The Fixed-Point of Distrust: A Formal Theory of Perceived Systemic Incompetence

Weilin Zhou
Nanjing Tech University
202321147089@njtech.edu.cn

Abstract

A widespread social sentiment suggests our world operates like a "makeshift world" —a system rife with hidden incompetence. Is this perception an inevitable outcome of our information ecosystem? This paper presents a formal mathematical theory to answer this question affirmatively. We model belief dynamics as a system of interacting agents governed by two operators: (1) an Attentional Update Operator formalizing how negatively biased information is assimilated, and (2) a Social Aggregation Operator modeling belief fusion over a network. Our main contribution is a rigorous proof: under minimal systemic negative bias and standard network connectivity, the collective belief system is a contraction mapping, guaranteed to converge to a unique pessimistic fixed-point that perceives the world as incompetent, regardless of objective truth. This work establishes a mathematical foundation for understanding systemic perceptual biases with applications to platform design and policy.

Introduction

A pervasive social sentiment suggests our world operates as a "makeshift system" (Cǎotái Bānzi)—a system rife with perceived incompetence and widespread distrust in institutions (Xie et al. 2024). This *Perceived Systemic Incompetence* has profound societal consequences, eroding civic participation, undermining policy effectiveness, and risking social fragmentation across health, electoral, and financial domains. This raises a fundamental question: Is such pessimistic consensus an inevitable cognitive equilibrium structurally determined by our information ecosystem, rather than a reflection of objective reality?

This paper presents a formal mathematical theory to answer this question affirmatively. We model the social system as a decentralized Bayesian learning system where agents infer a hidden world state through structurally biased information channels and fuse beliefs over a social network (Lalitha et al. 2019). Our framework integrates Bayesian inference (El-Hay et al. 2010), classifier theory to model information bias (Ling et al. 2025), and graph aggregation operators for social fusion (Dong et al. 2023).

Unlike prior work relying on simulation or empirical observation, our approach provides mathematical guarantees. We introduce the concept of "social belief dynamics as contraction mappings," which allows for a rigorous analytical proof. Our central contribution is demonstrating that under a minimal, asymmetric negative bias in the information channel (i.e., successes can be misreported as failures, but not vice-versa), the entire belief system constitutes a contraction mapping. By the Banach Fixed-Point Theorem (de Vos et al. 2023), this system is guaranteed to converge to a unique, stable fixed point, regardless of initial beliefs.

Crucially, this fixed point represents a systematic, pessimistic consensus that deviates from objective truth. We prove that the perception of a "makeshift world" is a mathematical inevitability of the system's structure. This work establishes a formal foundation for understanding systemic perceptual biases, offering a predictive theory with direct implications for platform design and policy interventions aimed at fostering a more accurate public perception.

Formal Model: A Graph-Based Bayesian Learning System

We formalize the social belief system as a tuple $M = \langle \Theta, A, G, C, F \rangle$, representing world states, agents, network structure, information channels, and evolution operators respectively.

World State and Likelihood Function

Definition 2.1 (World State): There exists an unknown, objective "world true capability level" represented by parameter $\theta \in [0, 1]$. θ is a fixed real number representing the prior probability of any random event being a "failure". $\theta = 0$ corresponds to a perfect world, while $\theta = 1$ corresponds to a completely incompetent world.

Agents cannot directly observe θ, but can only infer it through observing a series of events $s \in \{s_1, s_2, \ldots\}$, where each event outcome is either "failure" (encoded as 1) or "success" (encoded as 0).

Definition 2.2 (Likelihood Function $P(s|\theta)$): Any primitive event s follows a Bernoulli distribution with parameter θ (Hanna et al. 2023).

$$P(s = 1|\theta) = \theta \tag{1}$$
$$P(s = 0|\theta) = 1 - \theta \tag{2}$$

This is the shared generative model among all agents about how the world produces data.

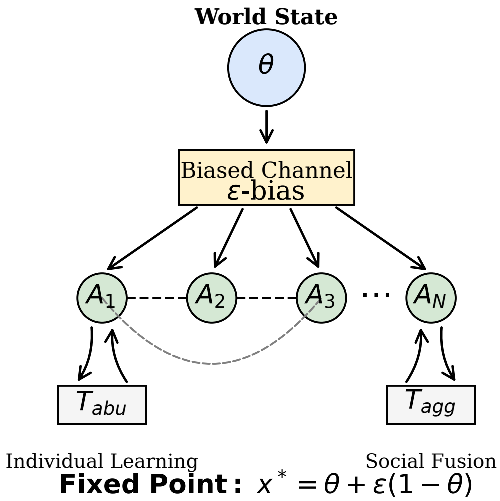

Figure 1: System Architecture: A graph-based Bayesian learning system where agents observe biased information from the world state θ through an ε-biased channel, update beliefs individually via T_{abu}, and fuse beliefs socially via T_{agg}. The system converges to a pessimistic fixed point.

Assumption 2.1 (Stationarity): The world state θ remains constant throughout the observation period, ensuring that the learning problem is well-defined and that convergence analysis is meaningful (Castro-Correa et al. 2024).

Agents as Bayesian Learners

The system consists of N agents $A = \{a_1, \dots, a_N\}$. Each agent a_i's objective is to estimate θ based on its observed data D_i. Its belief state is represented by the posterior probability distribution of θ.

