Dynamic Norm Evolution in Social Media: A Cognitive-Evolutionary Framework for Modeling Emergent Behavioral Patterns

Angela Li
Applied Mathematics & Physics
Stony Brook University
Stony Brook, NY
angela.li.4@stonybrook.edu

David Li
Katz School of Science and Health
Yeshiva University
New York, NY
david.li@yu.edu

***Abstract*—Understanding how social norms emerge and evolve on digital platforms is central to social computing. We present the Dynamic Norm Evolution (DNE) model, a cognitively grounded dynamical framework that integrates social reinforcement, confirmation bias, prestige influence, novelty-seeking, and memory decay into an evolutionary process of norm adoption. The model is formulated as a replicator–mutator system with cognitively modulated payoffs, yielding nonlinear differential equations that capture frequency-dependent selection among competing norms. Through fixed point analysis and bifurcation theory, we show that even simple two-norm systems exhibit bistable monopolization, tipping thresholds, and bias-driven transitions between dominant norms. Stochastic extensions reveal bimodal long-run adoption distributions and noise-induced switching between locked-in states. The DNE model provides a principled foundation for interpreting norm convergence, polarization, and sudden behavioral shifts in online ecosystems, offering a versatile theoretical lens for studying opinion dynamics, coordination, and collective behavior in socially networked environments.**

***Index Terms*—Social norm emergence, Social behavior modeling, Cultural evolution, Cognitive bias, Evolutionary game theory, Nonlinear dynamical systems, Social computing**

I. Introduction

The proliferation of social media platforms has radically transformed how individuals engage with information, form opinions, and shape collective behavior. In this digitally mediated environment, emergent social norms—informal, shared expectations about appropriate behavior—can arise spontaneously from millions of decentralized interactions. These norms influence everything from discourse tone and hashtag use to political expression and group identity, often evolving at a pace and scale not seen in offline settings. Understanding how such norms form, compete, and stabilize is a foundational challenge in social computing and data mining, with implications for information diffusion, moderation policy, and the design of socially intelligent systems.

While prior research in social contagion, opinion dynamics, and network influence has provided key insights into how behaviors spread online, most existing models assume relatively homogeneous agents and linear diffusion mechanisms. These frameworks often neglect the complex interplay between individual cognitive biases, evolving group-level feedback, and the selective pressures introduced by platform algorithms. In reality, social media users are not passive recipients of information—they are cognitively bounded agents influenced by social reinforcement, prestige, prior beliefs, novelty-seeking, and memory effects. Moreover, the norms they adopt are subject to nonlinear dynamics, cultural drift, and evolutionary selection.

To address these gaps, we introduce a new theoretical framework: the Dynamic Norm Evolution (DNE) model. This model integrates tools from dynamical systems theory, cultural evolution, and cognitively informed agent modeling to capture how behavioral norms emerge and evolve in large-scale social media ecosystems. Agents in our model evaluate candidate norms based on a utility function shaped by social signals, cognitive alignment, influencer prestige, novelty, and forgetting. We formalize these processes using replicator-mutator dynamics, leading to a system of nonlinear differential equations that govern norm frequency over time.

Our analysis demonstrates that even simple configurations of the DNE model can produce rich phenomena observed empirically in social platforms: spontaneous norm convergence, persistent norm diversity, polarization, and bifurcations between dominant norms. We also show how cognitive and structural parameters shape the stability and volatility of these outcomes. By providing a mathematically rigorous and cognitively grounded model of norm formation, this work contributes to a deeper theoretical understanding of online social behavior and offers a flexible foundation for future empirical studies and simulations.

II. Related Work

The emergence and evolution of social norms in online environments have been explored through diverse theoretical lenses, including social contagion, opinion dynamics, and evolutionary modeling. The Dynamic Norm Evolution (DNE) framework builds on and extends this foundational literature by incorporating cognitively biased decision processes and multi-channel feedback dynamics.

Social contagion theory models the spread of behaviors through interpersonal exposure, often drawing analogies to epidemiological diffusion. Classical models distinguish between simple contagions, which spread via single exposures, and complex contagions, which require reinforcement from multiple sources [1]. A decade-long review of complex contagion theory emphasizes how reinforcement thresholds, network structure, and content features jointly shape diffusion patterns [3]. More recent research integrates population genetics perspectives to explain how network heterogeneity influences contagion reach and speed [4]. Empirical studies have also demonstrated cooperative behavior cascades in large-scale social networks [5], revealing how individual actions propagate nonlinearly via structural influence.

