Applied Mathematics and Computation 526 (2026) 130072

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Full Length Article

Spatial eviction in attraction-Repulsion opinion dynamics: From polarized enclaves to moderate consensus

Hong Zhang

Center for Economic Behavior & Decision-Making (CEBD), and School of Economics, Zhejiang University of Finance and Economics, Hangzhou, 310018, China

ARTICLE INFO

Keywords:
Opinion dynamics
Polarization
Attraction-repulsion model
Spatial eviction
Segregation
Phase transition

ABSTRACT

Exclusionary social processes—from neighborhood ostracism to online deplatforming—can reshape collective opinion, yet existing models of opinion polarization rarely incorporate the forced relocation of dissenters. We address this gap with an agent-based model that couples Attraction–Repulsion opinion dynamics with a spatial eviction mechanism on a two-dimensional lattice. Agents interact with nearby neighbors: like-minded pairs converge, while those exceeding a tolerance threshold repel each other and may expel the most dissimilar neighbor to a distant location. Across the parameter space of eviction frequency and tolerance, three dynamical phases emerge. Without eviction, mutual repulsion drives the population into two opposing ideological camps—yet, counterintuitively, these camps remain spatially intermixed. At high eviction rates, constant reshuffling of dissenters prevents extremist clusters from forming and drives the population toward moderate consensus. Between these extremes lies a fragile pluralistic regime in which moderate and extreme subpopulations coexist but are easily destabilized by stochastic perturbations. The central, policy-relevant insight is that exclusion plays a dual, frequency-dependent role: infrequent eviction entrenches polarization by creating homogeneous enclaves, whereas frequent eviction dissolves them. These findings provide a mechanistic framework for understanding how the rate—not merely the presence—of exclusionary processes shapes the trajectory from polarization to consensus.

1. Introduction

Understanding how societies become polarized requires models that capture both attraction to like-minded views and repulsion from opposing views. The Attraction–Repulsion Model (ARM) provides a parsimonious yet powerful framework for this phenomenon: agents tend to interact preferentially with similar others, converging in opinion when viewpoints fall within a tolerance threshold, but diverging when differences exceed that threshold [1–4]. ARM-type models belong to a broader family of coevolutionary games where strategies and network structure co-evolve [5,6]—an approach rooted in the social-physics tradition that brings statistical-mechanics and game-theoretic tools to bear on collective human behavior [7]. Complementary game-theoretical treatments of opinion dynamics on social networks further support this framing by deriving when social influence propagates or stalls under strategic interaction [8]. Recent extensions demonstrate that success-driven opinion formation—where agents preferentially adopt views associated with higher payoffs—can generate persistent social tensions even in the absence of explicit repulsion mechanisms [9]. This repulsive mechanism draws on Social Judgment Theory, which posits that messages falling within an individual's "latitude of rejection" produce contrast effects—shifting attitudes away from the source rather than toward it [10–12]. By incorporating this psychologically plausible

E-mail address: poyeker@gmail.com

https://doi.org/10.1016/j.amc.2026.130072
Received 6 February 2026; Received in revised form 14 March 2026; Accepted 21 March 2026
Available online 31 March 2026

mechanism, ARM-type models generate opinion polarization—a state where a population splits into distinct, internally homogeneous camps with growing inter-group disagreement [13,14]. This contrasts with classic bounded-confidence models, such as those of Deffuant et al. [15] and Hegselmann and Krause [16], which permit only convergence or indifference and thus cannot readily explain the emergence of extreme discord from moderate initial conditions [3,17,18].

Recent ARM-based studies have identified key drivers of polarization and potential interventions. Axelrod, Daymude, and Forrest [1] present a minimal ARM parameterized by *tolerance* (the opinion difference threshold separating attraction from repulsion), *responsiveness* (the magnitude of opinion shifts per interaction), and *exposure* (the probability of interacting with dissimilar others). Despite its simplicity, this model reproduces fundamental patterns of ideological polarization—though the adequacy of such simplified representations remains debated [19,20]. Notably, low tolerance emerges as a critical catalyst for runaway extremism, particularly when paired with high cross-group exposure: agents who are both intolerant and frequently encounter opposing views enter a positive feedback loop that drives opinions toward the extremes [1,21]. Counterintuitively, reducing exposure between hostile factions can slow or prevent extreme polarization by minimizing inflammatory encounters. This insight aligns with prior negative-influence models demonstrating that initial segregation of viewpoints can, under certain conditions, hinder rather than promote polarization [3,22]. ARM simulations have also revealed promising remedies: introducing even modest self-interest incentives for moderate positions substantially reduces polarization, suggesting that policy interventions emphasizing common interests could mitigate ideological fragmentation [1]. The evolution of ARM-based models—from early negative-influence theories to recent multi-factor experiments—has yielded non-intuitive findings that underscore both the power of simple interaction rules and the need to incorporate real-world complexities [23,24].