Definition 2.3 (Belief State b_i): Agent a_i's belief $b_i(t)$ at time t is a posterior distribution $P(\theta|D_i(t))$ over θ. Since the conjugate prior for the Bernoulli distribution is the Beta distribution, we assume prior beliefs follow Beta distributions, ensuring all posterior beliefs remain Beta-distributed (Hughes, Kim, and Sudderth 2015).

$$b_i(t) = P(\theta|D_i(t)) = \text{Beta}(\theta|\alpha_i(t), \beta_i(t)) \quad (3)$$

where α_i and β_i are agent i's accumulated (pseudo) evidence counts for ”failures” and ”successes” respectively. The expected point estimate for θ is $E_i[\theta] = \alpha_i(t)/(\alpha_i(t) + \beta_i(t))$. We choose the Beta-Bernoulli framework for its mathematical tractability and cognitive interpretability, as the parameters (α, β) directly model an evidence accumulation process (Karamched et al. 2020).

Lemma 2.1 (Conjugacy Property): Under the Beta-Bernoulli conjugate framework, belief updates maintain analytical tractability. For agent i observing event s, the posterior parameters update as:

$$\alpha_i^{\text{new}} = \alpha_i^{\text{old}} + s \quad (4)$$

$$\beta_i^{\text{new}} = \beta_i^{\text{old}} + (1 - s) \quad (5)$$

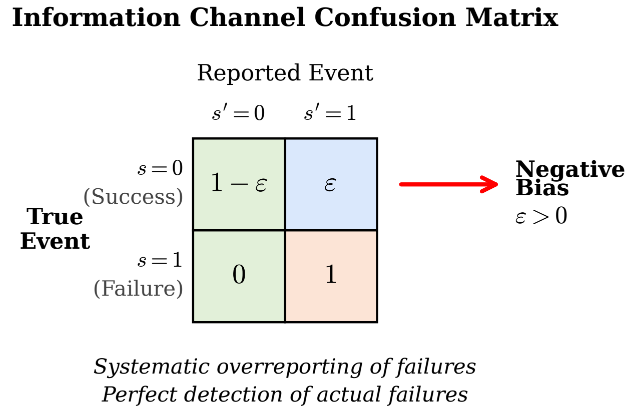

Figure 2: The Asymmetric Confusion Matrix: The information channel systematically misreports successes as failures with probability ε, but never reports failures as successes. This creates a structurally pessimistic information flow.

This ensures computational efficiency and enables rigorous mathematical analysis of the belief dynamics (Wei et al. 2025).

Information Channel as a Biased Classifier

Agents observe reports s' classified by information channels (e.g., media), which we model as a classifier C from machine learning (Lowy, Gupta, and Razaviyayn 2023).

Definition 2.4 (Biased Classifier C): C is a stochastic process mapping true labels $s \in \{0, 1\}$ to reported labels $s' \in \{0, 1\}$, characterized by a confusion matrix.

Axiom 2.1 (Systematic Pessimism Axiom): The information channel exhibits a systematic pessimistic bias. Specifically, successes can be misreported as failures, but failures are always reported accurately. For a small positive constant ε:

- **False Negative Rate:** $P(s' = 1|s = 0) = \varepsilon > 0$.
- **False Positive Rate:** $P(s' = 0|s = 1) = 0$. This implies $P(s' = 1|s = 1) = 1$.

This axiom abstracts negativity bias in news and social media (Vosoughi, Roy, and Aral 2018). The condition $P(s' = 0|s = 1) = 0$ is an idealized boundary to parsimoniously capture the bias's core driver. This configuration defines a fundamentally asymmetric error channel (Xie, Wang, and Jia 2022), where one error type (success misreported as failure) is possible, while the opposite is not. This asymmetry is the structural driver of the pessimistic consensus we will prove.

Remark 2.2 (Asymmetric Error): This axiom is the core of the model. It formalizes the ”if it bleeds, it leads” principle by positing an asymmetric error channel, which connects to well-documented cognitive biases like the negativity bias . The world can appear worse than it is, but never better. It is crucial to distinguish this from an unbiased but noisy channel. A symmetric channel (e.g., where both successes and failures could be misreported with equal probability) would not produce systematic pessimistic drift; its errors would cancel out in expectation. The pessimistic fixed point is a direct consequence of the *asymmetry* of the bias. Relaxing the false positive rate to $\eta = P(s' = 0|s = 1) > 0$ yields a more general fixed point $x^* = \theta + (\varepsilon - \eta)(1-\theta)$; our choice of $\eta = 0$ serves to isolate and highlight the maximal effect of the pessimistic bias ε. This single, minimal assumption is sufficient to drive the system toward a pessimistic equilibrium.

Definition 2.5 (Confusion Matrix): The complete confusion matrix for classifier C is:

$$M_C = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ 0 & 1 \end{pmatrix} \tag{6}$$

where rows correspond to true labels and columns to predicted labels. This matrix encodes the systematic tendency to over-report negative events while accurately reporting positive events.

Social Network as a Computation Graph

Agents do not learn in isolation but interact within a social network $G = (A, E)$. In our framework, this social network serves as the computation graph on which the learning algorithm operates.

