

Social Learning with Coarse Communication

Jonghyun Park*

Hector Tzavellas†

January 2026

Abstract

We study social learning in networks when communication is restricted to a finite vocabulary. Because messages are generated by a coarse quantizer, social information is piecewise constant and evolves only through threshold crossings. This creates a bandwidth-mixing mismatch: faster network mixing shrinks cross-sectional dispersion, yet dispersion is required to trigger threshold crossings. We show that sufficiently fast-mixing networks exhibit homogenization traps: with non-vanishing probability, the society enters an absorbing message region where a large cohesive core becomes permanently locked in, generating an information ceiling. By contrast, in segregated networks, correct "seeds" can form locally and propagate through weak ties. Consequently, learning is non-monotone in connectivity: very sparse networks fail to diffuse information, while very dense networks suppress the dispersion needed for communication to evolve.

JEL Classifications:

Keywords: Social learning, Communication constraints, Coarse learning, Information ceiling

*Department of Economics, Virginia Tech, Blacksburg, USA, Email: jongh2419@vt.edu

†Department of Economics, Virginia Tech, Blacksburg, USA, Email: hectortz@vt.edu

1 Introduction

1.1 Overview

Many real-world communication channels are designed to be *coarse*. Inside organizations, vast amounts of dispersed information are funneled into a small number of categories: a credit committee assigns an internal rating, a hiring committee produces a short-list rather than a distribution of opinions. In markets and platforms, public communication is similarly compressed: sellers and products are summarized by star ratings, borrowers by credit grades, or recommendation labels.

This paper studies repeated learning on a network when communication is restricted to a *finite vocabulary*. Agents observe bounded private evidence about a fixed binary state and update a belief statistic over time. In each period, agents exchange publicly interpretable messages that are generated by a coarse quantizer.

The key implication is that socially transmissible content is piecewise constant: aggregate exposure changes only when some agent crosses a threshold. This creates a fundamental tension we call the bandwidth–mixing mismatch. While standard intuition suggests connectivity aids learning, here it destroys it: fast mixing shrinks cross-sectional dispersion, but dispersion is the fuel for threshold crossings. If agents become too similar too quickly, the society "runs out" of boundary agents, and public messages freeze.

Our first result is a benchmark: Under unconstrained (continuous) communication—when the communication rule is continuous rather than stepwise—the cross-sectional mean belief drifts in the correct direction and dispersion remains uniformly controlled, implying that beliefs become state-aligned.

The second result shows that a finite vocabulary fundamentally changes learning quality. When communication has absorbing regions, sufficiently fast-mixing networks exhibit homogenization traps. With non-vanishing probability, the society enters a message region where a large cohesive core becomes permanently locked in. This generates an information ceiling: even as private evidence accumulates, the public channel goes silent, and a positive fraction of agents remain bounded away from the truth.

Our third result identifies a geometric solution. In networks with a segregated structure, local groups can incubate a correct consensus within an absorbing region. These local consensuses act as "seeds" that are robust to external noise. Weak ties then allow the formed seed to replicate across communities.

Combining these forces yields a non-monotone relationship between connectivity and learn-

ing. If connectivity is too low, information cannot propagate. If connectivity is too high and mixing is too fast, dispersion is destroyed so quickly that threshold crossings cease, and public content becomes trapped. Between these extremes, community structure and weak ties can support both local information production and global diffusion.

The results point to a substantive distinction between environments with rich communication and those with coarse public language. When communication is coarse, the welfare value of connectivity depends on whether additional exposure primarily accelerates *diffusion* or instead suppresses the *dispersion* needed for public content to update. This distinction helps rationalize why organizations and platforms sometimes rely on comparatively insulated groups (committees, desks, teams, or subcommunities) that form stable local consensus before broadcasting outward, and why indiscriminate “global mixing” can be counterproductive when public reports are categorical.

1.2 Related literature

This paper contributes to three literatures: learning on networks, communication constraints in information aggregation, and the role of clustered social structure and weak ties in diffusion.

Learning and information aggregation on networks. A large literature studies how network structure shapes learning when agents repeatedly exchange beliefs or signals. In canonical models of non-Bayesian updating and linear social learning, connectivity tends to accelerate consensus and facilitate aggregation (e.g., DeGroot (1974), Golub and Jackson (2010)). In Bayesian environments, richer observation or broader exposure often improves asymptotic learning, although speeds and transient dynamics depend sharply on network geometry and information structure (e.g., Acemoglu et al. (2011), Mossel et al. (2015), Jadbabaie et al. (2012)). Our contribution is to identify a distinct channel through which additional connectivity can be harmful: when public communication is restricted to a finite alphabet, fast mixing can eliminate the cross-sectional dispersion needed for public content to change, leading to a **homogenization trap** and an information ceiling even though private evidence continues to arrive.

Communication constraints, coarse messages, and information ceilings. The paper is also related to work emphasizing limits on what can be communicated, processed, or encoded. In economics, coarse communication arises in models of cheap talk and finite-message communication (e.g., Crawford and Sobel (1982), and subsequent work on quantized/finite-

message reporting), as well as in models where attention or information processing is constrained (e.g., rational inattention Sims (2003); and related discrete-choice RI frameworks such as Matejka and McKay (2015)). Our focus is different: agents are not strategically misreporting and the constraint is not cognitive in the sense of optimizing an information cost; rather, the public channel itself is discrete. The resulting step-function social transmission rule makes aggregate exposure *piecewise constant*, so that social learning requires threshold crossings. The paper therefore connects to a broader set of ideas in distributed control known as “quantization consensus” (e.g., Kashyap et al. (2007), Carli et al. (2010)), where quantized communication is shown to generate dead-zones that limit accuracy. Our emphasis, however, is on the economic comparative static: how mixing speed interacts with message resolution to determine whether these traps emerge.

Clustered networks, weak ties, and the production of public content. A third literature studies how segregation and clustered interaction affect information diffusion and polarization. Classic arguments emphasize the informational and diffusion value of weak ties (Granovetter (1973)). More recent work analyzes how homophily or clustering shapes learning, belief polarization, and diffusion patterns by changing who interacts with whom and what is observed. Our results complement these perspectives by highlighting a mechanism through which clustered social structure can be beneficial under coarse public communication: insulated groups can generate stable, state-aligned local consensus (**robust seeds**) that subsequently spread through weak ties. In this sense, segregation can raise the *rate of public content production* by preserving dispersion where it is needed for threshold crossings, while limited cross-group exposure then supports diffusion.

Diffusion on networks and threshold-type dynamics. Finally, the paper relates to diffusion models in which behavior or adoption spreads through network interactions, often via thresholds or complementarities. While our object is belief dynamics rather than adoption, the stepwise nature of public communication makes learning hinge on threshold events, creating a close formal analogy. In particular, our work is complementary to diffusion and word-of-mouth frameworks that study how network structure shapes propagation, incentives, and welfare (e.g., Campbell (2013), Shin (2017), Ajorlou et al. (2018), Leduc et al. (2017)). The key distinction is that in our setting thresholds are not primitive behavioral rules; they are induced by finite-vocabulary communication, and the main force behind failure is the interaction between fast mixing and limited message resolution.

Section 2 presents the environment, the exposure matrix, and the finite-vocabulary communication technology. Section 3 formalizes the bandwidth-mixing mismatch by showing that

mixing contracts dispersion while threshold crossings require it. Section 4 states and proves the main comparative-statics results: the continuous benchmark, homogenization traps on fast-mixing networks, learning via locally formed seeds and weak ties, and the resulting non-monotonicity in connectivity. Appendix A contains proofs of all numbered main-text results. Appendix B collects supporting derivations and justification for informal claims used in the main text. Appendix C collects other technical lemmas and auxiliary results.

2 Model

This section introduces the primitives: the state, private evidence, the fixed exposure matrix, and the belief dynamics. These objects are independent of the communication technology. Subsection 2.2 then imposes finite-vocabulary communication and formalizes the outcome notions used in the results.

2.1 Environment and belief dynamics

State, agents, and time

Time is discrete, $t = 0, 1, 2, \dots$. There is an unknown binary state

$$\theta \in \{0, 1\},$$

drawn at $t = 0$ from the common prior $\Pr(\theta = 1) = \mu_0 \in (0, 1)$ and held fixed thereafter. There are n agents indexed by $i \in N^{(n)} := \{1, \dots, n\}$. When n is fixed, we write $N := N^{(n)}$ and suppress the superscript (n) on network objects.

Network exposure and social aggregates

Social exposure is summarized by an $n \times n$ row-stochastic matrix

$$P^{(n)} = (P_{ij}^{(n)})_{i,j \in N^{(n)}}, \quad P_{ij}^{(n)} \geq 0, \quad \sum_{j=1}^n P_{ij}^{(n)} = 1 \quad \forall i.$$

We interpret P_{ij} as the attention/trust weight that agent i assigns to agent j ; thus P encodes the network topology and is fixed over time.

Each period, agents decode messages from a network. Let $y_{j,t} \in \mathbb{R}$ denote the decoded content produced by agent j at the end of period t . Agent i aggregates these decoded

contents according to

$$S_{i,t} := \sum_{j=1}^n P_{ij} y_{j,t}.$$

Under perfect (unconstrained) communication one may take $y_{j,t} = \ell_{j,t}$, agent j 's current log-odds belief. Under finite vocabulary, Subsection 2.2 specifies an encoder/decoder pair that maps beliefs to a finite alphabet and back to a real-valued decoded statistic.

Private evidence

At the beginning of each period $t \geq 1$, agent i receives a private signal $x_{i,t}$ with conditional density (or mass function) $f_i(\cdot | \theta)$. Because the state is binary, the informational content of $x_{i,t}$ is summarized by its log-likelihood ratio (LLR),

$$\Delta_{i,t} := \log \frac{f_i(x_{i,t} | \theta = 1)}{f_i(x_{i,t} | \theta = 0)}.$$

A positive $\Delta_{i,t}$ favors $\theta = 1$ and a negative $\Delta_{i,t}$ favors $\theta = 0$. The LLR is the natural statistic in a binary-state model because it is *additive*: for an isolated Bayesian learner, posterior log-odds equal prior log-odds plus the signal LLR.

Assumption 1. *For the signal process $\Delta_{i,t}$:*

- (i) *Conditional independence:* Conditional on θ , the family $\{\Delta_{i,t}\}_{i \in N, t \geq 1}$ is independent across agents and time.
- (ii) *Bounded private evidence:* There exists $\bar{\Delta} \in (0, \infty)$ such that for all i and all $t \geq 1$,

$$|\Delta_{i,t}| \leq \bar{\Delta} \quad a.s.$$

- (iii) *Directional informativeness:* There exists $m > 0$ such that for all i and all $t \geq 1$,

$$\mathbb{E}[\Delta_{i,t} | \theta = 1] \geq m, \quad \mathbb{E}[\Delta_{i,t} | \theta = 0] \leq -m.$$

Assumption 1(ii) rules out one-shot revelation about the state. If it is not bounded, a one-shot private signal may reveal the true state. Assumption 1(iii) ensures that, in expectation, private evidence drifts toward the true state.

Beliefs, log-odds, and information sets

Let $\mathcal{I}_{i,t}$ denote agent i 's information at the *start* of period t . Under the timing specified below, a convenient representation with σ -algebra is

$$\mathcal{I}_{i,t} := \sigma\left(\ell_{i,0}, \{\Delta_{i,s}\}_{s=1}^t, \{y_{j,s}\}_{j \in N, 0 \leq s \leq t-1}\right),$$

i.e., agent i knows her own private LLR history and the past decoded social contents.

Agent i 's belief is $\mu_{i,t} := \Pr(\theta = 1 \mid \mathcal{I}_{i,t})$. We work with the associated log-odds (logit) transform

$$\ell_{i,t} := \log \frac{\mu_{i,t}}{1 - \mu_{i,t}} \in \mathbb{R}, \quad \mu_{i,t} = \frac{1}{1 + e^{-\ell_{i,t}}}.$$

The logit scale maps probabilities in $(0, 1)$ to \mathbb{R} and makes private evidence additive: absent social interaction, $\ell_{i,t+1} = \ell_{i,t} + \Delta_{i,t+1}$. This alignment is central under finite vocabulary (Subsection 2.2): coarse communication makes the social component piecewise constant in ℓ , so social information moves only through threshold crossings.

We initialize $\ell_{i,0} = \log(\mu_0 / (1 - \mu_0))$ for all i .

Timing, non-strategic communication, and anchored updating

Timing. For each $t \geq 0$:

- (i) *Communication.* Each agent j sends a message based on her current log-odds $\ell_{j,t}$ with encoded. Receivers decode it into $y_{j,t} \in \mathbb{R}$ and compute the social aggregate $S_{i,t} = \sum_j P_{ij} y_{j,t}$.
- (ii) *Private evidence.* Each agent i observes a new private signal $x_{i,t+1}$, summarized by $\Delta_{i,t+1}$.
- (iii) *Update.* Agent i updates from $\ell_{i,t}$ to $\ell_{i,t+1}$ using $(S_{i,t}, \Delta_{i,t+1})$.

Communication is *myopic and non-strategic*: agents do not manipulate reports to influence others. They transmit their current learning outcome, with the only distortion arising from the exogenous communication constraint (finite vocabulary) introduced in Subsection 2.2. This assumption isolates the paper's mechanism: the interaction between coarse transmission and network mixing.

Beliefs evolve according to a linear constant-gain (“anchored”) update: for a responsiveness parameter $\alpha \in (0, 1)$,

$$\ell_{i,t+1} = (1 - \alpha)\ell_{i,t} + \alpha(S_{i,t} + \Delta_{i,t+1}). \tag{1}$$

We show later that this learning itself does not imply an information ceiling.

Decision problem and performance metric

To give the accuracy metric a decision-theoretic interpretation, suppose that at each period t , agent i chooses an action $a_{i,t} \in [0, 1]$ to minimize the myopic quadratic loss $\mathbb{E}[(a_{i,t} - \theta)^2 | \mathcal{I}_{i,t}]$. The optimal action is the posterior mean,

$$a_{i,t} = \mathbb{E}[\theta | \mathcal{I}_{i,t}] = \mu_{i,t} = \frac{1}{1 + e^{-\ell_{i,t}}}.$$

Accordingly, we evaluate performance by the cross-sectional mean-squared error

$$\text{MSE}_t^{(n)} := \frac{1}{n} \sum_{i=1}^n \mathbb{E}[(\mu_{i,t} - \theta)^2],$$

where the expectation is taken over the realization of the state and signal sequences. We study the behavior of $\text{MSE}_t^{(n)}$ in the long run ($t \rightarrow \infty$) and in large societies ($n \rightarrow \infty$).

2.2 Finite-vocabulary communication and threshold transmission

This subsection introduces the paper's core friction: agents cannot transmit arbitrary real-valued beliefs. Instead, each period they encode their current log-odds into one of K messages and receivers apply a common decoder. The implication is structural: the *socially transmissible* mapping from log-odds beliefs to decoded content is a *step function*. Hence, the social aggregate $S_{i,t}$ can change only through *threshold crossings*. Section 3 shows how this discreteness interacts with network mixing.

Encoder, decoder, and the socially transmissible map

Fix an integer $K \geq 2$ and message alphabet $\mathcal{M} := \{1, \dots, K\}$. An *encoder* is a measurable map $Q : \mathbb{R} \rightarrow \mathcal{M}$ and a *decoder* is a map $\Psi : \mathcal{M} \rightarrow \mathbb{R}$. When agent j holds log-odds $\ell_{j,t}$ at the communication stage of period t , she broadcasts

$$m_{j,t} := Q(\ell_{j,t}) \in \mathcal{M},$$

which receivers interpret as the decoded real-valued content

$$y_{j,t} := \Psi(m_{j,t}) \in \mathbb{R}.$$

Equivalently, only the composition

$$\psi := \Psi \circ Q : \mathbb{R} \rightarrow \mathbb{R}, \quad y_{j,t} = \psi(\ell_{j,t}),$$

matters for social learning. Under the anchored update (1), agent i 's social aggregate can therefore be written as

$$S_{i,t} = \sum_{j=1}^n P_{ij} \psi(\ell_{j,t}),$$

i.e., agents aggregate *decoded reconstructions* of others' log-odds, not the raw beliefs themselves.

Threshold form, bins, and bandwidth

Throughout, we focus on the empirically relevant and analytically convenient case of a *threshold encoder*. Let

$$-\infty = b_0 < b_1 < \cdots < b_{K-1} < b_K = +\infty$$

be cutpoints and define bins $B_k := (b_{k-1}, b_k]$. The threshold encoder assigns message k whenever $\ell \in B_k$, and the decoder assigns a reconstruction value $r_k := \Psi(k)$ to message k . In this case,

$$\psi(\ell) = r_k \quad \text{for all } \ell \in (b_{k-1}, b_k] \quad (k = 1, \dots, K). \quad (2)$$

We maintain the natural monotonicity restriction $r_1 \leq r_2 \leq \cdots \leq r_K$, so ψ is weakly increasing¹. The parameter K measures the *bandwidth* of the communication layer: a larger K yields a finer partition of the belief-statistic space.

The key implication is immediate: because ψ is piecewise constant, social information can move only when some agent crosses a bin boundary.

Lemma 1. *If $Q(\ell_{j,t}) = Q(\ell_{j,t-1})$ for all j , then $S_{i,t} = S_{i,t-1}$ for all i . Equivalently, social aggregates change between periods only if at least one agent crosses a threshold.*

Lemma 1 reduces learning under coarse communication to the dynamics of boundary mass: to change social influence, the system must keep some agents sufficiently close to a cutpoint.

3 Mechanism: mixing and dispersion destruction

Finite-vocabulary communication implies that the socially transmissible map $\psi = \Psi \circ Q$ is *step-valued*: it is constant on intervals and can change only when some agent's underlying belief statistic crosses a cutpoint (Lemma 1). As a result, aggregate social content is not continuously adjustable; it updates only through *threshold crossings* driven by the agents who lie close to the cutpoints.

¹We provide the micro-foundation for this threshold idea in Appendix B.1.

The tension is that many networks generate rapid *disagreement compression*. Repeated exposure averaging by P contracts cross-sectional dispersion in beliefs at an exponential rate (formalized below via the second-largest eigenvalue). When dispersion collapses quickly, the society quickly runs out of boundary agents, and threshold crossings become rare or impossible. This is the *bandwidth-mixing mismatch*: the faster the network homogenizes beliefs, the harder it becomes for a coarse vocabulary to generate new transmissible content.