One such complexity is spatial structure—how opinions are embedded in geographic or network space. While some models assume well-mixed populations, others impose explicit spatial topologies where only neighbors influence each other. A growing literature examines how such structure affects polarization dynamics, building on foundational work by Schelling [25], which showed how individual preferences for like-minded neighbors can generate large-scale residential segregation. Adaptive network models, in which agents rewire social ties based on opinion agreement, demonstrate that co-evolution of opinions and interaction patterns often leads to echo chambers—clusters of like-minded agents that can either stabilize subcultures or, under some conditions, preserve moderate bridges [26–29]. Chu et al. [27] incorporated geographic distance into an adaptive voter model and showed that political shocks amplify regional homogenization in areas with initial local majorities while maintaining diversity in initially divided regions, thereby exacerbating inter-regional polarization. Beyond network rewiring, models incorporating physical agent relocation reveal that even modest mobility can dramatically alter collective outcomes. Baumgaertner et al. [30] demonstrated that periodic shuffling of agents in a spatial opinion model breaks up entrenched opinion clusters and hastens consensus; migration-based coevolutionary studies likewise show that relocation frequency can reshape collective outcomes by modifying adverse local environments [31]; recent work by Okada [32] has further explored how positive and negative spatial influences interact. However, the effects of spatial mixing are nuanced: Feliciani et al. [3] found that under negative-influence dynamics, greater initial mixing actually intensifies polarization, whereas segregation reduces cross-group hostility. These contrasting findings highlight that the timing, type, and direction of spatial mixing can have opposite impacts on polarization trajectories.

Despite these advances, current ARM-based models face important limitations. Most assume fixed population positions with static networks or global mixing, allowing only opinions to evolve and thereby abstracting away explicit spatial or network geometry [1]. This precludes representing phenomena such as clustering, migration, or the formation of physically segregated enclaves over time—processes that reinforce polarization as individuals re-sort into like-minded neighborhoods. Recent work integrating Schelling-type residential dynamics with opinion formation demonstrates that mobility and social influence can create self-reinforcing echo chambers [33–35], yet such feedbacks between opinion dynamics and residential patterns remain largely unexplored in ARM frameworks [27,30]. A second limitation is the omission of institutional or enforced exclusion mechanisms. Real societies exhibit not merely voluntary segregation but institutionalized expulsion: dissenters may be ostracized from communities, removed from media platforms, or exiled by authoritarian regimes. Empirical research confirms that social exclusion heightens susceptibility to radicalization [36]. Recent work on social media deplatforming provides striking evidence of these dynamics: Mekacher et al. [37] found that removing influential accounts produces systemic ripple effects throughout platform ecosystems, while Buntain and Snegovaya [38] documented long-term shifts in ideological polarization following the January 2021 deplatforming wave. Notably, Ribeiro et al. [39] showed that deplatformed users often migrate to fringe platforms rather than moderating their views, suggesting that spatial exclusion in digital spaces may parallel physical eviction dynamics. Such processes could amplify polarization by purging moderating voices and isolating extremes, or occasionally dampen it if removal of incendiary actors enables compromise. Yet prevailing opinion models rarely incorporate agent removal or forced relocation triggered by opinion conflicts. The ARM framework has not included mechanisms for "neighborhood eviction," where agents are compelled to leave due to irreconcilable disagreement with their neighbors. This gap precludes studying how ostracism or spatial banishment might itself drive polarization patterns [23].

This paper addresses these limitations by proposing a novel agent-based model that combines ARM opinion dynamics with spatially explicit eviction rules. Agents hold continuous opinions and interact under the standard attraction–repulsion dynamic [2]. We introduce an additional mechanism: when an agent's local environment becomes sufficiently hostile—specifically, when opinion differences with neighbors exceed a threshold—an eviction event relocates the dissenting agent to an empty location elsewhere in the spatial domain. This rule captures, in abstract form, phenomena such as neighborhood ostracism, group expulsion, or banishment observed in polarized societies. We systematically investigate how this spatial eviction mechanism affects polarization trajectories across parameter space, identifying distinct dynamical phases—extremist, mixed, and consensus—and characterizing their dependence on population density, neighborhood geometry, initial conditions, and stochastic noise. The following sections detail the model

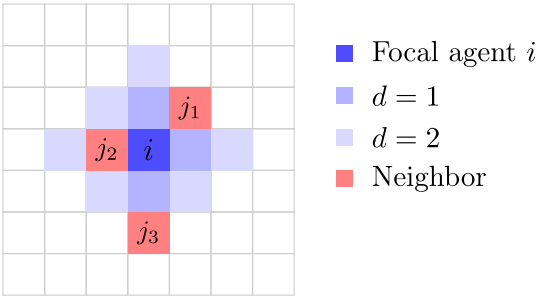

Fig. 1. Manhattan neighborhood with radius $r = 2$. The focal agent i (dark blue) can interact with agents within Manhattan distance 2. Cells are shaded by distance: $d = 1$ (medium blue) and $d = 2$ (light blue). Agents j_1, j_2, j_3 (red) occupy cells within this neighborhood. The tunable radius r provides multi-scale social horizons: $r = 1$ restricts interactions to four immediate neighbors, while larger radii progressively widen the pool of potential interaction partners.

formulation, analyze its behavior under various parameter regimes, and discuss how polarization outcomes change when agents can not only repel each other ideologically but also rearrange spatially.

2. Model

We present an agent-based model that synthesizes three distinct but interrelated processes: spatially constrained social interaction, opinion dynamics governed by attraction and repulsion, and institutionalized eviction of ideological dissenters. Building upon the Attraction–Repulsion Model (ARM) introduced by Axelrod et al. [1], our framework embeds agents in explicit two-dimensional space where geographic proximity determines who can influence whom. This spatial grounding—absent from the original ARM—enables us to investigate how the interplay between ideological divergence and physical relocation shapes polarization trajectories. The model's key innovation lies in coupling opinion-based repulsion with a concrete spatial consequence: agents sufficiently intolerant of their neighbors can expel them to distant locations, a mechanism that abstracts real-world phenomena ranging from neighborhood ostracism to platform deplatforming.