Opinion dynamics models, such as the DeGroot averaging process [2], capture how individuals iteratively revise their beliefs by integrating signals from peers. Although effective at modeling convergence and consensus, these models often assume scalar beliefs and linear averaging, limiting their expressiveness for norm competition. Related work on dynamic social influence [13] and viral content propagation [14] highlights the importance of message content, source prestige, and emotional salience in determining transmission strength.

Evolutionary game theory provides a mathematical foundation for modeling strategic behavior and norm competition in populations. Core formulations such as the replicator dynamic [8], [9] describe how behaviors grow or decline based on fitness differentials, aligning naturally with the frequency-dependent learning observed in digital communities. Extensions of these models have incorporated cultural evolution, showing how boundedly rational agents adapt via imitation and reinforcement. Our DNE framework adopts this foundation while expanding it to account for cognitive biases such as confirmation [12], novelty-seeking, and prestige-weighted influence.

Social media-specific findings further motivate the DNE model. Exposure to ideologically diverse content on Facebook has been shown to influence opinion formation and belief reinforcement [6]. Similarly, moralized content spreads more rapidly in online networks, driven by affective intensity rather than factual accuracy [7]. These effects are mediated not only by individual cognitive predispositions but also by algorithmic curation, which amplifies certain signals while suppressing others. Meta-analyses of social contagion theory and network influence underscore the need for models that account for algorithmic feedback and emotional resonance [10], [11].

Finally, recent advances in norm psychology explore how digital platforms reshape normative expectations at scale. Brady and Crockett [15] argue that online environments alter both the visibility and valence of norm signals, accelerating norm evolution in ways not observed offline. This complements the DNE model's treatment of novelty and memory decay as essential forces in the life cycle of digital norms.

Together, these strands of research affirm the need for a formal, multi-bias, feedback-sensitive model of norm emergence—an objective the Dynamic Norm Evolution framework seeks to fulfill.

III. Dynamic Norm Evolution Model

We formalize the dynamics of norm competition using a cognitively enriched evolutionary framework. Let $\mathcal{A} = \{a_1, a_2, \ldots, a_K\}$ be a finite set of candidate norms and let $x_k(t) \in [0,1]$ denote the fraction of the population adopting norm a_k at time t, subject to $\sum_{k=1}^{K} x_k(t) = 1$. Each agent evaluates candidate norms according to a utility function shaped by social exposure, cognitive similarity, prestige, novelty, and forgetting.

A. Cognitive Utility Function

The utility that agent i assigns to norm a_k at time t is given by:

$$U_i(a_k, t) = \beta_1 S_k(t) + \beta_2 C_{i,k}(t) + \beta_3 P_k(t) + \beta_4 N_k(t) - \delta_k(t) \tag{1}$$

where:

- $S_k(t)$: social signal strength (fraction of peers adopting a_k),
- $C_{i,k}(t)$: alignment with agent i's cognitive priors (confirmation bias),
- $P_k(t)$: prestige bias (influencer-weighted popularity),
- $N_k(t)$: novelty score (based on recency of peak usage),
- $\delta_k(t)$: forgetting penalty (decay if norm is unused).

The vector $\vec{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$ controls the relative strengths of these effects.

B. Mean-Field Evolution Equation

The expected fitness of norm a_k is:

$$\begin{aligned}\phi_k(t) &= \mathbb{E}[U_i(a_k, t)] \\ &= \beta_1 x_k(t) + \beta_2 \bar{C}_k(t) + \beta_3 \tilde{x}_k(t) + \beta_4 n_k(t) - \delta_k(t)\end{aligned} \tag{2}$$

where, with more details in III-C, the notations are:

- $\bar{C}_k(t) = \mathbb{E}[b_{i,k}]$: average alignment across agents,
- $\tilde{x}_k(t)$: prestige-weighted frequency of a_k,
- $n_k(t) = e^{-\gamma(t - t_k^{\text{peak}})}$: novelty decay, where t_k^{peak} is the time of last maximum adoption for a_k, and γ controls novelty decay.
- $\delta_k(t) = \alpha e^{-\lambda f_k(t)}$: memory-based decay.