We organize the mechanism in two steps:

- (M1) *Mixing compresses dispersion.* Iterated averaging by P shrinks cross-sectional disagreement at a geometric rate controlled by the second-largest eigenvalue. In high-connectivity societies, dispersion becomes small quickly.
- (M2) *Threshold crossings require dispersion.* Because ψ is a step function, a change in social content can occur only if some agents lie sufficiently close to a cutpoint. The fraction (or mass) of such boundary agents is quantitatively controlled by dispersion: when dispersion is small, very few agents are near any cutpoint, so message changes are unlikely.

If dispersion collapses while the mean belief remains inside a bin, the message profile freezes and coarse communication stops producing new public content. Section 4 formalizes how fast mixing yields homogenization traps (Theorem 1) and how modular structure can instead sustain seed formation and propagation (Theorem 2).

3.1 Mixing regularity and the π -geometry

Connectivity enters our model only through the exposure matrix P : each period, agents partially average what they have heard through P . To formalize the mechanism “mixing destroys dispersion” we therefore need a quantitative sense in which repeated application of P shrinks disagreement. To ensure the convergence under p , we impose a mild condition.

Assumption 2 (Mixing regularity). *The matrix P is row-stochastic and irreducible². Moreover, P is reversible with respect to π :*

$$\pi_i P_{ij} = \pi_j P_{ji} \quad \forall i, j.$$

Irreducibility guarantees that the stationary distribution π exists and is unique. Reversibility is the key benchmark: it makes P self-adjoint in $L^2(\pi)$, so the speed at which P eliminates

²Therefore, there exists a unique distribution π such that $\pi^\top P = \pi^\top$ and $\pi_i > 0, \forall i$,

disagreement is governed by eigenvalues (spectral gap). This is exactly what we need to state and prove the “dispersion destruction” part of the mismatch mechanism³.

For $u, v \in \mathbb{R}^n$, define the π -weighted inner product and norm by

$$\langle u, v \rangle_\pi := \sum_{i=1}^n \pi_i u_i v_i, \quad \|u\|_\pi := \sqrt{\langle u, u \rangle_\pi}.$$

For any $v \in \mathbb{R}^n$, define its π -mean and the centering operator

$$\bar{v} := \sum_{i=1}^n \pi_i v_i = \pi^\top v, \quad \text{CO}(v) := v - \bar{v} \mathbf{1},$$

where $\mathbf{1}$ is the all-ones vector.

Definition 1 (Dispersion). For $v \in \mathbb{R}^n$, define its π -dispersion by

$$\text{Var}_\pi(v) := \|\text{CO}(v)\|_\pi^2 = \sum_{i=1}^n \pi_i (v_i - \bar{v})^2.$$

For the belief profile $\ell_t = (\ell_{i,t})_{i \in N}$, write

$$D_t := \|\text{CO}(\ell_t)\|_\pi \quad \text{so that} \quad \text{Var}_\pi(\ell_t) = D_t^2.$$

Dispersion $\text{Var}_\pi(\ell_t)$ measures cross-sectional disagreement in log-odds beliefs. In our setting, disagreement is not merely ‘noise’: under finite-vocabulary communication, message changes occur only when some agent reaches a cutpoint first (Lemma 1), so a society with negligible dispersion has essentially no ‘boundary’ agents capable of triggering threshold crossings.

The next lemma shows how many agents can lie far from (or close to) the mean. We will use it to argue that when dispersion is small, only a small π -mass of agents can be near any cutpoint.

Lemma 2. *For any $v \in \mathbb{R}^n$ and any $\eta > 0$,*

$$\pi(\{i : |v_i - \bar{v}| \geq \eta\}) := \sum_{i \in N} \pi_i \mathbb{I}\{|v_i - \bar{v}| \geq \eta\} \leq \frac{\text{Var}_\pi(v)}{\eta^2}.$$

Equivalently, $\pi(\{i : |v_i - \bar{v}| < \eta\}) \geq 1 - \text{Var}_\pi(v)/\eta^2$.

³Non-reversible chains can be handled using singular values; we avoid that notation to keep the focus on the economic mechanism.

Under Assumption 2, reversibility implies that P is self-adjoint in $L^2(\pi)$.⁴ Let $1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_n$ be the eigenvalues of P (counted with multiplicity), where $\lambda_1 = 1$ corresponds to $\mathbf{1}$. Define the *disagreement contraction factor*

$$\lambda^* := \max\{|\lambda_2|, |\lambda_n|\} \in [0, 1].$$

Smaller λ^* means *faster homogenization*: any mean-zero heterogeneity is averaged out quickly.

The key implication is that P contracts centered vectors at rate λ^* : for any $x \in \mathbb{R}^n$,

$$\|\text{CO}(Px)\|_\pi \leq \lambda^* \|\text{CO}(x)\|_\pi, \quad \text{equivalently} \quad \text{Var}_\pi(Px) \leq (\lambda^*)^2 \text{Var}_\pi(x). \quad (3)$$

Iterating yields $\|\text{CO}(P^t x)\|_\pi \leq (\lambda^*)^t \|\text{CO}(x)\|_\pi$. Formal statements and proofs are in Appendix C.⁵

3.2 The mismatch principle: drift versus dispersion destruction

To separate what the network does to *average beliefs* from what it does to *disagreement*, fix t and decompose the log-odds profile $\ell_t \in \mathbb{R}^n$ into its π -mean and deviations from that mean:

$$\bar{\ell}_t := \pi^\top \ell_t, \quad d_t := \ell_t - \bar{\ell}_t \mathbf{1}.$$

This is an identity: $\ell_t = \bar{\ell}_t \mathbf{1} + d_t$, and d_t is mean-zero in the sense that $\pi^\top d_t = 0$. We interpret $\bar{\ell}_t$ as the *mean belief* and d_t as *disagreement*.

The key distinction lies in how the network operator P affects the different components of belief. When agents communicate, they effectively multiply the belief vector ℓ_t by the matrix P . To see how this affects the global average, we apply the stationarity condition $\pi^\top P = \pi^\top$:

$$\pi^\top (P\ell_t) = \pi^\top \ell_t = \bar{\ell}_t.$$

This shows that averaging redistribution among agents preserves the global weight mean. Consequently, network mixing itself does not pull society toward the truth or toward zero; it only reduces disagreement. The evolution $\bar{\ell}_t$ is driven primarily by the aggregate drift in private evidence, and under directional informativeness the mean drifts systematically toward the true state (Lemma 3). This is why, under continuous communication, learning is possible even in highly connected societies (Proposition 1).

⁴Self-adjointness is with respect to $\langle \cdot, \cdot \rangle_\pi$; equivalently, $D^{1/2}PD^{-1/2}$ is symmetric for $D = \text{diag}(\pi)$.

⁵Appendix C collects the operator tools used throughout.

Disagreement evolves differently. Because d_t is mean-zero, it lies in the subspace on which P contracts. Equation (3) implies

$$\|Pd_t\|_\pi \leq \lambda^* \|d_t\|_\pi,$$

so disagreement decays geometrically, faster when connectivity is higher (smaller λ^*). Private signals inject idiosyncratic shocks each period, but bounded evidence keeps these injections uniformly controlled, yielding a uniform envelope on dispersion in fast-mixing regimes (Lemma 4).

Whether rapid dispersion destruction is beneficial depends on the communication technology. With continuous communication, the social message varies smoothly with beliefs, so eliminating disagreement does not prevent the mean from being transmitted. With finite vocabulary, ψ is step-valued: decoded content changes only through threshold crossings (Lemma 1). Threshold crossings require *boundary agents* near cutpoints, and the supply of such agents is controlled by dispersion (Lemma 2). Hence fast mixing can choke off further content production. If dispersion collapses while the mean is buffered inside an absorbing bin (Assumption 4), bounded evidence prevents escape, yielding lock-in and an information ceiling (Theorem 1). In modular networks, communities can instead form robust seeds that propagate through weak ties (Theorem 2), reversing the effect of connectivity.

4 Information ceiling

This section states the paper’s main results. The key friction is finite-vocabulary communication (Section 2.2), which makes the socially transmissible map $\psi = \Psi \circ Q$ piecewise constant. Hence, socially observable information moves only through discrete threshold crossings of the underlying belief statistic. We begin with a continuous benchmark in which the ceiling disappears.

4.1 Continuous benchmark

We begin with a benchmark in which communication is unconstrained: agents can transmit their real-valued *log-odds* belief statistic. Formally, replace the finite-alphabet technology by the identity map

$$\psi(\ell) = \ell \quad (\text{equivalently, } \mathcal{M} = \mathbb{R} \text{ and } y_{j,t} = \ell_{j,t}).$$

Then the social aggregate is simply $S_t = P\ell_t$ (in vector form), and the anchored recursion (1) becomes linear:

$$\ell_{t+1} = (1 - \alpha)\ell_t + \alpha(P\ell_t + \Delta_{t+1}) = \underbrace{((1 - \alpha)I + \alpha P)}_{=:A}\ell_t + \alpha\Delta_{t+1}. \quad (4)$$

In this benchmark, transmission is continuous. Unlike the finite-vocabulary case, there are no thresholds that require dispersion to cross: even if mixing collapses the population toward consensus, the social aggregate continues to move smoothly with beliefs. As a result, the mean belief drifts toward the truth and all agents are dragged along.

Let π denote the stationary distribution of P from Assumption 2. For any $x \in \mathbb{R}^n$ write $\bar{x} := \pi^\top x$ for its π -mean.

Lemma 3. *Suppose Assumptions 1–2 hold. Then for all $t \geq 1$,*

$$\mathbb{E}[\bar{\Delta}_t \mid \theta = 1] \geq m, \quad \mathbb{E}[\bar{\Delta}_t \mid \theta = 0] \leq -m,$$

and moreover

$$\bar{\ell}_t \rightarrow +\infty \text{ a.s. under } \theta = 1, \quad \bar{\ell}_t \rightarrow -\infty \text{ a.s. under } \theta = 0.$$

Lemma 3 shows that the population average log-odds learns the correct state almost surely. The next proposition shows that, under mixing, individual deviations from the mean remain uniformly bounded, so every agent learns.

Proposition 1. *Suppose Assumptions 1–2 hold. Then for every fixed n ,*

$$\ell_{i,t} \rightarrow +\infty \text{ a.s. under } \theta = 1, \quad \ell_{i,t} \rightarrow -\infty \text{ a.s. under } \theta = 0, \quad \forall i.$$

Equivalently, $a_{i,t} = \mu_{i,t} \rightarrow \theta$ a.s. for every i , and therefore $\lim_{t \rightarrow \infty} \text{MSE}_t^{(n)} = 0$.

Proposition 1 shows that anchored updating and fast mixing alone do not generate an information ceiling. The reason is that mixing directly shrinks cross-sectional disagreement. Let $d_t := \text{CO}(\ell_t)$ denote the centered belief profile. Under the anchored update, the averaging step acts through $A := \alpha I + (1 - \alpha)P$, so applying (3) gives the one-step contraction⁶

$$\|\text{CO}(Ax)\|_\pi \leq (\alpha + (1 - \alpha)\lambda^*) \|\text{CO}(x)\|_\pi \quad \forall x \in \mathbb{R}^n.$$

Thus, apart from new idiosyncratic private-evidence shocks, disagreement is multiplied each

⁶Note CO is a linear operator.

period by at most $\rho := \alpha + (1 - \alpha)\lambda^* \in (0, 1)$, and ρ is increasing in λ^* . Bounded evidence implies that the dispersion injected each period is uniformly controlled, so $D_t = \|d_t\|_\pi$ satisfies a recursion of the form

$$D_{t+1} \leq \rho D_t + C,$$

for a constant C depending only on primitives. Iterating yields a uniform envelope

$$\sup_t D_t \leq \frac{C}{1 - \rho},$$

which tightens as mixing becomes faster (smaller λ^* , hence smaller ρ). In the continuous benchmark, this homogenization is beneficial: the social message moves smoothly with beliefs, so faster mixing pools information more quickly and does not obstruct learning.

The learning failures identified later therefore arise from the *interaction* between mixing and finite vocabulary: when ψ is step-valued, dispersion is needed to generate threshold crossings, and fast mixing can eliminate the boundary agents that would otherwise trigger further content production.

4.2 Homogenization traps on fast-mixing networks

We now return to a finite vocabulary. The central tension is a *bandwidth–mixing mismatch*. Fast mixing contracts cross-sectional dispersion in beliefs, while coarse communication requires dispersion for threshold crossings (Lemma 1). If dispersion collapses before the social layer reaches a cutpoint, the public message profile freezes and society stops producing new public content.

Throughout this subsection we maintain the primitives and outcome notions in Sections 2–2.2 and the $L^2(\pi)$ geometry in Section 3. Recall that $\psi = \Psi \circ Q$ is piecewise constant with reconstructions $\{r_k\}_{k=1}^K$ and cutpoints $\{-\infty = b_0 < b_1 < \dots < b_K = +\infty\}$. Let

$$\underline{r} := \min_k r_k, \quad \bar{r} := \max_k r_k, \quad R := \bar{r} - \underline{r}, \quad \pi_{\min} := \min_i \pi_i.$$

Assumption 3 (Fast mixing). *Along the society sequence, there exists $\bar{\lambda} \in (0, 1)$ such that $\lambda^* \leq \bar{\lambda}$ uniformly in n .*

Assumption 4 (Absorbing bin). *There exists an index $\hat{k} \in \{1, \dots, K\}$ such that*

$$[r_{\hat{k}} - \bar{\Delta}, r_{\hat{k}} + \bar{\Delta}] \subset (b_{\hat{k}-1}, b_{\hat{k}}).$$

Equivalently, defining the absorption margin

$$\delta_{\text{abs}} := \min\{r_{\hat{k}} - \bar{\Delta} - b_{\hat{k}-1}, b_{\hat{k}} - (r_{\hat{k}} + \bar{\Delta})\} > 0,$$

we have

$$[r_{\hat{k}} - \bar{\Delta}, r_{\hat{k}} + \bar{\Delta}] \subset (b_{\hat{k}-1} + \delta_{\text{abs}}, b_{\hat{k}} - \delta_{\text{abs}}).$$

Comment. Appendix B.3 provides convenient sufficient conditions for Assumption 4 (and clarifies when it fails under very fine vocabularies).

Assumption 4 is an irreversibility condition: if all agents transmit \hat{k} , then $S_{i,t} = r_{\hat{k}}$ for all i , and bounded private innovations cannot push any agent out of bin \hat{k} at the next update. The rest of the section shows that on fast-mixing networks, the process enters such an absorbing configuration with non-vanishing probability.

4.2.1 Absorption

We start with a *deterministic* lock-in statement. If, at some time t_0 , the entire belief profile fits inside a single absorbing bin, then coarse communication freezes the social signal and bounded evidence prevents any subsequent escape. The only nontrivial step is upgrading an $L^2(\pi)$ dispersion bound into a uniform (agent-by-agent) guarantee.

Proposition 2. *Fix n . Maintain Assumptions 1 and 4 for the bin index \hat{k} . Let $\pi_{\min} := \min_{i \in N} \pi_i$ and fix $\eta > 0$. If at some time t_0 :*

- (i) (Buffered mean) $\bar{\ell}_{t_0} \in (b_{\hat{k}-1} + 2\eta, b_{\hat{k}} - 2\eta)$;
- (ii) (π_{\min} -scale dispersion) $\text{Var}_{\pi}(\ell_{t_0}) \leq \pi_{\min}\eta^2$,

then $Q(\ell_{i,t}) \equiv \hat{k}$ for all i and all $t \geq t_0$.

Condition (i) ensures that the *mean* belief is deep inside the bin. Condition (ii) ensures that the *entire population* remains close enough to the mean to stay inside that same bin.

In large societies, this deterministic condition (ii) becomes stringent. For example, under uniform weights with $\pi_{\min} = 1/n$, so Condition (ii) requires $\text{Var}_{\pi}(\ell_{t_0}) \leq \eta^2/n$. This implies that for strict containment to hold deterministically, cross-sectional variance must vanish as $n \rightarrow \infty$. Since this is too restrictive for a general large- n sufficient condition, this motivates the next step: rather than trapping *everyone* at once, we will trap a *large cohesive core* (a set carrying most of the π -mass) and show that this is already enough to generate persistent lock-in and, ultimately, an information ceiling.

4.2.2 Probabilistic lock-in of a cohesive core

The key observation is that coarse communication does not require unanimity to create persistence. If a large subset of agents lies in the same bin and places only limited weight on outsiders, then outsiders may fluctuate, but they cannot move the core across thresholds. In this sense, cohesiveness turns a statistically concentrated set into an *absorbing coalition*.

Fix a time t_0 and tolerance $\eta > 0$. Define the *candidate core*

$$C_0(t_0, \eta) := \left\{ i \in N : |\ell_{i,t_0} - \bar{\ell}_{t_0}| < \eta \right\}. \quad (5)$$

When dispersion is small, Lemma 2 implies that $C_0(t_0, \eta)$ carries large π -mass.

Definition 2 (ζ -cohesive subset). Let $C \subseteq N$ and $\zeta \in (0, 1)$. We say C is ζ -cohesive if for every $i \in C$,

$$\sum_{j \notin C} P_{ij} \leq \zeta.$$

Definition 3 (ζ -robustness of a bin for the core). Fix $\hat{k} \in \{1, \dots, K\}$ and $\zeta \in (0, 1)$. Let $\underline{r} = \min_k r_k$ and $\bar{r} = \max_k r_k$. Define the *core shock interval*

$$\mathcal{A}_{\hat{k}}^{\text{core}}(\zeta) := \left[(1 - \zeta)r_{\hat{k}} + \zeta\underline{r} - \bar{\Delta}, (1 - \zeta)r_{\hat{k}} + \zeta\bar{r} + \bar{\Delta} \right]. \quad (6)$$

We say bin \hat{k} is ζ -robust for the core if $\mathcal{A}_{\hat{k}}^{\text{core}}(\zeta) \subset (b_{\hat{k}-1}, b_{\hat{k}})$.

The interval $\mathcal{A}_{\hat{k}}^{\text{core}}(\zeta)$ represents the worst-case set of posterior beliefs an agent in a ζ -cohesive group could possibly hold next period. It is constructed by assuming three things happen simultaneously:

- Internal Consensus: The agent hears the message $r_{\hat{k}}$ from all their friends inside the cohesive group (weight $1 - \zeta$).
- External Hostility: The agent hears the most contradictory possible message (either \underline{r} or \bar{r}) from all outsiders (weight ζ).
- Private Contradiction: The agent receives the strongest possible opposing private signal (pushing the belief by $\pm \bar{\Delta}$).

