Learning Who to Trust:
A Joint Inference Theory of Persistent Disagreement

Working Draft

Jacob VanNattan
Department of Economics, Washington University in St. Louis
j.d.vannattan@wustl.edu

August 2026

Abstract

We study belief formation on networks when agents simultaneously learn the state and the precision of available information, endogenizing the weights in the canonical DeGroot framework. Beliefs generically converge into socially supported limiting classes, and these dynamics allow us to classify all finite directed networks. Separate closed listening groups generically force disagreement; networks with at most one vertex-disjoint directed cycle generically force consensus; and the remaining networks form an indeterminate class where both outcomes occur on open sets of initial beliefs. Because the classification cannot select among these outcomes, we develop a model of belief-based network formation whose generated networks fall in the indeterminate class with probability approaching one. In large societies the formation environment then selects the outcome. Belief-independent formation yields consensus at the truth, while any strict preference for similar-belief sources produces persistent disagreement with a strictly positive asymptotic separation. Sorting sets the magnitude of that disagreement; connectivity determines how tightly it is locked in.

1 Introduction

Formal work on opinion formation in networks has largely assumed that agents know the precision of their sources. In the canonical framework of DeGroot (1974), agents average their neighbors' opinions with fixed weights. DeMarzo et al. (2003) microfound these weights as Bayesian updating when sources provide independent, unbiased signals with known precision. Once source quality is unknown, however, the weights themselves become objects of inference. DeGroot himself flagged exactly this point: "individual i might wish to change the weights that he assigns to the other individuals after he has learned their initial opinions" (DeGroot, 1974). We take up that possibility and provide a microfoundation for changing weights as inference about source precision.

The underlying information problem is broader than literal uncertainty about statistical precision. In practice, people question whether a source is informed, noisy, biased, manipulated, or simply unreliable; in political discourse, contradictory information is now routinely dismissed as "fake news." We do not separately model each of these channels. Precision is a one-dimensional measure of informativeness that makes the inference problem tractable: a high-precision report should move beliefs substantially, while a low-precision report should not. Persistent directional bias would require a richer signal model, but the economic object at issue is the same one agents confront here—how much evidentiary weight should a report receive when its quality is not known?

An adjacent literature generates belief-dependent influence by imposing it directly. Bounded-confidence models (Deffuant et al., 2000; Hegselmann and Krause, 2002) give agents an exogenous tolerance for disagreement; Dandekar et al. (2013) introduce a biased-assimilation parameter that makes agents overweight confirming opinions; and Polanski and Vega-Redondo (2023) endogenize an influence network through a homophily condition tying link strength to similarity in beliefs. In each case the relation between agreement and influence is a behavioral or equilibrium primitive. Here it is a posterior object. Agents begin with the same prior over source reliability, and the dependence of trust on beliefs is generated by Bayesian inference from the reports they observe.

This distinction matters for the role of homophily. Dandekar et al. (2013) show that homophily alone cannot generate polarization under repeated averaging; their polarizing result requires sufficiently strong biased assimilation in addition to a homophilous network. Our model has no separate biased-assimilation parameter. Once reliability is uncertain, belief-based homophily by itself changes the evidence agents receive, and joint inference converts that difference in evidence into endogenous differences in trust. Our large-network result makes this contrast sharp as we find that homophily need not be paired with an exogenous bias in information processing to sustain separated long-run beliefs in our framework.

Other related frameworks also infer latent features of other agents, but the object being learned and the source of disagreement differ. In Sethi and Yildiz (2016), agents face a sequence of states, learn the unobservable *perspectives*—subjective priors—of potential information sources, and choose whom to observe; long-run communication reflects a tradeoff between well-informed and well-understood targets. Gentzkow et al. (2025) instead study sources whose accuracies are ex ante uncertain across a sequence of states. Agents learn those accuracies by comparing source reports to their own reasoning, and small ideological biases in that reasoning can generate large long-run differences in trust. We hold both channels fixed. Agents reason symmetrically, share the same prior over source reliability, and learn about a single persistent state on a fixed observation network. Heterogeneous trust arises because agents with different evidence jointly infer the state and the precision of the reports they currently observe. This makes the topology of the observation network itself central to long run beliefs.

Our benchmark forms the exact within-period Bayesian posterior when agents treat current reports as conditionally independent and then applies the bounded-memory variational projection defined below between periods. This is the same independence stance that underlies the fixed-weight benchmark and that DeMarzo et al. (2003) call persuasion bias, so the comparison hinges on whether reliability is known rather than how dependence is treated. Section 6 later relaxes this restriction showing that our results extend across the class of updating rules generated by joint reliability inference, including the pairwise rule that is maximally robust to unobserved dependence across information sources.

Our first result disciplines limiting beliefs in this system. Individual beliefs generically converge; no agent can asymptotically dismiss all social information; and every limiting belief is shared by at least one source. This allows us to classify every finite directed network according to the potential for *consensus*: whether all agents reach the same limiting belief. Two or more minimal closed listening groups generically prevent consensus. If the graph contains at most one vertex-disjoint directed cycle, consensus is generic. Between these extremes is the *indeterminate class*: networks with a single minimal closed listening group but at least two vertex-disjoint cycles. Fixed-weight models still reach consensus throughout this class; our framework admits both consensus and persistent disagreement on open sets of initial beliefs.

The finite classification tells us what a network can sustain but not which outcome should be expected or how much disagreement should survive. To answer those questions we introduce a simple model of network generation. After initial beliefs are realized, agents are more likely to connect to agents with nearby beliefs than to agents farther away. In a dense network, each rank then has a deterministic source distribution after one round of learning. Our large-network theorem shows that the first round makes beliefs converge to a deterministic rank profile while agents' confidence in their first round positions grows proportionally with degree. Belief-independent formation collapses that profile of beliefs to the truth, while any fixed belief-based sorting creates positive-mass cohorts whose limiting beliefs remain strictly separated.

This result gives a sharp role to both sorting and connectivity. When link formation is independent of beliefs, growing degree exposes every agent to an increasingly representative sample of initial information and beliefs converge to the truth. With belief-based sorting, dense first-round communication leaves systematic differences across ranks while making agents increasingly confident in those beliefs. The separated cohorts then become locked in to their views and dismiss outside information. Any strict preference for similar-belief sources therefore produces macroscopic persistent disagreement in sufficiently large dense societies.

We close with a related observation about guaranteed consensus. In this information environment, no protocol is simultaneously evidence-responsive, guaranteed to reach consensus throughout the indeterminate class, and immune to a single non-compliant agent choosing the entire society's limiting belief. The capacity to reject sufficiently implausible information is therefore not only a source of disagreement; some failure of guaranteed consensus is also necessary for protection against unilateral influence.

Section 2 formalizes the model. Section 3 establishes convergence and source anchoring. Section 4 classifies finite networks. Section 5 studies large networks under a model of belief-based network formation. Section 6 shows robustness of the results to alternative updating rules and modeling choices. Section 7 concludes. While proof sketches are included in the main body, technical lemmas are collected in the appendices.

2 Model

Network, state, and beliefs

A finite set of agents $V = \{1, \ldots, n\}$ is arranged on a directed graph $G = (V, E)$, taken weakly connected throughout. This is without loss, disconnected components exchange no information and can be treated separately. Agent i's out-neighborhood $\mathcal{N}_i \subseteq V$ is her set of information *sources*. Links need not be reciprocal. If $\mathcal{N}_i \neq \emptyset$, then $i \notin \mathcal{N}_i$: agent i's own belief enters her update through her prior. If i has no external sources, set $\mathcal{N}_i = \{i\}$ by convention. The corresponding one-source recursion leaves her mean unchanged while her precision can continue to grow; she counts as a one-cycle for classification purposes, and we call her *self-sourced*.

There is an unknown state $\theta \in \mathbb{R}$. Agent i's belief at time t is normal,

$$\theta \sim \mathcal{N}\big(\mu_i^t, (\tau_i^t)^{-1}\big),$$

and, following DeMarzo et al. (2003), the pair (μ_i^t, τ_i^t) is the primitive object on which the dynamics operate. Beliefs remain normal by construction. Each period's update closes by replacing the posterior with its best normal summary so (μ_i^t, τ_i^t) is the complete carried state. This compression is discussed further in Assumption 2 below.

Initial beliefs are drawn independently across agents,

$$\tau_i^0 \sim F = \mathrm{Gamma}(a_0, b_0), \qquad \mu_i^0 \mid \tau_i^0 \sim \mathcal{N}\big(\theta, (\tau_i^0)^{-1}\big),$$

where F is the conjugate family for the precision of a normal signal, normalized so $\mathbb{E}_F[\tau] = a_0/b_0 = 1$.[1] Equivalently, each agent begins with an improper diffuse prior over θ and privately observes one signal of known precision; her initial belief is the resulting posterior. The network and beliefs are the entire set of initial conditions.

Each period, agent i observes her sources' current means, $\{\mu_j^t : j \in \mathcal{N}_i\}$, and regards each as a noisy signal of the state,

$$\mu_j^t \mid \theta, \tau_j^t \sim \mathcal{N}\big(\theta, (\tau_j^t)^{-1}\big),$$

with the precision τ_j^t unobserved.

The Normal–Gamma specification is chosen for mathematical convenience, not because either distribution is economically essential. Normal unbiased signals let a single precision parameter summarize how informative a report is, and the Gamma prior makes posterior source assessment available in closed form. Section 6 shows that this convenience is not essential to our results.

[1] The normalization is without loss: scaling every precision, F included, by a common factor λ while shrinking every belief difference by $\sqrt{\lambda}$ leaves every resulting weight W_{ij}^t unchanged.

Key Assumptions

The assumptions separate three objects: what the agent observes, what she carries across periods, and how she treats dependence among her sources' reports.

Assumption 1 (Information)**.** (i) Agent i knows her own initial precision τ_i^0, and thereafter tracks τ_i^t as defined by her own update in Equation (2) below.

(ii) No agent observes another agent's precision. Each agent's period-0 prior over every source's precision is the population distribution F, independently across sources.[2]

(iii) Agent i knows her own neighborhood $\mathcal{N}_i$ but not the structure of G beyond it, and G is not an object of inference. The agent's stance toward the dependence her sources' reports may inherit from it is fixed by Assumption 3 below.[3]

Assumption 2 (Compression)**.** (i) Each agent's carried state is the pair (μ_i^t, τ_i^t): at the end of each period she replaces her posterior over the state with its best normal summary and retains nothing else.

(ii) Agent i keeps no per-source records. Let $\hat{c}_i^t = \tau_i^t/\tau_i^0$ denote her realized growth factor; her period-t prior over each source's precision is the pushforward of F under multiplication by $\hat{c}_i^t$, independently across $j \in \mathcal{N}_i$: under the Gamma form, $\text{Gamma}(a_0, b_i^t)$ with $b_i^t = b_0\tau_i^0/\tau_i^t$.

Assumption 2 is a bounded-memory restriction. Under the maintained within-period model, the exact Bayesian posterior is well defined, but it does not close in the finite-dimensional family the agent carries forward. Integrating out unknown precisions produces mixture posteriors after one period, and the sufficient statistics thereafter grow with the history. The agent therefore forms the exact posterior each period and then compresses it to a normal summary of the state which captures both her point estimate and her confidence in this estimate rather than carrying forward the exact posterior and a source-by-source ledger. Scaling the prior on source precision by one's own learning is analogous to the equal learning assumption of DeMarzo et al. (2003).

Given the two-dimensional carried state (μ_i^t, τ_i^t), the compression step chooses the global mean-field variational projection of that exact posterior onto the normal-times-independent-Gammas family, minimizing $\text{KL}(q\|p)$ subject to the memory constraint. Relative to DeMarzo et al. (2003), the agent carries one additional statistic (τ_i) and that statistic supplies the yardstick needed to infer source quality. A one-number agent can average but cannot measure surprise; a two-number agent can. A statistic capturing the agent's confidence is thus required for DeGroot's fixed weights to become an object of inference. Because no source-specific reputation is retained, the compression is conservative for our disagreement results. Every source is reassessed from the same per-period prior rather than benefiting from a history of favorable assessments.

[2]No result uses the correctness of F: the theorems require only that agents share the prior F, not that the realized draws come from it, so the analysis is robust to a misspecified prior over others' precisions.

[3]Nor could her data identify G. Her observations depend only on what lies inside the closure of her sources, the agents reachable from $\mathcal{N}_i$, so infinitely many networks generate identical data. Furthermore, identifying G would not be enough; properly accounting for the correlation of reports requires the weight on each link in each period, not just the presence of the links alone.

Assumption 3 (Independence of sources)**.** Each period, agent i evaluates her sources' current reports as conditionally independent signals of the state given their precisions. Her joint posterior over the state and the precisions is computed under that stance, and whatever dependence the unobserved network induces among the reports is not part of her inference.

Assumption 3 carries the fixed-weight literature's treatment of repeated social information into the unknown-reliability setting. DeMarzo et al. (2003) call the same independence stance persuasion bias; related work studies it as naive learning or correlation neglect (Golub and Jackson, 2010; Enke and Zimmermann, 2019; Chandrasekhar et al., 2020; Murayama et al., 2026). Holding this stance fixed isolates this paper's departure from the existing literature: reliability is inferred from the available information rather than known ex ante. Section 6 relaxes Assumption 3.

