Why Polarization Becomes Slow and Deep: A System Dynamics Model of Perception Delay in Opinion Networks

Qinyang Yu*

Abstract

This paper introduces perception delay–a cognitive lag in processing information–into an opinion dynamics model to explain why polarization forms slowly yet becomes deeply entrenched. Using a stochastic block model with bounded-confidence updating, I embed a first-order information delay structure from system dynamics to represent gradual adjustment of perceived information. Simulations show that perception delay weakens fast reinforcing feedback while creating strong inertia through a delayed balancing loop, leading to a distinct "slow–deep" pattern of polarization. Once polarization emerges, the accumulated perception stock makes the system resistant to short-term interventions, while long-term interventions induce only gradual depolarization. These results identify perception delay as a structural mechanism underlying the persistence and irreversibility of polarization in social networks.

1 Introduction

Opinion dynamics models have been widely used to understand how individuals in social networks update their opinions over time. Following early models, such as the DeGroot model (DeGroot, 1974), recent opinion dynamics models have increasingly incorporated *perception bias*—the idea that individuals process social information through cognitive filters such as

*Department of Economics and Computer Science, Duke University. E-mail: qinyang.yu@duke.edu

confirmation bias and motivated reasoning (Allahverdyan and Galstyan, 2014; Sobkowicz, 2018; Yu et al., 2023).

Perception bias leads individuals to process information in a way that reinforces their pre-existing beliefs, rather than integrating new information in a balanced way. Research in opinion dynamics has shown that perception bias can affect the rate and direction of opinion change. For example, individuals in an echo chamber may be less likely to adopt opinions that deviate from their own, even when presented with overwhelming evidence to the contrary. These biases can thus lead to the *polarization* of opinions, where individuals cluster around extreme positions and become less willing to engage with opposing viewpoints (Bail et al., 2018; Lazer et al., 2018; Sunstein, 2009; Tucker et al., 2018).

However, most existing models that capture biased perception assume instantaneous updating. Rare studies have considered *perception delay*—the temporal inertia of information updating rooted in cognitive and psychological processes. This perception delay is conceptually distinct from the well-studied *time delay* in control or consensus theory. For instance, recent work such as Choi et al. (2025) extends the classic Hegselmann-Krause (HK) model by incorporating time delay, showing that even with fixed delays, consensus can still be achieved. Similarly, Neirotti (2023) explores the effects of interaction delay on stable attitudes in multi-agent opinion dynamics. Additionally, Li et al. (2025) combines delay with individual stubbornness to model opinion evolution in social networks. Furthermore, Yao and Li (2025) investigates the convergence of time-delayed opinion dynamics particularly under complex interaction types.

In contrast to these time delay models, which reflect external communication lags, perception delay captures internal cognitive lag in how individuals gradually integrate new information. This cognitive inertia stems from the psychological processes by which individuals revise their beliefs in response to new social signals, a process that does not happen instantaneously but rather unfolds over time. As such, perception delay accounts for the fact that individuals often require time to adjust their opinions based on evolving social contexts, reflecting a more gradual and internalized process of belief updating. Unlike external delays that stem from communication channels or network structure, perception delay is a product of the cognitive biases, prior experiences, and mental frameworks that shape how individuals process incoming information.

To fill the research gap, this study proposes a new framework that incorporates perception delay to understand opinion dynamics and polarization in social networks. To formalize this cognitive inertia, system dynamics (SD) provides a useful framework for modeling feedback loops and delays. A commonly used and well-established approach in SD is the use of *first-order information delays* (Forrester, 1997; Sterman, 2002), where the information updating is determined by a weighted difference between the current information and the perceived new information, moderated by a time constant of adaptation. I thus model perception delay using a first-order information delay structure, which naturally represents the gradual integration of new social signals and the slow adjustment of beliefs.

