] Abhijit V. Banerjee. 1992. A Simple Model of Herd Behavior. *The Quarterly Journal of Economics* (1992).

[5] Abhijit V. Banerjee, Emily Breza, Arun G. Chandrasekhar, and Markus Mobius. 2018. Naive Learning with Uninformed Agents. (2018).

[6] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. 1992. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy* (1992).

[7] Bogachan Celen and Shachar Kariv. 2004. Observational learning under imperfect information. *Games and Economic Behavior* (2004).

[8] Peter M. DeMarzo, Dimitri Vayanos, and Jeffrey Zwiebel. 2003. Persuasion Bias, Social Influence, and Unidimensional Opinions. *The Quarterly Journal of Economics* (2003).

[9] Erik Eyster and Matthew Rabin. 2010. Naive Herding in Rich-Information Settings. *American Economic Journal: Microeconomics* (2010).

[10] Erik Eyster and Matthew Rabin. 2014. Extensive Imitation is Irrational and Harmful. *The Quarterly Journal of Economics* (2014).

[11] Douglas Gale and Shachar Kariv. 2003. Bayesian Learning in Social Networks. *Games and Economic Behavior* (2003).

[12] Benjamin Golub and Matthew O. Jackson. 2010. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics* (2010).

[13] Ali Jadbabaie, Pooya Molavi, Alvaro Sandroni, and Alireza Tahbaz-Salehi. 2012. Non-Bayesian social learning. *Games and Economic Behavior* (2012).

[14] Wei Li and Xu Tan. 2018. Learning in Local Networks. (2018).

[15] Ilan Lobel and Evan Sadler. 2015. Information diffusion in networks through social learning. *Theoretical Economics* (2015).

[16] Carl D. Meyer. 2001. *Matrix Analysis and Applied Linear Algebra.* Society for Industrial and Applied Mathematics, Philadelphia, PA.

[17] Pooya Molavi, Alireza Tahbaz-Salehi, and Ali Jadbabaie. 2018. A Theory of Non-Bayesian Social Learning. *Econometrica* (2018).

[18] Manuel Mueller-Frank. 2013. A general framework for rational learning in social networks. *Theoretical Economics* (2013).

[19] Mohammad Amin Rahimian, Pooya Molavi, and Ali Jadbabaie. 2014. (Non-)Bayesian learning without recall. *IEEE Annual Conference on Decision and Control* (2014).

[20] Mohammad Amin Rahimian, Pooya Molavi, and Ali Jadbabaie. 2015. Learning without Recall: A Case for Log-Linear Learning. *IFAC Workshop on Distributed Estimation and Control in Networked Systems* (2015).

[21] Dinah Rosenberg, Eilon Solan, and Nicolas Vieille. 2009. Informational externalities and emergence of consensus. *Games and Economic Behavior* (2009).

[22] Lones Smith and Peter Sorensen. 2000. Pathological Outcomes of Observational Learning. *Econometrica* (2000).