Dynamics

Each period, agent i updates in two steps.

Step 1: Assessment and Pooling. The agent selects the global Kullback–Leibler-optimal projection of her exact within-period posterior onto the normal-times-independent-Gammas family. The parameters of that projection satisfy the joint first-order conditions

$$\hat{\tau}_{ij}^t \;=\; \frac{a_0 + \frac{1}{2}}{b_i^t + \frac{1}{2}\Big[(\mu_j^t - m_i^t)^2 + 1/s_i^t\Big]}, \qquad m_i^t \;=\; \frac{\tau_i^t \mu_i^t + \sum_j \hat{\tau}_{ij}^t \mu_j^t}{s_i^t}, \qquad s_i^t \;=\; \tau_i^t + \sum_j \hat{\tau}_{ij}^t, \tag{1}$$

with b_i^t the prior scale, $\hat{\tau}_{ij}^t$ the *trust* i places in j, m_i^t the pooled posterior mean, and s_i^t the pooled precision. Each trust is the mean of the corresponding Gamma factor in the variational projection, evaluated at squared distance $(\mu_j^t - m_i^t)^2 + 1/s_i^t$ from the pooled estimate.[4]

Step 2: Belief Update. The agent adopts the point estimates as her next normal belief:

$$\mu_i^{t+1} \;=\; m_i^t, \qquad \tau_i^{t+1} \;=\; s_i^t \;=\; \tau_i^t + \sum_{j \in \mathcal{N}_i} \hat{\tau}_{ij}^t. \tag{2}$$

This is clause (i) of Assumption 2 in operation. The agent resolves the inference to its globally selected projection and carries forward only the mean and precision of the state. For an externally sourced agent, write $W_{ij}^t = \hat{\tau}_{ij}^t/\tau_i^{t+1}$ for $j \in \mathcal{N}_i$ and $W_{ii}^t = \tau_i^t/\tau_i^{t+1} > 0$ for her own-prior weight; these coefficients form a row-stochastic average. For a self-sourced agent set the total self-weight to one. Thus μ_i^{t+1} is always a convex combination of current beliefs, and the dynamics are deterministic given the network and initial beliefs.

[4] Writing p for the exact within-period posterior and q for the normal-times-independent-Gammas approximation, the objective is $\mathrm{KL}(q\|p)$. It is lower semicontinuous and coercive, so a global minimizer exists for every configuration. Equation (1) gives its first-order conditions. These conditions can also possess nonglobal stationary solutions; the model selects the global minimizer. If several global minimizers tie, fix any deterministic tie-breaking rule. Such ties are nongeneric and none of the generic statements below depend on the tie-breaker.

3 Individual Convergence and Source Anchoring

Our primary object of interest is the limiting profile of beliefs. With fixed weights, convergence follows from powers of a fixed stochastic matrix. Here the weights and precisions evolve jointly with beliefs, so convergence is itself a substantive property of the model. The result below establishes that the long run is nevertheless simple: beliefs generically settle, and every limiting position is socially anchored.

Theorem 1 (Convergence and source anchoring)**.** *On any finite network G:*

(i) For every initial condition and every agent i,

$$\tau_i^t \to \infty, \qquad \sum_{t=0}^{\infty} \sum_{j \in \mathcal{N}_i} W_{ij}^t = \infty.$$

Consequently some source $j \in \mathcal{N}_i$ satisfies

$$\sum_{t=0}^{\infty} W_{ij}^t = \infty.$$

(ii) Outside a Lebesgue-null set of initial beliefs, every mean converges,

$$\mu_i^t \to \mu_i^\infty \in [\min_j \mu_j^0, \max_j \mu_j^0],$$

and every agent shares her limiting mean with at least one source:

$$\forall i \quad \exists j \in \mathcal{N}_i \qquad \mu_i^\infty = \mu_j^\infty.$$

Proof sketch

Appendix A develops the proof in three steps.

Step 1: total social influence does not vanish. Lemma 1 gives a uniform positive lower bound on every raw source assessment and a uniform upper bound on each agent's one-period proportional precision growth. The first bound implies

$$\tau_i^t \to \infty.$$

Writing

$$\rho_i^t = \sum_{j \in \mathcal{N}_i} \frac{\hat{\tau}_{ij}^t}{\tau_i^t}, \qquad \frac{\tau_i^T}{\tau_i^0} = \prod_{t<T} (1 + \rho_i^t),$$

the upper bound on ρ_i^t implies that divergence of precision requires

$$\sum_t \rho_i^t = \infty.$$

Since

$$\sum_{j\in\mathcal{N}_i} W_{ij}^t = \frac{\rho_i^t}{1+\rho_i^t},$$

Lemma 3 then gives part (i).

Step 2: persistent movement is generically impossible. Lemma 2 establishes the tail bound

$$\sum_{s=t}^{\infty}(\mu_i^{s+1}-\mu_i^s)^2 \le \frac{C_i}{\tau_i^t}.$$

Thus a late movement of fixed size must take an interval whose length is proportional to the agent's current precision.

Suppose some beliefs nevertheless fail to converge. Lemma 4 first shows that a nonconvergent agent must have a nonconvergent source. Lemma 5 strengthens this point: a fixed-size late movement by an observer forces one of their sources to make a fixed-size movement on the same time block. Finiteness then gives, by Lemma 6, arbitrarily late blocks containing a directed cycle whose members all move by nonvanishing amounts.

The remaining issue is whether such movement can persist indefinitely through the network. Lemma 7 gives the required precision-epoch dichotomy. Fast epochs move the observer only $o(1)$ and concentrate their social influence on near reports, so fixed-size movement must be inherited from a lower-precision source unless the agents are already moving together. Slow epochs put a fixed coefficient on every source in the observer's endpoint belief. Tracing fixed-size movement through the sources either contracts separated moving classes or merges them into one synchronized class.

A synchronized class cannot translate back and forth through internal convex averaging. Its remaining first-order motion is generated by settled outside reports, whose far-tail force has positive derivative. Lemma 8 rules out recurrent crossings outside nongeneric exact-cancellation configurations. Hence every mean converges outside a null set.

Step 3: unsupported limiting beliefs are nongeneric. Once all beliefs converge, Lemma 4 implies that an agent whose limit differs from every source limit must satisfy

$$\sum_{j\in\mathcal{N}_i} \frac{1}{\mu_j^\infty-\mu_i^\infty} = 0.$$

These are exact far-tail balance conditions. Lemma 9 studies all agents satisfying such conditions jointly. Their own precisions grow linearly, and the leading relative-belief dynamics have a repelling component in every non-translation direction. The initial conditions that converge to an unsupported balance therefore lie on positive-codimension sets. Their finite union is Lebesgue-null, which gives the source-anchoring conclusion in part (ii).

This is the most technical result in the current draft. Standard convergence tools, including set-balance arguments, do not apply because endogenous trust need not satisfy their balance conditions. Lemmas 1 through 7 and Lemma 9 are proved in full; Lemma 8 is given in outline.

Discussion of Theorem 1

The central message of Theorem 1 is that dismissal of information requires external support. An agent can sufficiently discount disagreeing information only by becoming confident in some alternative account of the state, and that confidence must itself be built from the reports she observes. No agent can eventually treat all social information as negligible, and a generic limiting belief cannot remain unsupported by every information source.

Theorem 1 also shows that generic disagreement is not cycling or arbitrary idiosyncrasy; agents individually converge to a finite collection of socially supported long-run beliefs. A trusted neighbor or set of neighbors is precisely the source of confidence which allows an agent to ignore contradictory reports in the limit. Additional connections can therefore promote mixing or entrenchment depending on the content of the information they supply.

4 Finite Network Classification

Theorem 1 further provides the foundations of a graph theoretic result. An agent almost surely belongs to a limiting class, and at least one of her sources belongs to the same class and shares the same limiting belief. Every limiting class therefore contains a directed cycle, and a cycle is the minimal network object capable of supporting a distinct limiting belief.

Call a nonempty set S a *closed listening group* if $\mathcal{N}_i \subseteq S$ for every $i \in S$. Outsiders may observe into S without opening it; influence runs from sources to observers. Distinct minimal closed listening groups are disjoint. Let $\nu_c(G)$ denote the maximum number of vertex-disjoint directed cycles, with self-sourced agents counted as one-cycles. Say the society reaches *consensus* if all limiting beliefs coincide and exhibits *persistent disagreement* otherwise.

Theorem 2 (Finite Network Classification)**.** *Let G be weakly connected.*

(i) If G contains two or more minimal closed listening groups, consensus occurs from only a null set of initial beliefs.

(ii) If $\nu_c(G) \leq 1$, consensus occurs from all but a null set of initial beliefs.

(iii) If G has a single minimal closed listening group and $\nu_c(G) \geq 2$, consensus occurs on a nonempty open set of initial beliefs, and for every $\Delta > 0$ there is a nonempty open set of initial beliefs with limiting spread at least Δ.

Proof sketch

For (i), two minimal closed groups evolve independently of each other. Holding fixed all within-group differences and shifting one group's initial means by a common offset shifts its entire trajectory and limit one-for-one, so equality of the two group limits occurs at only one value of a continuously distributed offset.

For (ii), if two limiting beliefs exist, the highest-limit class contains a directed cycle and the lowest-limit class contains another by Theorem 1; the classes are disjoint, so the cycles must be vertex-disjoint yielding a contradiction.

Part (iii) is constructive. Scale-separated initial beliefs can successively align a directed cycle and then the source paths feeding into it, producing an open consensus basin. Running the same construction around two widely separated vertex-disjoint cycles makes cross-cycle influence arbitrarily small and produces an open disagreement basin with any prescribed limiting spread. Appendix B gives the constructions.

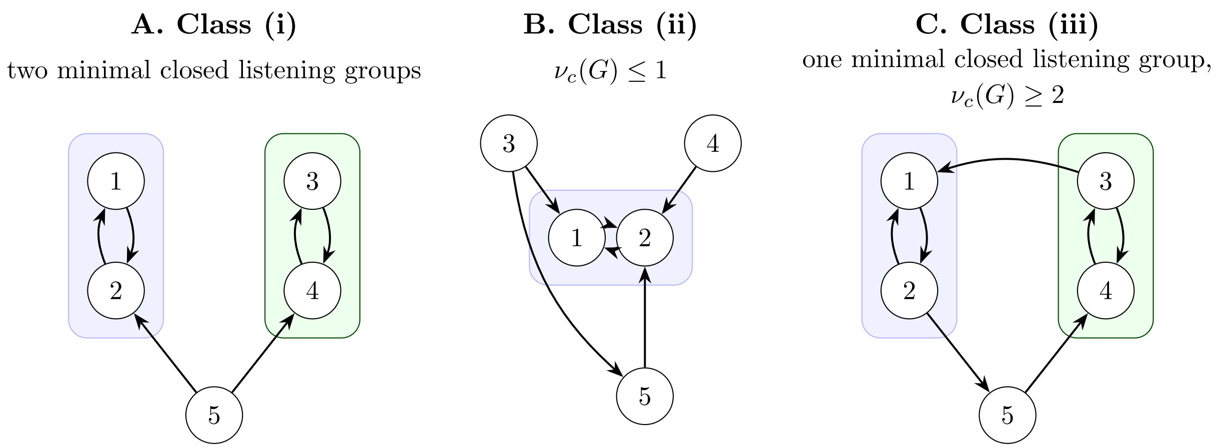

Figure 1: Representative finite networks for the three cases of Theorem 2. An arrow $i \to j$ means that i observes j. In panel A, the shaded two-cycles are the two minimal closed listening groups. Panel B contains only one directed cycle. Panel C contains two vertex-disjoint cycles, but neither shaded cycle is closed; all five nodes form the unique minimal closed listening group.

Discussion of Theorem 2

Topology determines the capacity of a network to sustain distinct limiting beliefs, but it does not necessarily determine the outcome. Under fixed weights with positive self-weight, a single minimal closed listening group is the classical boundary for consensus (Berger, 1981). Separate closed groups therefore prevent consensus under both this model and DeGroot, while a graph with only one possible cycle anchor cannot support two limiting beliefs under either.

The departure comes in (iii) which we call the *indeterminate class*. In such networks, the society has a single closed listening group and is therefore not informationally partitioned, yet it contains enough internal support for several long-run beliefs to coexist in the limit. Fixed weight models such as DeGroot (1974) and DeMarzo et al. (2003) still average these networks to consensus; joint inference lets initial beliefs select whether the same network integrates or separates. This indeterminate class is also the relevant one for modeling richly connected societies. Two closed listening groups require literal informational separation, while $\nu_c(G) \leq 1$ rules out even two disjoint reciprocated pairs. Most dense networks satisfy neither restriction.

Illustrative Simulations

Figures 2–4 illustrate the three cases of Theorem 2 and compare the known-precision benchmark of DeMarzo et al. (2003) with the unknown-precision model of this paper. Each panel contains $N = 100$ agents with $a_0 = b_0 = 2$ and $\theta = 0$. The plots show the first 50 periods; the dashed horizontal line is the truth, and each trajectory is colored by that agent's final simulated belief using the same color scale across panels.

For Classes (i) and (ii), initial beliefs are generated from the model. For Class (iii), the network and initial precision realization are held fixed while two initial mean profiles are used to display both consensus and persistent disagreement. Simulation code is available upon request.