In a high modularity community network, two groups update their beliefs based on the bounded-confidence model (Hegselmann, 2015), where individuals adopt opinions that are within a certain threshold of their current opinion, leading to polarization. By introducing perception delay, I examine the qualitative behavioral shift of the polarization formation.

The result is straightforward yet insightful: perception delay leads to slow but deep polarization. This result aligns with existing observations and research, such as Biondi et al. (2023), who explore how social relationships and individual susceptibility contribute to the gradual accumulation of polarization, and Curiel (2021), who discusses the inevitability of polarization due to homophily and information spreading. Furthermore, Zafeiris (2022) shows that once polarization sets in, it becomes deeply entrenched and hard to reverse, a phenomenon that is further deepened by confirmation bias as illustrated in Chen et al. (2022). Building on this empirical evidence, this study uses perception delay and feedback loops to offer an explanation about why polarization occurs slowly over time and becomes increasingly difficult to reverse once it has taken root.

The remainder of the paper is organized as follows. Section 2 introduces the model formulation. Section 3 presents the simulation results. Section 4 concludes and discusses directions for future work.

2 Model Formulation

2.1 Model Setup

I model and simulate the opinion dynamics of a networked system in which individuals update their opinions over time based on interactions with their neighbors.

To better fit the nature of polarization, the network structure is modeled using a stochastic block model (SBM), where individuals are divided into two communities, each characterized by different interaction strengths within and across communities. Specifically, the network consists of $N = 100$ agents, divided into two groups A and B of equal size. The probability of forming a connection between two agents within the same community is $p_{\text{in}} = 0.5$, while the probability of forming a connection between agents from different communities is $p_{\text{out}} = 0.05$. These values reflect stronger within-community connections and weaker cross-community connections. The network construction produces an adjacency matrix $\mathbf{A} = [a_{ij}]$, where $a_{ij} = 1$ indicates an edge between i and j.

Each agent i holds two main stocks: the opinion $x_i(t)$ at time t, which represents the agent's current belief, and the perceived information $p_i(t)$, which accumulates the influence of neighbors' opinions over time. The initial opinions of the agents are assigned as follows: agents in group A have opinions randomly distributed between 0 and 0.5, while agents in group B have opinions randomly distributed between 0.5 and 1.0. The initial perceived information $p_i(0)$ is set to be the same as the self opinion $x_i(0)$ for all agents, representing their initial belief. The visualization of the initial network and opinions is displayed in Figure 1.