If this worst-case interval fits entirely inside the bin boundaries $(b_{\hat{k}-1}, b_{\hat{k}})$, then the bin is a trap (or an invariant set) for the cohesive group. Even if the outside world screams the opposite message and private evidence contradicts the consensus, no agent in the group can accumulate enough evidence to cross the threshold and leave the bin.

Proposition 3. Fix n and $\hat{k} \in \{1, \dots, K\}$. Maintain Assumption 1. Let $\eta > 0$ and suppose that at time t_0 :

- (i) (Buffered mean) $\bar{\ell}_{t_0} \in (b_{\hat{k}-1} + 2\eta, b_{\hat{k}} - 2\eta)$;
- (ii) (Constant-scale variance) $\text{Var}_\pi(\ell_{t_0}) \leq \varepsilon \eta^2$ for some $\varepsilon \in (0, 1)$.

Let $C_0 := C_0(t_0, \eta)$ be as in (5). Fix $\zeta \in (0, 1)$ and assume:

- (iii) (Cohesive core) C_0 contains a subset $C \subseteq C_0$ that is ζ -cohesive (Definition 2);
- (iv) (Robust bin) bin \hat{k} is ζ -robust for the core (Definition 3).

Then:

- (a) $\pi(C_0) \geq 1 - \varepsilon$ and $Q(\ell_{i,t_0}) = \hat{k}$ for all $i \in C_0$;
- (b) for all $t \geq t_0$ and all $i \in C$, $\ell_{i,t} \in (b_{\hat{k}-1}, b_{\hat{k}})$ and hence $Q(\ell_{i,t}) = \hat{k}$.

Proposition 3 significantly relaxes the requirements for lock-in by abandoning the need for uniform control over every agent. Specifically, Condition (ii) imposes a variance bound that is independent of π_{\min} , thereby avoiding the requirement that dispersion vanishes in large networks (as was necessary in Proposition 2). Condition (iii) ensures this mass contains a subset that is structurally insulated (ζ -cohesive); and Condition (iv) guarantees that this insulation is sufficient to withstand worst-case shocks. The result is a permanent *partial* lock-in: the core's message profile freezes forever, even if agents outside the core continue to cross thresholds.

4.2.3 Large cores under fast mixing: buffering the mean buffers many agents

The absorption logic above is deterministic: if beliefs (or a sufficiently cohesive set of beliefs) lie *strictly inside* an absorbing bin, then coarse communication freezes decoded content and bounded private evidence cannot push the system across a cutpoint. To turn this geometry into a *probabilistic* trap result, we need to show that fast mixing makes it easy for a *large core* of agents to fall into the same bin.

The argument has three steps, which we will reuse throughout the information ceiling:

- (S1) *Fast mixing controls dispersion.* Under finite vocabulary, decoded messages are bounded and private evidence is bounded, so disagreement cannot explode. Moreover, when P mixes quickly (Assumption 3), averaging compresses disagreement, yielding a uniform (and mixing-dependent) upper bound on dispersion (Lemma 4).
- (S2) *A buffered mean implies a large core.* If dispersion is uniformly bounded and the mean belief $\bar{\ell}_t$ lies safely away from bin boundaries, then almost all agents must lie in the

same bin. This is a quantitative ‘‘mean buffers many’’ statement: if many agents were in other bins, dispersion would have to be large (Lemma 5).

- (S3) *Entry with non-vanishing probability.* A mild one-step alignment condition guarantees that the mean enters the buffered region at the first update with probability bounded away from zero (Lemma 6). Combining (S2) with the deterministic absorption logic yields lock-in of a large cohesive core.

We now formalize (S1)–(S3). Recall $D_t := \|\text{CO}(\ell_t)\|_\pi$ denote π -dispersion (Definition 1).

Lemma 4. *Maintain Assumptions 1–3. Then for all $t \geq 0$,*

$$D_t \leq (1 - \alpha)^t D_0 + (1 - (1 - \alpha)^t) D_{\sup}, \quad D_{\sup} := \bar{\lambda}R + 2\bar{\Delta}.$$

In particular, if priors are common so that $D_0 = 0$, then $D_t \leq D_{\sup}$ for all t .

The constant D_{\sup} is a mixing-dependent upper bound on disagreement. Faster mixing (smaller $\bar{\lambda}$) tightens this envelope by compressing the centered component each period; bounded evidence prevents the idiosyncratic shocks from undoing this compression.

Fix $\varepsilon \in (0, 1)$ and define the buffering radius

$$\eta := \frac{D_{\sup}}{\sqrt{\varepsilon}}.$$

Because D_{\sup} decreases with $\bar{\lambda}$, faster mixing makes η smaller.

Lemma 5. *Fix $\varepsilon \in (0, 1)$ and let $\eta := D_{\sup}/\sqrt{\varepsilon}$. Maintain Assumptions 1–3, and suppose priors are common so that $D_0 = 0$. Then for every t and every bin index k ,*

$$\bar{\ell}_t \in (b_{k-1} + 2\eta, b_k - 2\eta) \implies \pi(\{i : Q(\ell_{i,t}) = k\}) \geq 1 - \varepsilon.$$

This lemma translates the bounded dispersion implied by fast mixing into a statement about message consensus. Since the fast-mixing assumption keeps the cross-sectional variance of beliefs uniformly bounded, the population is statistically concentrated near the mean. The condition that the mean lies at least 2η from the bin edges ensures that this concentration occurs strictly within the bin boundaries. As a result, a proportion of at least $1 - \varepsilon$ of the population must fall inside the bin, implying they all generate the same quantized message k .

Lemma 5 is conditional on the mean being buffered. The next lemma provides a verifiable one-step sufficient condition under which this event occurs with probability bounded away from zero.

Define the $t = 0$ decoded level

$$y_0 := \psi(\ell_0) \in [\underline{r}, \bar{r}], \quad \bar{\Delta}_1 := \sum_{i=1}^n \pi_i \Delta_{i,1}.$$

Lemma 6. *Maintain Assumptions 1 and 4. Fix $\varepsilon \in (0, 1)$ and set $\eta := D_{\text{sup}}/\sqrt{\varepsilon}$ as above. Fix $\theta \in \{0, 1\}$ and write $\mu_\theta := \mathbb{E}[\bar{\Delta}_1 | \theta]$. Assume*

(i) *For some $\rho > 0$,*

$$\mathbb{E}[\bar{\ell}_1 | \theta] = (1 - \alpha)\ell_0 + \alpha(y_0 + \mu_\theta) \in (b_{\hat{k}-1} + 2\eta + \rho, b_{\hat{k}} - 2\eta - \rho) \quad (7)$$

(ii) *there exists $\kappa_\pi < \infty$ (independent of n) such that $\pi_{\max} := \max_i \pi_i \leq \frac{\kappa_\pi}{n}$.*

Then, there exist $c_{\text{entry}} \in (0, 1)$ and n_0 (depending only on primitives, ρ , κ_π) such that for all $n \geq n_0$,

$$\Pr(\bar{\ell}_1 \in (b_{\hat{k}-1} + 2\eta, b_{\hat{k}} - 2\eta) \mid \theta) \geq c_{\text{entry}}.$$

Condition (7) ensures that the *expected* mean update at $t = 1$ lands strictly inside the target bin with safety margin ρ . The second non-dominant agent condition guarantees the realized mean stays close to its expectation. Consequently, the realized mean enters the buffered region with probability bounded away from zero. Lemma 5 then implies that this placement of the mean forces a large $(1 - \varepsilon)$ -mass core into the same quantization cell, triggering the deterministic absorption mechanism and establishing permanent lock-in.

4.2.4 Homogenization trap and the information ceiling

The preceding lemmas isolate the race condition behind the ceiling. Under finite vocabulary, the social layer is *event-driven*: decoded content changes only when some agent reaches a cutpoint. Fast mixing is harmful precisely because it suppresses the cross-sectional dispersion needed for such threshold crossings. Lemma 4 shows that, when mixing is sufficiently fast, disagreement admits a *uniform envelope*. Consequently, if the mean log-odds $\bar{\ell}_t$ becomes buffered inside a bin, then many agents must lie in that same bin (Lemma 5), leaving too few “boundary agents” to trigger further message changes. The remaining steps are: (i) the mean enters a buffered region with non-vanishing probability already at $t = 1$ (Lemma 6), and (ii) once a large set is inside a ζ -robust absorbing bin, cohesiveness makes it self-confirmed forever (Proposition 3). Putting these pieces together yields a homogenization trap and, hence, an information ceiling.

We assume the scaling conditions used to convert π -mass bounds into cardinality bounds and to control leakage from large sets. Also, recall $\eta = \frac{D_{\sup}}{\sqrt{\varepsilon}}$.

Assumption 5 (Diffuse stationary weights and diffuse rows). *Along the society sequence there exist constants $\kappa_\pi, \kappa_P \geq 1$ (independent of n) such that for all n ,*

$$\frac{1}{\kappa_\pi n} \leq \pi_i \leq \frac{\kappa_\pi}{n} \quad \forall i, \quad \max_{i,j} P_{ij} \leq \frac{\kappa_P}{n}.$$

Comment. Appendix B.5 records canonical constructions under which the diffuse-weight bounds are satisfied.

Theorem 1. *Fix $K \geq 2$ and maintain Assumptions 1–5. Fix $\varepsilon \in (0, 1)$ with $\varepsilon < (\kappa_\pi \kappa_P)^{-1}$ and define $\zeta := \kappa_\pi \kappa_P \varepsilon \in (0, 1)$, where D_{\sup} is as in Lemma 4. Suppose the bin \hat{k} is ζ -robust in the sense that*

$$\mathcal{A}_{\hat{k}}^{\text{core}}(\zeta) \subset (b_{\hat{k}-1}, b_{\hat{k}}),$$

where $\mathcal{A}_{\hat{k}}^{\text{core}}(\zeta)$ is defined in (6). Finally, assume the entry gap (7) holds for some $\rho > 0$ under $\theta = 1$.

Then, there exist constants $c \in (0, 1)$ and n_0 (depending only on primitives and $\bar{\lambda}, \rho, \kappa_\pi, \kappa_P$) such that for all $n \geq n_0$,

$$\Pr\left(\exists C \subseteq N^{(n)} \text{ with } |C| \geq (1 - \kappa_\pi \varepsilon)n \text{ s.t. } Q(\ell_{i,t}) = \hat{k} \ \forall i \in C, \ \forall t \geq 1 \mid \theta = 1\right) \geq c.$$

Moreover, an information ceiling obtains under $\theta = 1$: there exists a constant $c_{\text{ceiling}} > 0$ (independent of n and t) such that for all $n \geq n_0$,

$$\liminf_{t \rightarrow \infty} \text{MSE}_t^{(n)} \geq c_{\text{ceiling}}.$$

Note the theorem states based on $\theta = 1$, but the same logic can be applied to $\theta = 0$. The theorem formalizes the mismatch between continuous aggregation and threshold transmission. Fast mixing implies a uniform envelope on dispersion (Lemma 4); hence, once the mean enters a buffered region of a bin, a $1 - \varepsilon$ fraction of agents must enter the same bin simultaneously (Lemma 5). This eliminates the boundary agents needed for further threshold crossings, so the social message profile freezes. If the bin is robust to small leakage from outside the core, absorption then makes the lock-in self-confirming, generating a persistent error floor despite ongoing private information.

5 Possibility of learning

5.1 Segregation enables learning via robust seeds

Proposition 3 highlighted a key asymmetry of finite-vocabulary communication: once a cohesive group's public content lies strictly inside a quantization bin, bounded private evidence cannot force a threshold crossing, so the group becomes *absorbed*. On fast-mixing networks this absorption is socially harmful because it freezes a large core (Section 4.2). Under segregation, however, absorption can become *constructive*: a sufficiently insulated community can incubate a stable local consensus (a *seed*), and weak ties can then replicate that message across communities. Learning follows when the state-aligned seed forms and spreads before any competing seed locks in elsewhere.

Fix cutpoints $-\infty = b_0 < b_1 < \dots < b_{K-1} < b_K = +\infty$ and reconstructions $r_k := \Psi(k)$. Let $\underline{r} := \min_k r_k$ and $\bar{r} := \max_k r_k$. To avoid confusion with other uses of ε , denote by $\varepsilon_{\text{fed}} \in (0, 1)$ the weak-tie bound used in this subsection.

Following (6), define for each bin k the robust band

$$\mathcal{A}_k^{\text{rob}} := \left[(1 - \varepsilon_{\text{fed}})r_k + \varepsilon_{\text{fed}}\underline{r} - \bar{\Delta}, (1 - \varepsilon_{\text{fed}})r_k + \varepsilon_{\text{fed}}\bar{r} + \bar{\Delta} \right]. \quad (8)$$

Intuitively, $\mathcal{A}_k^{\text{rob}}$ is the range of social terms consistent with (i) at most ε_{fed} weight on out-of-community messages, and (ii) bounded private-evidence increments of magnitude at most $\bar{\Delta}$ (cf. Assumption 1).

We work with a community partition $N^{(n)} = G_1^{(n)} \dot{\cup} \dots \dot{\cup} G_{M(n)}^{(n)}$, where $\dot{\cup}$ denotes sets are partitioned.

Assumption 6 (Structural assumptions: federation and connectivity). *For each n , the following hold.*

(S1) (Federation / weak ties.) *Each agent assigns total out-of-community weight at most ε_{fed} , i.e., for every community $g \leq M(n)$ and every agent $i \in G_g^{(n)}$, $\sum_{j \in G_g^{(n)}} P_{ij} \geq 1 - \varepsilon_{\text{fed}}$.*

(S2) (Many communities.) *Along the sequence $n \rightarrow \infty$, $M(n) \rightarrow \infty$.*

(S3) (Community-level influence and connectivity.) *For communities $g \neq h$, define the minimal row-wise influence weight*

$$W_{h \leftarrow g} := \min_{i \in G_h} \sum_{j \in G_g} P_{ij}.$$

There exists $\omega > 0$ such that the directed graph on communities with edge condition $g \rightarrow h \iff W_{h \leftarrow g} \geq \omega$ is strongly connected for every n . Let $\mathcal{C}^{(n)}$ denote this directed community graph.

(S4) (Bounded community diameter.) There exists $\bar{D}_{\mathcal{C}} < \infty$ such that the directed diameter $D_{\mathcal{C}}(n)$ of $\mathcal{C}^{(n)}$ satisfies $D_{\mathcal{C}}(n) \leq \bar{D}_{\mathcal{C}}$ for all n .

Assumption 7 (Informational assumptions: robust bins and incubation). *The following hold.*

(I1) (Robust consensus bins.) There exist indices $k^A, k^B \in \{1, \dots, K\}$ such that

$$\mathcal{A}_{k^A}^{\text{rob}} \subset (b_{k^A-1}, b_{k^A}), \quad \mathcal{A}_{k^B}^{\text{rob}} \subset (b_{k^B-1}, b_{k^B}), \quad r_{k^A} > r_{k^B}.$$

(I2) (Nondegenerate evidence blocks, with asymmetry.) There exist $\delta > 0$ and constants $p_+, p_- \in (0, 1)$ such that for all agents i and times t ,

$$\Pr(\Delta_{i,t} \geq \delta \mid \theta = 1) \geq p_+, \quad \Pr(\Delta_{i,t} \leq -\delta \mid \theta = 0) \geq p_+,$$

and wrong-direction tails are controlled:

$$\Pr(\Delta_{i,t} \leq -\delta \mid \theta = 1) \leq p_-, \quad \Pr(\Delta_{i,t} \geq \delta \mid \theta = 0) \leq p_-.$$

(I3) (Incubation feasibility.) There exists an integer $L \geq 1$ such that for every n , every time t , and all communities g, h :

(a) (Within-community incubation.) If $\Delta_{i,t+s} \geq \delta$ for all $i \in G_g$ and all $s = 1, \dots, L$, then G_g is a k^A -seed at time $t + L$. If $\Delta_{i,t+s} \leq -\delta$ for all $i \in G_g$ and all $s = 1, \dots, L$, then G_g is a k^B -seed at time $t + L$.

(b) (Edge incubation.) If G_g is a k^A -seed at time t , $W_{h \leftarrow g} \geq \omega$, and $\Delta_{i,t+s} \geq \delta$ for all $i \in G_h$ and all $s = 1, \dots, L$, then G_h is a k^A -seed at time $t + L$. If G_g is a k^B -seed at time t , $W_{h \leftarrow g} \geq \omega$, and $\Delta_{i,t+s} \leq -\delta$ for all $i \in G_h$ and all $s = 1, \dots, L$, then G_h is a k^B -seed at time $t + L$.

Given k^A, k^B from (I1), call a community G_g a k -seed at time t if $Q(\ell_{i,t}) = k$ for all $i \in G_g$. We refer to a k^A -seed as an A -seed and a k^B -seed as a B -seed.

Lemma 7 (Seed absorption). *Under Assumptions 6(S1) and 7(I1), if G_g is a k -seed at time t for $k \in \{k^A, k^B\}$, then it remains a k -seed for all $t' \geq t$.*

The next theorem separates two statements. Part (i) establishes *formation*: extreme local evidence blocks can create absorbed local consensus. Part (ii) establishes *propagation*: once

a seed exists, it can replicate along weak ties *provided* that communities encountered along a propagation path do not become absorbed into a conflicting bin first.

Theorem 2 (Robust seeds and conditional propagation). *Suppose Assumption 1 and Assumptions 6–7 hold. Then:*

- (i) (Formation.) *Conditional on $\theta = 1$, the probability that at least one community becomes an A-seed at some finite time converges to 1 as $n \rightarrow \infty$.*
- (ii) (Propagation along receptive paths.) *Conditional on $\theta = 1$, fix a time t at which some community g_0 is an A-seed. Let $g_0 \rightarrow g_1 \rightarrow \dots \rightarrow g_m$ be a directed path in $\mathcal{C}^{(n)}$. If for each $\ell = 1, \dots, m$ the community g_ℓ does not become a B-seed before it becomes an A-seed, then g_m eventually becomes an A-seed almost surely.*

Consider a counterfactual scenario where bin k^A is state-aligned when $\theta = 1$. Then, Theorem 2 shows that segregation removes the *structural* barrier to learning: unlike in fast-mixing networks, absorbed local consensus can arise within communities and can be transmitted across the federation along weak ties, rather than forcing the society into a homogenization trap.