2.1. Spatial structure

Spatial structure profoundly shapes opinion dynamics. Schelling's seminal work [25] demonstrated that even mild preferences for like-minded neighbors can generate striking patterns of residential segregation; subsequent research has shown that such spatial sorting interacts with opinion formation to create self-reinforcing echo chambers [28,33]. Motivated by these insights, we situate agents on a discrete spatial substrate where interactions are constrained by physical proximity.

Agents occupy cells on an $L \times L$ square lattice with periodic (toroidal) boundary conditions, eliminating edge effects that might otherwise distort dynamics near boundaries. Each cell accommodates at most one agent, and the total population consists of $N \leq L^2$ agents distributed across the grid. At initialization, agents are placed uniformly at random on distinct cells, yielding a spatially homogeneous starting configuration with density $\rho = N/L^2$.

The neighborhood structure determines who can interact with whom. For agent i located at position (x_i, y_i), the neighborhood $\mathcal{N}_i$ comprises all agents within Manhattan distance r:

$$\mathcal{N}_i = \{j : |x_i - x_j| + |y_i - y_j| \leq r,\ j \neq i\}, \tag{1}$$

where coordinate differences are computed with periodic wrapping to respect toroidal geometry. The Manhattan metric—counting horizontal and vertical steps rather than Euclidean distance—yields a characteristic diamond-shaped neighborhood (Fig. 1). This choice reflects the discrete nature of grid-based movement while providing a tunable interaction radius: setting $r = 1$ restricts influence to the four cardinal neighbors, whereas larger values progressively expand the social horizon. We adopt a default of $r = 4$, which yields a diamond-shaped neighborhood containing up to $2r(r + 1) = 40$ cells; at the default density $\rho = 0.6$ this provides approximately 24 expected neighbors—large enough for meaningful social sampling yet small enough that interactions remain spatially local. Section 3.3 systematically varies $r \in \{1, 2, 4, 8\}$ and confirms that the three-phase structure is qualitatively preserved across this range, so the specific default choice does not drive the main results.

2.2. Opinion dynamics

The opinion update mechanism draws on Social Judgment Theory [10], which posits that individuals possess latitudes of acceptance and rejection surrounding their current position. Messages falling within the latitude of acceptance produce assimilation—attitude shift toward the source—while those in the latitude of rejection trigger contrast effects, pushing attitudes away [11]. The Attraction–Repulsion Model operationalizes this insight through a tolerance threshold that separates convergent from divergent influence.

Each agent i holds a continuous opinion $o_i \in [0, 1]$, representing position along a single ideological dimension. While real-world attitudes span multiple dimensions, this unidimensional representation suffices to capture the essential polarization dynamics of interest [13] and permits tractable analysis. Initial opinions are drawn independently from a truncated normal distribution $\mathcal{N}(\mu, \sigma^2)$ clipped to $[0, 1]$, allowing control over the initial degree of consensus (σ small) or diversity (σ large).

The model evolves through discrete time steps. During each step, N activations are performed by sampling agents uniformly at random (with replacement), so each agent is activated once on average per step. When agent i is activated, the opinion update unfolds as follows.

First, agent i selects a random neighbor $j \in \mathcal{N}_i$. If the neighborhood is empty ($\mathcal{N}_i = \emptyset$), the activation ends and neither opinion updating nor eviction is attempted. This stochastic partner selection introduces noise that prevents deterministic lock-in and reflects the unpredictability of real social encounters.

Second, the probability that i and j engage in substantive interaction depends on their opinion distance $d_{ij} = |o_i - o_j|$ and an *exposure* parameter E:

$$p_{\text{interact}} = \left(\frac{1}{2}\right)^{d_{ij}/E}. \tag{2}$$

This formulation captures selective exposure—the tendency to engage more readily with like-minded others. When E is small, the interaction probability decays rapidly with opinion distance, creating strong homophily; agents effectively inhabit echo chambers where they rarely encounter dissenting views. Conversely, large E yields nearly uniform interaction probabilities regardless of opinion difference, modeling forced exposure or highly diverse social environments. The exponential form ensures $p_{\text{interact}} = 1$ when $d_{ij} = 0$ (identical opinions always interact) and decays smoothly toward zero as d_{ij} grows.

Third, conditional on interaction, agent i updates its opinion according to the attraction–repulsion rule:

$$o_i' = o_i + s \cdot R \cdot (o_j - o_i), \tag{3}$$

where $R \in (0, 1]$ is the *responsiveness* parameter and s encodes attraction versus repulsion. Responsiveness controls the step size of each opinion update: small R produces gradual nudges that slow both convergence and divergence, while large R amplifies every interaction, accelerating both assimilation among like-minded agents and polarization between dissimilar ones. Consequently, R modulates the *speed* of opinion dynamics without altering the *direction* set by the tolerance threshold—but because faster dynamics leave less time for spatial rearrangement, R also shifts the phase boundaries in the (p_m, τ) plane (Section 3.1). The sign s depends on whether j's opinion falls within i's latitude of acceptance:

$$s = \begin{cases} +1 & \text{if } d_{ij} \le \tau \quad \text{(attraction)}, \\ -1 & \text{if } d_{ij} > \tau \quad \text{(repulsion)}. \end{cases} \tag{4}$$

The *tolerance* threshold τ demarcates this boundary: when the opinion gap is small ($d_{ij} \le \tau$), agent i moves toward j, reflecting the natural tendency to converge with similar others. When the gap exceeds tolerance ($d_{ij} > \tau$), however, agent i moves *away* from j—a repulsion effect that can drive opinions toward opposite extremes. The updated opinion o_i' is clipped to $[0, 1]$ to maintain the bounded domain.