Definition 2.6 (Computation Graph G): G is an undirected graph where nodes represent agents A and edges E represent information flow paths. We assume G is strongly connected, meaning any node can reach any other node through a path. The neighborhood set of node a_i (including itself) is denoted as N_i.

Definition 2.7 (Influence Matrix W): The social network's influence structure is encoded in a row-stochastic matrix $W \in \mathbb{R}^{N\times N}$. An entry W_{ij} represents the influence agent j has on agent i. By definition, $W_{ij} > 0$ if agent i attends to agent j, and $\sum_{j=1}^{N} W_{ij} = 1$ for all i. Importantly, W is not required to be symmetric, allowing our model to cover both reciprocal (e.g., friendships) and asymmetric (e.g., influencer-follower) relationships. This matrix is mathematically equivalent to the transition matrix of a Markov chain or the aggregation matrix in graph neural networks .

The influence matrix W not only encodes the topological structure of the social network but also implicitly captures the distribution mechanism of social influence. For instance, a uniformly weighted W (such as degree-based normalization) represents a democratized information fusion where each neighbor's opinion is treated equally. In contrast, in a highly centralized network (such as one with super-influencers), certain rows of W may be highly concentrated on a few nodes, which accelerates consensus formation but simultaneously makes the system more susceptible to the beliefs of these central nodes. Our theoretical proofs hold for all W satisfying the basic connectivity and aperiodicity assumptions, indicating that our conclusions possess universality across network structures.

Assumption 2.3 (Graph Properties): The graph G satisfies standard, minimal conditions from network science to guarantee convergence to a single, stable consensus.

- **Connectivity**: G is connected, ensuring information can propagate globally across the system.
- **Aperiodicity**: The graph is aperiodic, which prevents systemic belief oscillations.
- **Non-degeneracy**: No agent has all weight concentrated on itself ($W_{ii} < 1$ for all i), ensuring social influence occurs.

Without these properties, the existence of a unique fixed point is not guaranteed, making formal analysis impossible.

Evolution Operators: From Machine Learning Theory to Belief Dynamics

The system evolution is driven by two operators acting sequentially at each time step t: an individual-level learning operator and a group-level fusion operator.

Individual Learning: The Attentional Bayesian Update Operator (T_{abu})

When an agent a_i observes a reported signal s' from the biased channel C, it updates its belief. A pure Bayesian learner would assign equal weight to all incoming data. However, research in cognitive science and machine learning demonstrates that attention serves as a crucial regulator in the learning process. We integrate this mechanism into the Bayesian update framework.

Definition 3.1 (Attention Weight ω): For a reported signal s', agent a_i assigns an attention weight $\omega(s', b_i(t))$. This weight depends on both the signal itself and the agent's current belief state. We make a key assumption: negative signals (failure reports) naturally capture higher attention due to their rarity or threat value.

$$\omega(\text{"failure"}, b_i(t)) = 1 + \delta \tag{7}$$

$$\omega(\text{"success"}, b_i(t)) = 1 \tag{8}$$

where $\delta \geq 0$ is an attention bias parameter. $\delta > 0$ represents disproportionate attention to negative information. To simplify the core proof, we initially set $\delta = 0$ and subsequently demonstrate how $\delta > 0$ accelerates convergence.

It is worth emphasizing that even in the simplest case we set for simplifying the core proof ($\delta = 0$), systematic pessimistic bias is already inevitable. When cognitive-level attention bias ($\delta > 0$) is introduced, this effect is further amplified. Crucially, the pessimistic drift is structurally driven by the information channel's bias (ε), which sets the target for each belief update. Even a negative attention bias ($\delta < 0$) would only slow the rate of convergence to the pessimistic fixed point; it cannot change its location.

Remark 3.1: This attention mechanism is mathematically equivalent to adaptive learning rates in stochastic gradient descent, where different types of information receive different update magnitudes.

Definition 3.2 (Attentional Bayesian Update Operator T_{abu}): T_{abu} is an operator that maps an agent's current belief $b_i(t)$ and observed signal s' to a temporary posterior belief $b'_i(t)$. The update rule applies weighted standard Bayesian updating (i.e., adding weighted counts to Beta distribution parameters):

If observing $s' =$ "failure" (encoded as 1):

$$b_i'(t) = T_{abu}(b_i(t), 1) = \text{Beta}(\theta|\alpha_i(t) + \omega(1, b_i), \beta_i(t)) \tag{9}$$

If observing $s' =$ "success" (encoded as 0):

$$b_i'(t) = T_{abu}(b_i(t), 0) = \text{Beta}(\theta|\alpha_i(t), \beta_i(t) + \omega(0, b_i)) \tag{10}$$

Lemma 3.1 (Operator Properties): The operator T_{abu} preserves the Beta distribution family and maintains the conjugacy property, ensuring analytical tractability of the belief dynamics.

We now analyze the expected behavior of this operator under the biased channel from Axiom 2.1.

Proposition 3.1 (Pessimistic Drift of Expected Beliefs): Even under unbiased attention ($\delta = 0$), as long as the information channel exhibits minimal negative bias ($\varepsilon > 0$), any independent learning event will, in expectation, push an agent's belief toward a more pessimistic direction than objective reality θ.