The theorem deliberately allows the possibility of competing seeds: an incorrect seed could form and become absorbing in some community before the state-aligned seed reaches it. Appendix C formalizes a race condition under which this is unlikely. The key force is that, under asymmetric evidence blocks, an L -period good-direction incubation event in a community is exponentially more likely (in community size) than an L -period wrong-direction incubation event. Under a mild thickness condition on community sizes, the state-aligned seed forms and propagates through the federation before any competing seed stabilizes, yielding state-aligned learning with probability approaching one.

5.2 Interior optimal connectivity: non-monotone learning in mixing

Sections 4.2 and 5.1 identify two opposing roles of connectivity under finite-vocabulary communication. On the one hand, stronger inter-community links help *propagate* any locally stable message: once a robust *seed* forms in some community, weak ties can replicate it across the federation (Section 5.1). On the other hand, stronger connectivity also contracts cross-sectional dispersion, making quantization threshold crossings—and hence the production of new public content—rare (Section 4.2). This subsection combines the two forces into a single comparative-statics conclusion: learning is *non-monotone* in connectivity, and long-run accuracy is maximized at an *interior* regime (“Goldilocks connectivity”).

Recall that the public message is $m_{i,t} := Q(\ell_{i,t}) \in \{1, \dots, K\}$. Let $k^1 := k^A$ and $k^0 := k^B$

denote the state-aligned seed bins from Assumption ???. For $\theta \in \{0, 1\}$ define the state-aligned absorption event

$$\mathcal{L}_\theta^{(n)} := \left\{ \exists T < \infty : m_{i,t} = k^\theta \quad \forall i \in N^{(n)} \quad \forall t \geq T \right\}. \quad (9)$$

Thus, $\mathcal{L}_\theta^{(n)}$ is the strongest learning benchmark: eventually the entire society sends the state-aligned message forever.

Theorem 3 compares three connectivity regimes. At the high-connectivity end, fast mixing generates homogenization traps and prevents state-aligned absorption. At the low-connectivity end, learning fails because a correct seed cannot propagate across communities, or because propagation is too slow relative to the opportunity for incorrect seeds to form. Between these extremes, a segregated federation can both incubate and replicate the correct seed with high probability (Corollary ??).

Theorem 3 (Goldilocks connectivity: non-monotone learning). *Fix $K \geq 2$. Consider three society sequences with the following connectivity regimes:*

- (H) High connectivity (fast mixing). *For each $\theta \in \{0, 1\}$ under consideration, the assumptions of Theorem 1 hold conditional on θ , yielding a ceiling bin $\hat{k}(\theta)$ with $\hat{k}(\theta) \neq k^\theta$.*
- (M) Intermediate connectivity (segregated federation). *The assumptions of Corollary ?? hold.*
- (L) Low connectivity (too little). *For the relevant state(s), at least one of the following holds:*
 - (L1) Propagation infeasible: *the hypotheses of Proposition 8 hold (for the state-aligned bin k^θ).*
 - (L2) Propagation too slow: *the hypotheses of Proposition 9 hold.*

Then, for each relevant θ ,

- (i) (Too much.) *Under (H),*

$$\Pr(\mathcal{L}_\theta^{(n)} | \theta) \not\rightarrow 1.$$

In particular, $\Pr(\mathcal{L}_\theta^{(n)} | \theta) \leq 1 - c$ for some $c \in (0, 1)$ and all large n .

- (ii) (Interior.) *Under (M),*

$$\Pr(\mathcal{L}_\theta^{(n)} | \theta) \rightarrow 1 \quad \text{for both } \theta \in \{0, 1\}.$$

(iii) (Too little.) Under (L), for the relevant state(s),

$$\Pr(\mathcal{L}_\theta^{(n)} \mid \theta) \rightarrow 0.$$

Consequently, long-run learning performance is maximized at an interior level of connectivity: both extreme sparsity and extreme mixing are detrimental under finite-vocabulary communication.

Under finite vocabulary, stronger connectivity has two opposing effects: it speeds up replication of any locally stable seed, but it also destroys dispersion and thereby shuts down threshold crossings that create new public content. Goldilocks connectivity balances these two forces.

A Appendix A: Proofs

A.1 Proof for Proofs of Main-Text Results

Proof of Lemma 1

If $Q(\ell_{j,t}) = Q(\ell_{j,t-1})$ for all j , then $\psi(\ell_{j,t}) = \psi(\ell_{j,t-1})$ for all j because $\psi = \Psi \circ Q$. Hence

$$S_{i,t} - S_{i,t-1} = \sum_{j=1}^n P_{ij} (\psi(\ell_{j,t}) - \psi(\ell_{j,t-1})) = 0$$

for every i .

■

Proof of Lemma 2

Let $S := \{i : |v_i - \bar{v}| \geq \eta\}$. Then

$$\text{Var}_\pi(v) = \sum_{i \in N} \pi_i (v_i - \bar{v})^2 \geq \sum_{i \in S} \pi_i (v_i - \bar{v})^2 \geq \eta^2 \sum_{i \in S} \pi_i = \eta^2 \pi(S),$$

which implies $\pi(S) \leq \text{Var}_\pi(v)/\eta^2$. The second inequality follows because $\{i : |v_i - \bar{v}| < \eta\} = N \setminus S$ and π is a probability measure.

■

Proof of Lemma 3

By definition, $\bar{\Delta}_t = \sum_i \pi_i \Delta_{i,t}$. Under $\theta = 1$,

$$\mathbb{E}[\bar{\Delta}_t \mid \theta = 1] = \sum_{i=1}^n \pi_i \mathbb{E}[\Delta_{i,t} \mid \theta = 1] \geq \sum_{i=1}^n \pi_i m = m,$$

and similarly under $\theta = 0$.

Next, take π^\top on both sides of (4). Using $\pi^\top P = \pi^\top$,

$$\bar{\ell}_{t+1} = \pi^\top \ell_{t+1} = \pi^\top \ell_t + \alpha \pi^\top \Delta_{t+1} = \bar{\ell}_t + \alpha \bar{\Delta}_{t+1}.$$

Iterating gives $\bar{\ell}_t = \bar{\ell}_0 + \alpha \sum_{s=1}^t \bar{\Delta}_s$.

Fix $\theta = 1$ and write $d_s := \mathbb{E}[\bar{\Delta}_s \mid \theta = 1] \geq m$. Let $Y_s := \bar{\Delta}_s - d_s$. By Assumption 1(ii), $|\bar{\Delta}_s| \leq \bar{\Delta}$ a.s., so $|Y_s| \leq 2\bar{\Delta}$ a.s. By Assumption 1(i), $\{Y_s\}_{s \geq 1}$ is independent over time with mean zero. Therefore, by Kolmogorov's strong law for independent uniformly bounded

variables,

$$\frac{1}{t} \sum_{s=1}^t Y_s \rightarrow 0 \quad \text{a.s.}$$

Hence

$$\liminf_{t \rightarrow \infty} \frac{\bar{\ell}_t}{t} = \alpha \liminf_{t \rightarrow \infty} \left(\frac{1}{t} \sum_{s=1}^t d_s + \frac{1}{t} \sum_{s=1}^t Y_s \right) \geq \alpha m > 0 \quad \text{a.s.,}$$

which implies $\bar{\ell}_t \rightarrow +\infty$ a.s. The case $\theta = 0$ is symmetric.

■

Proof of Proposition 1

Fix $\theta = 1$. By Lemma 3, $\bar{\ell}_t \rightarrow +\infty$ a.s. It remains to show that individual deviations from the mean are uniformly bounded.

Let $u_t := \text{CO}(\ell_t) = \ell_t - \bar{\ell}_t \mathbf{1}$ be the centered belief profile. Since $\overline{Px} = \bar{x}$ under stationarity, centering commutes with P and therefore with $A = (1 - \alpha)I + \alpha P$: $\text{CO}(Ax) = A\text{CO}(x)$. Applying CO to (4) yields

$$u_{t+1} = Au_t + \alpha \text{CO}(\Delta_{t+1}).$$

Under Assumption 2, P contracts centered vectors in $L^2(\pi)$ at rate λ^* , so for any centered v ,

$$\|Av\|_\pi = \|(1 - \alpha)v + \alpha Pv\|_\pi \leq (1 - \alpha)\|v\|_\pi + \alpha\|Pv\|_\pi \leq ((1 - \alpha) + \alpha\lambda^*)\|v\|_\pi.$$

Define the contraction bound $\tilde{\lambda} := (1 - \alpha) + \alpha\lambda^* \in (0, 1)$. Then

$$\|u_{t+1}\|_\pi \leq \tilde{\lambda}\|u_t\|_\pi + \alpha\|\text{CO}(\Delta_{t+1})\|_\pi.$$

By bounded evidence, $\|\Delta_{t+1}\|_\pi \leq \bar{\Delta}$ and $|\bar{\Delta}_{t+1}| \leq \bar{\Delta}$, hence $\|\text{CO}(\Delta_{t+1})\|_\pi \leq 2\bar{\Delta}$ a.s. Iterating the inequality gives

$$\sup_{t \geq 0} \|u_t\|_\pi \leq \|u_0\|_\pi + \sum_{s=0}^{\infty} \tilde{\lambda}^s \cdot 2\alpha\bar{\Delta} = \|u_0\|_\pi + \frac{2\alpha\bar{\Delta}}{1 - \tilde{\lambda}} < \infty \quad \text{a.s.}$$

Therefore, for each i , $\pi_i u_{i,t}^2 \leq \sum_j \pi_j u_{j,t}^2 = \|u_t\|_\pi^2$, so $|u_{i,t}| \leq \|u_t\|_\pi / \sqrt{\pi_i}$ is uniformly bounded in t . In particular, for each fixed i ,

$$|\ell_{i,t} - \bar{\ell}_t| = |u_{i,t}| \leq \frac{\|u_t\|_\pi}{\sqrt{\pi_i}} \leq \frac{1}{\sqrt{\pi_i}} \left(\|u_0\|_\pi + \frac{2\alpha\bar{\Delta}}{1 - \tilde{\lambda}} \right) < \infty \quad \text{a.s. for all } t.$$

Since $\ell_{i,t} = \bar{\ell}_t + u_{i,t}$ and $\bar{\ell}_t \rightarrow +\infty$ (Lemma 3), it follows that $\ell_{i,t} \rightarrow +\infty$ a.s. for every i .

The case $\theta = 0$ is symmetric.

Finally, $a_{i,t} = \mu_{i,t} = (1 + e^{-\ell_{i,t}})^{-1} \rightarrow 1$ a.s. under $\theta = 1$ (and $\rightarrow 0$ under $\theta = 0$). Because $(a_{i,t} - \theta)^2 \leq 1$, bounded convergence yields $\mathbb{E}[(a_{i,t} - \theta)^2] \rightarrow 0$ for each i , and hence $\text{MSE}_t^{(n)} \rightarrow 0$.

■

Proof of Proposition 2

Note for each i and t ,

$$\text{Var}_\pi(\ell_{i,t}) = \sum_j \pi_j (\ell_{i,t} - \bar{\ell}_t)^2 \geq \pi_i (\ell_{i,t} - \bar{\ell}_t)^2 \geq \pi_{\min} (\ell_{i,t} - \bar{\ell}_t)^2,$$

implying

$$\max_{i \in N} |\ell_{i,t_0} - \bar{\ell}_{t_0}| \leq \frac{\|\text{CO}(\ell_{t_0})\|_\pi}{\sqrt{\pi_{\min}}} = \frac{\sqrt{\text{Var}_\pi(\ell_{t_0})}}{\sqrt{\pi_{\min}}} \leq \eta,$$

where the last inequality uses the condition (ii). Combining with the condition (i) yields, for every i ,

$$\ell_{i,t_0} \in (\bar{\ell}_{t_0} - \eta, \bar{\ell}_{t_0} + \eta) \subset (b_{\hat{k}-1}, b_{\hat{k}}),$$

hence $Q(\ell_{i,t_0}) = \hat{k}$ for all i and therefore $\psi(\ell_{i,t_0}) = r_{\hat{k}}$.

Fix any agent i . Since all agents send the same decoded content $r_{\hat{k}}$ at time t_0 ,

$$S_{i,t_0} = \sum_j P_{ij} \psi(\ell_{j,t_0}) = \sum_j P_{ij} r_{\hat{k}} = r_{\hat{k}}.$$

By bounded evidence (Assumption 1(ii)),

$$S_{i,t_0} + \Delta_{i,t_0+1} \in [r_{\hat{k}} - \bar{\Delta}, r_{\hat{k}} + \bar{\Delta}].$$

Assumption 4 implies this interval is contained in $(b_{\hat{k}-1}, b_{\hat{k}})$. Since also $\ell_{i,t_0} \in (b_{\hat{k}-1}, b_{\hat{k}})$, the anchored update

$$\ell_{i,t_0+1} = (1 - \alpha) \ell_{i,t_0} + \alpha (S_{i,t_0} + \Delta_{i,t_0+1})$$

is a convex combination of two points in $(b_{\hat{k}-1}, b_{\hat{k}})$, hence $\ell_{i,t_0+1} \in (b_{\hat{k}-1}, b_{\hat{k}})$ and $Q(\ell_{i,t_0+1}) = \hat{k}$. Iterating yields $Q(\ell_{i,t}) = \hat{k}$ for all $t \geq t_0$ and all i .

■

Proof of Proposition 3

(a) By Lemma 2 applied to $v = \ell_{t_0}$,

$$\pi(C_0^c) = \pi(\{i : |\ell_{i,t_0} - \bar{\ell}_{t_0}| \geq \eta\}) \leq \frac{\text{Var}_\pi(\ell_{t_0})}{\eta^2} \leq \varepsilon,$$

so $\pi(C_0) \geq 1 - \varepsilon$. If $i \in C_0$, then

$$\ell_{i,t_0} \in (\bar{\ell}_{t_0} - \eta, \bar{\ell}_{t_0} + \eta) \subset (b_{\hat{k}-1}, b_{\hat{k}})$$

by (i), hence $Q(\ell_{i,t_0}) = \hat{k}$.

(b) Fix $i \in C$ and suppose inductively that $\ell_{j,t} \in (b_{\hat{k}-1}, b_{\hat{k}})$ for all $j \in C$. Then $\psi(\ell_{j,t}) = r_{\hat{k}}$ for all $j \in C$, while for $j \notin C$ we only know $\psi(\ell_{j,t}) \in [\underline{r}, \bar{r}]$. By ζ -cohesiveness,

$$S_{i,t} = \sum_{j \in C} P_{ij} r_{\hat{k}} + \sum_{j \notin C} P_{ij} \psi(\ell_{j,t}) \in (1 - \zeta)r_{\hat{k}} + \zeta[\underline{r}, \bar{r}].$$

By bounded evidence,

$$S_{i,t} + \Delta_{i,t+1} \in \mathcal{A}_{\hat{k}}^{\text{core}}(\zeta),$$

which lies in $(b_{\hat{k}-1}, b_{\hat{k}})$ by ζ -robustness. Since also $\ell_{i,t} \in (b_{\hat{k}-1}, b_{\hat{k}})$, the anchored update keeps $\ell_{i,t+1}$ in $(b_{\hat{k}-1}, b_{\hat{k}})$. This completes the induction.

■

Proof of Lemma 4

Apply CO(\cdot) to $\ell_{t+1} = (1 - \alpha)\ell_t + \alpha(P\psi(\ell_t) + \Delta_{t+1})$ and use linearity of CO and Lemma 13:

$$\text{CO}(\ell_{t+1}) = (1 - \alpha)\text{CO}(\ell_t) + \alpha P \text{CO}(\psi(\ell_t)) + \alpha \text{CO}(\Delta_{t+1}).$$

Taking $\|\cdot\|_\pi$ and using the triangle inequality gives

$$D_{t+1} \leq (1 - \alpha)D_t + \alpha \|P \text{CO}(\psi(\ell_t))\|_\pi + \alpha \|\text{CO}(\Delta_{t+1})\|_\pi.$$

Because CO($\psi(\ell_t)$) has π -mean zero, Lemma 14 implies

$$\|P \text{CO}(\psi(\ell_t))\|_\pi \leq \lambda^* \|\text{CO}(\psi(\ell_t))\|_\pi.$$

Moreover, since $\psi(\ell_t)$ takes values in $[\underline{r}, \bar{r}]$, we have the uniform bound

$$\|\text{CO}(\psi(\ell_t))\|_\pi = \sqrt{\sum_{i=1}^n \pi_i (\psi(\ell_{i,t}) - \overline{\psi(\ell_t)})^2} \leq \sqrt{\sum_{i=1}^n \pi_i R^2} = R.$$

Finally, bounded evidence implies $\|\text{CO}(\Delta_{t+1})\|_\pi \leq 2\bar{\Delta}$. Combining the bounds yields, for all t ,

$$D_{t+1} \leq (1 - \alpha)D_t + \alpha \lambda^* R + 2\alpha \bar{\Delta},$$

which is the first claim.

Now impose Assumption 3, so $\lambda^* \leq \bar{\lambda}$, and define the constant

$$D_{\sup} := \bar{\lambda}R + 2\bar{\Delta}.$$

Then the recursion becomes

$$D_{t+1} \leq (1 - \alpha)D_t + \alpha D_{\sup}. \quad (10)$$

Iterating (10) yields a geometric series:

$$D_t \leq (1 - \alpha)^t D_0 + \alpha D_{\sup} \sum_{s=0}^{t-1} (1 - \alpha)^s = (1 - \alpha)^t D_0 + (1 - (1 - \alpha)^t) D_{\sup},$$

which is the displayed bound. In particular, if priors are common so that $D_0 = 0$, then

$$D_t \leq (1 - (1 - \alpha)^t) D_{\sup} \leq D_{\sup} \quad \forall t,$$

as claimed. ■

Proof of Lemma 5

Let $C_0(t) := \{i : |\ell_{i,t} - \bar{\ell}_t| < \eta\}$. By Lemma 2,

$$\pi(C_0(t)^c) \leq \frac{\text{Var}_\pi(\ell_t)}{\eta^2} = \frac{D_t^2}{\eta^2} \leq \frac{D_{\sup}^2}{\eta^2} = \varepsilon,$$

where we used Lemma 4 (and $D_0 = 0$) in the last inequality. Hence $\pi(C_0(t)) \geq 1 - \varepsilon$. If $\bar{\ell}_t \in (b_{k-1} + 2\eta, b_k - 2\eta)$ and $i \in C_0(t)$, then

$$\ell_{i,t} \in (\bar{\ell}_t - \eta, \bar{\ell}_t + \eta) \subset (b_{k-1}, b_k)$$

and thus $Q(\ell_{i,t}) = k$. ■

Proof of Lemma 6

Step 1: Dynamics of the mean belief.