Assume no births or deaths, constant population, and a discrete-time replicator-mutator dynamic:

$$x_k(t+1) = \sum_{j=1}^{K} x_j(t) Q_{jk}(t) \cdot \frac{\phi_k(t)}{\bar{\phi}(t)} \tag{3}$$

where:

- $Q_{jk}(t)$: mutation probability (agent switches from norm j to k)
- $\bar{\phi}(t) = \sum_j x_j(t)\phi_j(t)$: average fitness across all norms

In the continuous-time limit, the dynamics of norm adoption follow a replicator-mutator structure:

$$\frac{dx_k}{dt} = \sum_{j=1}^{K} x_j(t) Q_{jk}(t) \phi_k(t) - x_k(t)\bar{\phi}(t) \tag{4}$$

In the no-mutation case ($Q_{jk} = \delta_{jk}$), this simplifies to:

$$\frac{dx_k}{dt} = x_k(t)\left[\phi_k(t) - \bar{\phi}(t)\right] \tag{5}$$

This core equation describes frequency-dependent selection: norms grow when their fitness exceeds the population average.

C. Component Dynamics

Each term in the utility function is modeled as follows:

- Social signal: $S_k(t) = x_k(t)$
- Prestige-weighted adoption: Let $w_i \in [0,1]$ be the prestige weight of agent i, normalized such that $\sum_i w_i = 1$. Let:

$$\tilde{x}_k(t) = \sum_{i:A_i(t)=a_k} w_i$$

  This is the prestige-weighted adoption frequency.
- Confirmation alignment: Let each agent i have belief vector $\vec{b}_i \in \mathbb{R}^K$, with $b_{i,k} \in [0,1]$, representing prior alignment with norm a_k.
  Then:

$$\bar{C}_k(t) = \frac{1}{N}\sum_{i=1}^{N} b_{i,k}$$

  Or as a scalar alignment score:

$$\bar{C}_k(t) = \mathbb{E}[b_{i,k}]$$

- Memory decay: Let:

$$\delta_k(t) = \alpha \cdot e^{-\lambda f_k(t)}$$

  where $f_k(t)$ is the frequency of seeing a_k in recent time window, e.g.:

$$f_k(t) = \int_{t-\tau}^{t} x_k(s)ds$$

This formulation yields a high-dimensional nonlinear ODE system that captures the coevolution of norms under cognitive constraints.

D. Summary of Parameters

Parameter	Interpretation
β_1	Social reinforcement weight
β_2	Confirmation bias strength
β_3	Prestige bias strength
β_4	Novelty bias strength
α	Forgetting penalty scale
λ	Memory retention rate
γ	Novelty decay rate

TABLE I
MODEL PARAMETERS AND BEHAVIORAL INTERPRETATIONS

IV. Fixed Points and Their Stability

We analyze the equilibrium behavior of the Dynamic Norm Evolution (DNE) model by identifying the fixed points of the system and assessing their stability. We begin with the full differential equation derived from the cognitive-modulated replicator model.

A. Full Differential System

The evolution of the adoption fraction $x_k(t)$ for each norm a_k is governed by the replicator equation:

$$\frac{dx_k}{dt} = x_k(t)\left[\phi_k(t) - \bar{\phi}(t)\right] \tag{6}$$

where

$$\phi_k(t) = \beta_1 x_k(t) + \beta_2 \bar{C}_k + \beta_3 \tilde{x}_k(t) + \beta_4 e^{-\gamma(t - t_k^{\text{peak}})} - \alpha e^{-\lambda f_k(t)} \tag{7}$$

and

$$\bar{\phi}(t) = \sum_{j=1}^{K} x_j(t)\phi_j(t) \tag{8}$$

Here we treat $\bar{C}_k$ as constant-in-time approximation for the equilibrium analysis. This system defines a set of K coupled nonlinear ODEs, with interactions mediated through social reinforcement, cognitive bias alignment, and evolutionary pressure.