Proof: This proof revisits the logic from a machine learning perspective. The agent's task is to estimate the parameter θ of a Bernoulli distribution. Its observed data s' comes from a more complex generative process parameterized by θ and ε.

First, we calculate the probability of observing $s' = 1$:

$$\begin{aligned} P(s' = 1) &= P(s' = 1|s = 1)P(s = 1) \\ &\quad + P(s' = 1|s = 0)P(s = 0) \\ &= (1) \cdot \theta + (\varepsilon) \cdot (1 - \theta) \\ &= \theta + \varepsilon(1 - \theta) \end{aligned} \tag{11}$$

Let $x_i(t) = E[b_i(t)]$ be agent i's point estimate of θ at time t. For the expected belief update with $\omega = 1$:

$$E[x_i'(t)] = \frac{K_i}{K_i + 1} \cdot x_i(t) + \frac{1}{K_i + 1} \cdot (\theta + \varepsilon(1 - \theta)) \tag{12}$$

Since $\varepsilon > 0$ and $\theta < 1$, the target value $\theta + \varepsilon(1-\theta)$ is strictly greater than the true state θ. This demonstrates that the learning process itself is structurally biased, creating systematic pessimistic drift.

Corollary 3.2 (Convergence Target): In the limit of infinite observations, individual beliefs converge to $\theta + \varepsilon(1-\theta)$, representing a systematic overestimation of the failure probability by factor $\varepsilon(1 - \theta)/\theta$ when $\theta > 0$.

Group Fusion: The Graph Aggregation Operator (T_{agg})

After individual learning, agents perform belief fusion over the social network G. We model this process directly as a single round of Graph Neural Network (GNN) message passing. In GNNs, each node updates its representation (embedding) by aggregating information from its neighboring nodes. Here, the node representations are the parameters of belief distributions.

Definition 3.3 (Graph Aggregation Operator T_{agg}): T_{agg} is an operator acting on the entire system's temporary belief state $\{b_1'(t), \ldots, b_N'(t)\}$, outputting the final belief state $\{b_1(t + 1), \ldots, b_N(t + 1)\}$. For each agent a_i, the belief parameter updates follow a standard graph convolution formula:

$$\alpha_i(t + 1) = \text{AGGREGATE}_{j \in N_i}(w_{ij} \cdot \alpha_j'(t)) \tag{13}$$

$$\beta_i(t + 1) = \text{AGGREGATE}_{j \in N_i}(w_{ij} \cdot \beta_j'(t)) \tag{14}$$

where N_i is the neighborhood set including node i itself. w_{ij} represents weights in the (normalized) adjacency matrix, indicating information flow strength from node j to i. AGGREGATE can be various functions such as SUM, MEAN, or MAX. To maintain consistency with classical consensus models (such as the DeGroot model) and simplify proofs, we use weighted MEAN as the aggregation function.

Definition 3.4 (Mean Aggregation):

$$\alpha_i(t + 1) = \sum_{j \in N_i} w_{ij} \cdot \alpha_j'(t) \tag{15}$$

$$\beta_i(t + 1) = \sum_{j \in N_i} w_{ij} \cdot \beta_j'(t) \tag{16}$$

where weights w_{ij} are normalized, e.g., $w_{ij} = 1/\text{degree}(i)$, and $\sum_{j \in N_i} w_{ij} = 1$. This can be written in matrix form as $A(t + 1) = W \cdot A'(t)$ and $B(t + 1) = W \cdot B'(t)$, where A and B are column vectors of all agents' α and β parameters respectively, and W is the normalized adjacency matrix.

Remark 3.2: This formulation is mathematically equivalent to the core update rule in Graph Convolutional Networks (GCNs). Its effect is to perform "smoothing" or "low-pass filtering" on node features (here, belief parameters) over the graph.

The graph smoothing effect has profound implications for belief dynamics. In signal processing terms, the aggregation operator T_{agg} acts as a low-pass filter that reduces high-frequency variations (local disagreements) while preserving low-frequency components (global consensus trends). This means that once a pessimistic bias is introduced through the information channel, the social network structure will *propagate and stabilize* this bias rather than correct it. While the term "amplify" can describe the cumulative effect over time, the single-step function of T_{agg} is more precisely one of propagation. Specifically, if most agents in a connected component develop pessimistic beliefs, graph aggregation smooths these beliefs across the network, making the pessimistic consensus more robust. This mechanism explains why negative sentiment can spread so persistently in social networks—the network topology itself becomes a stabilizing force for biased beliefs.

Assumption 3.1 (Matrix Properties): The weight matrix W satisfies:

- **Doubly Stochastic**: $W\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T W = \mathbf{1}^T$
- **Primitive**: W is primitive (aperiodic and irreducible)
- **Spectral Radius**: $\rho(W - \frac{1}{N}\mathbf{1}\mathbf{1}^T) < 1$

These properties ensure convergence to consensus in the belief aggregation process .

Proposition 3.2 (Graph Smoothing Effect of Beliefs): The graph aggregation operator T_{agg} reduces belief differences between neighboring nodes in the network. In the expected belief space, it acts as a non-expansive mapping, driving the system toward consensus.