At time $t = 0$, all agents share the common prior ℓ_0 . Therefore, every agent j transmits the same message $m_0 = Q(\ell_0)$, which decodes to $y_0 = \psi(\ell_0)$. The social exposure of any agent i is:

$$S_{i,0} = \sum_{j=1}^n P_{ij} y_{j,0} = \sum_{j=1}^n P_{ij} y_0 = y_0 \left(\sum_{j=1}^n P_{ij} \right) = y_0,$$

since P is row-stochastic. The belief update for agent i at $t = 1$ is therefore:

$$\ell_{i,1} = (1 - \alpha)\ell_0 + \alpha(y_0 + \Delta_{i,1}),$$

where $\Delta_{i,1} = \sum_{j \in N} P_{ij} \Delta_{j,1}$. We define the π -weighted mean belief $\bar{\ell}_1 := \sum_{i=1}^n \pi_i \ell_{i,1}$. Multiplying the update equation by π_i and summing over i :

$$\bar{\ell}_1 = (1 - \alpha)\ell_0 \sum_{i=1}^n \pi_i + \alpha y_0 \sum_{i=1}^n \pi_i + \alpha \sum_{i=1}^n \pi_i \Delta_{i,1}.$$

Using $\sum \pi_i = 1$ and the definition $\bar{\Delta}_1 := \sum \pi_i \Delta_{i,1}$, we obtain:

$$\bar{\ell}_1 = (1 - \alpha)\ell_0 + \alpha y_0 + \alpha \bar{\Delta}_1. \quad (11)$$

Taking the expectation conditional on θ , and using $\mathbb{E}[\bar{\Delta}_1 | \theta] = \mu_\theta$, the expected mean belief is:

$$\mathbb{E}[\bar{\ell}_1 | \theta] = (1 - \alpha)\ell_0 + \alpha y_0 + \alpha \mu_\theta.$$

Step 2: The geometric condition for entry.

Let I_{safe} denote the target buffered interval:

$$I_{\text{safe}} := (b_{\hat{k}-1} + 2\eta, b_{\hat{k}} - 2\eta).$$

The condition (7) states that

$$\mathbb{E}[\bar{\ell}_1 | \theta] \in (b_{\hat{k}-1} + 2\eta + \rho, b_{\hat{k}} - 2\eta - \rho).$$

For the realized variable $\bar{\ell}_1$ to fall into I_{safe} , it suffices that the distance between the realization and the expectation is less than ρ :

$$|\bar{\ell}_1 - \mathbb{E}[\bar{\ell}_1 | \theta]| \leq \rho \implies \bar{\ell}_1 \in I_{\text{safe}}.$$

Using (11), the distance is:

$$|\bar{\ell}_1 - \mathbb{E}[\bar{\ell}_1 | \theta]| = |\alpha\bar{\Delta}_1 - \alpha\mu_\theta| = \alpha|\bar{\Delta}_1 - \mu_\theta|.$$

Thus, the sufficient condition for entry is:

$$|\bar{\Delta}_1 - \mu_\theta| \leq \frac{\rho}{\alpha}. \quad (12)$$

Step 3: Concentration of measure.

We define the random variables $X_i := \pi_i \Delta_{i,1}$. Then $\bar{\Delta}_1 = \sum_{i=1}^n X_i$.

By Assumption 1, the private signal is bounded: $\Delta_{i,1} \in [-\bar{\Delta}, \bar{\Delta}]$. Therefore, each term X_i is bounded within an interval of length:

$$\text{range}(X_i) = \pi_i \bar{\Delta} - (-\pi_i \bar{\Delta}) = 2\pi_i \bar{\Delta}.$$

We apply Hoeffding's Inequality to the sum $\bar{\Delta}_1$. For any deviation $u > 0$:

$$\Pr(|\bar{\Delta}_1 - \mu_\theta| \geq u | \theta) \leq 2 \exp\left(-\frac{2u^2}{\sum_{i=1}^n [\text{range}(X_i)]^2}\right).$$

We now bound the sum of squares using the assumption that no single agent dominates ($\pi_{\max} \leq \kappa_\pi/n$):

$$\sum_{i=1}^n \pi_i^2 \leq \pi_{\max} \sum_{i=1}^n \pi_i = \pi_{\max} \cdot 1 \leq \frac{\kappa_\pi}{n}.$$

Substituting this back into the Hoeffding bound with $u = \rho/\alpha$:

$$\Pr\left(|\bar{\Delta}_1 - \mu_\theta| \geq \frac{\rho}{\alpha} \mid \theta\right) \leq 2 \exp\left(-\frac{2(\rho/\alpha)^2}{4\bar{\Delta}^2(\kappa_\pi/n)}\right) = 2 \exp\left(-\frac{\rho^2 n}{2\alpha^2 \bar{\Delta}^2 \kappa_\pi}\right).$$

The exponent is proportional to $-n$. Define the strictly positive decay constant (independent of n) by

$$\Lambda := \frac{\rho^2}{2\alpha^2 \bar{\Delta}^2 \kappa_\pi} > 0.$$

The upper bound on the failure probability is the function $g(n) := 2 \exp(-\Lambda n)$. Since $g(0) = 2$ and $\lim_{n \rightarrow \infty} g(n) = 0$, for any arbitrary target failure probability $\delta \in (0, 1)$, there exists a finite threshold $n_0(\delta)$ such that for all $n \geq n_0(\delta)$, we have $g(n) \leq \delta$.

Consequently, for all $n \geq n_0(\delta)$, the success probability satisfies

$$\Pr\left(\left|\bar{\Delta}_1 - \mu_\theta\right| \leq \frac{\rho}{\alpha} \mid \theta\right) \geq 1 - g(n) \geq 1 - \delta.$$

We set $c_{\text{entry}} := 1 - \delta$.

■

Proof of Theorem 1

Work conditional on $\theta = 1$ and assume common priors so that $D_0 = 0$.

Step 1 (Entry and a large candidate core). By Lemma 6, there exist $c_{\text{entry}} \in (0, 1)$ and n_1 such that for all $n \geq n_1$,

$$\Pr(\bar{\ell}_1 \in (b_{\hat{k}-1} + 2\eta, b_{\hat{k}} - 2\eta) \mid \theta = 1) \geq c_{\text{entry}}.$$

On this entry event, Lemma 5 implies that the set

$$C_0 := \{i : |\ell_{i,1} - \bar{\ell}_1| < \eta\}$$

satisfies $\pi(C_0) \geq 1 - \varepsilon$ and $Q(\ell_{i,1}) = \hat{k}$ for all $i \in C_0$.

Step 2 (Size and cohesiveness of the core). Since $\pi_i \geq 1/(\kappa_\pi n)$, the mass bound implies

$$\varepsilon \geq \pi(C_0^c) = \sum_{j \in C_0^c} \pi_j \geq |C_0^c| \cdot \frac{1}{\kappa_\pi n}, \quad \text{so} \quad |C_0| \geq (1 - \kappa_\pi \varepsilon)n.$$

Moreover, by Assumption 5,

$$\sum_{j \notin C_0} P_{ij} \leq \sum_{j \notin C_0} \frac{\kappa_P}{n} = \kappa_P \frac{|C_0^c|}{n} \leq \kappa_P \kappa_\pi \varepsilon = \zeta, \quad \forall i \in C_0.$$

Thus C_0 is ζ -cohesive (Definition 2).

Step 3 (Permanent lock-in of the core). On the entry event, the hypotheses of Proposition 3 hold at $t_0 = 1$ with candidate core C_0 and leakage level ζ . (Variance control is built into Lemma 5 via the choice $\eta = D_{\text{sup}}/\sqrt{\varepsilon}$ and the envelope $D_1 \leq D_{\text{sup}}$ from Lemma 4 when $D_0 = 0$.) Hence $Q(\ell_{i,t}) = \hat{k}$ for all $i \in C_0$ and all $t \geq 1$. Combining with the entry probability yields the first claim with $c := c_{\text{entry}}$ and $n_0 := n_1$.

Step 4 (Information ceiling). Define deterministic bounds

$$\underline{\ell} := \min\{\ell_0, \underline{r} - \bar{\Delta}\}, \quad \bar{\ell} := \max\{\ell_0, \bar{r} + \bar{\Delta}\}.$$

Note $S_{i,t} + \Delta_{i,t+1} \in [\underline{r} - \bar{\Delta}, \bar{r} + \bar{\Delta}]$ for any $t \geq 0$ and $\ell_{i,0} = \ell_0$ when agents have the same prior. Since $\ell_{i,1}$ is the convex combination of $\ell_{i,0}$ and $S_{i,0} + \Delta_{i,1}$, $\ell_{i,1} \in [\underline{\ell}, \bar{\ell}]$.

Assume at t , $\ell_{i,t} \in [\underline{\ell}, \bar{\ell}]$. Since $S_{i,t} + \Delta_{i,t+1} \in [\underline{r} - \bar{\Delta}, \bar{r} + \bar{\Delta}]$ for any $t \geq 0$ and $\ell_{i,t+1}$ is the convex combination of $\ell_{i,t}$ and $S_{i,t} + \Delta_{i,t+1}$, we can know $\ell_{i,t+1} \in [\underline{\ell}, \bar{\ell}]$ for all i, t .

Let $\sigma(x) = 1/(1 + e^{-x})$. Since $\ell_{i,t} \leq \bar{\ell}$, monotonicity implies:

$$a_{i,t} = \sigma(\ell_{i,t}) \leq \sigma(\bar{\ell}) < 1.$$

The last inequality comes from the fact that $\bar{\ell}$ is finite.

Let $\widehat{\text{MSE}}_t^{(n)} := \frac{1}{n} \sum_{i=1}^n (a_{i,t} - 1)^2$, so that $\text{MSE}_t^{(n)} = \mathbb{E}[\widehat{\text{MSE}}_t^{(n)} \mid \theta = 1]$. Let E be the lock-in event from Steps 1–3, i.e., $Q(\ell_{i,t}) = \hat{k}$ for all $i \in C_0$ and all $t \geq 1$. On E , we have $|C_0| \geq (1 - \kappa_\pi \varepsilon)n$ and for every $t \geq 1$,

$$\widehat{\text{MSE}}_t^{(n)} \geq \frac{1}{n} \sum_{i \in C_0} (1 - a_{i,t})^2 \geq \frac{|C_0|}{n} (1 - \sigma(\bar{\ell}))^2 \geq (1 - \kappa_\pi \varepsilon) (1 - \sigma(\bar{\ell}))^2.$$

Taking expectations and using $\Pr(E \mid \theta = 1) \geq c$ yields, for all $t \geq 1$,

$$\text{MSE}_t^{(n)} \geq c (1 - \kappa_\pi \varepsilon) (1 - \sigma(\bar{\ell}))^2.$$

Thus $\liminf_{t \rightarrow \infty} \text{MSE}_t^{(n)}$ is bounded below by the same strictly positive constant; set

$$c_{\text{ceiling}} := c (1 - \kappa_\pi \varepsilon) (1 - \sigma(\bar{\ell}))^2 > 0.$$

■

Proof of Lemma 7.

Fix $k \in \{k^A, k^B\}$. Suppose G_g is a k -seed at time t , i.e., $Q(\ell_{j,t}) = k$ for all $j \in G_g$. This implies $\ell_{j,t} \in (b_{k-1}, b_k]$ and $y_{j,t} = r_k$ for all $j \in G_g$.

Fix $i \in G_g$ and write the within-community weight $w_i := \sum_{j \in G_g} P_{ij}$. By Assumption (S1), $w_i \geq 1 - \varepsilon_{\text{fed}}$. Decompose the social term:

$$S_{i,t} = \sum_{j \in G_g} P_{ij} y_{j,t} + \sum_{j \notin G_g} P_{ij} y_{j,t} = w_i r_k + (1 - w_i) \tilde{y}_{i,t},$$

where $\tilde{y}_{i,t}$ is the weighted average of out-of-community decoded messages. Since each decoded message is a reconstruction value, $\tilde{y}_{i,t} \in [\underline{r}, \bar{r}]$. Using the bound $w_i \geq 1 - \varepsilon_{\text{fed}}$, the social term satisfies

$$S_{i,t} \in \left[(1 - \varepsilon_{\text{fed}})r_k + \varepsilon_{\text{fed}}\underline{r}, (1 - \varepsilon_{\text{fed}})r_k + \varepsilon_{\text{fed}}\bar{r} \right].$$

By Assumption 1(ii), the private evidence is bounded by $|\Delta_{i,t+1}| \leq \bar{\Delta}$. Combining these bounds yields

$$S_{i,t} + \Delta_{i,t+1} \in \mathcal{A}_k^{\text{rob}}.$$

By Assumption (I1), $\mathcal{A}_k^{\text{rob}}$ is a strict subset of the open interval (b_{k-1}, b_k) . Thus, the “new information” term satisfies $S_{i,t} + \Delta_{i,t+1} \in (b_{k-1}, b_k)$.

Finally, consider the belief update (1):

$$\ell_{i,t+1} = (1 - \alpha)\ell_{i,t} + \alpha(S_{i,t} + \Delta_{i,t+1}).$$

Since $\ell_{i,t} \in (b_{k-1}, b_k]$ and the new information is strictly interior to (b_{k-1}, b_k) , the convex combination (with $\alpha > 0$) must lie strictly inside the open interval:

$$\ell_{i,t+1} \in (b_{k-1}, b_k).$$

Therefore $Q(\ell_{i,t+1}) = k$ for all $i \in G_g$, and G_g remains a k -seed at $t + 1$. Iterating this argument yields the claim for all $t' \geq t$.

■

Proof of Theorem 2.

We prove the claims under $\theta = 1$; the case $\theta = 0$ is symmetric.

Part (i) (Formation). Fix a community G_g . Consider the event

$$E_g := \{\Delta_{i,s} \geq \delta \text{ for all } i \in G_g \text{ and all } s = 1, \dots, L\}.$$

By Assumption 1(i) (conditional independence across agents and time) and Assumption (I2),

$$\Pr(E_g \mid \theta = 1) \geq \prod_{i \in G_g} \prod_{s=1}^L \Pr(\Delta_{i,s} \geq \delta \mid \theta = 1) \geq p_+^{L|G_g|}.$$

By Assumption (I3a), on the event E_g , the community G_g becomes an A -seed at time L .

Using the uniform bound on community sizes ($|G_g| \leq \bar{s}$) from Assumption 6, we have $\Pr(E_g \mid \theta = 1) \geq p_+^{L\bar{s}} =: q > 0$ uniformly in n and g . Moreover, the events $\{E_g\}_{g \leq M(n)}$

are independent because they depend on disjoint sets of signals and Assumption 1(i) implies independence across agents. Hence

$$\begin{aligned} \Pr(\text{no } A\text{-seed by time } L \mid \theta = 1) &\leq \Pr\left(\bigcap_{g \leq M(n)} E_g^c \mid \theta = 1\right) \\ &= \prod_{g \leq M(n)} (1 - \Pr(E_g \mid \theta = 1)) \leq (1 - q)^{M(n)} \rightarrow 0, \end{aligned}$$

since $M(n) \rightarrow \infty$. This proves (i).

Part (ii) (Propagation along receptive paths). Fix a time t at which some community g_0 is an A -seed, and fix a directed path $g_0 \rightarrow g_1 \rightarrow \dots \rightarrow g_m$ in $\mathcal{C}^{(n)}$. Assume the receptiveness condition: for each $\ell = 1, \dots, m$, community g_ℓ does not become a B -seed before becoming an A -seed.

We show by induction that each g_ℓ becomes an A -seed almost surely. The base case $\ell = 0$ holds by assumption; by Lemma 7, g_0 remains an A -seed forever.

Fix $\ell \geq 1$ and suppose $g_{\ell-1}$ is an A -seed from some time onward. Because $g_{\ell-1} \rightarrow g_\ell$ is an edge, $W_{g_{\ell-1} \rightarrow g_\ell} \geq \omega$. For integers $r \geq 0$, define the disjoint L -block events (starting after the neighbor stabilizes):

$$F_r := \{\Delta_{i,t+rL+s} \geq \delta \text{ for all } i \in G_{g_\ell} \text{ and } s = 1, \dots, L\}.$$

By Assumption 1(i) and (I2), $\Pr(F_r \mid \theta = 1) \geq p_+^{L|G_{g_\ell}|} =: q_\ell > 0$, and the events $\{F_r\}_{r \geq 0}$ are independent. Therefore $\sum_{r \geq 0} \Pr(F_r \mid \theta = 1) = \infty$ and by the second Borel–Cantelli lemma, F_r occurs for some finite r almost surely.

Let r^* be the first index such that F_{r^*} occurs. By the inductive hypothesis, $g_{\ell-1}$ is an A -seed at time $t + r^*L$. By the receptiveness assumption, g_ℓ has not become a B -seed prior to this block. Hence Assumption (I3b) applies and implies that g_ℓ becomes an A -seed at time $t + r^*L + L$. Lemma 7 then gives permanence. This completes the induction and proves (ii). ■

Proof of Theorem 3

Part (i) follows from Proposition 7. Part (ii) follows from Corollary 2. Part (iii) follows from Proposition 8 in case (L1) and from Proposition 9 in case (L2). ■

A.2 Auxiliary Results

B Appendix B: Online Appendix

B.1 Microfoundation: rational quantization under quadratic loss

This subsection provides a microfoundation for the threshold structure in Section 2.2. Consider a single sender–receiver problem. The sender observes a scalar statistic $\ell \in \mathbb{R}$ and sends a message $m \in M := \{1, \dots, K\}$. Upon observing m , the receiver chooses an action $a \in \mathbb{R}$. Both parties share the same quadratic-loss objective: conditional on (ℓ, m, a) the loss is $(a - \ell)^2$. Hence the receiver chooses a to minimize expected loss given m , and the sender chooses m anticipating the receiver’s best response.

Let F denote the distribution of ℓ and assume $\mathbb{E}[\ell^2] < \infty$. An encoder–decoder pair consists of a measurable message rule $Q : \mathbb{R} \rightarrow M$ and an action rule $\Psi : M \rightarrow \mathbb{R}$. Let $\psi := \Psi \circ Q$ denote the induced (decoded) action as a function of ℓ . The expected distortion is

$$D(Q, \Psi) := \mathbb{E}[(\ell - \Psi(Q(\ell)))^2] = \mathbb{E}[(\ell - \psi(\ell))^2].$$

An optimal quantizer minimizes $D(Q, \Psi)$ over all measurable Q and all Ψ .