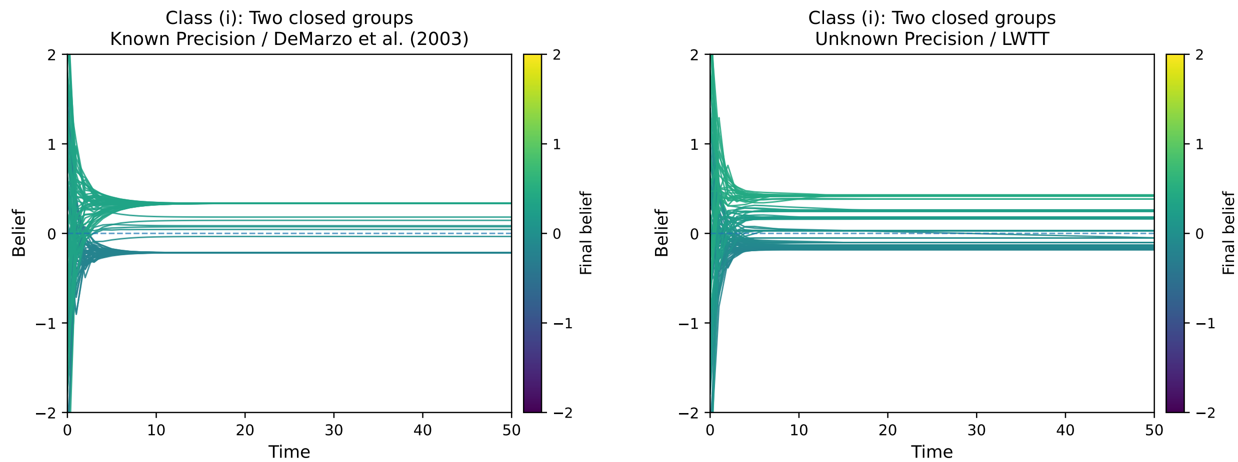

Figure 2: **Class (i): persistent disagreement.** The network has two minimal closed listening groups. The left panel uses known precision as in DeMarzo et al. (2003); the right panel uses our unknown precision model.

Figure 2 illustrates that while persistent disagreement is guaranteed under either model, its qualitative structure is quite different. The network contains two 45-agent closed reciprocal regular groups, four one-sided transient listeners, and six brokers who observe sources in both closed groups. With known precision, each closed group internally reaches consensus. The brokers then converge to convex combinations of the two recurrent-group beliefs, with the particular combination determined by their fixed observation weights. As a result, they can occupy distinct intermediate limiting positions even though no source shares those beliefs.

Under unknown precision, the closed groups themselves can retain internal disagreement because each contains multiple independent cycles capable of supporting distinct limiting classes. Endogenous source assessment also changes the limiting belief possibilities for the brokers who have connections to both closed listening groups. An agent whose belief drifts away from all of her sources eventually assigns vanishing relative weight to those reports, so an unsupported intermediate position cannot generically persist. Instead, the brokers are absorbed into belief classes supported by sources they continue to trust. More generally, Theorem 1 implies that an externally sourced agent generically cannot occupy a singleton limiting belief.

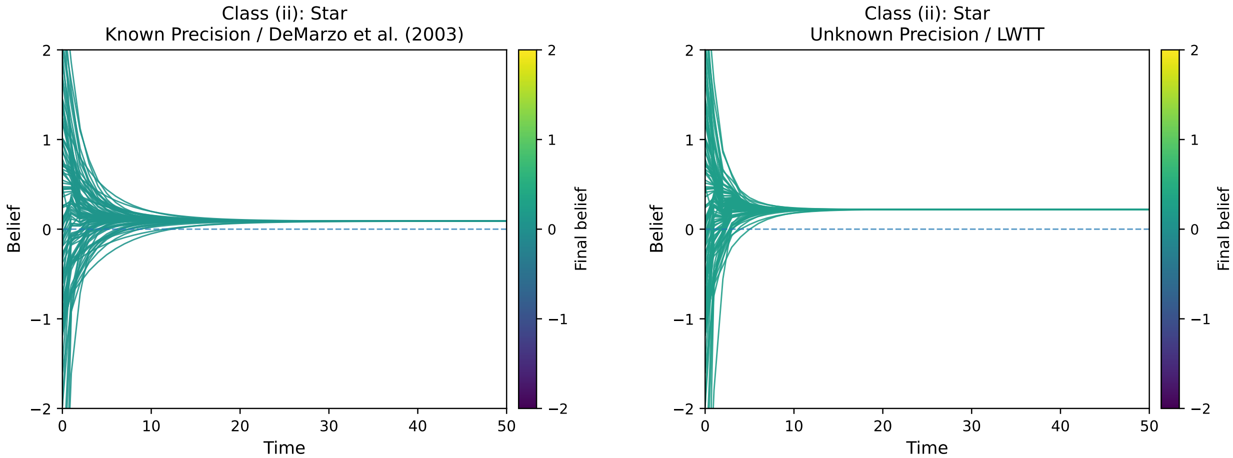

Figure 3: **Class (ii): guaranteed consensus.** The network is a reciprocal star. The left panel uses known precision as in DeMarzo et al. (2003); the right panel uses our unknown precision model.

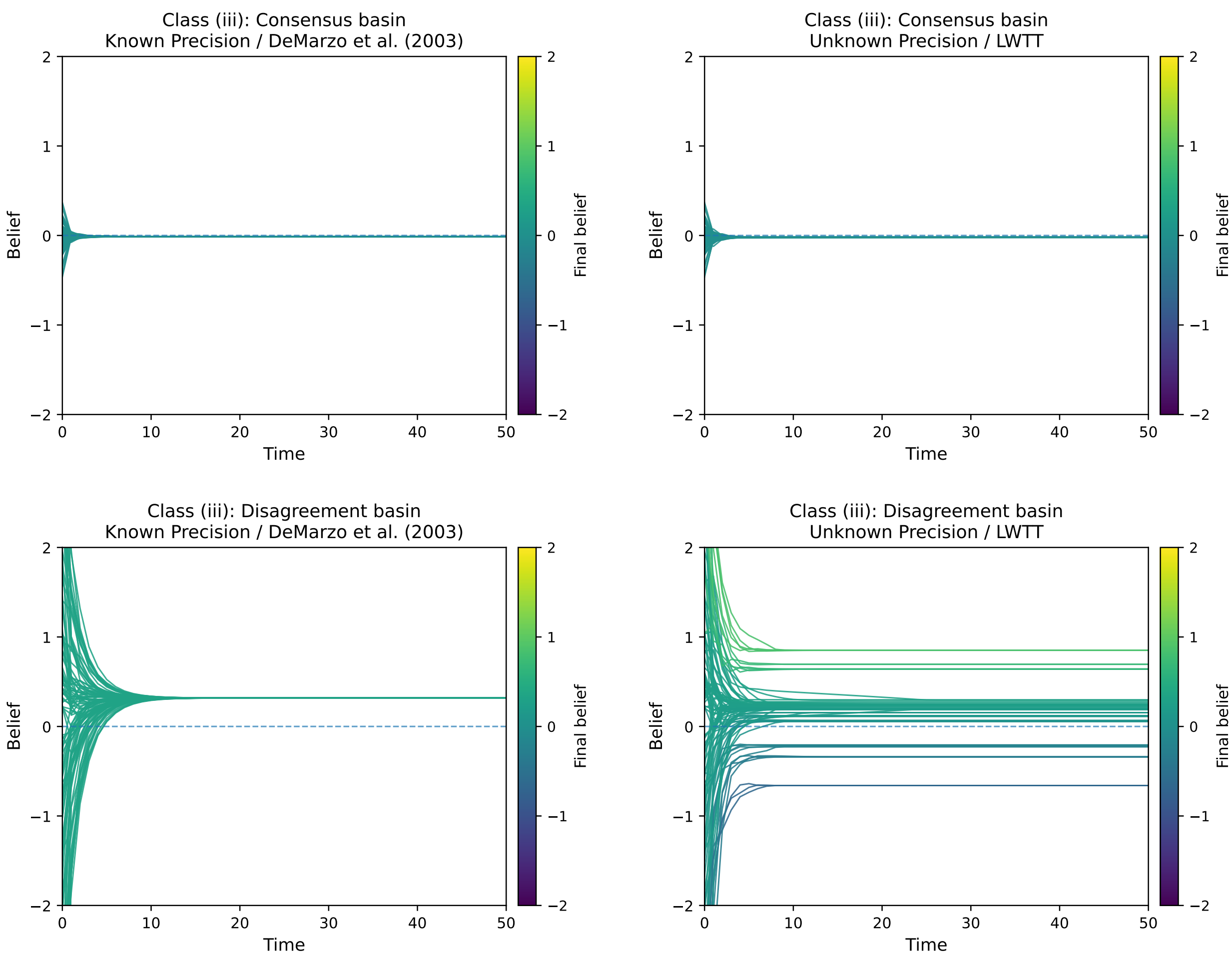

Figure 4: **Class (iii): indeterminacy.** All four panels use the same connected degree-12 reciprocal random-regular network. The initial precision realization is also held fixed. The top row uses an initial mean profile in the consensus basin; the bottom row uses a profile in a disagreement basin.

Figure 3 emphasizes that uncertainty about precision does not guarantee persistent disagreement. The network is a reciprocal star, so every directed cycle contains the hub and therefore $\nu_c(G) = 1$. Theorem 2 then places the network in the generic-consensus class. Under known precision, the usual averaging forces all agents toward a common limit. Under unknown precision, agents continually reassess the reliability of the reports they receive, but the network provides no second independent cycle on which a distinct limiting belief can be anchored. Both frameworks consequently converge to consensus, even though the path of adjustment and the endogenous weights placed on different sources may differ substantially.

Figure 4 illustrates the indeterminacy in part (iii) of Theorem 2. The same connected degree-12 reciprocal network is used in all four panels, and the initial precision realization is held fixed. Under known precision, the fixed averaging matrix reaches consensus from both initial mean profiles. Changing the initial beliefs can alter the common limiting value but not the qualitative outcome. Under unknown precision, the network can support either consensus or fragmentation. When initial beliefs are sufficiently concentrated, agents continue to regard cross-cutting reports as plausible and the society converges toward a common belief. When the same network begins from a more separated configuration, endogenous precision assessment progressively weakens cross-group influence and allows distinct source-supported belief classes to persist.

The network by itself does not produce persistent disagreement, but it does create the structural possibility of more than one long-run belief. A unique minimal closed listening group prevents disagreement from being forced, while multiple vertex-disjoint cycles make it possible for separate belief classes to become self-supporting. Which possibility is selected depends on the initial belief configuration. This is the basin-selection problem left unresolved by the finite-network classification and taken up in the large-network analysis below, where the random formation process determines which basin is selected with high probability as the network grows large.

5 Asymptotics of the Indeterminate Class

The finite classification tells us which outcomes a network can support but not which one is selected. This section introduces a simple process of belief-based network formation and asks what happens in large societies generated according to this model. The generated networks fall in the indeterminate class with probability approaching one. Dense communication then produces a deterministic first-round belief profile while making first-round own precision grow with degree. Our main result shows that this combination selects sharply between the two benchmark formation environments: belief-independent formation leads to consensus at the truth, while any fixed belief-based sorting leaves a strictly positive amount of persistent disagreement. We then state a stronger conjecture that the entire first-round range is asymptotically locked in.

Network formation model

The network formation model has three primitives: a within-window propensity π_{in}, an outside propensity $\pi_{\text{out}} \leq \pi_{\text{in}}$, and a similarity breadth $\gamma \in (0, 1/2]$. Formation occurs once, after initial beliefs are realized and before any belief updating occurs.

Sort the population of N agents by initial mean belief. Agent i's *window* is the $w = \lceil \gamma(N-1) \rceil$ agents nearest her in that ranking, a contiguous rank interval that becomes one-sided near the extremes. She initiates an observation link to each member of her window independently with probability π_{in}, and to each agent outside her window independently with probability π_{out}. Links are directed, although reciprocal links may arise.

When $\pi_{\text{in}} = \pi_{\text{out}}$, links are independent of beliefs. When $\pi_{\text{in}} > \pi_{\text{out}}$, agents are more likely to observe others with similar initial beliefs. The parameter γ determines how broadly "nearby" is defined, while the difference between π_{in} and π_{out} determines the strength of sorting. Thus the model generates assortative structure without imposing groups ex ante, consistent with evidence on belief-based homophily (McPherson et al., 2001; Baldassarri and Gelman, 2008; DellaPosta et al., 2015; Mosleh et al., 2021).