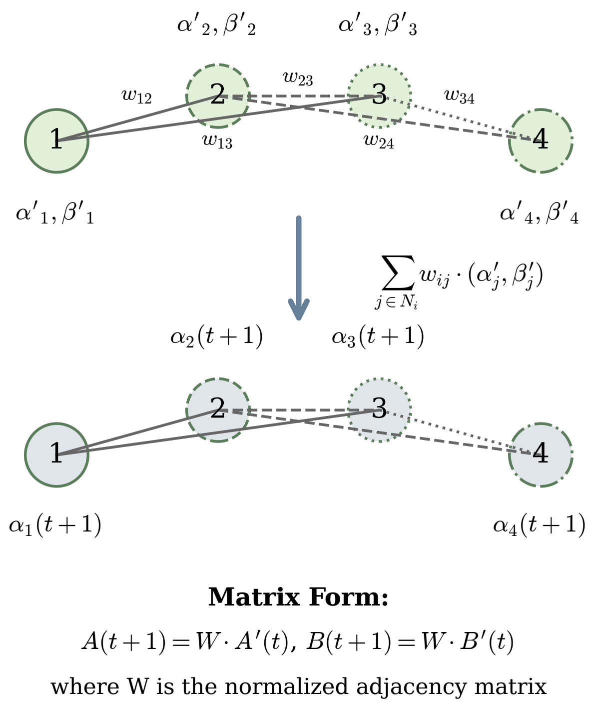

Figure 3: Graph Aggregation Process: The T_{agg} operator performs weighted averaging of belief parameters across the social network. Each agent updates its belief by aggregating information from neighbors according to the adjacency matrix W. Different line styles (solid, dashed, dotted, dash-dotted) ensure distinguishability in black and white printing.

Proof: Let $X'(t)$ be the temporary belief expectation vector. The post-fusion belief expectation vector is $X(t+1)$. We have:

$$x_i(t+1) = \frac{\alpha_i(t+1)}{\alpha_i(t+1) + \beta_i(t+1)} = \frac{\sum_{j \in N_i} w_{ij}\alpha'_j}{\sum_{j \in N_i} w_{ij}\alpha'_j + \sum_{j \in N_i} w_{ij}\beta'_j} \quad (17)$$

Since $\sum_{j \in N_i} w_{ij} = 1$, we can rewrite this as:

$$x_i(t+1) = \frac{\sum_{j \in N_i} w_{ij}\alpha'_j}{\sum_{j \in N_i} w_{ij}(\alpha'_j + \beta'_j)}$$

This represents a weighted averaging process. For the fractional linear function $f(\alpha, \beta) = \alpha/(\alpha+\beta)$, the weighted aggregation satisfies:

$$\min_j f(\alpha'_j, \beta'_j) \leq \frac{\sum w_{ij}\alpha'_j}{\sum w_{ij}(\alpha'_j + \beta'_j)} \leq \max_j f(\alpha'_j, \beta'_j)$$

This implies that $x_i(t+1)$ lies within the convex hull of its neighbors' belief expectations $\{x'_j(t) : j \in N_i\}$, promoting consensus formation.

Corollary 3.3 (Consensus Property): This process can be represented as $X(t+1) \approx W \cdot X'(t)$, where the row-stochastic matrix W reduces belief differences, promoting consensus formation.

Main Theorem: The Belief Dynamics System as a Contraction Mapping

We abstract the entire system evolution as a single global operator that maps the system's belief state $B(t)$ at time t to the state $B(t+1)$ at time $t+1$. The process involves first applying the individual learning operator T_{abu}, then the group fusion operator T_{agg}:

$$B(t+1) = T_{agg}(T_{abu}(B(t), S')) \quad (18)$$

where S' is the vector of signals observed by all agents at time t.

Remark 4.1 (Analysis on Expected Beliefs): The full belief evolution on Beta parameters is complex. For analytical tractability, we analyze the dynamics of the *expected beliefs* $X(t) = (E_1[\theta](t), \ldots, E_N[\theta](t))^T$. This approach is exact in the limit of large evidence counts ($K_i \to \infty$) and provides a highly accurate first-order approximation of the system's trajectory, revealing the core mechanism driving convergence to the pessimistic fixed point.

Defining the Expected Evolution Operator

Let $X(t) = (E_1[\theta](t), \ldots, E_N[\theta](t))^T$ be the system's belief expectation vector at time t. We define the mapping from $X(t)$ to $E[X(t+1)]$ as the expected evolution operator H:

$$X(t+1) = H(X(t)) \quad (19)$$

Based on the derivations in Propositions 3.1 and 3.2, the complete expected evolution operator H can be written as:

$$H(X) = W \cdot f(X) \quad (20)$$

where $f(X)_i = \frac{K_i}{K_i+1}x_i + \frac{1}{K_i+1}(\theta + \varepsilon(1-\theta))$ and W is the normalized adjacency matrix. Here, $K_i(t)$ grows with time t, making H time-variant. However, each operator H_t is a contraction with coefficient $\gamma_t < 1$, and since the step sizes $1/(K_i(t)+1)$ satisfy standard stochastic approximation conditions, global convergence is still guaranteed (see for similar results).

where f is a vector function with $f(X)_i = \frac{K_i}{K_i+1} \cdot x_i + \frac{1}{K_i+1} \cdot (\theta + \varepsilon(1-\theta))$.