B. Fixed Points in the Two-Norm Case

Consider a simplified case with $K = 2$ norms, a_1 and a_2. Let $x(t) = x_1(t)$ and $1 - x(t) = x_2(t)$. The fitness functions become:

$$\phi_1 = \beta_1 x + b_1 \tag{9}$$

$$\phi_2 = \beta_1(1 - x) + b_2 \tag{10}$$

where b_1 and b_2 are constants aggregating the cognitive and prestige terms.

Substituting into the replicator equation yields:

$$\frac{dx}{dt} = x(1-x)\left[\beta_1(2x-1) + (b_1 - b_2)\right] \tag{11}$$

C. Equilibrium Analysis

The fixed points satisfy $\frac{dx}{dt} = 0$, which occurs when:

$$x = 0, \quad x = 1, \quad \text{or} \quad x^* = \frac{1}{2}\left(1 - \frac{b_1 - b_2}{\beta_1}\right) \tag{12}$$

The interior fixed point $x^* \in (0,1)$ exists only if:

$$|b_1 - b_2| < \beta_1 \tag{13}$$

Otherwise, the system admits only boundary equilibria.

D. Stability Conditions

Let $f(x) = \frac{dx}{dt} = x(1-x)\left[\beta_1(2x-1) + \Delta b\right]$ where $\Delta b = b_1 - b_2$. Taking the derivative:

$$f'(x) = (1-2x)[\beta_1(2x-1) + \Delta b] + 2\beta_1 x(1-x) \tag{14}$$

Evaluating at the boundary points:

$$f'(0) = -(\beta_1 - \Delta b) \tag{15}$$

$$f'(1) = -(\beta_1 + \Delta b) \tag{16}$$

For the interior fixed point $x^* = \frac{1}{2}(1 - \Delta b/\beta_1)$, the derivative is

$$f'(x^*) = \frac{\beta_1^2 - \Delta b^2}{2\beta_1}.$$

The interior equilibrium x^* is unstable when $f'(x^*) > 0$, which occurs under:

$$|\Delta b| < \beta_1 \tag{17}$$

E. Bifurcation Structure and Hysteresis

We now examine how the equilibrium structure of the two-norm system changes as the bias differential $\Delta b = b_1 - b_2$ varies, holding $\beta_1 > 0$ fixed.

For $|\Delta b| < \beta_1$, we have $f'(0) < 0$ and $f'(1) < 0$, so both boundary equilibria $x = 0$ and $x = 1$ are locally stable. At the same time, $f'(x^*) > 0$, so the interior fixed point is unstable. In this regime the system is bistable: each norm can become nearly monopolistic, and the unstable interior equilibrium x^* acts as a separatrix dividing their basins of attraction. Small perturbations near x^* push the population toward one of the two dominant norms.

As Δb increases past β_1, the interior equilibrium collides with the left boundary at $x = 0$ (since $x^* \to 0$ as $\Delta b \to \beta_1$). At $\Delta b = \beta_1$, we have $x^* = 0$ and $f'(0) = 0$; for $\Delta b > \beta_1$, x^* moves outside the admissible interval and $x = 0$ becomes unstable ($f'(0) > 0$), while $x = 1$ remains stable. Symmetrically, as Δb decreases past $-\beta_1$, the interior equilibrium collides with the right boundary at $x = 1$, after which $x = 1$ becomes unstable and $x = 0$ is the unique stable equilibrium.

This pattern is best interpreted as a boundary-anchored bistability with saddle-node–like transitions at $\Delta b = \pm\beta_1$. For moderate bias differentials ($|\Delta b| < \beta_1$), the system exhibits hysteresis and path dependence: which norm ultimately dominates depends on initial conditions and transient shocks. For large bias ($\Delta b > \beta_1$ or $\Delta b < -\beta_1$), the system is effectively monostable, converging to the advantaged norm regardless of initial composition.

F. Generalization to K Norms

For $K > 2$, a vector $\vec{x}^*$ is an equilibrium if and only if:

$$x_k^* > 0 \Rightarrow \phi_k = \bar{\phi}, \quad x_k^* = 0 \Rightarrow \phi_k \leq \bar{\phi} \tag{18}$$

Local stability can be assessed by analyzing the Jacobian:

$$J_{ij} = \frac{\partial}{\partial x_j}\left[x_i(\phi_i - \bar{\phi})\right] \tag{19}$$

Eigenvalue analysis of J at $\vec{x}^*$ reveals the nature of the equilibrium—whether it is stable, unstable, or a saddle.