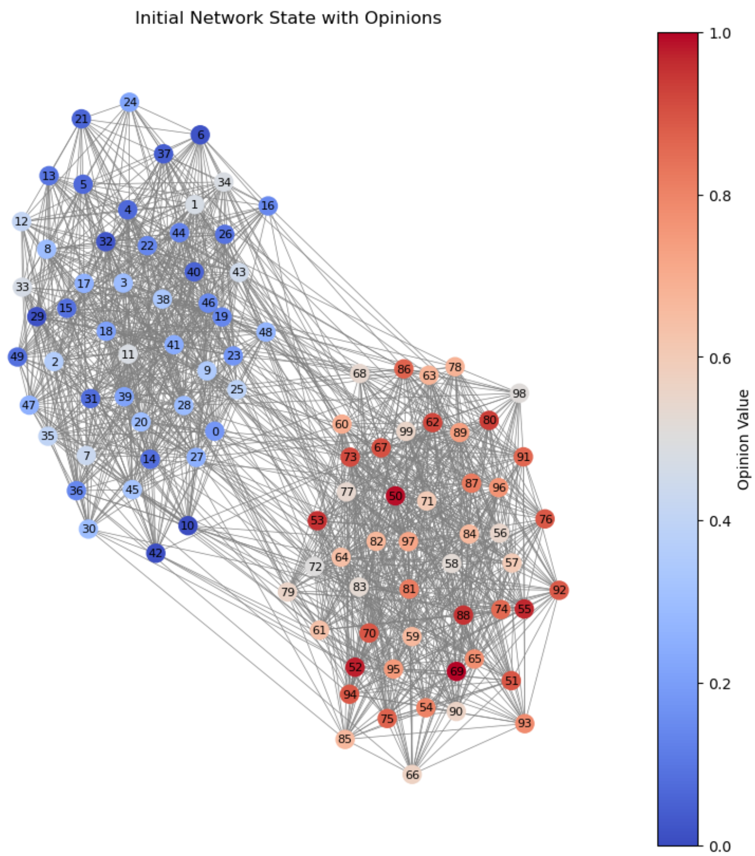

Figure 1: The network structure and initial opinion distribution

The interactions between agents are governed by a bounded-confidence model (BCM), in which individuals update their opinions based on the opinions of their neighbors, but only if the difference between their opinions is smaller than a predefined threshold $\epsilon > 0$. Based on this threshold, the effective neighbor set is

$$\mathcal{N}_i(t) = \Big\{ j \mid a_{ij} = 1,\ |x_i(t) - x_j(t)| < \epsilon \Big\}. \tag{1}$$

The true information available to agent i is defined as the average opinion of its effective neighbors:

$$s_i(t) = \begin{cases} \dfrac{1}{|\mathcal{N}_i(t)|} \sum_{j \in \mathcal{N}_i(t)} x_j(t), & |\mathcal{N}_i(t)| > 0, \\ x_i(t), & |\mathcal{N}_i(t)| = 0. \end{cases} \tag{2}$$

After observing the true information $s_i(t)$, agent i forms its perceived information $p_i(t)$, which may differ from $s_i(t)$ depending on the perception mechanism. Regardless of the perception mechanism, opinions update as

$$x_i(t+1) = x_i(t) + \alpha \left(p_i(t) - x_i(t)\right), \tag{3}$$

with $\alpha \in (0, 1]$ representing the opinion update rate. This parameter controls the speed at which an agent updates its opinion based on its perceived information.

2.2 Modeling Perceptions

In this subsection, I consider two different models of how the perceived information $p_i(t)$ responds to true information $s_i(t)$: one with instantaneous perception and one with perception delay. In both models, agents update their opinions based on interactions with their neighbors, but the way the perceived information $p_i(t)$ is updated differs between the two models.

Model 1: Instantaneous Perception. In the benchmark model, agents perfectly and immediately perceive the true information:

$$p_i(t) = s_i(t). \tag{4}$$

In this case, the opinion update rule reduces to

$$x_i(t+1) = x_i(t) + \alpha\big(s_i(t) - x_i(t)\big). \tag{5}$$

This is the standard bounded-confidence averaging mechanism, where each agent adopts a weighted average of their neighbors' opinions, within the bounds defined by ϵ.

Model 2: Perception Delay. In the model of primary interest, each agent tracks the true information $s_i(t)$ and forms a delayed perception. The internal perception of agent i evolves gradually over time according to a first-order information delay, as defined in System

Dynamics:

$$p_i(t+1) = p_i(t) + \frac{1}{\tau}\left(s_i(t) - p_i(t)\right), \tag{6}$$

where $\tau > 1$ is the perception delay time constant. This parameter controls how quickly an individual's perception adapts to the information received from their neighbors. Larger τ implies slower adjustment of $p_i(t)$, and therefore slower and more inertial opinion dynamics. When $\tau = 1$, Model 2 collapses to Model 1 except for a one-step time delay due to the discrete-time formulation. This shift has no behavioral effect because $x_i(t)$ always reacts to the most recently updated p_i.

To provide intuitions, consider how people process political information. If a person strongly identifies with a particular political party, they may be slow to update their opinions when exposed to new information that contradicts their beliefs. Even after learning new facts, their beliefs might change only gradually, as they might need time to reassess their views or may dismiss conflicting information due to cognitive biases. In this case, the perception delay reflects the time it takes for an individual's opinion to catch up with the new information they receive, illustrating the slow and often resistant nature of belief adjustment.