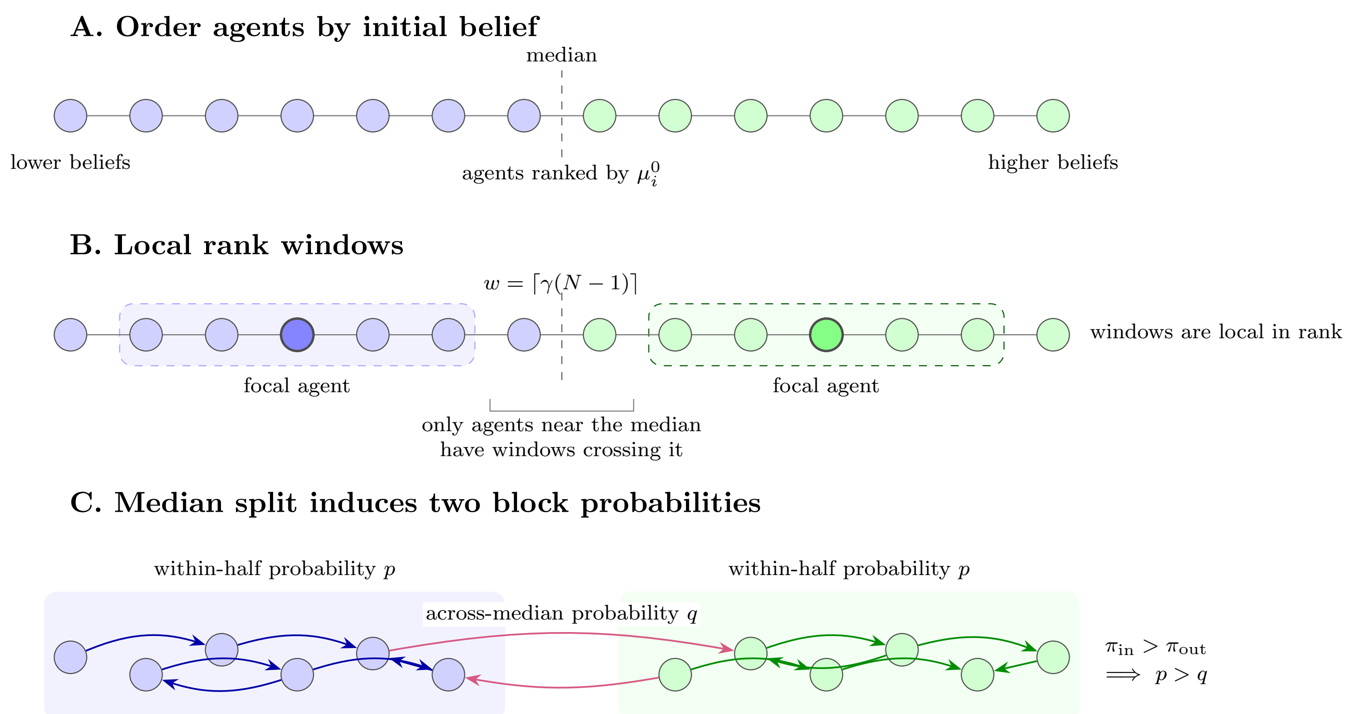

Figure 5: **Rank-window network formation.** Agents are ordered by initial belief and link with higher probability inside a local rank window. Aggregating the resulting network at the median therefore produces more within-half than across-half links when $\pi_{\text{in}} > \pi_{\text{out}}$.

Figure 5 illustrates the mechanism. Windows away from the median lie entirely on one side of it, while only agents sufficiently near the median have windows that cross it. Aggregating the realized network at the median therefore produces a higher expected within-half link probability than across-half link probability when $\pi_{\text{in}} > \pi_{\text{out}}$. The apparent two-block structure is an implication of local rank sorting, not an assumption of the model; the analysis below continues to use the full rank-window source distribution.

Expected out-degree is

$$d = w\pi_{\text{in}} + (N - 1 - w)\pi_{\text{out}}.$$

Degree is then determined by the formation primitives as opposed to set exogenously, and realized out-degrees concentrate around d. Because the formation primitives are fixed,

$$d = (N-1)\big[\gamma\pi_{\text{in}} + (1-\gamma)\pi_{\text{out}}\big] + O(1),$$

so expected degree grows linearly with the population and $d/\log N \to \infty$.

Proposition 1 (The generated networks land in the indeterminate class)**.** *Fix $\pi_{\text{in}} \geq \pi_{\text{out}} > 0$ and $\gamma \leq 1/2$. With probability tending to one, the realized network is strongly connected and contains two vertex-disjoint directed cycles, and is therefore in the indeterminate class of Theorem 2.*

Proof. See Appendix C.

The networks generated by this model typically have enough connectivity to admit consensus but enough independent cycles to support persistent disagreement as well, so the finite classification alone cannot select the outcome. Asymptotic outcomes are instead determined by what belief-based sorting does to the information agents pool.

Dense first-round learning and persistent disagreement

Normalize $\theta = 0$. Alongside expected degree d, define the finite-population *sorting share.*

$$\eta_N := \frac{w(\pi_{\text{in}} - \pi_{\text{out}})}{d}, \qquad \eta := \frac{\gamma(\pi_{\text{in}} - \pi_{\text{out}})}{\gamma\pi_{\text{in}} + (1-\gamma)\pi_{\text{out}}}. \tag{3}$$

Because the formation primitives are fixed, $\eta_N = \eta + o(1)$; the subscript on η_N reflects only the integer rounding in the window size w. Degree measures how much social information an agent receives, while η measures how selectively that information is drawn from nearby initial beliefs. Under strict sorting $\pi_{\text{in}} > \pi_{\text{out}}$, we have $\eta > 0$; when formation is independent of beliefs, $\eta = 0$.

To make this distinction precise, index agents by initial-belief rank $x \in [0,1]$. Let $\mathcal{W}_\gamma(x)$ denote the length-γ rank window centered at x and shifted inward near the endpoints, let H be the population distribution of initial beliefs, and let $H_{\mathcal{W}_\gamma(x)}$ be its restriction to that rank window. The limiting source distribution for rank x is

$$H_{\eta,x} = (1-\eta)H + \eta H_{\mathcal{W}_\gamma(x)}. \tag{4}$$

Thus each agent's source pool is a mixture of representative population information and information drawn locally in the belief ranking.

Let $M_\eta(x)$ denote the population first-round belief produced by applying the model's pooling rule to $H_{\eta,x}$. This is the belief to which an agent of rank x converges in the first round when the number of sources becomes arbitrarily large; its formal population objective is given in Appendix C. Define the spread of this profile by

$$R(\eta) := \max_{x\in[0,1]} M_\eta(x) - \min_{x\in[0,1]} M_\eta(x). \tag{5}$$

The quantity $R(\eta)$ is therefore the systematic disagreement left by sorting after an infinitely dense first round of communication.

Two parts of this profile will be useful below. For every rank $x \in [0, \gamma/2]$, the shifted endpoint window is exactly $[0, \gamma]$, so $M_\eta(x)$ is constant on that interval; write the common value as $m_-(\eta)$. Likewise, for $x \in [1 - \gamma/2, 1]$, write the common upper-endpoint value as $m_+(\eta)$.

The main large-network theorem combines the first-round mean-field approximation with the subsequent lock-in created by high precision. As agents accumulate more sources they become less willing to consider alternative reports after even one period of learning.

Theorem 3 (Dense communication and persistent disagreement)**.** *Fix $\gamma \in (0, 1/2]$ and fixed propensities $\pi_{\text{in}} \geq \pi_{\text{out}} > 0$.*

(i) Deterministic first round. *If x_i is agent i's empirical initial-belief rank, then*

$$\max_i \left|\mu_i^1 - M_{\eta_N}(x_i)\right| = O_p\left(\sqrt{\frac{\log N}{d}}\right). \tag{6}$$

First-round precision grows proportionally with d uniformly across agents, and consequently

$$\max_i \mu_i^1 - \min_i \mu_i^1 = R(\eta_N) + O_p\left(\sqrt{\frac{\log N}{d}}\right). \tag{7}$$

(ii) Belief-independent formation. *If $\pi_{\text{in}} = \pi_{\text{out}}$, then $\eta_N = 0$, $M_0(x) = 0$ for every rank, and*

$$\max_i |\mu_i^\infty| = O_p\left(\sqrt{\frac{\log N}{d}}\right) \longrightarrow 0.$$

(iii) Strict sorting. *If $\pi_{\text{in}} > \pi_{\text{out}}$, then $m_-(\eta) < 0 < m_+(\eta)$. Moreover, there exist limiting beliefs ℓ_N^- and ℓ_N^+ satisfying*

$$\ell_N^- = m_-(\eta_N) + o_p(1), \qquad \ell_N^+ = m_+(\eta_N) + o_p(1).$$

Hence, with

$$\Delta(\eta) := m_+(\eta) - m_-(\eta) > 0,$$

$$\max_i \mu_i^\infty - \min_i \mu_i^\infty \geq \Delta(\eta) + o_p(1).$$

Theorem 3 both presents the large-network result and isolates the mechanism behind it. Dense communication averages away finite-sample noise in the first round while simultaneously making agents highly confident in the resulting deterministic profile. Without sorting, that profile is constant at the truth. Under strict sorting, the lower and upper endpoint windows each contain a positive mass of ranks with distinct population estimates; the confidence created by first-round corroboration locks those separated cohorts into their first-round beliefs.

The endpoint cohorts need not span the full range $R(\eta)$. The nonlinear trust response can generate more extreme interior values of M_η, including isolated extrema. The following stronger statement is therefore left as a conjecture.

Conjecture 1 (Full one-round lock-in)**.** *Under fixed strict sorting,*

$$\frac{\max_i \mu_i^\infty - \min_i \mu_i^\infty}{R(\eta_N)} \xrightarrow{p} 1. \tag{8}$$

Conjecture 1 says that the same confidence mechanism established for the positive-mass endpoint cohorts also preserves any more extreme first-round positions. Figure 6 below is consistent with this stronger conclusion.

Formation independent of beliefs

The belief-independent conclusion in Theorem 3 is not special to our network generation process. The same logic applies whenever network formation is independent of initial beliefs.

Proposition 2 (Formation independent of beliefs)**.** *Let the network be drawn by any process independent of initial beliefs, with minimum out-degree d satisfying $d/\log N \to \infty$. Then the spread of limiting beliefs converges to zero in probability, and every limiting belief converges to θ.*

Proof. See Appendix C.

Connectivity by itself does not inherently support consensus nor persistent disagreement. When source selection is independent of beliefs, growing degree gives every agent an increasingly representative sample of initial information and the first pooled estimates collapse to the truth. The persistent disagreement in Theorem 3 comes from the interaction of dense communication with belief-based sorting.

Illustrative Simulations

Figure 6 illustrates the large-network comparative statics in finite networks. We set $a_0 = b_0 = 2$, $\theta = 0$, and $\gamma = 0.25$, and vary population size and the network formation parameters. Within a population size, the initial state and underlying link-formation draws are held fixed across formation environments. The simulations are used to illustrate the dynamics of the system and are not relied upon for proof evidence. Simulation code is available upon request.

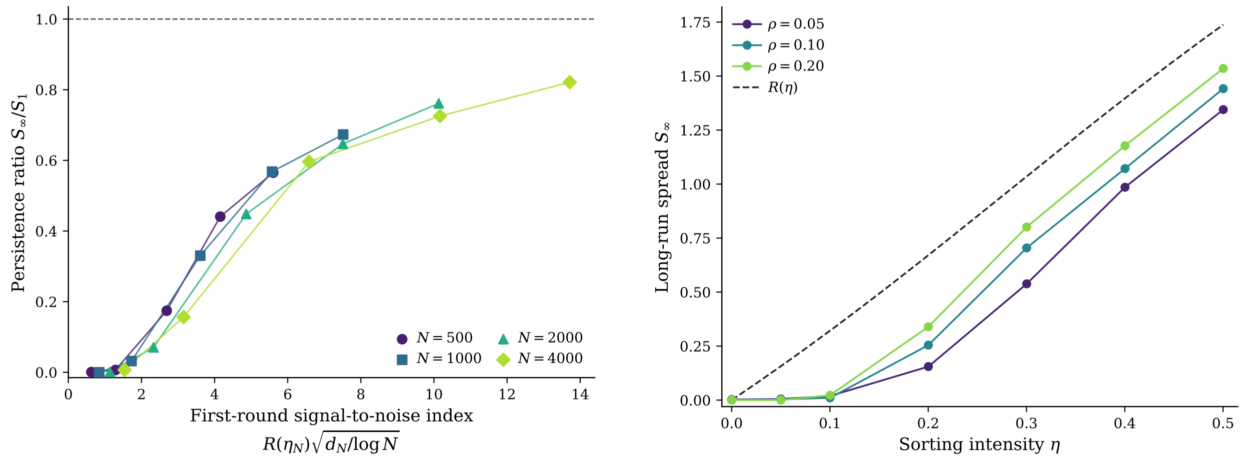

Figure 6: **Finite-network illustration of the full lock-in conjecture.** The left panel plots the persistence ratio S_∞/S_1 against $R(\eta_N)\sqrt{d_N/\log N}$. The right panel plots long-run spread against sorting intensity at several network densities, together with $R(\eta)$.

The left panel shows that first-round disagreement becomes increasingly persistent as the sorting-induced spread grows relative to finite-sample noise. The right panel separates the two underlying forces. Stronger sorting creates a larger first-round spread, while greater connectivity makes the long-run spread track that first-round prediction more closely. These patterns illustrate the current results and are consistent with Conjecture 1.

6 Robustness and Guaranteed Consensus

The preceding results use the pooled joint-inference benchmark obtained under the assumption that neighbors' reports are conditionally independent. We first ask whether these results depend on that particular assumption or instead reflect the broader logic of learning unknown source reliability. We find that our results hold for a larger class of joint-inference rules, including the pairwise rule that is maximally robust to dependence across reports.