Proposition 4. *Suppose $\mathbb{E}[\ell^2] < \infty$ and an optimal quantizer exists. Then there exists an optimal pair (Q^*, Ψ^*) such that:*

- (i) *Q^* is a threshold partition: there exist cutpoints $-\infty = b_0 < b_1 < \dots < b_K = +\infty$ such that*

$$Q^*(\ell) = k \iff \ell \in (b_{k-1}, b_k].$$

- (ii) *Ψ^* is centroid decoding:*

$$\Psi^*(k) = \mathbb{E}[\ell \mid \ell \in (b_{k-1}, b_k)] \quad \text{for all } k \text{ with } \Pr(\ell \in (b_{k-1}, b_k)) > 0.$$

- (iii) *The boundaries satisfy midpoint indifference:*

$$b_k = \frac{\Psi^*(k) + \Psi^*(k+1)}{2} \quad \text{for } k = 1, \dots, K-1.$$

Consequently, the induced map $\psi^* = \Psi^* \circ Q^*$ is piecewise constant with K bins and is monotone nondecreasing.

Proof of Proposition 4

Let (Q, Ψ) be any feasible encoder–decoder pair. Write the message cells as $C_k := \{\ell : Q(\ell) = k\}$.

Step 1 (Optimal decoder is centroid decoding). Fix Q and consider minimizing $D(Q, \Psi)$ over Ψ . Conditional on the event $\{Q(\ell) = k\}$, the contribution to the expected loss is

$$\mathbb{E}[(\ell - \Psi(k))^2 \mid Q(\ell) = k].$$

As a function of the scalar $\Psi(k)$, this conditional expectation is minimized uniquely at $\Psi(k) = \mathbb{E}[\ell \mid Q(\ell) = k]$ whenever $\Pr(Q(\ell) = k) > 0$ (a standard projection property of conditional expectation under quadratic loss). If $\Pr(Q(\ell) = k) = 0$, $\Psi(k)$ is irrelevant for D . Therefore, for any Q , there exists an optimal decoder Ψ_Q given by

$$\Psi_Q(k) = \mathbb{E}[\ell \mid Q(\ell) = k] \quad \text{for all } k \text{ with } \Pr(Q(\ell) = k) > 0,$$

and any optimal pair can be chosen to satisfy this property. This proves (ii) once we identify the cells as intervals.

Step 2 (For fixed decoder values, optimal encoder is nearest-neighbor). Fix any decoder Ψ and define $c_k := \Psi(k)$. For each realized ℓ , the pointwise loss from sending message k is $(\ell - c_k)^2$. Hence, holding Ψ fixed, an encoder Q minimizes $D(Q, \Psi)$ if and only if, for F -almost every ℓ ,

$$Q(\ell) \in \arg \min_{k \in \{1, \dots, K\}} (\ell - c_k)^2.$$

In words: an optimal Q assigns each ℓ to one of the closest reconstruction points.

Now relabel messages so that the reconstruction points are weakly increasing: $c_1 \leq c_2 \leq \dots \leq c_K$ (this relabeling does not change the achievable distortion). For $k \in \{1, \dots, K-1\}$, define the midpoint

$$b_k := \frac{c_k + c_{k+1}}{2}.$$

A direct comparison shows that for any ℓ ,

$$(\ell - c_k)^2 \leq (\ell - c_{k+1})^2 \iff \ell \leq \frac{c_k + c_{k+1}}{2} = b_k.$$

Thus, under the ordered codebook, the nearest-neighbor regions are intervals separated by midpoints: there exists an optimal encoder Q such that

$$Q(\ell) = k \iff \ell \in (b_{k-1}, b_k],$$

where we set $b_0 = -\infty$ and $b_K = +\infty$. This establishes the threshold form in (i) and the midpoint indifference condition in (iii), for the ordered reconstruction points (c_k) .

Step 3 (Combine the two best responses). Let (Q^*, Ψ^*) be an optimal pair. By Step 1, we may choose Ψ^* to be a best response to Q^* , i.e., centroid decoding on each cell. By Step 2, we may choose Q^* to be a best response to Ψ^* , i.e., nearest-neighbor assignment with threshold boundaries at midpoints between adjacent reconstruction points. Therefore there exists an optimal pair (Q^*, Ψ^*) satisfying (i)–(iii).

Step 4 (Monotonicity of the induced map). Under (i), $\psi^*(\ell) = \Psi^*(k)$ for all $\ell \in (b_{k-1}, b_k]$. For any k with $\Pr(\ell \in (b_{k-1}, b_k]) > 0$, the centroid satisfies $\Psi^*(k) \in [b_{k-1}, b_k]$ because it is the mean of a random variable supported on $(b_{k-1}, b_k]$. Hence, whenever consecutive bins are non-null,

$$\Psi^*(k) \leq b_k < \Psi^*(k+1),$$

which implies $\Psi^*(1) \leq \Psi^*(2) \leq \dots \leq \Psi^*(K)$ after ordering and discarding any null bins (which do not affect D). Therefore ψ^* is monotone nondecreasing and piecewise constant with K bins.

■

B.2 Federation geometry

Lemma 8 (Exposure bounds under federation). *Suppose Assumption ?? holds. Fix a community G_g and an index k . If at time t we have $Q(\ell_{j,t}) = k$ for all $j \in G_g$, then for every $i \in G_g$,*

$$S_{i,t} := \sum_{j=1}^n P_{ij} \psi(\ell_{j,t}) \in \left[(1 - \varepsilon_{\text{fed}})r_k + \varepsilon_{\text{fed}}r, (1 - \varepsilon_{\text{fed}})r_k + \varepsilon_{\text{fed}}\bar{r} \right].$$

Proof of Lemma 8

Fix $i \in G_g$ and split the exposure into in- and out-of-community parts:

$$S_{i,t} = \sum_{j \in G_g} P_{ij} \psi(\ell_{j,t}) + \sum_{j \notin G_g} P_{ij} \psi(\ell_{j,t}).$$

By hypothesis, $Q(\ell_{j,t}) = k$ for all $j \in G_g$, hence $\psi(\ell_{j,t}) = r_k$ for all $j \in G_g$. Let

$$w_i := \sum_{j \in G_g} P_{ij} \in [1 - \varepsilon_{\text{fed}}, 1]$$

be the total in-community weight (Assumption ??), so $\sum_{j \notin G_g} P_{ij} = 1 - w_i$. Then

$$S_{i,t} = w_i r_k + \sum_{j \notin G_g} P_{ij} \psi(\ell_{j,t}).$$

Since $\psi(\cdot) \in [\underline{r}, \bar{r}]$, the out-of-community contribution satisfies

$$\sum_{j \notin G_g} P_{ij} \psi(\ell_{j,t}) \in (1 - w_i)[\underline{r}, \bar{r}].$$

Therefore,

$$S_{i,t} \in [w_i r_k + (1 - w_i)\underline{r}, w_i r_k + (1 - w_i)\bar{r}]. \quad (13)$$

We now convert (13) into bounds that depend only on ε_{fed} . For the lower endpoint, note that

$$w \mapsto wr_k + (1 - w)\underline{r} = \underline{r} + w(r_k - \underline{r})$$

is weakly increasing in w because $r_k \geq \underline{r}$. Hence the smallest possible lower endpoint over $w \in [1 - \varepsilon_{\text{fed}}, 1]$ is attained at $w = 1 - \varepsilon_{\text{fed}}$, giving

$$w_i r_k + (1 - w_i)\underline{r} \geq (1 - \varepsilon_{\text{fed}})r_k + \varepsilon_{\text{fed}}\underline{r}.$$

For the upper endpoint, note that

$$w \mapsto wr_k + (1 - w)\bar{r} = \bar{r} + w(r_k - \bar{r})$$

is weakly *decreasing* in w because $r_k \leq \bar{r}$. Hence the largest possible upper endpoint over $w \in [1 - \varepsilon_{\text{fed}}, 1]$ is also attained at $w = 1 - \varepsilon_{\text{fed}}$, yielding

$$w_i r_k + (1 - w_i)\bar{r} \leq (1 - \varepsilon_{\text{fed}})r_k + \varepsilon_{\text{fed}}\bar{r}.$$

Combining these bounds with (13) gives the claimed interval.

■

Lemma 9 (Exposure pull along a weak-tie edge). *Fix communities $g \neq h$ and suppose $W_{h \leftarrow g} \geq \omega$. If at time t we have $Q(\ell_{j,t}) = k$ for all $j \in G_g$, then for every $i \in G_h$,*

$$S_{i,t} \in [\omega r_k + (1 - \omega)\underline{r}, \omega r_k + (1 - \omega)\bar{r}].$$

Proof of Lemma 9

Fix $i \in G_h$ and set $w_i := \sum_{j \in G_g} P_{ij} \geq W_{h \leftarrow g} \geq \omega$. Split exposure into the contribution from G_g and the rest:

$$S_{i,t} = \sum_{j \in G_g} P_{ij} \psi(\ell_{j,t}) + \sum_{j \notin G_g} P_{ij} \psi(\ell_{j,t}).$$

By hypothesis, $\psi(\ell_{j,t}) = r_k$ for all $j \in G_g$, hence

$$S_{i,t} = w_i r_k + \sum_{j \notin G_g} P_{ij} \psi(\ell_{j,t}).$$

Since $\psi(\cdot) \in [\underline{r}, \bar{r}]$ and $\sum_{j \notin G_g} P_{ij} = 1 - w_i$, we obtain

$$S_{i,t} \in w_i r_k + (1 - w_i)[\underline{r}, \bar{r}].$$

Because $w_i \geq \omega$, we have $(1 - w_i) \leq (1 - \omega)$ and thus

$$w_i r_k + (1 - w_i)[\underline{r}, \bar{r}] \subseteq \omega r_k + (1 - \omega)[\underline{r}, \bar{r}],$$

which is the stated interval.

■

B.3 Non-vacuity of absorbing and robust-bin conditions

The main text imposes two “robust inclusion” conditions that make lock-in and seed dynamics irreversible under bounded evidence: the *absorbing-bin* condition (Assumption 4) and the *robust-consensus* conditions (Assumption ?? and the core-robustness requirement in Theorem 1). This subsection records simple sufficient conditions that (i) make these assumptions easy to check and (ii) clarify when they must fail as the vocabulary becomes fine.

Lemma 10 (Wide bins guarantee an absorbing cell). *Maintain Assumption 1. Fix a bin index k and write its width as $w_k := b_k - b_{k-1} \in (0, \infty]$. If the reconstruction satisfies*

$$r_k \in (b_{k-1} + \bar{\Delta}, b_k - \bar{\Delta}),$$

then Assumption 4 holds for $\hat{k} = k$. In particular, if $r_k = (b_{k-1} + b_k)/2$ (midpoint reconstruction) and $w_k > 2\bar{\Delta}$, then Assumption 4 holds and the absorption margin is $\delta_{\text{abs}} = w_k/2 - \bar{\Delta}$.

Proof of Lemma 10

If $r_k \in (b_{k-1} + \bar{\Delta}, b_k - \bar{\Delta})$, then $r_k - \bar{\Delta} > b_{k-1}$ and $r_k + \bar{\Delta} < b_k$, so $[r_k - \bar{\Delta}, r_k + \bar{\Delta}] \subset (b_{k-1}, b_k)$, which is Assumption 4 with $\hat{k} = k$.

For midpoint reconstruction, $r_k - b_{k-1} = b_k - r_k = w_k/2$. Hence $r_k \in (b_{k-1} + \bar{\Delta}, b_k - \bar{\Delta})$ is equivalent to $w_k/2 > \bar{\Delta}$, i.e. $w_k > 2\bar{\Delta}$, and then $\delta_{\text{abs}} = \min\{r_k - \bar{\Delta} - b_{k-1}, b_k - (r_k + \bar{\Delta})\} = w_k/2 - \bar{\Delta}$.

■

Corollary 1 (Equal-width designs and the “fine vocabulary” failure). *Maintain Assumption 1. Suppose the designer uses midpoint reconstructions and chooses K interior bins of equal width w covering a relevant belief envelope of length $L_\ell > 0$ (e.g. an interval containing $[\underline{\ell}, \bar{\ell}]$ from Lemma 15), so that $w = L_\ell/K$. If $w > 2\bar{\Delta}$ then there exists an absorbing bin. Conversely, if $w \leq 2\bar{\Delta}$ then no interior bin can satisfy Assumption 4 under midpoint reconstruction. Equivalently, under equal-width midpoint designs, Assumption 4 is feasible only if*

$$K < \frac{L_\ell}{2\bar{\Delta}}.$$

Proof of Corollary 1

Under midpoint reconstruction, Lemma 10 reduces feasibility of an absorbing bin to $w_k > 2\bar{\Delta}$. Under equal widths, $w_k = w$ for all interior bins, so the claim follows immediately.

■

The robust-consensus and core-robustness requirements are analogous, except that the relevant “worst-case” interval is enlarged by possible contamination from outsiders. The next lemma provides a single sufficient margin condition that implies the various inclusions used in the main text.

Lemma 11 (A sufficient margin condition for robust-bin inclusions). *Maintain Assumption 1 and let $R := \bar{r} - \underline{r}$. Fix $\zeta \in [0, 1]$ and define the worst-case interval*

$$\mathcal{A}_k(\zeta) := \left[(1 - \zeta)r_k + \zeta\underline{r} - \bar{\Delta}, (1 - \zeta)r_k + \zeta\bar{r} + \bar{\Delta} \right].$$

If for some bin k ,

$$[r_k - (\zeta R + \bar{\Delta}), r_k + (\zeta R + \bar{\Delta})] \subset (b_{k-1}, b_k), \quad (14)$$

then $\mathcal{A}_k(\zeta) \subset (b_{k-1}, b_k)$. In particular, (14) with $\zeta = \varepsilon_{\text{fed}}$ implies the corresponding robust-consensus inclusion in Assumption ??, and (14) with ζ equal to the leakage parameter in Definition 3 implies ζ -robustness of the ceiling bin.

Proof of Lemma 11

Because $\underline{r} \geq r_k - R$ and $\bar{r} \leq r_k + R$, we have

$$(1 - \zeta)r_k + \zeta\underline{r} \geq (1 - \zeta)r_k + \zeta(r_k - R) = r_k - \zeta R,$$

and

$$(1 - \zeta)r_k + \zeta\bar{r} \leq (1 - \zeta)r_k + \zeta(r_k + R) = r_k + \zeta R.$$

Therefore

$$\mathcal{A}_k(\zeta) \subset [r_k - (\zeta R + \bar{\Delta}), r_k + (\zeta R + \bar{\Delta})],$$

and the inclusion (14) implies $\mathcal{A}_k(\zeta) \subset (b_{k-1}, b_k)$.

■

B.4 Incubation feasibility: primitive inequalities

Assumption ?? packages two incubation requirements: (i) a community can move into a robust bin under a run of favorable private evidence, and (ii) once a seed exists elsewhere, a weak-tie edge combined with favorable evidence can replicate the seed in a neighboring community. The auxiliary lemmas in Appendix C provide primitive, checkable inequalities that imply these claims.

Proposition 5 (Primitive sufficient conditions for Assumption ??). *Maintain Assumptions 1, ??, ??, ??, and ??.* Fix ω from Assumption ?? . Suppose there exists an integer $L \geq 1$ such that:

- (i) L satisfies the seed-incubation inequalities (15)–(16) for k^A and also the symmetric seed-incubation inequalities for k^B ;
- (ii) L satisfies the edge-incubation inequalities (17)–(18) for k^A and also the symmetric edge-incubation inequalities for k^B .

Then Assumption ?? holds.

Proof of Proposition 5

Fix L as in the statement.

Within-community incubation. If $\Delta_{i,t+1}, \dots, \Delta_{i,t+L} \geq \delta$ for all $i \in G_g$, Lemma 17 implies $Q(\ell_{i,t+L}) = k^A$ for all $i \in G_g$. The analogous statement for k^B follows from the symmetric seed-incubation inequalities.

Edge incubation. Fix communities $g \neq h$ with $W_{h \leftarrow g} \geq \omega$. If G_g is an A -seed at time t then it is an H -seed from time t onward by Lemma 16. If, in addition, $\Delta_{i,t+1}, \dots, \Delta_{i,t+L} \geq \delta$ for all $i \in G_h$, then Lemma 19 implies $Q(\ell_{i,t+L}) = k^A$ for all $i \in G_h$. The analogous statement for k^B follows from the symmetric edge-incubation inequalities.

Combining the within-community and edge cases yields Assumption ??.

■

Lemma 12 (Existence of a finite incubation horizon). *Fix a bin k and numbers $T^{\min}, T^{\max} \in \mathbb{R}$ satisfying*

$$T^{\min} > b_{k-1}, \quad T^{\max} < b_k.$$

Then there exists $L \geq 1$ such that

$$T^{\min} + (1 - \alpha)^L (\underline{\ell} - T^{\min}) \geq b_{k-1}, \quad T^{\max} + (1 - \alpha)^L (\bar{\ell} - T^{\max}) \leq b_k,$$

where $[\underline{\ell}, \bar{\ell}]$ is the uniform belief envelope from Lemma 15.

Proof of Lemma 12

Because $0 < 1 - \alpha < 1$, the term $(1 - \alpha)^L$ decreases to 0 as $L \rightarrow \infty$. Hence

$$T^{\min} + (1 - \alpha)^L (\underline{\ell} - T^{\min}) \rightarrow T^{\min} \quad \text{and} \quad T^{\max} + (1 - \alpha)^L (\bar{\ell} - T^{\max}) \rightarrow T^{\max}.$$

By the strict inequalities $T^{\min} > b_{k-1}$ and $T^{\max} < b_k$, choose L large enough that the two displayed inequalities hold simultaneously.

■

Remark 1 (How to apply Lemma 12). For within-community incubation into bin k^A , Lemma 17 uses $T^{\min} = \underline{r} + \delta$ and $T^{\max} = \bar{r} + \bar{\Delta}$. For edge incubation along a weak-tie link of strength ω into the same bin, Lemma 19 uses $T^{\min} = \omega r_{k^A} + (1 - \omega)\underline{r} + \delta$ and $T^{\max} = \omega r_{k^A} + (1 - \omega)\bar{r} + \bar{\Delta}$. Thus, a simple (though conservative) sufficient condition for the existence of a finite incubation horizon is that these limiting values lie strictly inside the relevant bin.