This formulation preserves the core ARM dynamics of Axelrod et al. [1] while embedding them in spatial structure. The critical difference is that interactions now occur only among spatial neighbors rather than in a well-mixed population, allowing geographic clustering to emerge and co-evolve with opinion dynamics.

2.3. Eviction mechanism

Beyond ideological divergence, real societies exhibit institutionalized forms of spatial exclusion. Dissenters may find themselves ostracized from their communities, deplatformed from online spaces, or—in extreme cases—exiled by authoritarian regimes. The recent wave of social media deplatforming provides a natural experiment in how enforced relocation affects polarization dynamics [37–39]. While empirical studies find mixed effects—deplatforming can reduce harmful content on mainstream platforms but often drives users to fringe alternatives where views may intensify—the phenomenon illustrates how ideological minorities can be forcibly relocated from one social space to another. Such processes may amplify polarization by purging moderating voices from local environments or, alternatively, dampen conflict by separating incompatible factions. Our eviction mechanism shares structural similarities with "adverse neighborhood" models in evolutionary game theory, where agents respond to hostile environments through relocation or link rewiring [40]. To investigate these dynamics in a controlled setting, we introduce an eviction mechanism whereby agents can expel ideologically distant neighbors.

Fig. 2 illustrates the complete activation sequence. After the interaction check (Eq. (2)), and regardless of whether an opinion update occurs, agent i attempts eviction with probability p_m by targeting the most dissimilar neighbor. This probabilistic triggering allows tuning the relative importance of eviction versus pure opinion dynamics. An additional *eviction noise* parameter η controls whether the tolerance check is bypassed (see below).

The eviction process operates as follows. Agent i first identifies the most ideologically dissimilar neighbor:

$$j^* = \arg\max_{j \in \mathcal{N}_i} |o_i - o_j|. \tag{5}$$

If multiple neighbors are equally most dissimilar, ties are broken by the scan order. This targeting rule reflects the intuition that the most deviant member of one's social environment draws the strongest exclusionary impulse. Whether eviction actually occurs depends

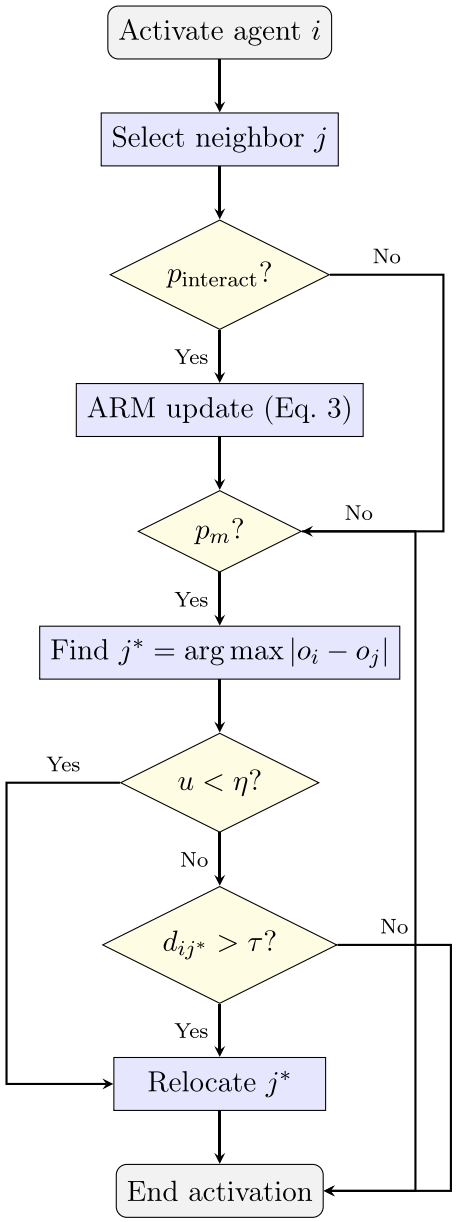

Fig. 2. Flowchart of a single agent activation. After selecting a neighbor, opinion updating occurs with probability p_{interact} (Eq. (2)). Subsequently, with probability p_m, the agent identifies the most dissimilar neighbor j^*. With probability η (*eviction noise*), j^* is evicted unconditionally; otherwise, eviction requires $|o_i - o_{j^*}| > \tau$. The sequential structure ensures that opinion updating precedes eviction, coupling ideological and spatial dynamics within each activation.

on the *eviction noise* parameter $\eta \in [0, 1]$: with probability η, eviction proceeds unconditionally, bypassing the tolerance check; with probability $1 - \eta$, eviction requires that the opinion distance to the target exceed the tolerance threshold:

$$|o_i - o_{j^*}| > \tau. \tag{6}$$

Formally, agent j^* is evicted if and only if a uniform draw $u < \eta$ or $|o_i - o_{j^*}| > \tau$. When $\eta = 0$ (the default), eviction is purely threshold-driven: if all neighbors fall within tolerance, no eviction occurs. When $\eta > 0$, a fraction of evictions bypass the ideological criterion, modeling "noisy" or impulsive exclusion—situations where eviction occurs not because a neighbor is genuinely intolerable but through stochastic social processes such as arbitrary enforcement, interpersonal friction, or institutional caprice. When the eviction condition is met, agent j^* is relocated to a uniformly random empty cell on the grid (if one exists), while agent i remains in place. If no empty site exists (e.g., $\rho = 1$), relocation fails and the eviction attempt has no effect.