We then ask whether a different updating protocol could eliminate persistent disagreement altogether. A final proposition shows that guaranteed consensus comes with a robustness cost. No protocol can simultaneously remain responsive to evidence, guarantee consensus throughout the indeterminate class, and rule out the possibility that a single non-compliant agent determines the common limiting belief of the entire society. Thus some capacity for consensus to fail is necessary to preserve both evidence-responsiveness and protection against dictatorship.

The joint inference class

The main model assesses source reliability relative to the pooled estimate $(m_i^t, 1/s_i^t)$ which is derived under the assumption that neighbors' reports are conditionally independent. More generally, a joint inference rule can assess each report relative to a *benchmark pair* (β_i^t, v_i^t), consisting of a benchmark belief and an associated variance. The other natural extreme is the *pairwise rule*, which scores each report against the observer's own current belief, $(\mu_i^t, 1/\tau_i^t)$. For $\lambda \in [0,1]$, these two rules are connected by

$$\beta_i^t = (1-\lambda)\mu_i^t + \lambda m_i^t, \qquad v_i^t = (1-\lambda)/\tau_i^t + \lambda/s_i^t. \tag{9}$$

Interior values of λ index how much corroboration across reports enters the assessment of source reliability. A solution exists by Brouwer's theorem on the compact set defined by the bounds in Section 3.

Definition 1 (The joint inference class)**.** An updating rule is a *joint inference rule* if, for every observer i, source j, and period t,

$$\hat{\tau}_{ij}^t = \tau_i^t \, \phi_{\xi_i^t}\left(\tau_i^t (\delta_{ij}^t)^2\right), \qquad \delta_{ij}^t := \mu_j^t - \beta_i^t,$$

where β_i^t lies in the convex hull of the observer's current belief and pooled estimate and ξ_i^t collects scale-free nuisance parameters, including the observer's initial type and the normalized benchmark variance $\tau_i^t v_i^t$. On every compact set of nuisance parameters, the family $\{\phi_\xi\}$ is positive, smooth, and strictly decreasing, its value at zero is bounded above and away from zero, and

$$z\phi_\xi(z) \to \kappa(\xi), \qquad 0 < \underline{\kappa} \le \kappa(\xi) \le \bar{\kappa} < \infty.$$

Importantly, disagreement is always measured relative to the observer's own uncertainty, not an exogenous metric.

Every rule in (9) belongs to this class, since

$$\tau_i^t v_i^t = (1-\lambda) + \lambda \frac{\tau_i^t}{s_i^t} \in [0,1].$$

Call a rule *dependence-invariant* if its source assessments are correct conditional expectations under every joint distribution of reports consistent with the marginal model of Section 2, allowing arbitrary dependence across sources.

Proposition 3 (The anchor rules)**.** *(i) The rule with benchmark pair $(m_i^t, 1/s_i^t)$ is the global mean-field variational projection of the exact Bayesian posterior when sources' reports are conditionally independent given the state.*

(ii) The unique dependence-invariant rule is the pairwise rule, with benchmark pair $(\mu_i^t, 1/\tau_i^t)$: the maximally correlation-robust member of the class.

Proof. See Appendix D.

The preceding definition varies how agents benchmark reports while retaining the same qualitative trust response. We next show that the main results are robust across this class of joint inference updating rules.

Proposition 4 (Robustness)**.** *The convergence and finite-network results extend to updating rules in the joint inference class. For large networks, let M_η^ϕ denote the corresponding population first-round profile. The first-round concentration conclusion of Theorem 3 continues with M_η replaced by M_η^ϕ. If under strict sorting the lower and upper endpoint cohorts have distinct population estimates $m_-^\phi(\eta) < m_+^\phi(\eta)$, then the limiting spread is at least*

$$m_+^\phi(\eta) - m_-^\phi(\eta) + o_p(1).$$

Proof. See Appendix D.

The joint-inference structure is also not specific to the Gamma prior. The Gamma specification gives a convenient closed form, but the next result shows that a broad class of reliability priors generates the same scale-free, declining trust response with inverse-square tails.

Proposition 5 (Representation)**.** *Fix a benchmark pair and let*

$$c_v^t := \tau_i^t v_i^t \in [0,1].$$

Under Assumption 2, suppose the reliability prior F has finite moments of all orders and a density that behaves like x^{a-1} near zero, in the sense of regular variation, for some $a > 0$. Then the posterior mean precision assigned to a report with benchmark gap δ can be written as

$$\hat{\tau}_{ij}^t = \tau_i^t \, \phi_{\tau_i^0, c_v^t}\left(\tau_i^t \delta^2\right),$$

where, for fixed initial type, $\phi_{\tau_i^0, c_v}$ is positive, smooth, and strictly decreasing, is finite at zero uniformly in $c_v \in [0,1]$, and satisfies

$$z\phi_{\tau_i^0, c_v}(z) \to 2a + 1$$

uniformly in $c_v \in [0,1]$ on compact sets of initial types. For the Gamma specification, the kernel is available in closed form and the tail constant is $2a_0 + 1 = \kappa_0$.

Proof. See Appendix D.

The prior over source precision therefore determines the particular trust function and its tail constant, but not the qualitative structure of the framework presented here.

Guaranteed consensus

The robustness results rule out one natural route to guaranteed consensus as changing the benchmark within the joint inference class does not alter the preceding findings. Another alternative is to abandon that class and impose a protocol designed to guarantee consensus. This is the objective of much of the distributed-consensus literature (Olfati-Saber and Murray, 2004), with related work in economics studying nonlinear opinion dynamics and the influence of stubborn agents (Cerreia-Vioglio et al., 2024; Acemoglu et al., 2013).

The next proposition identifies the cost of doing so in this information environment of unknown source precision. On the indeterminate class, no protocol can guarantee consensus while remaining both responsive to evidence and protected against unilateral control. Call an agent *non-compliant* if her reports need not follow the prescribed update rule. She may instead report any history-measurable sequence while all other agents remain compliant.

Proposition 6 (Impossibility)**.** *Fix any protocol in which each agent updates on her sources' reports: any report space, deterministic or randomized. Consider three conditions.*

(i) Evidence-responsiveness*: an agent who observes no one retains her mean belief.*

(ii) Guaranteed consensus*: on every indeterminate-class society of Theorem 2, fully compliant play reaches consensus almost surely.*

(iii) Non-dictatorship*: on no indeterminate-class society can a single non-compliant agent make her chosen position the limiting belief of every compliant agent.*

No protocol satisfies all three.

Proof. See Appendix D.

Evidence-responsiveness rules out the trivial consensus protocol in which agents converge to a fixed value regardless of the information they receive. Once that possibility is excluded, Proposition 6 shows that guaranteeing consensus throughout the indeterminate class necessarily creates a vulnerability somewhere in that class: on some network, a single non-compliant agent can choose the common limit. The result does not imply that disagreement is desirable. It instead identifies the cost of ruling disagreement out altogether. Some capacity for consensus to fail is necessary if the protocol is also to remain responsive to evidence and protected against unilateral control.

7 Conclusion

This framework starts from an information environment that is common in the real world; people rarely know exactly how precise their sources are. A report may come from someone poorly informed, from a noisy measurement, from a source thought to have an agenda, or from information whose provenance is simply uncertain. Public disagreement therefore often concerns not only the state of the world but whether the evidence offered by the other side is credible at all. The language ranges from "unreliable" and "biased" to "fake news." Our Normal specification compresses those concerns into uncertainty over precision. This is a tractable representation of source informativeness, not the substantive claim of the model.

Once source quality is uncertain, learning the state and deciding which information to trust are inseparable. Putting that joint inference inside the canonical DeGroot environment changes what connectivity implies. Beliefs generically settle into socially supported limiting beliefs, and network structure determines how many such positions the society can sustain. At the structural extremes the model agrees with fixed-weight learning. On the broad indeterminate class, however, the same connected network can reach consensus or persistent disagreement depending on the profile of agents' initial beliefs.

The large-network results identify what selects between those outcomes. With links independent of beliefs, growing degree exposes each agent to a representative sample of initial information and the society converges to the truth. With belief-based sorting, the same increase in degree supplies more corroborating information within each side, raises confidence, and makes cross-cutting reports easier to dismiss. Connectivity therefore magnifies the informational consequences of source selection rather than pushing beliefs toward agreement or disagreement on its own.

Modern communication technologies plausibly move both margins at once. They increase the number of people and information sources an individual can reach, while also loosening geographic, institutional, and linguistic constraints. Existing evidence documents both the growing importance of online matching and assortative tie formation by political similarity (Rosenfeld et al., 2019; Mosleh et al., 2021). We do not treat those facts as tests of the model. Their relevance is that a more connected society need not be a more randomly mixed society. If a larger menu of possible sources is used to select disproportionately like-minded contacts, the model predicts that the informational benefit of greater connectivity can backfire.

The mechanism behind our framework is particularly suited to settings in which opposing groups are not literally isolated. They may read the same news, observe the same public events, and hear the same arguments while disagreeing over which sources deserve their attention. In such a world, an additional confirming source does more than add another opinion; it makes the receiver more confident in the interpretation she already finds plausible, which in turn makes contradictory reports look less reliable. Echo chambers in this model are not caused by an absence of outside information but by an abundance of confirmatory evidence. This framework shows how greater access to information can deepen rather than resolve disagreement when agents must simultaneously learn both what is true and who to trust.

A Proof of Theorem 1

This appendix collects the dynamical arguments behind Theorem 1.

Theorem 1. *On any finite network G:*

(i) For every initial condition and every agent i,

$$\tau_i^t \to \infty, \qquad \sum_{t=0}^{\infty} \sum_{j\in\mathcal{N}_i} W_{ij}^t = \infty.$$

Consequently some source $j \in \mathcal{N}_i$ satisfies

$$\sum_{t=0}^{\infty} W_{ij}^t = \infty.$$

(ii) Outside a Lebesgue-null set of initial beliefs, every mean converges,

$$\mu_i^t \to \mu_i^\infty \in [\min_j \mu_j^0, \max_j \mu_j^0],$$

and every agent shares her limiting mean with at least one source:

$$\forall i \quad \exists j \in \mathcal{N}_i \qquad \mu_i^\infty = \mu_j^\infty.$$

A.1 Exact kernel and finite movement energy

Set

$$c_i^t := 2b_i^t + \frac{1}{\tau_i^{t+1}} = \frac{2b_0\tau_i^0}{\tau_i^t} + \frac{1}{\tau_i^{t+1}}, \qquad \kappa_0 := 2a_0 + 1.$$

Because $\mu_i^{t+1} = m_i^t$ and $s_i^t = \tau_i^{t+1}$, the trust equation is exactly

$$\hat{\tau}_{ij}^t = \frac{\kappa_0}{c_i^t + (\delta_{ij}^t)^2}. \tag{10}$$

For the remainder of the appendix, let

$$R := \max_k \mu_k^0 - \min_k \mu_k^0, \qquad \delta_{ij}^t := \mu_j^t - \mu_i^{t+1}, \qquad \rho_i^t := \sum_{j\in\mathcal{N}_i} \frac{\hat{\tau}_{ij}^t}{\tau_i^t}.$$

Row stochasticity keeps every mean belief in the initial hull.

Lemma 1 (Basic trust bounds)**.** *For every agent i there are constants*

$$\underline{\tau}_i > 0, \qquad \bar{\rho}_i < \infty$$

such that for every source $j \in \mathcal{N}_i$ and every period t,

$$\hat{\tau}_{ij}^t \geq \underline{\tau}_i, \qquad \rho_i^t \leq \bar{\rho}_i.$$

Proof. Because precision is nondecreasing,

$$c_i^t = \frac{2b_0\tau_i^0}{\tau_i^t} + \frac{1}{\tau_i^{t+1}} \leq 2b_0 + \frac{1}{\tau_i^0}.$$

Also $|\delta_{ij}^t| \leq R$. Equation (10) therefore gives

$$\hat{\tau}_{ij}^t \geq \frac{\kappa_0}{2b_0 + 1/\tau_i^0 + R^2} =: \underline{\tau}_i > 0.$$

For the upper bound, the denominator in (10) is at least

$$\frac{2b_0\tau_i^0}{\tau_i^t}.$$

Hence

$$\frac{\hat{\tau}_{ij}^t}{\tau_i^t} \leq \frac{\kappa_0}{2b_0\tau_i^0},$$

and summing over the finite set $\mathcal{N}_i$ gives

$$\rho_i^t \leq \frac{|\mathcal{N}_i|\kappa_0}{2b_0\tau_i^0} =: \bar{\rho}_i.$$

□

The update also gives

$$\tau_i^t(\mu_i^{t+1} - \mu_i^t) = \sum_{j \in \mathcal{N}_i} \hat{\tau}_{ij}^t \delta_{ij}^t. \tag{11}$$

Since

$$\hat{\tau}_{ij}^t (\delta_{ij}^t)^2 \leq \kappa_0,$$

Cauchy–Schwarz yields a stronger form of the displacement bound.