G. Summary of Norm Dynamics in the Two-Norm Case

The results from this section show a clear and internally consistent pattern of behavior for the two-norm system. Key conclusions are:

- **Equilibria:** The system admits three fixed points on $[0, 1]$ when $|\Delta b| < \beta_1$: the boundaries $x = 0$ and $x = 1$ and an interior point $x^* = \frac{1}{2}(1 - \Delta b/\beta_1)$. The interior equilibrium is always *unstable*, while both boundaries are *stable*.
- **Bistable monopolization:** For moderate bias differentials ($|\Delta b| < \beta_1$), the system exhibits two locally stable outcomes corresponding to near-complete dominance by either norm. Which norm prevails depends on initial conditions and perturbations relative to the unstable threshold x^*.
- **Bias-driven tipping:** As Δb crosses $\pm\beta_1$, one boundary equilibrium loses stability and the system becomes monostable, converging to the norm favored by the bias. These transitions generate hysteresis and path dependence.
- **Generalization:** Although the analysis is presented for two norms, the same structure—multiple stable norm-dominant states separated by unstable manifolds—naturally extends to systems with more than two competing norms.

V. Numerical Simulation Results

We now present numerical simulations to illustrate the dynamic behavior of the DNE model under both deterministic and stochastic conditions. These results validate the model's ability to reproduce empirically relevant phenomena, including convergence to equilibrium, bifurcations, random fluctuations, and adoption variability across time.

A. Deterministic Trajectories and Basins of Attraction

We first examine deterministic trajectories under moderate bias, where the analysis predicts bistability and an unstable interior threshold. Consider $\beta_1 = 1.0$, $b_1 = 0.5$, and $b_2 = 0.0$, so that $\Delta b = 0.5$ and

$$x^* = \frac{1}{2}\left(1 - \frac{\Delta b}{\beta_1}\right) = \frac{1}{2}(1 - 0.5) = 0.25.$$

Since $|\Delta b| = 0.5 < \beta_1$, there are three fixed points in $[0, 1]$: $x = 0$, $x = 1$, and $x^* = 0.25$; both boundaries are stable while the interior point is unstable.

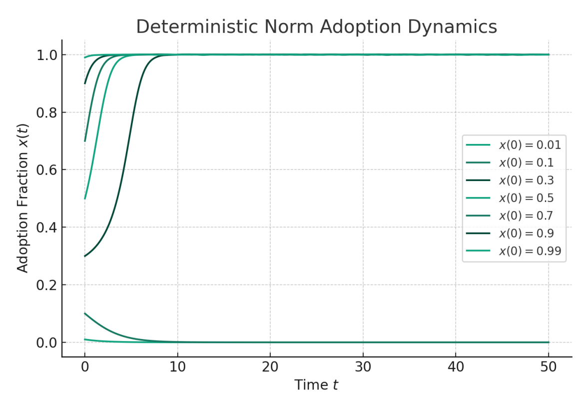

Fig. 1. Deterministic norm adoption dynamics for varying initial conditions.

Simulating trajectories Figure 1 from different initial conditions confirms this picture. Initial conditions with $x(0) < x^*$ converge to $x = 0$, representing eventual monopolization by norm 2. Initial conditions with $x(0) > x^*$ converge to $x = 1$, representing monopolization by norm 1. The unstable interior equilibrium at x^* acts as a tipping threshold: small differences in the initial adoption level around x^* are amplified over time, leading to qualitatively different long-run outcomes.

From a behavioral perspective, this regime captures situations where neither norm has an overwhelming structural advantage, yet the system still tends to lock into a single dominant norm. Early fluctuations and interventions that move the state across the threshold x^* have persistent effects, while later small perturbations around a stable boundary equilibrium are damped out.