The update rule for agent i's opinion remains the same as in Model 1, except that it now uses the delayed perception, as shown in Equation 3.

Overall, Model 2 simulates how an agent's beliefs update more gradually due to perception delay, reflecting a slower response to new information received from the social network compared to Model 1. Section 3 further details how these two models lead to different patterns of polarization formation.

2.3 Intervention Mechanism

To explore the effect of an external intervention and the reversibility of polarization in the system, I introduce an intervention active during the time window

$$t_{\text{int}} \leq t < t_{\text{int}} + T_{\text{int}}, \tag{7}$$

where t_{int} is the start time of the intervention and T_{int} is the duration of the intervention, with the indicator function

$$\mathbf{I}(t) = \begin{cases} 1, & t_{\text{int}} \leq t < t_{\text{int}} + T_{\text{int}}, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

T_{int} can be small or large, representing short-term and long-term interventions, respectively.

During the intervention period, a public signal $u(t)$ is provided, which is the average opinion of the entire population at time t:

$$u(t) = \bar{x}(t) = \frac{1}{N} \sum_{i=1}^{N} x_i(t). \tag{9}$$

For example, $u(t)$ can be thought of as a public signal generated by a government campaign, media influence, or public policy announcements, where the goal is to shift the collective opinion of the population.

The intervention affects individual opinions through the perceived information $p_i(t)$. During the intervention, the perceived information is determined by both the intervention and the information from the agent's neighbors. This means that the external signal $u(t)$ enters the system not by directly changing opinions, but by shaping how agents perceive and interpret social information. While the intervention exposure, including the external signal $u(t)$, its strength η, and duration T_{int}, is identical across models, the mechanism through which it influences opinions—the specific formulation of $p_i(t)$—differs between the two models, depending on whether perception is instantaneous or delayed.

For Model 1, since perception is instantaneous, the intervention directly replaces a fraction of the information:

$$p_i(t) = (1 - \eta \mathbf{I}(t))\, s_i(t) + \eta \mathbf{I}(t)\, u(t), \tag{10}$$

where $\eta \in [0, 1]$ represents the strength of the intervention. The higher η is, the more influence the global opinion has on individual perceptions.

For Model 2, intervention enters through the perception stock by the same delay process.

The perception update becomes

$$p_i(t+1) = p_i(t) + \frac{1}{\tau}\big(s_i(t) - p_i(t)\big) + \frac{\eta}{\tau}\,\mathbf{I}(t)\big(u(t) - p_i(t)\big). \tag{11}$$

In both models, the opinion update rule follows Equation 3. Because the intervention must pass through the delay operator in Model 2, its effective impact is expected to be much weaker and slower than in Model 1. The next section empirically verifies this prediction through simulation.

3 Simulation Results

I simulate both Model 1 and Model 2 under three scenarios: one without intervention, one with short-term intervention, and one with long-term intervention. At each time step, agents update their opinions according to the corresponding model equations described in Section 2. Across all scenarios, the simulation runs for $T = 200$ time steps. The bounded-confidence threshold is $\epsilon = 0.15$, the opinion update rate is $\alpha = 0.5$, the perception delay constant in Model 2 is $\tau = 10$, and the intervention strength is $\eta = 0.8$.

To quantify the extent of opinion divergence between the two communities, I measure the *polarization gap* over time. The polarization gap is defined as the absolute difference between the mean opinions of the two groups:

$$\text{Polarization Gap} = \big|\text{mean}_A - \text{mean}_B\big|. \tag{12}$$

This metric quantifies how polarized the system is at any given time. A larger value of this metric indicates stronger polarization, while smaller values reflect greater convergence of opinions between groups.

Figure 2 shows the evolution of opinions and the corresponding polarization gap over time for both models without external intervention. In Model 1, where perception is instantaneous, opinions within each community rapidly converge to internal consensus, forming two sharply divided opinion clusters. This dynamic reflects a dominant *reinforcing feedback loop* driven by network homophily and bounded confidence: individuals preferentially interact with similar others, reinforcing within-group agreement and amplifying differences between groups. As a result, polarization emerges quickly and stabilizes at a high level.