Lemma 2 (Finite movement energy)**.** *There is $C_i = |\mathcal{N}_i|\kappa_0(1 + \bar{\rho}_i)$ such that*

$$\sum_{s=t}^{\infty} (\mu_i^{s+1} - \mu_i^s)^2 \leq \frac{C_i}{\tau_i^t}.$$

Consequently $\mu_i^{t+1} - \mu_i^t \to 0$. If i moves distance at least d on $[p, q]$, then

$$q - p \geq \frac{d^2\tau_i^p}{C_i}.$$

Proof. Let $\Delta_i^t = \mu_i^{t+1} - \mu_i^t$ and $H_i^t = \tau_i^{t+1} - \tau_i^t = \sum_j \hat{\tau}_{ij}^t$. From (11),

$$(\tau_i^t \Delta_i^t)^2 \le H_i^t \sum_j \hat{\tau}_{ij}^t (\delta_{ij}^t)^2 \le |\mathcal{N}_i| \kappa_0 H_i^t.$$

Hence

$$(\Delta_i^t)^2 \le |\mathcal{N}_i| \kappa_0 \frac{\tau_i^{t+1} - \tau_i^t}{(\tau_i^t)^2}.$$

Lemma 1 gives

$$\frac{\tau_i^{t+1} - \tau_i^t}{(\tau_i^t)^2} \le (1 + \bar{\rho}_i) \left(\frac{1}{\tau_i^t} - \frac{1}{\tau_i^{t+1}} \right).$$

Summing telescopes. The movement-time bound follows from Cauchy–Schwarz. □

A.2 Total source influence and settled sources

Lemma 3 (Total trust diverges)**.** *For every sourced agent,*

$$\tau_i^t \to \infty, \qquad \sum_t \sum_{j \in \mathcal{N}_i} W_{ij}^t = \infty.$$

Hence some source receives infinite cumulative normalized weight.

Proof. Lemma 1 gives

$$\tau_i^{t+1} - \tau_i^t = \sum_j \hat{\tau}_{ij}^t \ge |\mathcal{N}_i| \underline{\tau}_i,$$

so precision diverges. Also

$$\frac{\tau_i^T}{\tau_i^0} = \prod_{t<T} (1 + \rho_i^t) \to \infty.$$

Since $\rho_i^t \le \bar{\rho}_i$, $\sum_t \rho_i^t = \infty$. Finally

$$\sum_j W_{ij}^t = \frac{\rho_i^t}{1 + \rho_i^t} \ge \frac{\rho_i^t}{1 + \bar{\rho}_i},$$

and finiteness of $\mathcal{N}_i$ implies that at least one link receives infinite cumulative weight. □

Lemma 4 (Settled sources imply a settled observer)**.** *If every source of i converges, $\mu_j^t \to v_j$, then μ_i^t converges. If its limit ℓ_i differs from every v_j, then*

$$\sum_{j \in \mathcal{N}_i} \frac{1}{v_j - \ell_i} = 0.$$

Proof. Suppose $\liminf_t \mu_i^t = a < b = \limsup_t \mu_i^t$. Define

$$F_i(x) = \sum_j \frac{1}{v_j - x}.$$

On every interval between source-limit poles,

$$F_i'(x) = \sum_j \frac{1}{(v_j - x)^2} > 0.$$

Choose $x \in (a, b)$ that is neither a pole nor a zero. Lemma 2 makes one-period increments vanish, so there are both upward and downward crossings of x along which $\mu_i^t, \mu_i^{t+1} \to x$. Using the exact force form of (11), upward crossings imply $F_i(x) \geq 0$ and downward crossings imply $F_i(x) \leq 0$, a contradiction. If ℓ_i differs from all source limits, all raw assessments converge to finite positive constants and $\tau_i^t \asymp t$. A nonzero $F_i(\ell_i)$ would then give eventually one-signed increments of order $1/t$, contradicting convergence. □

Let

$$O = \{i : \mu_i^t \text{ does not converge}\}.$$

Lemma 4 implies every member of O has a source in O. Hence, if $O \neq \varnothing$, the induced graph $G[O]$ contains a sink strongly connected component.

For $I = [p, q]$, write

$$\mathrm{osc}_I(\mu_j) = \max_{p \leq t \leq q} \mu_j^t - \min_{p \leq t \leq q} \mu_j^t.$$

We say that j has *fixed-size movement* on a sequence of blocks I_n if

$$\liminf_n \mathrm{osc}_{I_n}(\mu_j) > 0.$$

This terminology will be used throughout the remainder of the proof.

Lemma 5 (Movement inheritance)**.** *If i makes complete crossings of a fixed interval $[a, b]$ on late intervals $I_n = [p_n, q_n]$, $p_n \to \infty$, then at least one source has fixed-size movement on those blocks:*

$$\liminf_n \max_{j \in \mathcal{N}_i} \mathrm{osc}_{I_n}(\mu_j) > 0.$$

Proof. If every source had vanishing movement on I_n, pass to a subsequence on which each converges uniformly on those blocks to a constant v_j. Choose a regular level $x \in (a, b)$ that is neither a pole nor a zero of $\sum_j (v_j - x)^{-1}$. Complete crossings give both upward and downward crossings of x, but vanishing increments and the exact force identity require the limiting force to have both signs, a contradiction. □

Lemma 6 (Common-block movement cycle)**.** *If O is nonempty, there are a directed cycle*

$$i_1 \to i_2 \to \cdots \to i_K \to i_1,$$

constants $\eta_m > 0$, and arbitrarily late blocks $I_n = [T_n, Q_n]$ such that

$$\mathrm{osc}_{I_n}(\mu_{i_m}) \geq \eta_m \qquad (m = 1, \ldots, K).$$

Proof. Choose for each $i \in O$ an interval strictly inside its liminf–limsup range. For every sufficiently late T_n, choose a finite Q_n by which every member of O completes a crossing. Lemma 5 assigns each observer a source with fixed-size movement on the same block; sufficiently late such sources cannot be convergent. Pass to a subsequence on which this finite source map is constant. A finite directed graph with out-degree one contains a cycle. □

If $L_n = Q_n - T_n$, Lemma 2 also gives

$$\tau_{i_m}^{T_n} \le \frac{C_{i_m}}{\eta_m^2} L_n. \tag{12}$$

A.3 Precision-doubling epochs and exclusion of persistent movement

Start a precision epoch for i at p, put $T = \tau_i^p$, and let

$$q = \min\{t > p : \tau_i^t \ge 2T\}.$$

Minimality and Lemma 1 imply

$$T \le \tau_i^t < 2T \quad (p \le t < q), \qquad \tau_i^q \le 2(1+\bar{\rho}_i)T.$$

Lemma 7 (Precision-epoch dichotomy)**.** *There are constants $\alpha_i, \beta_i > 0$ such that every epoch $[p, q]$ satisfies*

$$\prod_{t=p}^{q-1} W_{ii}^t = \frac{\tau_i^p}{\tau_i^q} \ge \alpha_i, \qquad \sum_{t=p}^{q-1} \sum_{j \in \mathcal{N}_i} W_{ij}^t \ge \beta_i. \tag{13}$$

Fix $\lambda > 0$.

(i) If $q - p \ge \lambda T$, then every source $j \in \mathcal{N}_i$ receives endpoint coefficient

$$\Gamma_{ij}(p, q) := \sum_{t=p}^{q-1} \left(\prod_{s=t+1}^{q-1} W_{ii}^s \right) W_{ij}^t \ge \gamma_i(\lambda) > 0. \tag{14}$$

(ii) Along any sequence with $(q-p)/T \to 0$,

$$\operatorname{osc}_{[p,q]}(\mu_i) \to 0,$$

and for every fixed $\varepsilon > 0$ the cumulative normalized weight on ε-far reports tends to zero.

Proof. The self-weight telescopes and the epoch doubles precision, so

$$\prod_{t=p}^{q-1} W_{ii}^t = \frac{\tau_i^p}{\tau_i^q} \ge \frac{1}{2(1+\bar{\rho}_i)} =: \alpha_i.$$

Also

$$\sum_j W_{ij}^t = \frac{\rho_i^t}{1+\rho_i^t} \geq \frac{\rho_i^t}{1+\bar{\rho}_i}, \qquad \sum_{t=p}^{q-1} \log(1+\rho_i^t) = \log \frac{\tau_i^q}{\tau_i^p} \geq \log 2,$$

so $\beta_i := \log 2/(1+\bar{\rho}_i)$ works. On a slow epoch,

$$W_{ij}^t \geq \frac{\underline{\tau}_i}{2(1+\bar{\rho}_i)T},$$

and every surviving self-weight product in (14) is at least α_i. Hence

$$\Gamma_{ij}(p,q) \geq \alpha_i \frac{\lambda \underline{\tau}_i}{2(1+\bar{\rho}_i)} =: \gamma_i(\lambda).$$

On a fast epoch, an ε-far report has $\hat{\tau}_{ij}^t \leq \kappa_0/\varepsilon^2$ and $\tau_i^{t+1} \geq T$, so its total normalized weight is $O((q-p)/T) = o(1)$. Finally Lemma 2 gives

$$\operatorname{osc}_{[p,q]}(\mu_i)^2 \leq (q-p)\frac{C_i}{T} \to 0.$$

$\square$

A slow epoch is a mixing event: unrolling the one-period recursion shows that every source contributes at least $\gamma_i(\lambda)$ to the endpoint belief. On a fast epoch the observer has only $o(1)$ movement, and all nonvanishing social influence comes from increasingly near reports. Suppose a source j has fixed-size movement at least $\eta_j > 0$ on a block $[p,q]$ with $q-p = o(\tau_i^p)$. Lemma 2 then gives

$$\tau_j^p \leq \frac{C_j}{\eta_j^2}(q-p) = o(\tau_i^p). \tag{15}$$

Lemma 8 (No persistent movement)**.** *Conditional on a fixed initial precision profile, the initial means for which the induced graph $G[O]$ contains a sink strongly connected component form a Lebesgue-null set.*

Proof sketch. Suppose S is a sink strongly connected component of $G[O]$. Choose compact crossing intervals inside the liminf–limsup range of each $i \in S$, away from the limits of convergent outside sources, and use Lemma 6 to obtain arbitrarily late common blocks on which the relevant agents have fixed-size movement.

Trace one such fixed-size movement backward through sources with fixed-size movement on the same block. Because S is a sink component of $G[O]$, any nonconvergent source of an agent in S also lies in S. On a fast epoch the observer itself has only $o(1)$ movement, so some near source in S must have fixed-size movement. Whenever that source moves on a block short relative to the observer's precision, (15) gives a strict contemporaneous precision descent. Such descents cannot form a cycle on a finite set: returning to the starting agent after finitely many steps would imply $\tau_i^p \leq o(1)\tau_i^p$. If no precision descent occurs, the relevant near sources have comparable precision and form an $o(1)$-diameter class with fixed-size movement.

A slow epoch either averages within that class or, by (14), assigns a fixed coefficient to a distinct source class. For an extremal class with fixed-size movement, this produces a fixed inward movement unless another source offsets it; in that case continue with the offsetting source. Because S is finite, this sequence of source classes either reaches a convergent outside source, incorporates all nonconvergent sources in S into one class, or revisits a previously visited class. If it revisits a class after passing through r separated classes, one traversal assigns at least a coefficient $\gamma^r > 0$ across their hull and therefore contracts its diameter to at most $(1-\gamma^r)D + o(1)$. Repeated fixed-size movement is therefore impossible unless all relevant nonconvergent agents synchronize.

It remains to rule out persistent movement of a synchronized class C. Internal convex averaging cannot translate its $o(1)$ hull, so first-order movement comes only from convergent outside sources with limits v_j. Along a subsequence on which the normalized internal weights converge, project on a stationary vector of the limiting internal matrix. After normalizing the row step sizes, the common coordinate has leading force

$$F_C(x) = \sum_{i \in C} \omega_{C,i} \sum_{j \in \mathcal{N}_i \setminus C} \frac{1}{v_j - x}, \qquad F'_C(x) > 0 \tag{16}$$

whenever a separated outside source has positive limiting weight. If there is no effective outside source, internal averaging alone cannot produce persistent movement and the class converges. Otherwise the class cannot make alternating crossings of a compact interval away from a zero of F_C.

Finally consider a zero $x^\star$ of F_C. Locally the projected coordinate satisfies

$$x_{t+1} - x_t = a_t\{F_C(x_t) + o(|x_t - x^\star|) + o(1)\}, \qquad a_t > 0.$$

If $\sum_t a_t < \infty$, outside-driven movement is summable and the class converges. If $\sum_t a_t = \infty$, then $F'_C(x^\star) > 0$ makes the zero repelling; convergence to it requires the leading translation component to vanish, a positive-codimension condition. There are only finitely many possible source partitions and graph faces and a countable localization by compact rational boxes, so the exceptional initial means form a null set.

The convergence part of Theorem 1 follows: if $O \neq \varnothing$, the finite induced graph $G[O]$ contains a sink strongly connected component, which Lemma 8 rules out generically.

A.4 Unmatched balances

Suppose all means converge and call i *unmatched* if no source shares its limit. Then

$$\hat{\tau}_{ij}^t \to \frac{\kappa_0}{(\mu_j^\infty - \mu_i^\infty)^2}, \qquad \frac{\tau_i^t}{t} \to a_i > 0,$$

and convergence requires

$$\sum_{j \in \mathcal{N}_i} \frac{1}{\mu_j^\infty - \mu_i^\infty} = 0.$$

If an agent has a same-limit source, by contrast, the corresponding raw assessment diverges and $\tau_i^t/t \to \infty$.