B.5 Examples for mixing and diffuse-weight conditions

This subsection records canonical constructions of exposure matrices that satisfy the regularity conditions used in the main text.

Proposition 6 (Dense reversible random walks satisfy Assumptions 2 and 5). *For each n , let $G^{(n)}$ be a connected undirected graph on n nodes with degrees satisfying*

$$cn \leq d_i \leq Cn \quad \forall i$$

for constants $0 < c \leq C < \infty$ independent of n . Let P be the simple random-walk matrix on $G^{(n)}$: $P_{ij} = 1/d_i$ if i is adjacent to j and $P_{ij} = 0$ otherwise. Then P is row-stochastic, irreducible, and reversible with respect to the stationary distribution $\pi_i = d_i / \sum_j d_j$. Moreover,

$$\frac{c}{C} \cdot \frac{1}{n} \leq \pi_i \leq \frac{C}{c} \cdot \frac{1}{n}, \quad \max_{i,j} P_{ij} \leq \frac{1}{c} \cdot \frac{1}{n}.$$

Hence Assumption 5 holds with $\kappa_\pi = C/c$ and $\kappa_P = 1/c$. If, in addition, the eigenvalue modulus satisfies $\lambda^* \leq \bar{\lambda} < 1$ uniformly in n , then the fast-mixing requirement in Assumption 3 holds.

Proof of Proposition 6

Row-stochasticity and irreducibility are immediate from construction and connectedness. For undirected graphs, the random walk is reversible with respect to $\pi_i \propto d_i$, i.e. $\pi_i P_{ij} = \pi_j P_{ji}$.

To bound π , note that $\sum_j d_j$ lies in $[cn^2, Cn^2]$. Therefore $\pi_i = d_i / \sum_j d_j$ lies in $[(cn)/(Cn^2), (Cn)/(cn^2)] = [(c/C)(1/n), (C/c)(1/n)]$. Finally, $P_{ij} \leq 1/d_i \leq 1/(cn)$, giving the diffuse-row bound.

■

Remark 2 (Federation as a block-perturbed exposure matrix). A simple way to construct federation matrices satisfying Assumption ?? is to start from a block-diagonal within-community matrix P^{in} (each block row-stochastic and irreducible within the community) and add weak ties via a convex perturbation

$$P = (1 - \varepsilon_{\text{fed}}) P^{\text{in}} + \varepsilon_{\text{fed}} P^{\text{out}},$$

where P^{out} is row-stochastic and spreads weight across other communities. The convex combination implies that every agent places total within-community weight at least $1 - \varepsilon_{\text{fed}}$. Strong connectivity of the induced community graph then depends on the support of P^{out} .

B.6 Interpretive notes

Remark 3 (Interpreting the divergence condition as “replication is too slow”). A natural benchmark horizon $T^{(n)}$ is the time by which the correct seed would saturate the federation (convert all communities) with high probability *absent* wrong-seed formation. For example, Lemma 21 implies that saturation occurs within $O(\log M(n))$ length- L windows once a correct seed exists, yielding a candidate

$$T^{(n)} := L + \bar{D}_{\mathcal{C}} L c \log M(n)$$

for an appropriate constant $c > 0$. If the inter-community links are so weak or so sparse that saturation requires horizons $T^{(n)}$ much larger than logarithmic in $M(n)$, then $J(n)$ becomes large and (20) captures the idea that the federation offers “too many chances” for a wrong seed to form before the correct message can replicate.

C Appendix C: Technical Lemmas and Auxiliary Results

C.1 Linear-algebra and mixing tools

Lemma 13 (Centering commutes with P). *If $\pi^\top P = \pi^\top$, then for every $x \in \mathbb{R}^n$,*

$$\text{CO}(Px) = P \text{CO}(x).$$

Proof of Lemma 13

Write $\bar{x} = \pi^\top x$ and note that $\overline{Px} = \pi^\top(Px) = (\pi^\top P)x = \pi^\top x = \bar{x}$. Then

$$\text{CO}(Px) = Px - \overline{Px}\mathbf{1} = Px - \bar{x}\mathbf{1}.$$

Also,

$$P \text{CO}(x) = P(x - \bar{x}\mathbf{1}) = Px - \bar{x}P\mathbf{1}.$$

Since P is row-stochastic, $P\mathbf{1} = \mathbf{1}$, hence $P \text{CO}(x) = Px - \bar{x}\mathbf{1} = \text{CO}(Px)$.

■

Lemma 14 ($L^2(\pi)$ contraction). *Suppose Assumption 2 holds. Then for every $x \in \mathbb{R}^n$,*

$$\text{Var}_\pi(Px) \leq (\lambda^*)^2 \text{Var}_\pi(x), \quad \text{equivalently} \quad \|\text{CO}(Px)\|_\pi \leq \lambda^* \|\text{CO}(x)\|_\pi.$$

Proof of Lemma 14

Let $z := \text{CO}(x)$, so $\bar{z} = 0$. By Lemma 13, $\text{CO}(Px) = Pz$ and hence

$$\text{Var}_\pi(Px) = \|Pz\|_\pi^2 = \langle Pz, Pz \rangle_\pi.$$

Because P is self-adjoint under $\langle \cdot, \cdot \rangle_\pi$,

$$\langle Pz, Pz \rangle_\pi = \langle z, P(Pz) \rangle_\pi = \langle z, P^2 z \rangle_\pi.$$

Expand z in an orthonormal eigenbasis $\{v_k\}_{k=1}^n$ of P in $L^2(\pi)$. Since $\bar{z} = 0$, z is orthogonal to $\mathbf{1} = v_1$, so

$$z = \sum_{k=2}^n c_k v_k, \quad Pv_k = \lambda_k v_k, \quad \langle v_k, v_{k'} \rangle_\pi = \mathbf{1}\{k = k'\}.$$

Then $Pz = \sum_{k=2}^n c_k \lambda_k v_k$ and $P^2 z = \sum_{k=2}^n c_k \lambda_k^2 v_k$, so

$$\langle z, P^2 z \rangle_\pi = \sum_{k=2}^n c_k^2 \lambda_k^2 \leq (\lambda^*)^2 \sum_{k=2}^n c_k^2 = (\lambda^*)^2 \|z\|_\pi^2.$$

Since $\|z\|_\pi^2 = \text{Var}_\pi(x)$, the claim follows. ■

C.2 Cohesion bounds

C.3 Uniform envelope bounds

Lemma 15 (Uniform belief envelope). *Maintain Assumption 1. Define*

$$\underline{\ell} := \min\{\ell_0, \underline{r} - \bar{\Delta}\}, \quad \bar{\ell} := \max\{\ell_0, \bar{r} + \bar{\Delta}\}.$$

Then for all agents i and all times t , $\ell_{i,t} \in [\underline{\ell}, \bar{\ell}]$ almost surely.

Proof of Lemma 15

At $t = 0$, $\ell_{i,0} = \ell_0 \in [\underline{\ell}, \bar{\ell}]$ for all i .

Assume inductively that $\ell_{i,t} \in [\underline{\ell}, \bar{\ell}]$ for all i . Then $\psi(\ell_{j,t}) \in [\underline{r}, \bar{r}]$ for all j , so the social exposure

$$S_{i,t} = \sum_j P_{ij} \psi(\ell_{j,t}) \in [\underline{r}, \bar{r}]$$

for every i , because it is a convex combination of values in $[\underline{r}, \bar{r}]$. By bounded evidence, $\Delta_{i,t+1} \in [-\bar{\Delta}, \bar{\Delta}]$ a.s., hence

$$S_{i,t} + \Delta_{i,t+1} \in [\underline{r} - \bar{\Delta}, \bar{r} + \bar{\Delta}] \subseteq [\underline{\ell}, \bar{\ell}] \quad \text{a.s.}$$

The update $\ell_{i,t+1} = (1 - \alpha)\ell_{i,t} + \alpha(S_{i,t} + \Delta_{i,t+1})$ is a convex combination of two points in $[\underline{\ell}, \bar{\ell}]$, hence remains in $[\underline{\ell}, \bar{\ell}]$ a.s. This closes the induction. ■

C.4 Federation geometry

The key exposure inequalities implied by federation structure (within-community leakage bounds and weak-tie pull along inter-community edges) are stated and proved in Appendix B. They are used repeatedly in the robust-seed auxiliary lemmas below.

C.5 Auxiliary results for robust-seed analysis

The following auxiliary lemmas are used in the robust-seeds analysis and are stated and proved here for completeness.

Lemma 16 (Robust seed absorption). *Suppose Assumptions 1, ??, and ?? hold. Fix a community G_g and a seed index $k \in \{k^A, k^B\}$. If at some time t we have $Q(\ell_{i,t}) = k$ for all $i \in G_g$, then for all $s \geq 0$ and all $i \in G_g$,*

$$Q(\ell_{i,t+s}) = k, \quad \psi(\ell_{i,t+s}) = r_k.$$

Proof of Lemma 16

Proceed by induction on s . The case $s = 0$ holds by hypothesis.

Fix $s \geq 0$ and assume $Q(\ell_{i,t+s}) = k$ for all $i \in G_g$. Then $\psi(\ell_{j,t+s}) = r_k$ for all $j \in G_g$. By Lemma 8, for each $i \in G_g$,

$$S_{i,t+s} = \sum_{j=1}^n P_{ij} \psi(\ell_{j,t+s}) \in [(1 - \varepsilon_{\text{fed}})r_k + \varepsilon_{\text{fed}}\underline{r}, (1 - \varepsilon_{\text{fed}})r_k + \varepsilon_{\text{fed}}\bar{r}].$$

By bounded evidence (Assumption 1), $\Delta_{i,t+s+1} \in [-\bar{\Delta}, \bar{\Delta}]$ a.s., hence

$$S_{i,t+s} + \Delta_{i,t+s+1} \in \mathcal{A}_k^{\text{rob}} \quad \text{a.s.}$$

By Assumption ??, $\mathcal{A}_k^{\text{rob}} \subset (b_{k-1}, b_k)$. Also, since $Q(\ell_{i,t+s}) = k$, we have $\ell_{i,t+s} \in (b_{k-1}, b_k)$ under the fixed bin convention. Therefore the anchored update

$$\ell_{i,t+s+1} = (1 - \alpha)\ell_{i,t+s} + \alpha(S_{i,t+s} + \Delta_{i,t+s+1})$$

is a convex combination of two points in (b_{k-1}, b_k) , hence $\ell_{i,t+s+1} \in (b_{k-1}, b_k)$ and $Q(\ell_{i,t+s+1}) = k$ for all $i \in G_g$. This completes the induction.

■

Lemma 17 (Seed incubation into the H -bin). *Maintain Assumptions 1 and ??, and let k^A be as in Assumption ???. Fix an integer $L \geq 1$ and define*

$$T_H^{\min} := \underline{r} + \delta, \quad T^{\max} := \bar{r} + \bar{\Delta}.$$

If

$$T_H^{\min} + (1 - \alpha)^L (\ell - T_H^{\min}) \geq b_{k^A-1}, \tag{15}$$

and

$$T^{\max} + (1 - \alpha)^L (\bar{\ell} - T^{\max}) \leq b_{k^A}, \quad (16)$$

then for any community G_g and any start time t , on the event

$$\Delta_{i,t+1}, \dots, \Delta_{i,t+L} \geq \delta \quad \text{for all } i \in G_g,$$

we have $Q(\ell_{i,t+L}) = k^A$ for all $i \in G_g$.

Proof of Lemma 17

Fix a community G_g and an agent $i \in G_g$. Work on the event

$$E := \bigcap_{i \in G_g} \bigcap_{m=1}^L \{\Delta_{i,t+m} \geq \delta\}.$$

Step 1 (Lower bound). For any $\tau \in \{t, \dots, t+L-1\}$, since $\psi(\ell_{j,\tau}) \in [\underline{r}, \bar{r}]$ for all j , the social exposure satisfies $S_{i,\tau} \geq \underline{r}$. On E we also have $\Delta_{i,\tau+1} \geq \delta$, hence $S_{i,\tau} + \Delta_{i,\tau+1} \geq T_H^{\min}$. Substituting into the update gives

$$\ell_{i,\tau+1} \geq (1 - \alpha)\ell_{i,\tau} + \alpha T_H^{\min}, \quad \tau = t, \dots, t+L-1.$$

Iterating yields

$$\ell_{i,t+L} \geq T_H^{\min} + (1 - \alpha)^L (\ell_{i,t} - T_H^{\min}) \geq T_H^{\min} + (1 - \alpha)^L (\underline{\ell} - T_H^{\min}),$$

where the last inequality uses $\ell_{i,t} \geq \underline{\ell}$ (Lemma 15). By (15), $\ell_{i,t+L} \geq b_{k^A-1}$.

Step 2 (Upper bound). Similarly, $S_{i,\tau} \leq \bar{r}$ and $\Delta_{i,\tau+1} \leq \bar{\Delta}$ imply $S_{i,\tau} + \Delta_{i,\tau+1} \leq T^{\max}$. Thus

$$\ell_{i,t+L} \leq T^{\max} + (1 - \alpha)^L (\ell_{i,t} - T^{\max}) \leq T^{\max} + (1 - \alpha)^L (\bar{\ell} - T^{\max}) \leq b_{k^A},$$

where we used $\ell_{i,t} \leq \bar{\ell}$ (Lemma 15) and (16).

Step 3 (Bin membership). Combining the bounds yields $b_{k^A-1} \leq \ell_{i,t+L} \leq b_{k^A}$, hence $Q(\ell_{i,t+L}) = k^A$ under the fixed bin convention. Since $i \in G_g$ was arbitrary, this holds for all $i \in G_g$.

■

Lemma 18 (seed formation w.h.p.). *Suppose Assumptions ??, ??, 1, and ?? hold. Fix L satisfying (15)–(16) for k^A (and the symmetric seed-incubation inequalities for k^B under $\theta = 0$). Then:*

(H) *Conditional on $\theta = 1$,*

$$\Pr\left(\exists g \leq M(n) \text{ such that } G_g \text{ is an } H\text{-seed by time } L\right) \geq 1 - (1 - p_+^{L\bar{s}})^{M(n)}.$$

(L) *Conditional on $\theta = 0$,*

$$\Pr\left(\exists g \leq M(n) \text{ such that } G_g \text{ is an } L\text{-seed by time } L\right) \geq 1 - (1 - p_+^{L\bar{s}})^{M(n)}.$$

Proof of Lemma 18

We prove (H); (L) is symmetric. Fix a community G_g and define

$$E_g := \left\{ \Delta_{i,1}, \dots, \Delta_{i,L} \geq \delta \text{ for all } i \in G_g \right\}.$$

Conditional on $\theta = 1$, Assumption ?? implies $\Pr(E_g \mid \theta = 1) \geq p_+^{L|G_g|} \geq p_+^{L\bar{s}}$. On E_g , Lemma 17 implies $Q(\ell_{i,L}) = k^A$ for all $i \in G_g$, and Lemma 16 implies G_g remains in bin k^A forever; hence G_g is an H -seed by time L .

Because the communities are disjoint and private signals are independent conditional on θ (Assumption ??), the events $\{E_g\}_{g \leq M(n)}$ are mutually independent conditional on $\theta = 1$. Therefore,

$$\Pr\left(\text{no community becomes an } H\text{-seed by time } L \mid \theta = 1\right) \leq \prod_{g=1}^{M(n)} (1 - p_+^{L\bar{s}}) = (1 - p_+^{L\bar{s}})^{M(n)}.$$

Taking complements yields (H). ■

Lemma 19 (Edge incubation into the H -bin). *Maintain Assumptions 1 and ??, and let k^A be as in Assumption ??.. Fix communities $g \neq h$ with $W_{h \leftarrow g} \geq \omega$. Fix an integer $L \geq 1$ and define*

$$T_{H,\omega}^{\min} := \omega r_{k^A} + (1 - \omega)\underline{r} + \delta, \quad T_{H,\omega}^{\max} := \omega r_{k^A} + (1 - \omega)\bar{r} + \bar{\Delta}.$$

If

$$T_{H,\omega}^{\min} + (1 - \alpha)^L (\underline{\ell} - T_{H,\omega}^{\min}) \geq b_{k^A-1}, \tag{17}$$

and

$$T_{H,\omega}^{\max} + (1 - \alpha)^L (\bar{\ell} - T_{H,\omega}^{\max}) \leq b_{k^A}, \quad (18)$$

then for any start time t , if G_g is an H -seed throughout $t, \dots, t + L - 1$ and

$$\Delta_{i,t+1}, \dots, \Delta_{i,t+L} \geq \delta \quad \text{for all } i \in G_h,$$

we have $Q(\ell_{i,t+L}) = k^A$ for all $i \in G_h$.

Proof of Lemma 19

Fix $i \in G_h$ and consider $\tau \in \{t, \dots, t + L - 1\}$. Work on the event that (i) G_g is an H -seed throughout $t, \dots, t + L - 1$, and (ii) $\Delta_{i,t+1}, \dots, \Delta_{i,t+L} \geq \delta$ for all $i \in G_h$.

Since G_g is an H -seed, $\psi(\ell_{j,\tau}) = r_{k^A}$ for all $j \in G_g$. Lemma 9 implies that for each $i \in G_h$ and each such τ ,

$$S_{i,\tau} \in [\omega r_{k^A} + (1 - \omega)\underline{r}, \omega r_{k^A} + (1 - \omega)\bar{r}].$$

Combining with $\Delta_{i,\tau+1} \in [\delta, \bar{\Delta}]$ on the block event yields

$$S_{i,\tau} + \Delta_{i,\tau+1} \in [T_{H,\omega}^{\min}, T_{H,\omega}^{\max}].$$

The rest follows exactly as in Lemma 17, using Lemma 15 and the inequalities (17)–(18) to conclude $b_{k^A-1} \leq \ell_{i,t+L} \leq b_{k^A}$ and hence $Q(\ell_{i,t+L}) = k^A$. Since i was arbitrary, this holds for all $i \in G_h$.

■

Lemma 20 (One-step propagation along a weak-tie edge). *Suppose Assumptions ?? and 1 hold. Fix $\theta = 1$ and let G_g be an H -seed from some time t_0 onward. Let G_h satisfy $W_{h \leftarrow g} \geq \omega$, and assume L satisfies (17)–(18). Then conditional on $\theta = 1$, with probability one there exists a finite (random) time $T \geq t_0$ such that G_h becomes an H -seed from time T onward.*

Proof of Lemma 20

Work conditional on $\theta = 1$ and on the event that G_g is an H -seed for all $t \geq t_0$.