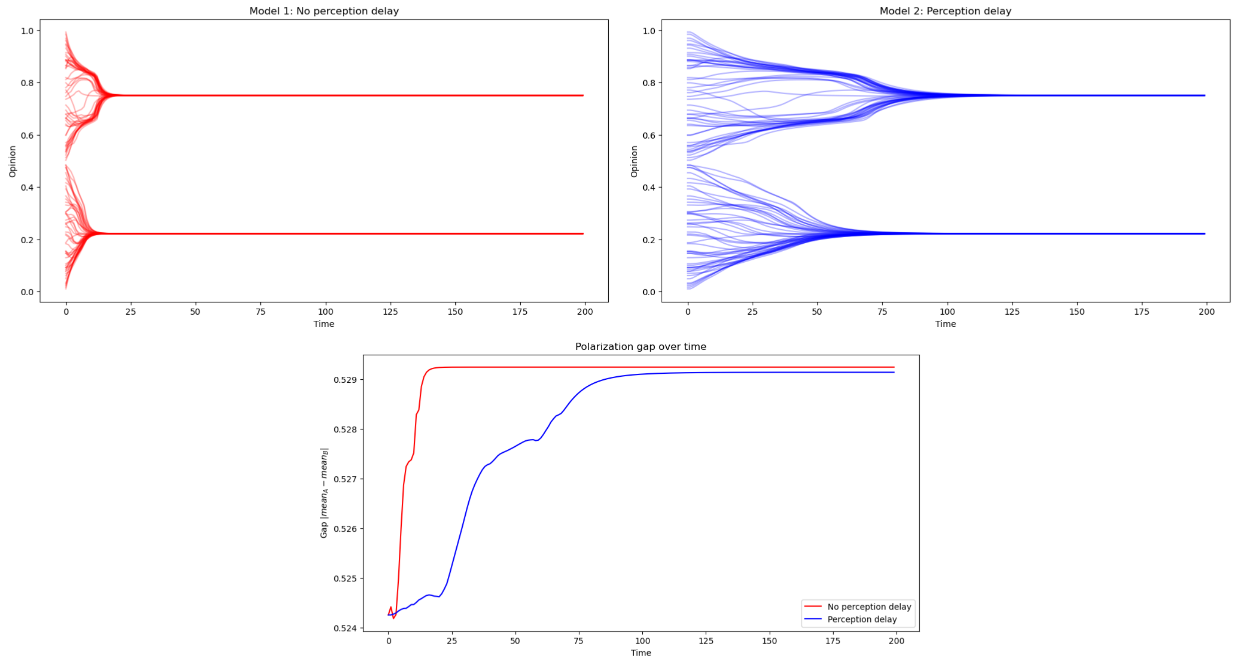

Figure 2: Opinion trajectories and polarization gap over time without intervention

In Model 2, the introduction of perception delay slows this reinforcing process. Because individuals integrate new social information gradually, the reinforcing loop operates more sluggishly, weakening the immediate amplification of opinion differences. Consequently, polarization still develops but at a slower pace, and opinions exhibit greater temporal inertia before reaching their final polarized state. This pattern illustrates the effect of cognitive inertia: perception delay slows short-term convergence within groups but ultimately results in equally entrenched divisions once opinions stabilize.

The intervention is introduced at time $t = 125$, after both systems have reached a stable polarized state. Two durations are tested to capture different temporal effects: a short-run intervention with $T_{\text{int}} = 5$ and a long-run intervention with $T_{\text{int}} = 25$.