For unmatched agents define

$$p_{ij} = \frac{(\mu_j^\infty - \mu_i^\infty)^{-2}}{\sum_k (\mu_k^\infty - \mu_i^\infty)^{-2}}.$$

The joint relative-belief linearization on the unmatched set U is

$$e_U^{t+1} = e_U^t + \frac{1}{t}(I - P_{UU})e_U^t + o\left(\frac{\|e^t\|}{t}\right),$$

with feedback through matched agents lower order.

Lemma 9 (Unmatched balances are nongeneric)**.** *Conditional on any fixed initial precision profile, initial means whose convergent trajectory has a nonempty unmatched set form a Lebesgue-null set.*

Proof. On an irreducible unmatched component with leakage, P_{UU} is substochastic with spectral radius below one. On a closed irreducible unmatched component it is stochastic; its Perron eigenvalue 1 is simple and represents only common translation. Hence every genuine difference mode of

$$B = I - P_{UU}$$

has strictly positive real part.

On the repelling difference subspace choose $H > 0$ solving

$$B^\top H + HB = I.$$

For

$$e^{t+1} = \left(I + \frac{B}{t} + o(t^{-1})\right)e^t,$$

the quadratic form $e_t^\top He_t$ grows at rate at least a positive constant times itself divided by t whenever the repelling component is nonzero. A trajectory can therefore converge to the balance only from the corresponding center-stable slice, which has positive codimension in relative-belief space. There are finitely many unmatched sets and graph decompositions; localizing limiting profiles by countably many compact rational boxes gives a countable union of null sets. □

Lemmas 3–9 prove Theorem 1.

B Construction Lemmas for Theorem 2

Only the two open-set claims in part (iii) require construction.

Theorem 2. *Let G be weakly connected.*

(i) If G contains two or more minimal closed listening groups, consensus occurs from only a null set of initial beliefs.

(ii) If $\nu_c(G) \leq 1$, consensus occurs from all but a null set of initial beliefs.

(iii) If G has a single minimal closed listening group and $\nu_c(G) \geq 2$, consensus occurs on a nonempty open set of initial beliefs, and for every $\Delta > 0$ there is a nonempty open set of initial beliefs with limiting spread at least Δ.

For agent i, define

$$\bar{c}_i := 2b_0 + \frac{1}{\tau_i^0}.$$

By Lemma 1, $c_i^t \leq \bar{c}_i$ for every t. The contribution of source j to the force in (11) is

$$\hat{\tau}_{ij}^t \delta_{ij}^t = \frac{\kappa_0 \delta_{ij}^t}{c_i^t + (\delta_{ij}^t)^2}.$$

Hence, if a designated source has

$$r_- \leq |\delta_{ij}^t| \leq r_+$$

for fixed $0 < r_- \leq r_+ < \infty$, then its force has magnitude at least

$$\frac{\kappa_0 r_-}{\bar{c}_i + r_+^2}.$$

By contrast, any source at distance at least R contributes at most

$$\frac{\kappa_0}{R}$$

in absolute value. Since every neighborhood is finite, choosing R sufficiently large makes the designated source dominate the combined force of all sources at distance at least R, uniformly while the designated gap remains in $[r_-, r_+]$.

We use this observation recursively below. Because the graph is finite, only finitely many such scale inequalities are required, and they can be imposed simultaneously by choosing successively larger belief separations. All inequalities can be chosen strict and therefore continue to hold on an open neighborhood of the constructed initial profile.

Lemma 10 (Finite absorption cascade)**.** *If G has a unique minimal closed listening group, there is a nonempty open set of initial beliefs from which every agent converges to the same limit.*

Proof. Let S be the unique minimal closed listening group. Under our edge convention, S is the unique sink strongly connected component of G. Choose a directed cycle

$$C = (i_1, \ldots, i_K)$$

inside S.

First construct a common limiting class on C. Place the cycle members in a bounded cluster and designate for each i_m its successor on the cycle as its relevant near source. Place every other source of a cycle member at a much larger belief scale. By the scale-separation bound above, the designated cycle links dominate the force generated by all non-designated sources. The induced averaging around the directed cycle contracts differences within C. As the cycle members become close, their mutual reports enter the near-report region and reinforce the common class, while the sources placed at the remote scale remain in the inverse-square tail. Choosing the remote scale sufficiently large makes their total effect on the location of the resulting cycle class arbitrarily small.

Now enlarge the class one agent at a time. Since S is strongly connected, every member of $S \setminus C$ has a source path into C. Order these agents by their shortest source-path distance to C. When agent i is added, choose a source j one step closer to C as its designated source. The source j has already been assigned to the existing class. Place i at the next construction scale and place any sources not yet incorporated into the class at a still larger scale. Scale separation then makes j dominate the movement of i until i enters the near-report region of the existing class. Once there, the same corroboration mechanism keeps i attached to that class.

After all of S has been incorporated, repeat the construction for $V \setminus S$. Because S is the unique sink component of the condensation graph, every agent outside S has a source path eventually entering S. Order agents by source distance to S and apply the same argument, choosing for each agent a designated source already incorporated into the common class.

There are finitely many agents, so the construction requires only finitely many scale separations. These scales can therefore be chosen recursively so that every designated-source dominance condition holds simultaneously. The resulting profile leads every agent into the same limiting class. Since all dominance conditions are strict, they continue to hold for all initial beliefs in a sufficiently small neighborhood of the constructed profile. Thus the consensus basin contains a nonempty open set. □

Lemma 11 (Two-cycle separation)**.** *If G contains two vertex-disjoint directed cycles, then for every $\Delta > 0$ there is a nonempty open set of initial beliefs whose limiting spread is at least Δ.*

Proof. Let C_L and C_H be two vertex-disjoint directed cycles. Construct a near-report class on each cycle using the same scale-separation argument as above. Center the first construction near L and the second near H, and let

$$D := H - L.$$

The internal construction scales can be fixed independently of D. Place all remaining agents at scales chosen so that their reports do not disturb the designated links used to form either cycle class.

Within each cycle, designated near reports have force bounded away from zero during the finite construction stages. A report from the other cycle, whose distance is of order D, contributes force of order D^{-1}. Since the graph contains only finitely many links and the construction has only finitely many stages, the total cross-cycle displacement during these stages can be made arbitrarily small by increasing D.

Once each cycle has entered its own near-report class, internal corroboration raises precision while reports from the other class remain separated by a distance of order D. Their normalized influence therefore becomes negligible, and the remaining cross-cycle movement can likewise be made arbitrarily small by choosing D sufficiently large.

Fix $\Delta > 0$. Choose $D > \Delta$ large enough that the total displacement of each cycle caused by sources associated with the other is less than

$$\frac{D - \Delta}{4}.$$

The two limiting classes are then separated by more than

$$D - 2\frac{D - \Delta}{4} = \frac{D + \Delta}{2} > \Delta.$$

Thus the limiting spread is at least Δ.

All scale-separation and displacement inequalities are strict. They therefore continue to hold on a sufficiently small neighborhood of the constructed initial profile, giving a nonempty open set of initial beliefs with limiting spread at least Δ. $\square$

C Proofs for the Large-N Results

This appendix contains the proofs for Section 5 encompassing the network-generator bridge, the deterministic first-round profile, persistent disagreement under strict sorting, and the belief-independent benchmark. The stronger full-range statement is Conjecture 1 and is not used in any theorem.

Proof of Proposition 1

Proposition 1. *Fix $\pi_{\text{in}} \geq \pi_{\text{out}} > 0$ and $\gamma \leq 1/2$. With probability tending to one, the realized network is strongly connected and contains two vertex-disjoint directed cycles, and is therefore in the indeterminate class of Theorem 2.*

Proof. Every ordered pair carries a link with probability at least the fixed constant $\pi_{\text{out}} > 0$, independently. For a nonempty proper set S, the probability that no link leaves S is at most

$$(1 - \pi_{\text{out}})^{|S||S^c|} \leq e^{-\pi_{\text{out}}|S||S^c|},$$

and the same bound holds for no link entering S. A union bound over cuts tends to zero, so the graph is strongly connected with probability tending to one. For two disjoint cycles, take the lowest and highest quarters of the belief ranking and partition each into disjoint pairs. Each pair is reciprocally linked with probability at least $\pi_{\text{out}}^2 > 0$, so each quarter contains a directed two-cycle with probability tending to one. $\square$

Dense first-round learning and persistent disagreement

Normalize $\theta = 0$. Under $b_0 = a_0$, the marginal initial-belief distribution is Student-t with $\nu = 2a_0$ degrees of freedom. Write F_ν, f_ν, and Q_ν for its cdf, density, and quantile function, and define

$$\psi(z) := \frac{(2a_0 + 1)z}{2a_0 + z^2}.$$

For rank x, the population first-round estimate $M_\eta(x)$ used in Section 5 is the selected global minimizer of

$$\mathcal{L}_{\eta,x}(m) := \frac{2a_0 + 1}{2} \int \log\bigl(2a_0 + (y - m)^2\bigr)\ dH_{\eta,x}(y),$$

with the same fixed tie-breaking convention as in the finite-agent problem. Generically the minimizer is unique, and its first-order condition is

$$\int \psi(y - M_\eta(x))\, dH_{\eta,x}(y) = 0. \tag{17}$$

Its range $R(\eta)$ is defined in (5).

Proof of Theorem 3

Proof. For part (i), at $t = 0$ we have $b_i^0 = b_0 = a_0$ for every observer. Since first-round source precision is

$$\hat{\tau}_{ij}^0 = \frac{2a_0 + 1}{2a_0 + (\mu_j^0 - \mu_i^1)^2 + 1/\tau_i^1},$$

the source force is bounded uniformly in the report. Conditional on ranks, source links are independent Bernoulli draws and each observer receives $d(1 + o_p(1))$ reports. Dividing the first-round pooling equation by degree therefore gives a bounded M-estimation problem whose population equation is (17). Bernstein bounds, a union over agents, and local identification of the selected population minimizer give (6).

The own-prior score is $\tau_i^0(\mu_i^0 - m)$. Under the Gamma–Normal initialization,

$$\max_i \tau_i^0 = O_p(\log N), \qquad \max_i |\tau_i^0 \mu_i^0| = O_p(\log N),$$

so uniformly for m in any fixed compact set the own-prior contribution is $O_p(\log N/d) = o_p(1)$. Thus idiosyncratic own precision washes out in the mean-field first round.

The same argument applied to the bounded positive trust function gives

$$\frac{\tau_i^1}{d} = T_{\eta_N}(x_i) + o_p(1)$$

uniformly, where T_η is continuous and bounded away from zero and infinity on the compact rank space. Own precision after the first round of updating is therefore proportional to degree uniformly, and (7) follows.

For part (ii), if $\pi_{\text{in}} = \pi_{\text{out}}$, then $H_{\eta,x} = H$ for every rank and symmetry gives $M_0(x) = 0$. Part (i) places the entire period-one belief hull in an $O_p(\sqrt{\log N/d})$ neighborhood of the truth. Every later update is a convex combination of current beliefs, so every subsequent and limiting belief remains in that hull.

For part (iii), suppose $\pi_{\text{in}} > \pi_{\text{out}}$. For $x \in [0, \gamma/2]$ the endpoint window is exactly $[0, \gamma]$, while for $x \in [1-\gamma/2, 1]$ it is $[1-\gamma, 1]$. Hence each endpoint cohort has a common population first-round estimate, $m_-(\eta)$ and $m_+(\eta)$ respectively.

The lower endpoint estimate is strictly negative. The contribution of the population component H to $\mathcal{L}_{\eta,x}(m)$ is symmetric in m. For the lower-window component, every report satisfies $y \le 0$, and for every $m > 0$,

$$|y + m| < |y - m|$$

for a set of reports of positive measure. Since $\log(2a_0 + r^2)$ is strictly increasing in $r \ge 0$, the lower-window objective is strictly smaller at $-m$ than at m. Thus no positive m can be a global minimizer. At $m = 0$, the population component has zero score by symmetry, while the lower-window score is

$$\frac{\eta}{\gamma} \int_0^\gamma \psi(Q_\nu(u))\, du < 0,$$

because $Q_\nu(u) < 0$ for $u < 1/2$ and ψ has the sign of its argument. Hence zero is not a minimizer either, so $m_-(\eta) < 0$. The mirror argument for the upper endpoint gives $m_+(\eta) > 0$.

It remains to show that these two positive-mass cohorts survive. By part (i), after the first round an asymptotically positive fraction of each endpoint cohort lies within $O_p(d^{-1/2})$ of its population estimate and has precision in $[cd, Cd]$ for fixed $0 < c < C < \infty$. With probability tending to one, every member of this endpoint core has $\Theta(d)$ links to other core members of the same endpoint cohort, and any two such observation rows overlap on $\Theta(d)$ core sources. At the first subsequent update, core belief gaps are $O_p(d^{-1/2})$, so the exact kernel assigns raw precision $\Theta(d)$ to core reports and the internal trust mass is $\Theta(d^2)$. Reports a fixed distance from the core contribute only $O(d)$ total precision, while reports within the shrinking core neighborhood cannot move the update outside that neighborhood.