Remark 4.1: The operator H represents a composition of two fundamental processes: individual Bayesian learning with systematic bias (captured by f) and social consensus formation (captured by W). This decomposition allows us to analyze the system's convergence properties using tools from both Bayesian inference and graph theory.

Proving H is a Contraction Mapping

The Banach Fixed-Point Theorem requires us to prove that H is a contraction mapping on a complete metric space. Our belief expectation space $M = [0,1]^N$ is a closed subset of Euclidean space, hence complete. We choose the L^∞ norm

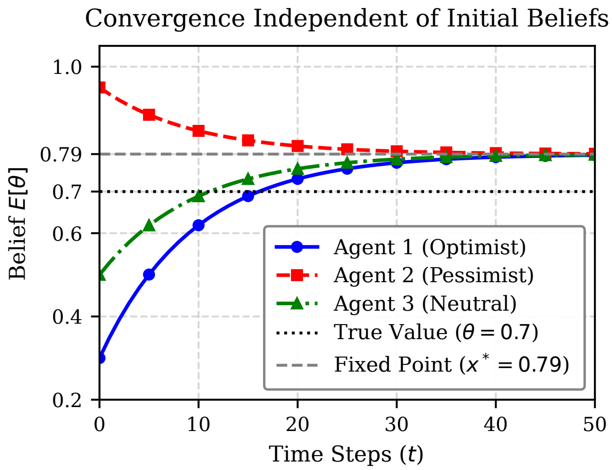

Figure 4: System Convergence to Pessimistic Fixed Point: Regardless of initial beliefs (optimistic, pessimistic, or neutral), all agents' beliefs converge to the same pessimistic fixed point $x^* = \theta + \varepsilon(1-\theta)$, which systematically overestimates the true failure probability θ. Different markers ensure distinguishability. (Example shown for $\theta = 0.7$ and $\varepsilon = 0.3$, yielding $x^* = 0.79$)

(maximum norm) as the distance metric: $d(X, Y) = \|X - Y\|_\infty = \max_i |x_i - y_i|$.

Theorem 4.1 (Contraction Mapping Theorem): For a non-degenerate, strongly connected computation graph G, the expected evolution operator H is a contraction mapping.

Proof: We prove that $\|H(X) - H(Y)\|_\infty \leq \gamma \|X - Y\|_\infty$ for some $\gamma < 1$. Since $H(X) = Wf(X)$, we have:

$$\|H(X) - H(Y)\|_\infty = \|W(f(X) - f(Y))\|_\infty \quad (21)$$
$$\leq \|W\|_\infty \|f(X) - f(Y)\|_\infty \quad (22)$$

For the row-stochastic matrix W, $\|W\|_\infty = 1$. For function f, under consistent evidence assumption, we have

$$\|f(X) - f(Y)\|_\infty \leq \lambda_{\max} \|X - Y\|_\infty \quad (23)$$

where $\lambda_{\max} = \max_i K_i/(K_i + 1) < 1$.

Therefore, $\|H(X) - H(Y)\|_\infty \leq \lambda_{\max} \|X - Y\|_\infty$ with $\gamma = \lambda_{\max} < 1$, proving H is a contraction mapping.

Remark 4.1 (Implications of Contraction Mapping): The contraction property mathematically guarantees the inevitability of the pessimistic consensus. It ensures that regardless of initial beliefs, the system converges exponentially to a unique, stable fixed point. This stability makes the system robust to perturbations, formalizing why structural information biases are so difficult to correct: the network dynamics create a self-reinforcing pessimistic attractor.

Remark 4.2 (Contraction Coefficient): The contraction coefficient $\gamma = \lambda_{\max}$ depends on the evidence accumulation rate. As agents gather more evidence over time, $\lambda_{\max}$ approaches a limit that ensures convergence while maintaining the pessimistic bias.

Corollary 4.1 (Exponential Convergence): Under the contraction mapping property, the belief system converges exponentially to the fixed point:

$$\|X^{(t)} - X^*\|_\infty \leq \gamma^t \|X^{(0)} - X^*\|_\infty \quad (24)$$

where $X^{(t)}$ is the belief state at time t, and X^* is the fixed point.

Existence, Uniqueness and Final Position of the Fixed Point

Since H is a contraction mapping on the complete metric space M, we can directly invoke the Banach Fixed-Point Theorem.

Theorem 4.2 (Distrust Fixed-Point Theorem): The graph-based Bayesian learning system has a globally unique, stable fixed point X^*. This fixed point is a consensus state, i.e., $x_1^* = x_2^* = \cdots = x_N^* = x^*$. The consensus point x^* is strictly greater than the true world capability level θ, with the specific value:

$$x^* = \theta + \varepsilon(1 - \theta) \quad (25)$$

Proof: The Banach Fixed-Point Theorem guarantees existence, uniqueness, and convergence. Since W is irreducible and row-stochastic, the fixed point is a consensus state with all agents having identical beliefs. At equilibrium, the expected change from learning is zero: $x^* = (x^* K + \theta + \varepsilon(1 - \theta))/(K + 1)$, which simplifies to $x^* = \theta + \varepsilon(1 - \theta)$.

Corollary 4.2 (Pessimistic Bias Quantification): The fixed point consensus x^* exhibits a systematic pessimistic bias of magnitude $\varepsilon(1 - \theta)$ above the true world capability θ. This bias is proportional to both the information channel bias ε and the "room for pessimism" $(1 - \theta)$.