B. Bifurcation Diagram and Hysteresis

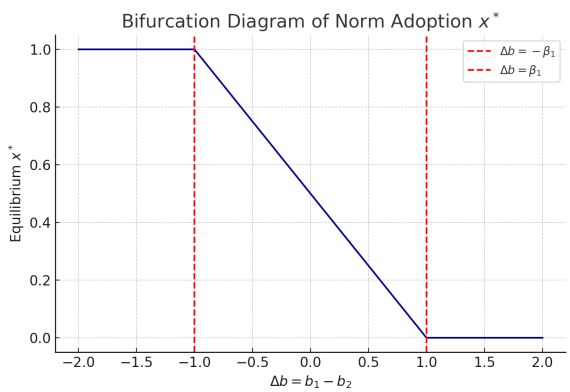

Fig. 2. Bifurcation diagram showing the equilibrium norm adoption x^* as Δb varies.

To visualize how long-run outcomes depend on the bias differential, we vary Δb while holding β_1 fixed. Figure 2 presents the bifurcation diagram of the equilibrium adoption fraction x^* as a function of $\Delta b = b_1 - b_2$. For each value of Δb, we integrate the deterministic dynamics from a grid of initial conditions and record the attracting equilibrium reached by each trajectory.

The resulting bifurcation diagram shows three qualitatively distinct regimes:

1. Strong bias for norm 1 ($\Delta b > \beta_1$): The left boundary $x = 0$ is unstable, and $x = 1$ is the unique stable equilibrium. All initial conditions converge to near-complete adoption of norm 1.

2. Bistable regime ($|\Delta b| < \beta_1$): Both boundaries $x = 0$ and $x = 1$ are stable, and the interior fixed point x^* exists but is unstable. The population converges either to $x = 0$ or $x = 1$, depending on whether the initial condition lies below or above x^*. As Δb varies within this interval, the threshold x^* shifts smoothly between 0 and 1, rebalancing the size of each basin of attraction but never becoming stable itself.

3. Strong bias for norm 2 ($\Delta b < -\beta_1$): The right boundary $x = 1$ is unstable, $x = 0$ is the unique stable equilibrium, and all initial conditions converge to near-complete adoption of norm 2.

This structure implies hysteresis. Suppose Δb slowly varies over time, for example due to exogenous changes in prestige or platform incentives. If the system is initially locked into a norm-dominant state at $x = 1$, reducing Δb toward zero does not immediately trigger a switch to $x = 0$; the state remains at $x = 1$ so long as $\Delta b > -\beta_1$. Only when the bias crosses the critical threshold $\Delta b = -\beta_1$ does the boundary equilibrium at $x = 1$ lose stability and the system tip toward dominance by norm 2. The reverse transition exhibits a different critical point at $\Delta b = \beta_1$, generating path dependence and history-dependent switching thresholds.

C. Stochastic Dynamics

To explore robustness under noise, we introduce stochastic perturbations via the Langevin-style extension:

$$dx = x(1 - x)(\phi_1 - \phi_2)dt + \sigma dW_t$$

which introduces idiosyncratic fluctuations in users' adoption decisions. In the bistable regime $|\Delta b| < \beta_1$, the deterministic dynamics possess two locally stable equilibria at 0 and 1, separated by an unstable threshold at x^*. Figure 3 displays sample stochastic trajectories with $\sigma = 0.05$ and varying initial conditions.

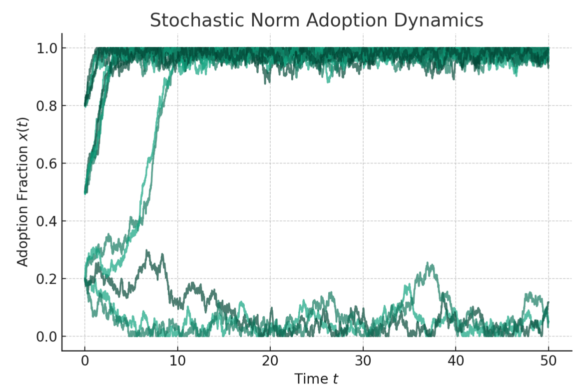

Fig. 3. Stochastic norm adoption trajectories under moderate noise ($\sigma = 0.05$).

Simulated sample paths under small noise (σ modest) largely follow one of the deterministic basins: trajectories that start above x^* are attracted toward (x=1), while those that start below x^* are attracted toward $x = 0$. Once near a stable boundary, the drift term dominates and keeps the trajectory close to that equilibrium, with only small fluctuations due to noise. Over long horizons, rare noise-driven escape events across the unstable threshold may occur, but for realistic parameter choices these are infrequent, emphasizing the resilience of locked-in norm dominance.