Partition time into disjoint windows of length L starting at t_0 :

$$W_m := \{t_0 + mL + 1, t_0 + mL + 2, \dots, t_0 + (m + 1)L\}, \quad m = 0, 1, 2, \dots$$

For each m , define the event

$$F_m := \bigcap_{i \in G_h} \bigcap_{\tau \in W_m} \{\Delta_{i,\tau} \geq \delta\}.$$

By Assumption ??, conditional on $\theta = 1$ we have $\Pr(F_m \mid \theta = 1) \geq p_+^{L|G_h|} =: q_h \in (0, 1)$. Because the windows are disjoint and private signals are independent conditional on θ (Assumption ??), the events $\{F_m\}_{m \geq 0}$ are independent conditional on $\theta = 1$. Hence

$$\Pr\left(\bigcap_{m=0}^{\infty} F_m^c \mid \theta = 1\right) = \lim_{M \rightarrow \infty} \prod_{m=0}^{M-1} \Pr(F_m^c \mid \theta = 1) \leq \lim_{M \rightarrow \infty} (1 - q_h)^M = 0.$$

Therefore, almost surely there exists m^* such that F_{m^*} occurs. Let $t^* := t_0 + m^*L$ and $T := t^* + L$. On F_{m^*} , community G_h receives a favorable L -block over $t^* + 1, \dots, t^* + L$, while G_g is an H -seed over $t^*, \dots, t^* + L - 1$. Lemma 19 implies $Q(\ell_{i,T}) = k^A$ for all $i \in G_h$, and then Lemma 16 implies G_h remains an H -seed forever.

■

Lemma 21 (Replication completes in $O(\log M)$ windows). *Maintain Assumptions ??, ??, ??, and ??.* Fix $\theta = 1$ and suppose an H -seed exists from some time t_0 onward. Let

$$q := p_+^{L\bar{s}} \in (0, 1),$$

where L is the incubation length used in Theorem 2. Then for every integer $m \geq 1$,

$$\Pr\left(\text{all communities are } H\text{-seeds by time } t_0 + \bar{D}_C m L \mid \theta = 1\right) \geq 1 - \bar{D}_C M(n) (1 - q)^m.$$

Proof of Lemma 21

Work conditional on $\theta = 1$ and on the event that at least one community is an H -seed from time t_0 onward. Let $\mathcal{S}_0 \subseteq \{1, \dots, M(n)\}$ be the nonempty set of source communities that are H -seeds from time t_0 onward.

Step 1 (Layers and phases). For each community index h , let $\text{dist}(h)$ be the length of the shortest directed path in $\mathcal{C}^{(n)}$ from some source $g \in \mathcal{S}_0$ to h . Define distance layers

$$\mathcal{L}_d := \{h : \text{dist}(h) = d\}, \quad d = 0, 1, \dots, D_C(n).$$

By Assumption ??, $D_C(n) \leq \bar{D}_C$.

We allocate one *phase* of length mL for each layer. Phase d (for $d = 1, \dots, \bar{D}_C$) covers

$$[t_0 + (d - 1)mL, \quad t_0 + dmL - 1],$$

which contains m disjoint windows of length L . During phase d , we attempt to convert all communities in \mathcal{L}_d using already-converted communities in \mathcal{L}_{d-1} as pull sources.

Step 2 (Failure probability for one community in one phase). Fix $d \geq 1$ and $h \in \mathcal{L}_d$. In phase d , define U_h as the event that community G_h experiences at least one uniformly favorable L -block (all $\Delta \geq \delta$ for all agents in G_h over the block). By Assumptions ?? and ??, each window succeeds with probability at least $q := p_+^{L\bar{s}}$, and disjoint windows are independent conditional on $\theta = 1$. Therefore,

$$\Pr(U_h^c \mid \theta = 1) \leq (1 - q)^m.$$

Step 3 (Union bound over layers and conclusion). Let $E := \bigcap_{h \notin \mathcal{S}_0} U_h$. If E occurs, then by induction over d every layer \mathcal{L}_d converts to H -seeds by the end of phase d : the inductive step uses that each $h \in \mathcal{L}_d$ has some in-neighbor $g \in \mathcal{L}_{d-1}$ with $W_{h \leftarrow g} \geq \omega$, and then applies Lemma 19 on the favorable-block window guaranteed by U_h . Hence, on E , all communities are H -seeds by time $t_0 + \bar{D}_C mL$.

Finally,

$$\Pr(E^c \mid \theta = 1) \leq \sum_{d=1}^{\bar{D}_C} \sum_{h \in \mathcal{L}_d} \Pr(U_h^c \mid \theta = 1) \leq \bar{D}_C M(n) (1 - q)^m,$$

which yields the claimed bound. ■

C.6 Supporting results for the non-monotonicity theorem

Too much connectivity: homogenization traps bound success away from one

The fast-mixing regime in Section 4.2 yields a nonvanishing-probability lock-in of a large core into an absorbing (ceiling) bin \hat{k} (Theorem 1). If \hat{k} differs from the state-aligned bin, then state-aligned absorption is impossible on that lock-in event.

Proposition 7 (Too much connectivity prevents state-aligned absorption). *Maintain the assumptions of Theorem 1 conditional on $\theta = 1$ (in particular, the fast-mixing regularity and the entry condition (7)). Suppose additionally that the ceiling bin \hat{k} differs from the*

state-aligned bin:

$$\hat{k} \neq k^A.$$

Then there exist constants $c \in (0, 1)$ and n_0 such that for all $n \geq n_0$,

$$\Pr(\mathcal{L}_1^{(n)} \mid \theta = 1) \leq 1 - c.$$

Analogously, if the hypotheses of Theorem 1 hold conditional on $\theta = 0$ with $\hat{k} \neq k^B$, then there exist $c \in (0, 1)$ and n_0 such that $\Pr(\mathcal{L}_0^{(n)} \mid \theta = 0) \leq 1 - c$ for all $n \geq n_0$.

Proof of Proposition 7

We prove the $\theta = 1$ statement; the $\theta = 0$ statement is identical.

Work conditional on $\theta = 1$. By Theorem 1, there exist constants $c \in (0, 1)$, a finite time $T < \infty$, and n_0 such that for all $n \geq n_0$,

$$\Pr\left(\exists C \subseteq N^{(n)} \text{ with } |C| \geq (1 - \kappa_\pi \varepsilon)n \text{ s.t. } m_{i,t} = \hat{k} \forall i \in C, \forall t \geq T \mid \theta = 1\right) \geq c.$$

Let $E^{(n)}$ denote the event in parentheses. On $E^{(n)}$, at least one agent (indeed, a linear fraction) sends message \hat{k} at all times $t \geq T$. Since $\hat{k} \neq k^A$, it is impossible that eventually *all* agents send k^A forever. Formally, $E^{(n)} \subseteq (\mathcal{L}_1^{(n)})^c$. Therefore,

$$\Pr(\mathcal{L}_1^{(n)} \mid \theta = 1) \leq 1 - \Pr(E^{(n)} \mid \theta = 1) \leq 1 - c,$$

for all $n \geq n_0$.

■

Interpretation: in sufficiently fast-mixing societies, dispersion collapses before the society can reliably generate threshold crossings. With nonvanishing probability, a large core becomes trapped in a ceiling bin that is not state-aligned, bounding the probability of full learning away from one.

Too little connectivity: infeasible propagation and slow replication

The “too little” extreme has two distinct failure modes.

(A) **Infeasible propagation (pull too weak).** Even if a correct seed exists somewhere, inter-community influence can be so weak that downstream communities can never reach the high bin: even the *best-case* anchored target remains below the high cutpoint on every sample path.

(B) Slow replication (saturation loses the race). Even when propagation is feasible in principle, replication can be so slow that a wrong seed forms somewhere before the correct seed saturates the federation. Since seeds are absorbing (Lemma 16), a single wrong seed rules out state-aligned absorption.

(A) Infeasible propagation under insufficient inter-community pull. Recall the minimal row-wise influence weight between communities $g \neq h$:

$$W_{h \leftarrow g} := \min_{i \in G_h} \sum_{j \in G_g} P_{ij}.$$

Proposition 8 (Propagation infeasible under insufficient inter-community pull). *Maintain bounded evidence (Assumption 1) and the uniform belief envelope (Lemma 15). Fix distinct communities $g \neq h$ and suppose that throughout all times $t \geq 0$, community G_g is an H -seed, i.e. $m_{j,t} = k^A$ for all $j \in G_g$.*

Let $w := W_{h \leftarrow g} = \min_{i \in G_h} \sum_{j \in G_g} P_{ij}$. Assume that the best-case one-step target faced by agents in G_h satisfies

$$T_H^{\max}(w) := w r_{k^A} + (1 - w) \bar{r} + \bar{\Delta} < b_{k^A-1}. \quad (19)$$

Here $T_H^{\max}(w)$ is the largest value the one-step target $S_{i,t} + \Delta_{i,t+1}$ can attain for agents in G_h , even if all non- G_g neighbors send the maximal reconstruction \bar{r} and the private shock equals $\bar{\Delta}$.

Then there exists a deterministic time $T^* < \infty$ such that for every $i \in G_h$ and every $t \geq T^*$,

$$\ell_{i,t} < b_{k^A-1} \quad \text{and hence} \quad m_{i,t} \neq k^A.$$

In particular, G_h cannot become an H -seed from any finite time onward.

Proof of Proposition 8

Fix $i \in G_h$.

Step 1 (Uniform one-step upper bound). Since G_g is an H -seed, $\psi(\ell_{j,t}) = r_{k^A}$ for all $j \in G_g$ and all t . Split $S_{i,t} = \sum_j P_{ij} \psi(\ell_{j,t})$:

$$S_{i,t} = \sum_{j \in G_g} P_{ij} r_{k^A} + \sum_{j \notin G_g} P_{ij} \psi(\ell_{j,t}) \leq \left(\sum_{j \in G_g} P_{ij} \right) r_{k^A} + \left(1 - \sum_{j \in G_g} P_{ij} \right) \bar{r},$$

using $\psi(\cdot) \leq \bar{r}$. Let $x_i := \sum_{j \in G_g} P_{ij} \geq w$. Since $r_{k^A} \leq \bar{r}$, the map $x \mapsto xr_{k^A} + (1-x)\bar{r}$ is weakly decreasing in x , hence

$$S_{i,t} \leq x_i r_{k^A} + (1-x_i)\bar{r} \leq wr_{k^A} + (1-w)\bar{r} \quad \text{for all } t.$$

By bounded evidence, $\Delta_{i,t+1} \leq \bar{\Delta}$ a.s. Hence for all t ,

$$S_{i,t} + \Delta_{i,t+1} \leq wr_{k^A} + (1-w)\bar{r} + \bar{\Delta} = T_H^{\max}(w) < b_{k^A-1},$$

where the last inequality is (19).

Step 2 (Solve the dominated recursion). Substituting the bound from Step 1 into the anchored update yields

$$\ell_{i,t+1} = (1-\alpha)\ell_{i,t} + \alpha(S_{i,t} + \Delta_{i,t+1}) \leq (1-\alpha)\ell_{i,t} + \alpha T_H^{\max}(w).$$

Iterating gives, for every $t \geq 0$,

$$\ell_{i,t} \leq (1-\alpha)^t \ell_{i,0} + (1 - (1-\alpha)^t) T_H^{\max}(w).$$

By Lemma 15, $\ell_{i,0} \leq \bar{\ell}$ for all i , hence

$$\ell_{i,t} \leq T_H^{\max}(w) + (1-\alpha)^t (\bar{\ell} - T_H^{\max}(w)).$$

Since $T_H^{\max}(w) < b_{k^A-1}$, define the gap $\gamma := b_{k^A-1} - T_H^{\max}(w) > 0$. Choose T^* such that

$$(1-\alpha)^{T^*} (\bar{\ell} - T_H^{\max}(w)) < \gamma.$$

Then for all $t \geq T^*$,

$$\ell_{i,t} < T_H^{\max}(w) + \gamma = b_{k^A-1}.$$

Step 3 (Conclude infeasibility of an H -seed in G_h). For $t \geq T^*$, every agent $i \in G_h$ satisfies $\ell_{i,t} < b_{k^A-1}$ and therefore $m_{i,t} \neq k^A$. Thus G_h cannot be an H -seed from any finite time onward.

■

(B) Slow replication loses the race to wrong seeds. To obtain a negative learning result from slow replication, we need a strictly positive lower bound on the probability of

wrong-direction blocks. Assumption ?? controls wrong-direction tails from above via p_- . We now add a matching lower bound.

Assumption 8 (Nondegenerate wrong-direction blocks (lower bound)). *In addition to Assumption ??, there exists $\underline{p}_- > 0$ such that for all i, t ,*

$$\Pr(\Delta_{i,t} \leq -\delta \mid \theta = 1) \geq \underline{p}_-, \quad \Pr(\Delta_{i,t} \geq \delta \mid \theta = 0) \geq \underline{p}_-.$$

Proposition 9 (Slow replication implies wrong seeds appear before saturation). *Maintain Assumptions ??, 1, ??, ??, ??, ??, and 8. Fix $\theta = 1$ and assume the symmetric seed-incubation inequalities hold for k^B under $\theta = 1$, with incubation length L .*

Let $T^{(n)}$ be any deterministic horizon and define the number of disjoint length- L windows

$$J(n) := \left\lfloor \frac{T^{(n)}}{L} \right\rfloor.$$

If

$$M(n) J(n) \underline{p}_-^{L\bar{s}} \rightarrow \infty, \tag{20}$$

then

$$\Pr\left(\text{some community becomes an } L\text{-seed by time } T^{(n)} \mid \theta = 1\right) \rightarrow 1,$$

and consequently

$$\Pr(\mathcal{L}_1^{(n)} \mid \theta = 1) \rightarrow 0.$$

A symmetric statement holds under $\theta = 0$ (with H and L interchanged).

Proof of Proposition 9

Work conditional on $\theta = 1$.

Step 1 (Independent wrong-direction trials). Partition $\{1, \dots, T^{(n)}\}$ into $J(n)$ disjoint windows of length L : for $m = 1, \dots, J(n)$, let

$$W_m := \{(m-1)L+1, \dots, mL\}.$$

For each community $g \leq M(n)$ and window $m \leq J(n)$ define

$$E_{g,m} := \left\{ \Delta_{i,t} \leq -\delta \text{ for all } i \in G_g \text{ and all } t \in W_m \right\}.$$

Because communities are disjoint and windows W_m are disjoint, the events $\{E_{g,m}\}_{g,m}$ depend on disjoint collections of private signals. Conditional independence in Assumption ??

therefore implies that the events $\{E_{g,m}\}$ are mutually independent conditional on $\theta = 1$.

Moreover, by Assumption 8, for each (g, m) ,

$$\Pr(E_{g,m} \mid \theta = 1) = \prod_{i \in G_g} \prod_{t \in W_m} \Pr(\Delta_{i,t} \leq -\delta \mid \theta = 1) \geq (\underline{p}_-)^{L|G_g|} \geq (\underline{p}_-)^{L\bar{s}},$$

using $|G_g| \leq \bar{s}$ (Assumption ??).

Step 2 (A wrong-direction block creates a permanent wrong seed). Fix (g, m) and suppose $E_{g,m}$ occurs. Then every agent in G_g receives $\Delta \leq -\delta$ for L consecutive periods during W_m . By the symmetric version of Lemma 17 for k^B under $\theta = 1$ (assumed to hold), the community enters bin k^B by the end of W_m . Lemma 16 then implies that G_g remains in bin k^B forever, i.e. G_g becomes an L -seed.

Therefore, on $\bigcup_{g \leq M(n)} \bigcup_{m \leq J(n)} E_{g,m}$, at least one community becomes an L -seed by time $T^{(n)}$.

Step 3 (High probability that some wrong seed appears, and misalignment). Let $E_{\text{none}} := \bigcap_{g \leq M(n)} \bigcap_{m \leq J(n)} E_{g,m}^c$. By independence and Step 1,

$$\Pr(E_{\text{none}} \mid \theta = 1) = \prod_{g \leq M(n)} \prod_{m \leq J(n)} (1 - \Pr(E_{g,m} \mid \theta = 1)) \leq (1 - (\underline{p}_-)^{L\bar{s}})^{M(n)J(n)} \leq \exp(-M(n)J(n)(\underline{p}_-)^{L\bar{s}}),$$

using $1 - x \leq e^{-x}$. Under (20), the exponent diverges, so $\Pr(E_{\text{none}} \mid \theta = 1) \rightarrow 0$ and hence

$$\Pr\left(\exists \text{ an } L\text{-seed by time } T^{(n)} \mid \theta = 1\right) \rightarrow 1.$$

If any community becomes an L -seed at a finite time, it remains an L -seed forever (Lemma 16). Under $\theta = 1$, $\mathcal{L}_1^{(n)}$ requires eventual absorption into k^A , which is impossible if any community is permanently in $k^B \neq k^A$. Therefore,

$$\Pr(\mathcal{L}_1^{(n)} \mid \theta = 1) \leq \Pr(E_{\text{none}} \mid \theta = 1) \rightarrow 0.$$

■

Intermediate connectivity: a segregated federation yields learning

Section 5.1 provides sufficient conditions under which (i) a correct seed forms quickly with high probability, (ii) it replicates throughout the federation, and (iii) the probability that

any wrong seed forms before saturation vanishes. We restate that conclusion directly in terms of $\mathcal{L}_\theta^{(n)}$.

Corollary 2 (A segregated federation yields state-aligned absorption w.h.p.). *Maintain the assumptions of Corollary ??.* Then conditional on $\theta = 1$,

$$\Pr(\mathcal{L}_1^{(n)} \mid \theta = 1) \rightarrow 1,$$

and symmetrically conditional on $\theta = 0$,

$$\Pr(\mathcal{L}_0^{(n)} \mid \theta = 0) \rightarrow 1.$$

Proof of Corollary 2

We prove the case $\theta = 1$; the case $\theta = 0$ is symmetric.

By Corollary ??, conditional on $\theta = 1$, with probability tending to 1 there exists a finite (random) time T such that $m_{i,t} = k^A$ for all agents i and all $t \geq T$. Since $k^A = k^1$ by definition, this event is precisely $\mathcal{L}_1^{(n)}$. Therefore, $\Pr(\mathcal{L}_1^{(n)} \mid \theta = 1) \rightarrow 1$.

■