Several design choices merit emphasis. First, the same tolerance parameter τ governs both opinion repulsion and eviction eligibility, creating a unified threshold of ideological intolerance: agents who repel in opinion space also exclude in physical space. This coupling reflects the psychological consistency whereby those most averse to dissenting views may also act to remove such views from their environment. Second, eviction is asymmetric: the evicting agent stays while the target moves. This captures scenarios where established residents expel newcomers or where majority factions banish dissenters. Third, relocation to a random empty cell—rather than a nearby location or a chosen destination—abstracts the often-arbitrary nature of forced displacement, where exiles land wherever circumstances permit rather than optimizing their new position.

2.4. Outcome measures

Polarization is a multifaceted concept encompassing dispersion of opinions, concentration at extremes, and spatial segregation [13]. No single metric captures all dimensions; accordingly, we employ three complementary measures that together characterize the system's state.

Polarization (opinion variance) quantifies the overall spread of the opinion distribution:

$$\mathrm{Var}(o) = \frac{1}{N}\sum_{i=1}^{N}(o_i - \bar{o})^2, \tag{7}$$

where $\bar{o} = N^{-1}\sum_i o_i$ is the population mean. Higher polarization indicates greater ideological diversity—opinions are more widely dispersed. However, variance alone cannot distinguish between unimodal dispersion (many moderates with some outliers) and bimodal polarization (two opposing camps with few moderates). To avoid confusion with the initial distribution's σ, we denote opinion variance as $\mathrm{Var}(o)$ (or σ_o^2 in figures).

Extremity fraction addresses this limitation by measuring the proportion of agents holding positions near the ideological poles in opinion space:

$$f_{\text{extreme}} = \frac{|\{i : o_i < \varepsilon \text{ or } o_i > 1-\varepsilon\}|}{N}, \tag{8}$$

where ε is a small threshold (we use $\varepsilon = 0.1$). A high extremity fraction combined with high variance signals bimodal polarization—the population has split into opposing camps clustered at the extremes. Low extremity with high variance suggests dispersed but moderate opinions; low extremity with low variance indicates consensus.

Spatial autocorrelation captures whether similar opinions cluster geographically. We employ Moran's I statistic [41], a standard measure of spatial association:

$$I = \frac{N}{W}\cdot\frac{\sum_i\sum_{j\in\mathcal{N}_i}(o_i-\bar{o})(o_j-\bar{o})}{\sum_i(o_i-\bar{o})^2}, \tag{9}$$

where $W = \sum_i |\mathcal{N}_i|$ is the total number of neighbor pairs. We use binary adjacency weights without row standardization. Positive I indicates positive spatial autocorrelation: similar opinions cluster together, as in residential segregation. Negative I indicates spatial mixing of opposing views, where neighbors tend to differ. Values near zero suggest spatial randomness—opinion and location are independent. When $\mathrm{Var}(o)$ becomes very small, I can be inflated; we therefore interpret I alongside $\mathrm{Var}(o)$ and spatial snapshots rather than as a standalone indicator.

Together, these three metrics characterize the ideological landscape: variance measures dispersion, extremity fraction detects bimodality, and Moran's I reveals spatial patterning. Their joint evolution over time traces the trajectory from initial conditions through any emergent polarization or segregation.

2.5. Parameter summary

Table 1 summarizes the model parameters. Default values are informed by prior ARM implementations [1] and adjusted to balance computational tractability with empirical plausibility for spatial dynamics.

3. Results

We map dynamical regimes, characterize temporal and spatial structure, assess sensitivity to density and neighborhood geometry, and test robustness to initial conditions and stochastic perturbations. Unless stated otherwise, simulations use the default parameters in Table 1 with $L = 50$, $\rho = 0.6$, and 10 000 time steps, averaged over ensembles of 20–100 independent realizations. Steady-state metrics are computed as the mean over the final 100 time steps of each run.

3.1. Phase identification

We sweep the eviction probability $p_m \in [0, 1]$ and tolerance $\tau \in [0, 0.5]$ on a 41×41 parameter mesh, measuring the steady-state extremity fraction f_{extreme} (Eq. (8)) for each parameter combination. Fig. 3 shows phase diagrams for four combinations of exposure $E \in \{0.1, 0.2\}$ and responsiveness $R \in \{0.2, 0.4\}$, with each point averaged over 50 runs.

Table 1
Model parameters and default values.

Symbol	Parameter	Description	Default
Spatial structure			
L	Grid size	Side length of square lattice	50
N	Population	Number of agents	1500
ρ	Density	Population density N/L^2	0.6
r	Neighborhood range	Manhattan distance for neighbors	4
Opinion dynamics			
μ	Initial mean	Mean of initial opinion distribution	0.5
σ	Initial SD	Std. dev. of initial opinions	0.2
τ	Tolerance	Attraction/repulsion threshold	0.2
R	Responsiveness	Opinion change magnitude	0.2
E	Exposure	Interaction selectivity	0.1
Eviction mechanism			
p_m	Eviction probability	Eviction attempt probability per activation	0.2
η	Eviction noise	Unconditional eviction probability	0
Metrics			
ϵ	Extremity threshold	Extreme opinion cutoff	0.1