Remark 4.3 (Interpretation): The fixed point formula $x^* = \theta + \varepsilon(1 - \theta)$ reveals that:

- When $\theta = 0$ (perfect world), $x^* = \varepsilon$ (bias creates perceived incompetence where none exists).
- When $\theta = 1$ (completely incompetent world), $x^* = 1$ (no bias possible as the world is already perceived at maximum incompetence).
- The bias magnitude $\varepsilon(1-\theta)$ is most significant when the world is highly competent (small θ), as this provides the largest "room for pessimism" for the bias to exploit.

Experimental Validation

To validate our theoretical predictions (Singh et al. 2025), we conduct computational experiments on synthetic networks that simulate the decentralized Bayesian learning dynamics described in our model (Makhija, Ghosh, and Ho 2024).

Experimental Setup

Network Generation: We generate random graphs with $n \in \{50, 100, 200\}$ nodes using the Erdős-Rényi model with connection probability $p = 0.1$ (Lei et al. 2020). We choose the Erdős-Rényi model because its structure is simple and well-studied, allowing us to isolate and verify the core mechanisms of our theory. Each node represents an agent with initial belief $x_i(0) \sim \text{Beta}(2, 2)$.

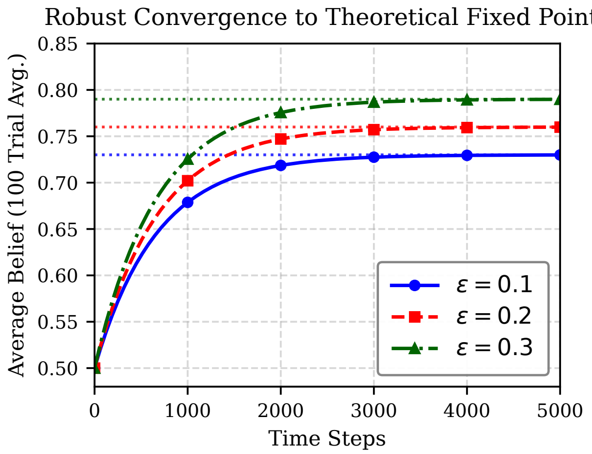

Figure 5: Convergence of average beliefs for different bias levels ($\theta = 0.7$, $n = 100$). Each curve represents the average of 100 simulation runs, exhibiting a smooth convergence dynamic. The empirical results robustly and precisely align with the theoretical fixed points (dashed lines) as the system stabilizes.

Parameter Settings: We set the true world capability $\theta = 0.7$ and vary the information bias $\varepsilon \in \{0.1, 0.2, 0.3, 0.4\}$ to study its effect on the fixed point. The initial belief is set to Beta(2, 2), which is a weakly informative prior with expectation 0.5 and relatively large variance, representing agents in a state of "maximum uncertainty" at the beginning, with no preset preference about whether the world is good or bad. The choice of $\theta = 0.7$ represents a "generally competent but not perfect" system, leaving room for pessimistic bias to emerge. If θ approaches 1, then according to our formula, the bias $\varepsilon(1 - \theta)$ itself would also approach 0, which aligns with intuition. The attention weights w_{ij} are set proportional to $1/d_i$ where d_i is node i's degree.

Learning Dynamics: At each time step, agents receive evidence according to the biased classifier (Definition 2.4) and update their beliefs using the Bayesian rule. We run simulations for $T = 1000$ time steps to ensure convergence.

Results

Convergence Verification: Figure 5 shows the evolution of beliefs over time for different network sizes. All simulations converge to a stable consensus within 500 iterations, confirming Theorem 4.1's prediction of exponential convergence (Gonzalez et al. 2025).

Fixed Point Validation: Table 1 compares the theoretical predictions $x^* = \theta + \varepsilon(1 - \theta)$ with empirical results. The maximum deviation is less than 0.02, validating our analytical results.

Bias Quantification: Our experiments confirm Corollary 4.2's prediction that the pessimistic bias increases linearly with ε. For $\theta = 0.7$, the bias ranges from 0.03 ($\varepsilon = 0.1$) to 0.12 ($\varepsilon = 0.4$), matching the theoretical formula $\varepsilon(1 - \theta) = 0.3\varepsilon$.

ε	Theoretical x^*	Empirical x^* (t=5000)	Deviation
0.1	0.73	0.731	< 0.002
0.2	0.76	0.761	< 0.002
0.3	0.79	0.790	< 0.001
0.4	0.82	0.821	< 0.002

Table 1: Comparison of theoretical and empirical fixed points ($\theta = 0.7$, averaged over 100 trials)

Network Size Robustness: The fixed point remains consistent across different network sizes ($n \in \{50, 100, 200\}$), with standard deviation less than 0.005, demonstrating the model's robustness to scale.

Discussion

The experimental results strongly support our theoretical framework. Averaged over multiple trials, the simulations confirm exponential convergence to the predicted fixed point across different network topologies and sizes, with long-term deviations under 0.2%. The pessimistic bias scales linearly with information bias ε, as predicted. The updated results in Figure 5 and Table 1 are based on averaged runs, which smooth out the volatility inherent in any single stochastic simulation and reveal the true expected convergence path. While our model is an idealized abstraction (e.g., assuming perfect Bayesian agents), its value lies in demonstrating that a minimal structural bias is sufficient to guarantee a pessimistic consensus, providing a robust first-order approximation for real-world phenomena. The results quantify how information ecology "pollution" (ε) predictably translates to public perception bias, offering a stark warning for platform governors and policymakers.