Figures 3 show the effects of short-term interventions on opinion trajectories and polarization dynamics in both models. In Model 1, where perception is instantaneous, even a short intervention rapidly collapses polarization: opinions in both communities converge almost immediately toward the public signal, and the polarization gap drops to near zero. Once the intervention ends, opinions remain close to the unified level, indicating a full recovery of consensus. This behavior reflects a direct and unlagged response to external information. In this instantaneous-perception model, opinions respond immediately to the current perceived information, and the balancing feedback caused by opinion updating operates without delay, allowing external intervention signals to propagate quickly throughout the system. As a result, polarization forms rapidly but remains easy to reverse, since external consensus cues are absorbed instantly and eliminate group differences within only a few time steps.

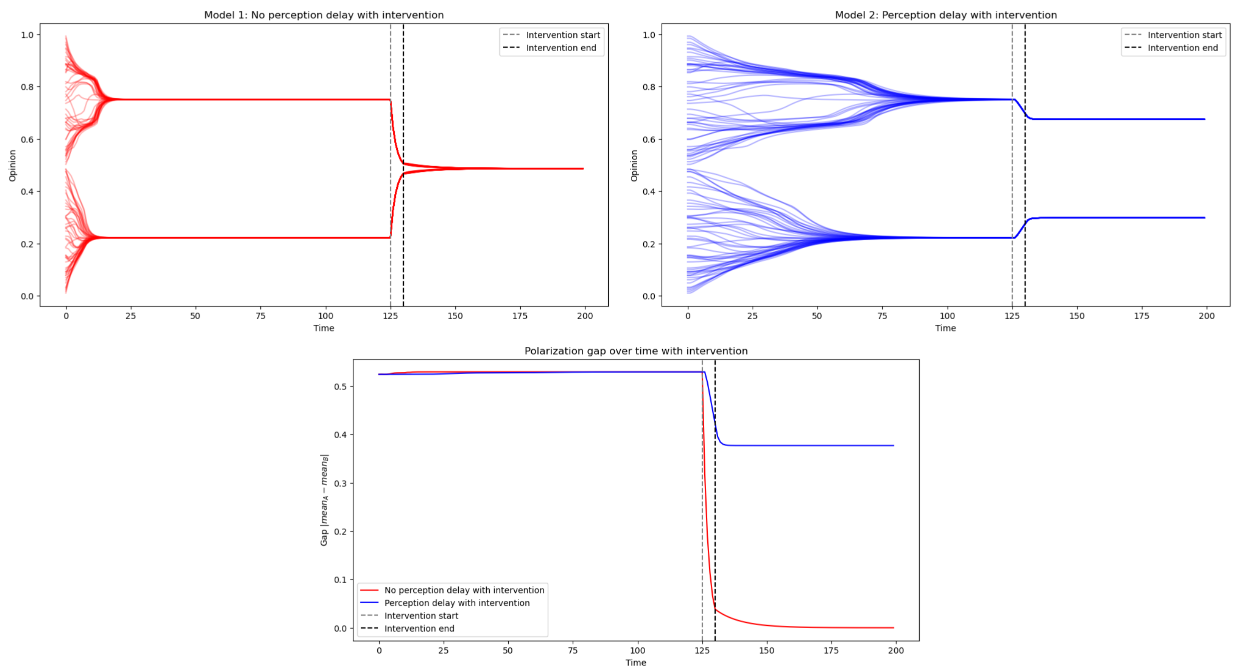

Figure 3: Opinion trajectories and polarization gap over time with short-term intervention

In contrast, Model 2 exhibits markedly different dynamics due to the presence of perception delay. When the same short intervention is applied, the reduction in polarization is partial and much weaker: the delayed perception prevents agents from fully internalizing the external signal before the intervention ends. After the intervention is withdrawn, opinions gradually drift back toward polarized positions rather than maintaining convergence. This behavior reflects a *delayed balancing feedback loop* induced by the first-order information delay, which slows the adjustment of perceptions and causes agents to respond to outdated, memory-like information. As a result, the system exhibits a sluggish and incomplete relaxation instead of immediate depolarization. Once polarization becomes established, it remains resistant to reversal, as the delayed perception stock absorbs external consensus signals only weakly and the homophilous network structure continues to reinforce division.