Three distinct phases emerge across all parameter panels. The *extremist* phase occupies the region of low p_m and low τ, where $f_{\text{extreme}} > 0.80$: the population splits into two opposing camps clustered at the ideological poles in opinion space. The *consensus* phase appears at high p_m, where frequent eviction disrupts nascent clusters and drives convergence toward a unimodal consensus ($f_{\text{extreme}} < 0.15$). Between these lies a *mixed* regime with intermediate f_{extreme}, with coexistence of moderate and extreme subpopulations. The phase boundaries, marked by contour lines at $f_{\text{extreme}} = 0.15$ and 0.80, shift systematically with exposure E and responsiveness R: higher exposure broadens the extremist region by increasing the frequency of cross-group encounters that trigger repulsion, while higher responsiveness amplifies both attraction and repulsion, sharpening the phase transitions. Because f_{extreme} varies smoothly across parameter space, modest threshold shifts move these contours without altering the overall three-regime topology; the continuous fields in Fig. 3 permit alternative cutoffs if desired.

The three phases differ not only in their opinion distributions but also in their spatial organization. Fig. 4 displays representative steady-state snapshots of the 50×50 lattice for each phase (with $\tau = 0.2$, $E = 0.1$, $R = 0.2$, and $r = 4$). In the extremist phase ($p_m = 0$), the spatial pattern remains mixed: red and blue agents are interspersed with no stable contiguous domains, and occasional weak anti-alignment is visible, consistent with Moran's I near zero. Gray cells mark unoccupied sites. The mixed regime ($p_m = 0.075$) exhibits partial spatial ordering, with localized same-opinion patches interspersed with moderate agents. In the consensus phase ($p_m = 0.8$), frequent eviction events continuously displace dissenting agents, preventing the formation of stable clusters; opinions converge toward a unimodal consensus, and the spatial distribution appears nearly homogeneous.

3.2. Temporal dynamics

To track how each phase emerges from identical initial conditions, we follow the opinion distribution and spatial autocorrelation over time. Fig. 5 presents kymographs—time–opinion density plots—alongside the evolution of Moran's I for the three representative phases.

In the extremist phase [Fig. 5(a,b)], the initially unimodal opinion distribution centered at $\mu = 0.5$ rapidly splits into two peaks migrating toward the extremes. This bifurcation is driven by the repulsion mechanism: agents with $|o_i - o_j| > \tau$ push each other apart, and in the absence of eviction ($p_m = 0$), this positive feedback loop proceeds unchecked. Because no eviction occurs, agents remain at their initial (random) positions throughout the simulation; Moran's I consequently stays near zero or drifts slightly negative, indicating that the emergent ideological polarization is not accompanied by spatial sorting—extreme and moderate agents remain random neighbors despite holding opposing views.

The mixed regime [Fig. 5(c,d)] exhibits a slower, more nuanced evolution. The opinion distribution broadens and develops partial bimodality, but a substantial moderate population persists. Moderate eviction ($p_m = 0.075$) begins to sort agents spatially: Moran's I rises to strongly positive values, indicating that like-minded agents cluster as dissimilar neighbors are gradually expelled. The consensus phase [Fig. 5(e,f)] converges rapidly: frequent eviction ($p_m = 0.8$) drives opinions toward a narrow consensus while simultaneously sorting agents spatially. Moran's I also rises to high positive values; however, because the opinion distribution narrows substantially, even weak residual spatial correlations are amplified by the shrinking denominator of Eq. (9). This highlights a known limitation of Moran's I: when global opinion variance approaches zero, the statistic becomes sensitive to small perturbations and should be interpreted with caution in conjunction with $\text{Var}(o)$ and spatial snapshots.

3.3. Density and neighborhood effects

We next examine how population density $\rho = N/L^2$ and neighborhood range r modulate behavior. Fixing $p_m = 0.8$ (consensus regime) and varying N from 100 to 2500 ($\rho \in [0.04, 1.0]$) with $r \in \{1, 2, 4, 8\}$ and $\tau \in \{0.1, 0.15, 0.2, 0.25, 0.3\}$, we measure steady-state polarization and spatial autocorrelation, each averaged over 100 realizations. At $\rho = 1$ (all cells occupied), eviction attempts fail