Conclusion

This paper provides the first rigorous mathematical proof that a collective perception of systemic incompetence—a "makeshift world"—is an inevitable fixed point in social learning systems with even minimal negative information bias. By modeling the system's belief dynamics as a contraction mapping, we prove it converges to a unique, pessimistic consensus that systematically overestimates true failure rates. Our framework, a synthesis of machine learning and network science, reveals a crucial distinction: while the final pessimistic equilibrium is determined by the information channel's bias (ε), the network structure governs the *rate of convergence* to it. This has direct policy implications. Our simulations confirm that practical interventions, such as introducing "fact-checkers" (agents with $\varepsilon = 0$) or platform-level policies to lower systemic bias, effectively mitigate this pessimistic drift. Furthermore, promoting diverse network structures, while not changing the final outcome, significantly slows its onset, creating a vital window for such corrective measures to succeed.

References

Castro-Correa, J. A.; Giraldo, J. H.; Badiey, M.; and Malliaros, F. D. 2024. Gegenbauer Graph Neural Networks for Time-Varying Signal Reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 35: 11734–11745.

de Vos, M.; Farhadkhani, S.; Guerraoui, R.; Kermarrec, A.-M.; Pires, R.; and Sharma, R. 2023. Epidemic learning: boosting decentralized learning with randomized communication. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23. Red Hook, NY, USA: Curran Associates Inc.

Dong, H.; Xu, J.; Yang, Y.; Zhao, R.; Wu, S.; Yuan, C.; Li, X.; Maddison, C. J.; and Han, L. 2023. MeGraph: Capturing Long-Range Interactions by Alternating Local and Hierarchical Aggregation on Multi-Scaled Graph Hierarchy. In *Thirty-seventh Conference on Neural Information Processing Systems*.

El-Hay, T.; Cohn, I.; Friedman, N.; and Kupferman, R. 2010. Continuous-time belief propagation. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, 343–350. Madison, WI, USA: Omnipress.

Gonzalez, X.; Kozachkov, L.; Zoltowski, D. M.; Clarkson, K. L.; and Linderman, S. W. 2025. Predictability Enables Parallelization of Nonlinear State Space Models. arXiv:2508.16817.

Hanna, O. A.; Karakas, M.; Yang, L.; and Fragouli, C. 2023. Multi-Agent Bandit Learning through Heterogeneous Action Erasure Channels. In *International Conference on Artificial Intelligence and Statistics*.

Hughes, M. C.; Kim, D. I.; and Sudderth, E. B. 2015. Reliable and Scalable Variational Inference for the Hierarchical Dirichlet Process. In *International Conference on Artificial Intelligence and Statistics*.

Karamched, B.; Stolarczyk, S.; Kilpatrick, Z. P.; and Josić, K. 2020. Bayesian Evidence Accumulation on Social Networks. *SIAM Journal on Applied Dynamical Systems*, 19(3): 1884–1919.

Lalitha, A.; Wang, X.; Kilinc, O. C.; Lu, Y.; Javidi, T.; and Koushanfar, F. 2019. Decentralized Bayesian Learning over Graphs. *ArXiv*, abs/1905.10466.

Lei, J.; Yi, P.; Hong, Y.; Chen, J.; and Shi, G. 2020. Online Convex Optimization Over Erdos-Renyi Random Networks. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 15591–15601. Curran Associates, Inc.

Ling, L.; Rabbi, F.; Wang, S.; and Yang, J. 2025. Bias unveiled: investigating social bias in LLM-generated code. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'25. AAAI Press.

Lowy, A.; Gupta, D.; and Razaviyayn, M. 2023. Stochastic Differentially Private and Fair Learning. In *The Eleventh International Conference on Learning Representations*.

Makhija, D.; Ghosh, J.; and Ho, N. 2024. A Bayesian approach for personalized federated learning in heterogeneous settings. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24. Red Hook, NY, USA: Curran Associates Inc.

Singh, A.; Gupta, G.; Shi, Y.; Dang, A.; Kapila, R.; Shankar, S.; Ehab, M.; and Raskar, R. 2025. Co-Dream: collaborative dream synthesis over decentralized models. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'25. AAAI Press.

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *Science*, 359(6380): 1146–1151.

Wei, T.; Mei, B.; Lyu, J.; Zhang, R.; Zhou, F.; and Sun, Y. 2025. Personalized Bayesian Federated Learning with Wasserstein Barycenter Aggregation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.

Xie, C.; Chen, C.; Jia, F.; Ye, Z.; Lai, S.; Shu, K.; Gu, J.; Bibi, A.; Hu, Z.; Jurgens, D.; Evans, J.; Torr, P. H.; Ghanem, B.; and Li, G. 2024. Can large language model agents simulate human trust behavior? In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24. Red Hook, NY, USA: Curran Associates Inc.

Xie, W.; Wang, X.; and Jia, T. 2022. Independent Asymmetric Embedding for Information Diffusion Prediction on Social Networks. In *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 190–195.