The dense overlap of the internal rows makes the induced averaging on the core a strict contraction. While the core diameter remains above its current trust bandwidth, its precision increases by at least order d^2 per round and its diameter contracts geometrically; after $O(\log d)$ rounds the diameter is below the trust bandwidth. The cumulative displacement caused by fixed-distance outside reports over this phase is $O(\log d/d) = o(1)$. Once the core lies inside the trust bandwidth, each internal report receives precision proportional to the observer's current precision. The $\Theta(d)$ internal reports then multiply precision by order d from one round to the next, so the remaining outside influence and the resulting tail movement are summable. Thus the core has limiting beliefs

$$\ell_N^- = m_-(\eta_N) + o_p(1), \qquad \ell_N^+ = m_+(\eta_N) + o_p(1).$$

Since $\eta_N = \eta + o(1)$ under fixed primitives, their separation is $\Delta(\eta) + o_p(1)$, where $\Delta(\eta) = m_+(\eta) - m_-(\eta) > 0$. The total limiting spread is at least this separation. $\square$

The proof above uses only the positive-mass endpoint cohorts. It does not show that a more extreme isolated value of M_η also survives. That additional step is the content of Conjecture 1.

Proof of Proposition 2

Proposition 2. *Let the network be drawn by any process independent of initial beliefs, with minimum out-degree d satisfying $d/\log N \to \infty$. Then the spread of limiting beliefs converges to zero in probability, and every limiting belief converges to θ.*

Proof. Because formation is independent of initial beliefs, each observer's sources are an independent sample from the common initial-belief law. The same bounded-score concentration argument as part (i) of Theorem 3, together with $d/\log N \to \infty$, places every first-round estimate uniformly near the unique population minimizer. In the baseline model that population objective is, up to an affine transformation,

$$\mathbb{E}[-\log f_\nu(Y - m)]\,,$$

where $Y \sim f_\nu(\cdot - \theta)$ is the Student-t initial-belief law. Its excess over the value at $m = \theta$ is the Kullback–Leibler divergence between $f_\nu(\cdot - \theta)$ and its translate $f_\nu(\cdot - m)$, so θ is the unique minimizer. Hence

$$\max_i |\mu_i^1 - \theta| = O_p\left(\sqrt{\frac{\log N}{d}}\right).$$

Every later update is a convex combination of current beliefs, so the period-one hull contains all subsequent and limiting beliefs. Its width vanishes and its center converges to θ. $\square$

D Proofs for Robustness and Guaranteed Consensus

This appendix contains the proofs for Section 6.

Proof of Proposition 3

Proposition 3.

(i) The rule with benchmark pair $(m_i^t, 1/s_i^t)$ is the global mean-field variational projection of the exact Bayesian posterior when sources' reports are conditionally independent given the state.

(ii) The unique dependence-invariant rule is the pairwise rule, with benchmark pair $(\mu_i^t, 1/\tau_i^t)$: the maximally correlation-robust member of the class.

Proof. For (i): under conditional independence the exact joint posterior over the state and the source precisions is well defined, and (1) with the pooled pair, solved jointly at the global variational minimum, characterizes its normal-times-independent-Gammas projection; this is the variational characterization in the footnote to (1).

For (ii): the fixed marginal laws are consistent with many joints, sources independent given the state or sharing a common shock of any size. Conditioning on a subset of variables uses only that subset's law, so the conditional law of τ_j given the assessor's state and μ_j alone is the same under every admissible joint: the pairwise rule is invariant. The conditional law given (μ_1, μ_2) is not: under independence, agreement of μ_2 with μ_1 is evidence for τ_1; under a common shock the same agreement is partly the shock's work. Invariance therefore forces trust to depend on the source's own report and the assessor's state alone; correctness there is Bayes on the single-report problem, which is (1) with $(\mu_i^t, 1/\tau_i^t)$. Every rule of (9) with $\lambda > 0$ conditions on the pooled estimate. □

Proof of Proposition 4

Proposition 4. *The convergence and finite-network results extend to updating rules in the joint inference class. For large networks, let M_η^ϕ denote the corresponding population first-round profile. The first-round concentration conclusion of Theorem 3 continues with M_η replaced by M_η^ϕ. If under strict sorting the lower and upper endpoint cohorts have distinct population estimates $m_-^\phi(\eta) < m_+^\phi(\eta)$, then the limiting spread is at least*

$$m_+^\phi(\eta) - m_-^\phi(\eta) + o_p(1).$$

Proof. The convergence and finite-network arguments consume only the uniform scale-free trust bounds of Definition 1, directly or through Lemmas 1 and 2. In the large-network argument, a different member of the joint-inference class changes the bounded first-round estimating function and therefore the population profile, but not the concentration argument in part (i) of Theorem 3. If the two endpoint population estimates are distinct, each is carried by a positive-mass rank cohort. Scale-free trust assigns precision of order current precision to sufficiently close reports, while inverse-square discounting makes reports a fixed distance away lower order. The same positive-mass corroboration argument as in part (iii) of Theorem 3 therefore locks in limiting beliefs near the two endpoint estimates and yields the stated separation. □

Proof of Proposition 5

Proposition 5. *Fix a benchmark pair and write its normalized variance as $c_v^t := \tau_i^t v_i^t \in [0, 1]$. Under Assumption 2, suppose the reliability prior F has finite moments of all orders and a density that behaves like x^{a-1} near zero, in the sense of regular variation, for some $a > 0$. Then the posterior mean of any source's precision is exactly*

$$\tau_i^t \, \phi_{\tau_i^0, c_v^t}\big(\tau_i^t \delta^2\big),$$

where, for fixed initial type τ_i^0, the family is positive, smooth, strictly decreasing, and finite at zero uniformly in $c_v \in [0,1]$, with

$$z\phi_{\tau_i^0, c_v}(z) \to 2a+1$$

uniformly in $c_v \in [0,1]$ on compact sets of initial types. The Gamma form gives exactly $2a_0+1 = \kappa_0$.

Proof. Write the source's period-t precision as $\hat{c}_i^t x$, with $x \sim F$, by Assumption 2. Holding the normalized benchmark variance c_v^t fixed, the report's predictive variance is

$$\frac{\tau_i^0/x + c_v^t}{\tau_i^t}.$$

Thus the likelihood of a benchmark deviation δ is proportional in x to

$$A(x)e^{-zh(x)}, \qquad z := \tau_i^t \delta^2, \qquad A(x) := \left(\frac{\tau_i^0}{x} + c_v^t\right)^{-1/2}, \qquad h(x) := \frac{1}{2(\tau_i^0/x + c_v^t)}.$$

The posterior mean of source precision is therefore

$$\tau_i^t\, \phi_{\tau_i^0, c_v^t}(z), \qquad \phi_{\tau_i^0, c_v^t}(z) = \frac{\mathbb{E}\big[xA(x)e^{-zh(x)}\big]}{\tau_i^0\, \mathbb{E}\big[A(x)e^{-zh(x)}\big]},$$

which gives the claimed scale-free representation.

Positivity is immediate. Finite moments of all orders ensure finiteness at $z = 0$ and justify repeated differentiation under the integral, uniformly for $c_v \in [0,1]$ and τ_i^0 in compact subsets of $(0,\infty)$. For monotonicity, let π_z denote the tilted law proportional to $A(x)e^{-zh(x)}\, dF(x)$. Then

$$\tau_i^0\, \partial_z \phi_{\tau_i^0, c_v^t}(z) = -\operatorname{Cov}_{\pi_z}\big(x, h(x)\big) < 0,$$

because $h(x)$ is strictly increasing in x.

For the tail, write the density near zero as

$$f(x) = x^{a-1} L(x),$$

where L is slowly varying at zero. As $z \to \infty$, the factor $e^{-zh(x)}$ localizes the integrals near $x = 0$, where, uniformly over $c_v \in [0,1]$ and compact sets of τ_i^0,

$$h(x) \sim \frac{x}{2\tau_i^0}, \qquad A(x) \sim \left(\frac{x}{\tau_i^0}\right)^{1/2}.$$

The standard Laplace asymptotics for regularly varying densities therefore give

$$\mathbb{E}\Big[Ae^{-zh}\Big] \sim (\tau_i^0)^{-1/2}\Gamma\left(a + \frac{1}{2}\right)\left(\frac{2\tau_i^0}{z}\right)^{a+1/2} L(2\tau_i^0/z)$$

and

$$\mathbb{E}\Big[xAe^{-zh}\Big] \sim (\tau_i^0)^{-1/2}\Gamma\left(a+\frac{3}{2}\right)\left(\frac{2\tau_i^0}{z}\right)^{a+3/2} L(2\tau_i^0/z),$$

with the common multiplicative factors cancelling in their ratio. Since $\Gamma(a+3/2)/\Gamma(a+1/2) = a+1/2$,

$$z\phi_{\tau_i^0, c_v^t}(z) \longrightarrow 2a+1.$$

The same asymptotics are uniform in $c_v \in [0,1]$ and on compact sets of initial types. □

Proof of Proposition 6

Proposition 6. *Fix any protocol in which each agent updates on her sources' reports: any report space, deterministic or randomized. Consider three conditions.*

(i) Evidence-responsiveness*: an agent who observes no one retains her mean belief.*

(ii) Guaranteed consensus*: on every indeterminate-class society of Theorem 2, fully compliant play reaches consensus almost surely.*

(iii) Non-dictatorship*: on no indeterminate-class society can a single non-compliant agent make her chosen position the limiting belief of every compliant agent.*

No protocol satisfies all three.

Proof. Suppose a protocol satisfies (i) and (ii), and fix any target position x. Start from a strongly connected compliant society containing two vertex-disjoint cycles. Adjoin one additional agent h who observes no one, and let at least one compliant agent observe h. Then $\{h\}$ is closed. No nonempty set of compliant agents is closed: within the original strongly connected society every proper subset has an observation edge leaving it, while the full compliant set is opened by the added observation of h. Hence $\{h\}$ is the unique minimal closed listening group of the augmented society. With the one-cycle at h and any directed cycle inside the compliant society, the augmented graph has two vertex-disjoint cycles and therefore lies in the indeterminate class.

Now consider the counterfactual fully compliant play on this augmented graph in which h starts at x. By evidence-responsiveness, h remains at x; by guaranteed consensus, every compliant agent converges to x almost surely. A single non-compliant agent can reproduce exactly the report history that h generates in this counterfactual play, using the same realization of any private randomization. The other agents then receive exactly the same histories as in the fully compliant counterfactual and therefore follow the same trajectories, even if the non-compliant agent's actual initial state was different. She can thus make the limiting belief of every compliant agent equal to the arbitrary target x, violating (iii). □

References

Daron Acemoglu, Giacomo Como, Fabio Fagnani, and Asuman Ozdaglar. Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research*, 38(1):1–27, 2013.

Delia Baldassarri and Andrew Gelman. Partisans without constraint: Political polarization and trends in american public opinion. *American Journal of Sociology*, 114(2):408–446, 2008.

Roger L. Berger. A necessary and sufficient condition for reaching a consensus using DeGroot's method. *Journal of the American Statistical Association*, 76(374):415–418, 1981.

Simone Cerreia-Vioglio, Roberto Corrao, and Giacomo Lanzani. Dynamic opinion aggregation: Long-run stability and disagreement. *Review of Economic Studies*, 91(3):1406–1447, 2024.

Arun G. Chandrasekhar, Horacio Larreguy, and Juan Pablo Xandri. Testing models of social learning on networks: Evidence from two experiments. *Econometrica*, 88(1):1–32, 2020.

Pranav Dandekar, Ashish Goel, and David T. Lee. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796, 2013.

Guillaume Deffuant, David Neau, Frédéric Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(01n04):87–98, 2000.

Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69 (345):118–121, 1974.

Daniel DellaPosta, Yongren Shi, and Michael Macy. Why do liberals drink lattes? *American Journal of Sociology*, 120(5):1473–1511, 2015.

Peter M. DeMarzo, Dimitri Vayanos, and Jeffrey Zwiebel. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*, 118(3):909–968, 2003.

Benjamin Enke and Florian Zimmermann. Correlation neglect in belief formation. *Review of Economic Studies*, 86(1):313–332, 2019.

Matthew Gentzkow, Michael B. Wong, and Allen T. Zhang. Ideological bias and trust in information sources. *American Economic Journal: Microeconomics*, 17(2):162–213, 2025.

Benjamin Golub and Matthew O. Jackson. Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–149, 2010.

Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.

Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.

Mohsen Mosleh, Cameron Martel, Dean Eckles, and David G. Rand. Shared partisanship dramatically increases social tie formation in a twitter field experiment. *Proceedings of the National Academy of Sciences*, 118(7):e2022761118, 2021.

Junya Murayama, Brian W. Rogers, Jacob VanNattan, and Xiannong Zhang. Belief formation on networks. Working paper, 2026.

Reza Olfati-Saber and Richard M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9):1520–1533, 2004.

Arnold Polanski and Fernando Vega-Redondo. Homophily and influence. *Journal of Economic Theory*, 207:105576, 2023.

Michael J. Rosenfeld, Reuben J. Thomas, and Sonia Hausen. Disintermediating your friends: How online dating in the united states displaces other ways of meeting. *Proceedings of the National Academy of Sciences*, 116(36):17753–17758, 2019.

Rajiv Sethi and Muhamet Yildiz. Communication with unknown perspectives. *Econometrica*, 84 (6):2029–2069, 2016.