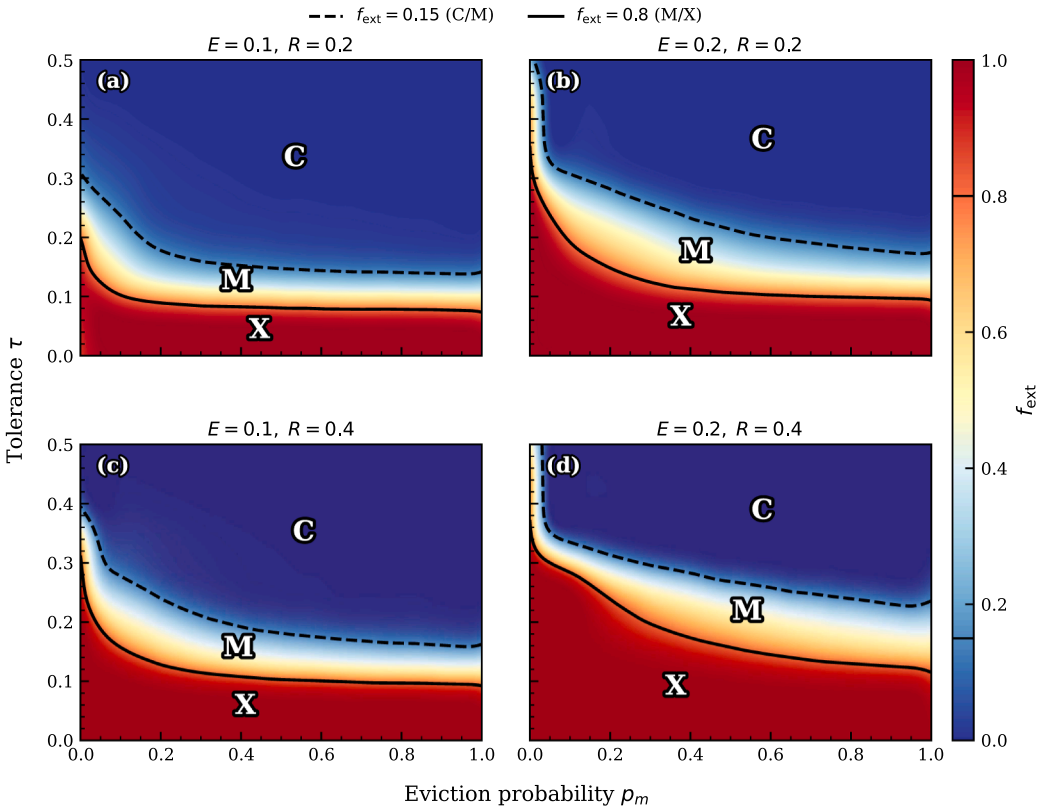

Fig. 3. Phase diagram of the extremity fraction f_{extreme} in the (p_m, τ) plane. Panels correspond to different combinations of exposure E and responsiveness R: (a) $E = 0.1$, $R = 0.2$; (b) $E = 0.2$, $R = 0.2$; (c) $E = 0.1$, $R = 0.4$; (d) $E = 0.2$, $R = 0.4$. Dashed contour: $f_{\text{extreme}} = 0.15$ (C/M boundary); solid contour: $f_{\text{extreme}} = 0.80$ (M/X boundary). Labels X, M, C denote extremist, mixed, and consensus phases. Color scale from blue (low f_{extreme}) through yellow to red (high f_{extreme}). Each point is averaged over 50 realizations with $\rho = 0.6$, $r = 4$, and 10 000 time steps. The three-phase topology is robust across all (E, R) combinations; higher exposure and responsiveness expand the extremist region by amplifying repulsive encounters.

because no empty cell is available; the model reduces to a spatial ARM on a fixed lattice topology (no relocation), so the $\rho = 1$ points are a boundary case.

Fig. 6 shows polarization Var(o) versus density for each neighborhood range. Two trends emerge. First, tolerance τ is the dominant control parameter: lower τ produces systematically higher polarization regardless of density or neighborhood size, consistent with the phase diagram results. Second, for a given τ, increasing density generally increases polarization, as higher ρ provides more neighbors within range r and thus more opportunities for repulsive interactions with dissimilar agents. This effect is strongest at small r [panel (a), $r = 1$], where the neighborhood contains at most four agents and additional neighbors have a larger marginal impact, and diminishes at large r [panel (d), $r = 8$], where even moderate densities already saturate the neighborhood.

Fig. 7 presents the corresponding Moran's I values. Spatial autocorrelation follows a similar pattern: lower τ and higher ρ promote spatial clustering. The effect of neighborhood range r is more nuanced. At $r = 1$, Moran's I can be high because the metric is computed over the same small neighborhood in which clustering occurs. As r increases, Moran's I averages over larger neighborhoods and becomes more sensitive to long-range disorder, yielding lower values. Thus local opinion clusters may form at any r, but large neighborhoods dilute the apparent spatial autocorrelation. Because Moran's I is computed using the same neighborhood definition $\mathcal{N}_i$ as interactions, varying r changes the weight matrix and hence the measurement scale; cross-r comparisons should therefore be interpreted as multi-scale diagnostics rather than a single fixed-scale statistic.

3.4. Robustness

To verify that the three-phase structure is not an artifact of specific initial conditions, we repeat the simulations with four initializations: a baseline ($\mu = 0.5$, $\sigma = 0.2$), a left-skewed distribution ($\mu = 0.2$, $\sigma = 0.1$), a right-skewed distribution ($\mu = 0.8$, $\sigma = 0.1$), and a broad uniform-like distribution ($\mu = 0.5$, $\sigma = 0.5$). Fig. 8 displays the steady-state opinion distributions $P(o)$, polarization Var(o), and Moran's $I(t)$ for each initial condition across the three phases.

In all three phases, the qualitative regime is robust to initialization: the extremist regime remains bimodal with mass near the extremes, the mixed regime retains an intermediate profile, and the high-p_m regime yields a narrow unimodal consensus. However, steady-state distributions need not be identical across initial conditions. In particular, in the consensus regime ($p_m = 0.8$) the distri-