In contrast, when $|\Delta b|$ is pushed beyond β_1, the noise only perturbs trajectories around the unique stable boundary equilibrium, and switching to the disfavored norm becomes extremely unlikely.

D. Distribution of Long-Run Adoption

We approximate the stationary distribution of the stochastic process by simulating long trajectories and constructing histograms of $x(t)$ in Figure 4. In the bistable regime, the stationary distribution is bimodal, with peaks concentrated near 0 and 1 and a trough around the unstable threshold x^*. The relative mass in the two modes reflects both the size of each basin of attraction (controlled by the position of x^*) and the noise level σ, which governs the rate of rare transitions between the two norm-dominant states.

As $|\Delta b|$ increases toward β_1, one of the modes grows and the other shrinks, reflecting the shrinking basin of the disfavored norm. Once $|\Delta b| > \beta_1$, the stationary distribution becomes unimodal, concentrating tightly around the unique stable equilibrium for the advantaged norm. This transition in the shape of the stationary distribution mirrors the underlying deterministic bifurcation structure, but also highlights that even when two stable equilibria coexist, stochastic shocks make long-run occupancy weights depend not only on basins but also on noise-induced escape dynamics.

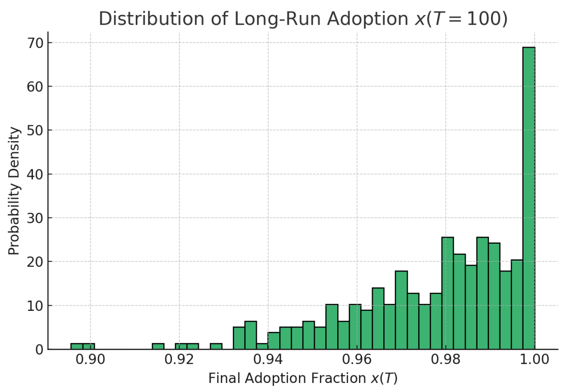

Fig. 4. Histogram of final adoption levels across 300 stochastic trials.

VI. Discussion and Implications

Our analysis reveals that, under the cognitive-utility and replicator dynamics specified in Sections III–IV, the dominant qualitative behavior of norm adoption is bistable monopolization rather than stable coexistence of competing norms. When the bias differential $|\Delta b|$ is small relative to the social reinforcement strength β_1, the system admits two stable equilibria corresponding to near-complete adoption of each norm, separated by an unstable interior threshold x^*. In this regime, which norm ultimately dominates is determined by early conditions and transient shocks that move the state across the tipping boundary.

This has several implications for the design and governance of social platforms:

1. Path dependence and lock-in. Once a platform crosses the unstable threshold in favor of a particular norm, social reinforcement and cognitive alignment mechanisms drive the system toward that norm's dominance. Retrospectively altering incentives or content ranking to favor an alternative norm may not suffice if the system remains within the same basin of attraction; substantial interventions are required to push the state back across the unstable threshold.

2. Asymmetric switching thresholds. Because stability changes occur at $\Delta b = \pm\beta_1$, the bias required to dislodge a currently dominant norm can differ markedly from the bias that originally enabled its ascent. This asymmetry produces hysteresis: platform changes that would have prevented the rise of a harmful norm ex ante may be insufficient to remove it ex post, once it has become entrenched.

3. Role of stochastic fluctuations. Noise in user behavior does not generically stabilize interior coexistence; instead, in the bistable regime it produces a bimodal stationary distribution concentrated near the two norm-dominant states. Rare, noise-induced transitions between these states are possible but typically slow, implying that platforms cannot rely on organic fluctuations alone to undo undesirable norm dominance.

4. Targeted interventions. In the context of our model, interventions can act either by shifting the bias differential Δb (e.g., via platform policies that alter effective prestige or cognitive costs) or by directly perturbing the state x (e.g., large-scale campaigns that temporarily boost adoption of a desired norm). Our results suggest that sustained changes to Δb that move the system beyond $\pm\beta_1$, or sufficiently strong direct shocks that push x across the unstable threshold, are required to permanently redirect norm trajectories.