When the intervention duration is extended (as shown in Figure 4), the longer intervention in Model 1 again produces a rapid and complete depolarization, and the system maintains a unified opinion state thereafter. In contrast, Model 2 eventually achieves a comparable reduction in polarization, but only after prolonged exposure, as the delayed perceptions gradually catch up with the external signal. This comparison underscores the asymmetry introduced by perception delay: while both systems can respond to sustained intervention, the delayed model does so with significant inertia and path dependence, implying that once polarization becomes entrenched, reversing it requires interventions that are stronger and longer-lasting.

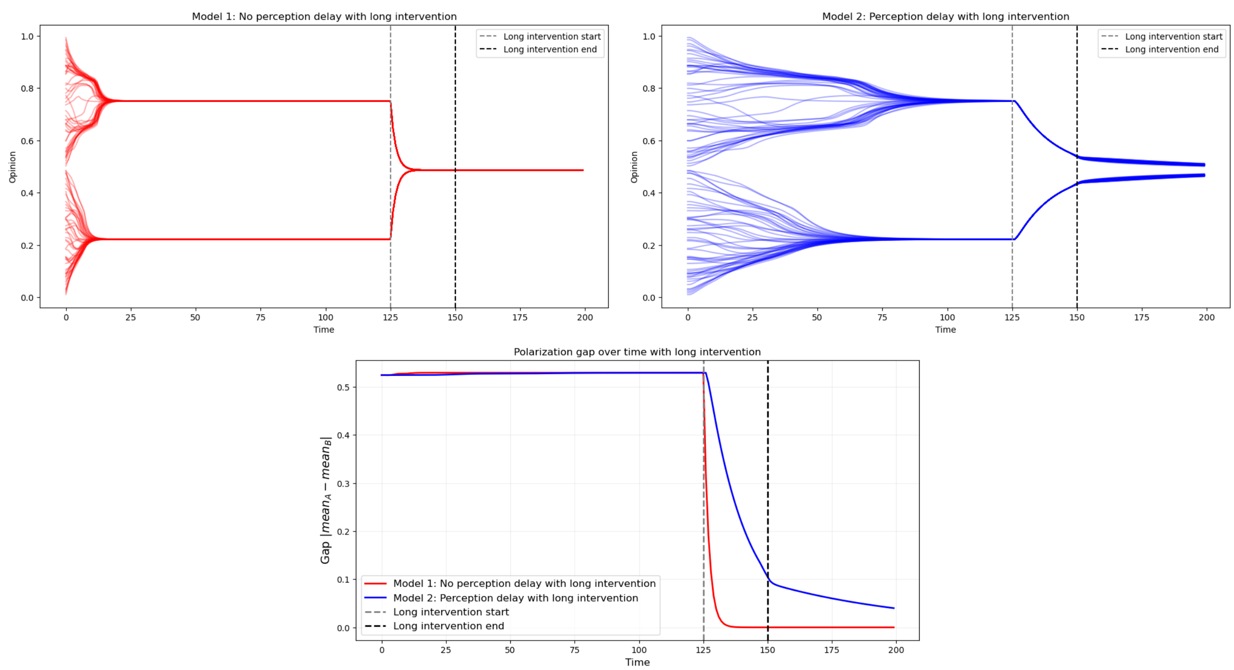

Figure 4: Opinion trajectories and polarization gap over time with long-term intervention

Overall, both models produce polarized steady states, yet the underlying dynamics differ fundamentally. Instantaneous perception leads to rapid polarization through fast reinforcing feedback, while delayed perception generates a distinctive “slow–deep” pattern: polarization unfolds gradually but becomes more persistent and resistant to change. This pattern arises from the interaction between a reinforcing loop driven by network homophily and bounded-confidence filtering, whose effective strength is dampened by delayed perception, and a delayed balancing loop formed by the opinion-updating and perception-adjustment process. Together, these dynamics demonstrate that perception delay functions as a structural mechanism explaining why polarization in opinion networks tends to evolve slowly yet remains difficult to reverse once established—a characteristic outcome of interacting reinforcing and delayed balancing feedbacks in system dynamics.

4 Conclusion and Discussion

This paper develops a formal model of opinion dynamics that incorporates perception delay. By embedding a first-order information delay structure inspired by system dynamics, the model captures how cognitive inertia shapes the slow formation and difficult reversal of polarization. The framework adds cognitive realism to classical opinion dynamics and helps explain why depolarization interventions often fail to achieve lasting effects.

Future work can further advance this framework by enriching the representation of perception delay itself. One promising direction is to make perception delay endogenous or adaptive, allowing individuals' responsiveness to evolve with their opinion extremity, emotional commitment, or social exposure. Another avenue is to introduce multi-timescale perception, distinguishing between fast information channels such as media and slower interpersonal influence, to capture the layered temporal structure of opinion formation. In addition, perception delay may interact with network adaptation, forming a feedback loop in which slow cognitive processing alters the pace of social tie reconfiguration and, in turn, reshapes exposure to information. Finally, future models could couple perception delay with other cognitive mechanisms such as confirmation bias or motivated reasoning, providing a richer account of how interacting biases and delays shape the persistence and evolution of polarization.

References

Allahverdyan, A. E. and Galstyan, A. (2014). Opinion dynamics with confirmation bias. *PloS one*, 9(7):e99557.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F., and Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221.

Biondi, E., Boldrini, C., Passarella, A., and Conti, M. (2023). Dynamics of opinion polarization. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(9):5381–5392.

Chen, T., Wang, X., and Tsourakakis, C. E. (2022). Polarizing opinion dynamics with confirmation bias. In *International Conference on Social Informatics*, pages 144–158. Springer.

Choi, Y.-P., Cicolani, C., and Pignotti, C. (2025). Time-delayed opinion dynamics with leader-follower interactions: consensus, stability, and mean-field limits. *arXiv preprint arXiv:2508.08157*.

Curiel, R. P. (2021). Opinion dynamics and the inevitability of a polarised and homophilic society. In *Theory of Complexity-Definitions, Models, and Applications*. IntechOpen.

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

Forrester, J. W. (1997). Industrial dynamics. *Journal of the Operational Research Society*, 48(10):1037–1041.

Hegselmann, R. (2015). Opinion dynamics and bounded confidence: models, analysis and simulation. *The Journal of Artificial Societies and Social Simulation*.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., et al. (2018). The science of fake news. *Science*, 359(6380):1094–1096.

Li, M., Zhang, J., and Jin, L. (2025). Opinion dynamics considering matthew effect with time delays and stubborn influence in social networks. *IEEE Transactions on Network Science and Engineering*.

Neirotti, J. (2023). Perturbed anisotropic opinion dynamics with delayed information. *Journal of Statistical Physics*, 190(8):141.

Sobkowicz, P. (2018). Opinion dynamics model based on cognitive biases of complex agents. *Journal of artificial societies and social simulation*, 21(4).

Sterman, J. (2002). System dynamics: systems thinking and modeling for a complex world.

Sunstein, C. R. (2009). *Going to extremes: How like minds unite and divide*. Oxford University Press.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., and Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)*.

Yao, L. and Li, A. (2025). Convergence of time-delayed opinion dynamics with complex interaction types. *arXiv preprint arXiv:2501.12219*.

Yu, H., Xue, B., Zhang, J., Liu, R.-R., Liu, Y., and Meng, F. (2023). Opinion cascade under perception bias in social networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(11).

Zafeiris, A. (2022). Opinion polarization in human communities can emerge as a natural consequence of beliefs being interrelated. *Entropy*, 24(9):1320.