While our two-norm reduction is stylized, the same mechanism naturally generalizes to higher-dimensional norm spaces: multiple locally stable norm configurations separated by unstable manifolds, with social reinforcement and cognitive alignment generating lock-in and hysteresis. Extending the analysis to richer network structures and multi-dimensional action spaces is a promising direction for future work.

VII. Conclusion

This paper introduced the Dynamic Norm Evolution (DNE) model, a cognitively informed and mathematically grounded framework for understanding how behavioral norms emerge and evolve in online social ecosystems. By combining social reinforcement, confirmation bias, prestige weighting, novelty seeking, and memory decay within a replicator–mutator system, the DNE model captures key feedback mechanisms that shape collective behavior on digital platforms.

Our analysis of the two-norm case revealed a clear equilibrium structure: when the bias differential is small relative to social reinforcement, the system becomes bistable, with each boundary equilibrium representing near-complete dominance by a single norm. The interior fixed point, although present for moderate bias levels, is always unstable and serves as a tipping threshold dividing the basins of attraction. As the bias differential crosses critical thresholds, the system undergoes boundary-anchored bifurcations that eliminate one stable state and lead to monostable dominance by the favored norm. Numerical simulations confirmed these dynamics, demonstrating path dependence, hysteresis, and the emergence of bimodal stationary distributions under stochastic fluctuations.

The DNE framework offers a principled tool for interpreting a wide range of online social phenomena, including lock-in effects, rapid norm shifts, polarization, and the persistence of entrenched behaviors. Its cognitive foundations and dynamical formulation also provide a basis for evaluating interventions—such as modifying prestige signals or altering content exposure—that aim to shift collective outcomes.

Future work will extend the model to incorporate realistic network structures, heterogeneous cognitive profiles, and empirical calibration using large-scale social data. These directions will enable the DNE framework to support predictive modeling and inform the design of platform mechanisms that foster constructive, stable, and transparent normative environments.

REFERENCES

[1] D. Centola, *How Behavior Spreads: The Science of Complex Contagions*. Princeton University Press, 2018.

[2] M. H. DeGroot, “Reaching a consensus,” *Journal of the American Statistical Association*, vol.69, no.345, pp.118–121, 1974.

[3] D. Guilbeault, J. Becker, and D. Centola, “Complex contagions: A decade in review,” arXiv preprint arXiv:1710.07606, 2017.

[4] J. Kates-Harbeck and M. M. Desai, “Social network structure and the spread of complex contagions from a population genetics perspective,” arXiv preprint arXiv:2208.02864, 2022.

[5] J. H. Fowler and N. A. Christakis, “Cooperative behavior cascades in human social networks,” arXiv preprint arXiv:0908.3497, 2009.

[6] E. Bakshy, S. Messing, and L. A. Adamic, “Exposure to ideologically diverse news and opinion on Facebook,” *Science*, vol.348, no.6239, pp.1130–1132, 2015.

[7] W. J. Brady, J. A. Wills, J. T. Jost, J. A. Tucker, and J. J. Van Bavel, “Emotion shapes the diffusion of moralized content in social networks,” *PNAS*, vol.114, no.28, pp.7313–7318, 2017.

[8] M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press, 2006.

[9] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

[10] N. A. Christakis and J. H. Fowler, “Social contagion theory: examining dynamic social networks and human behavior,” *Statistics in Medicine*, 2013.

[11] D. Centola and M. Macy, “Complex contagions and the weakness of long ties,” *American Journal of Sociology*, vol.113, no.3, pp. 702–734, 2007.

[12] H. A. Simon, “A behavioral model of rational choice,” *Quarterly Journal of Economics*, vol.69, no.1, pp.99–118, 1955.

[13] B. Latané, “Dynamic social impact theory: The creation of culture by communication,” *Journal of Communication*, vol.46, no.4, pp.13–25, 1996.

[14] J. Berger and K. Milkman, “What makes online content viral?” *Journal of Marketing Research*, vol.49, no.2, pp.192–205, 2012.

[15] W. J. Brady and M. Crockett, “Norm psychology in the digital age: how social media shapes the cultural evolution of normativity,” *Perspectives on Psychological Science*, 2024.