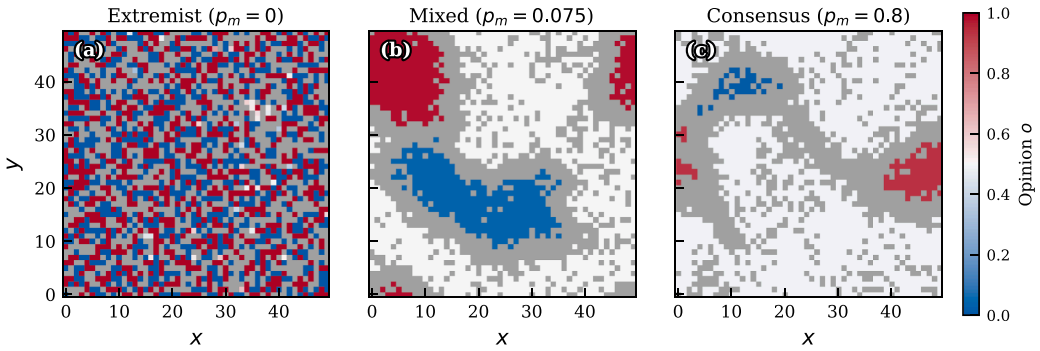

Fig. 4. Steady-state spatial snapshots for three representative phases. Each panel shows the 50×50 lattice after 10 000 time steps at density $\rho = 0.6$. Cell color encodes opinion o on a diverging blue–white–red scale ($o = 0$ blue, $o = 0.5$ white, $o = 1$ red); gray cells are unoccupied. (a) Extremist phase ($p_m = 0$): spatially mixed pattern with no persistent contiguous domains. (b) Mixed regime ($p_m = 0.075$): fragmented clusters with moderate agents interspersed. (c) Consensus phase ($p_m = 0.8$): spatially homogeneous distribution near $o \approx 0.5$. Fixed parameters: $\tau = 0.2$, $E = 0.1$, $R = 0.2$, $r = 4$. Spatial patterning differs qualitatively across phases: ideological polarization without spatial sorting under no eviction gives way to fragmented clustering under moderate eviction and near-uniform consensus under frequent eviction.

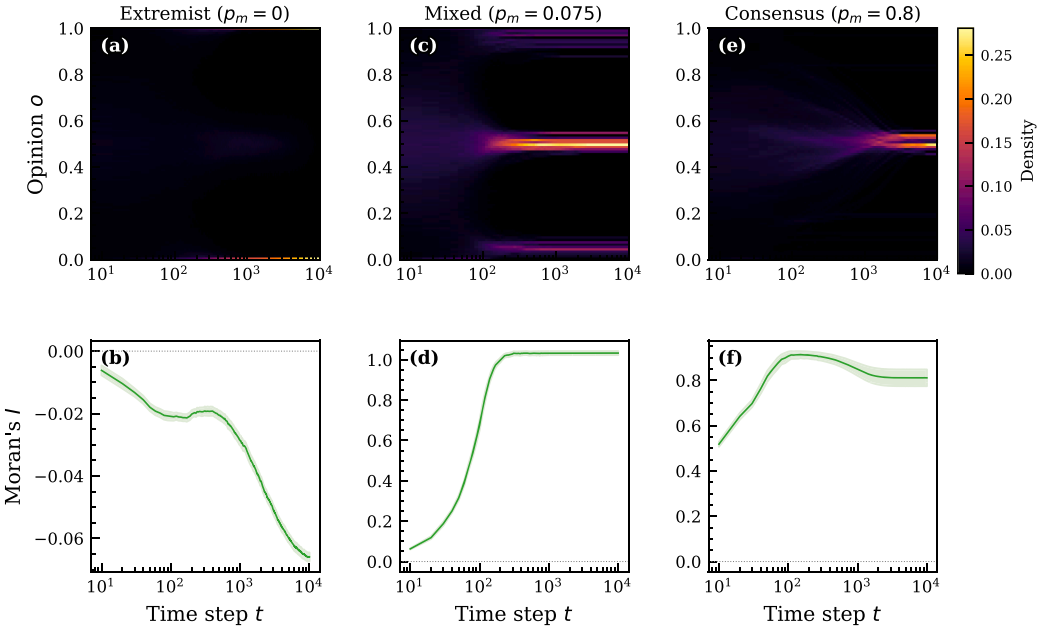

Fig. 5. Temporal evolution of opinion distributions and spatial autocorrelation. Top row: kymographs showing the ensemble-averaged opinion density $P(o, t)$ on a logarithmic time axis. Bottom row: Moran's $I(t)$ (mean ± SEM over 20 realizations). Columns correspond to (a,b) extremist ($p_m = 0$), (c,d) mixed ($p_m = 0.075$), and (e,f) consensus ($p_m = 0.8$) phases. Parameters: $\tau = 0.2$, $E = 0.1$, $R = 0.2$, $r = 4$, $\rho = 0.6$. The extremist phase shows rapid opinion bifurcation without spatial sorting; the consensus phase achieves both opinion convergence and spatial homogenization; the mixed regime exhibits the slowest, most nuanced trajectory.

bution concentrates but its location depends on the initialization (in our experiments it closely tracks the initial mean), whereas the mixed regime exhibits initialization-dependent peak structure. Transients differ across initializations, but the regime classification and phase structure in Fig. 3 remain stable.

Finally, we test sensitivity to stochastic perturbations by varying the eviction noise parameter η (Section 2): when an eviction attempt is triggered (with probability p_m), a fraction η of attempts bypass the tolerance check and unconditionally evict the most dissimilar neighbor. Fig. 9 shows the response to $\eta \in \{0, 0.01, 0.05, 0.1\}$.

Because eviction attempts occur with probability p_m, η affects dynamics only when $p_m > 0$; at the extremist reference point ($p_m = 0$) the system is unaffected by η by construction. The consensus phase is largely insensitive to η, as eviction is already frequent and additional random evictions produce negligible marginal effects. The mixed regime is the most sensitive, as its intermediate structure—maintained by a delicate balance between attraction, repulsion, and eviction—is readily destabilized by even modest noise levels. These results suggest that the extremist and consensus phases are structurally stable attractors, while the mixed regime occupies a more fragile transitional region in parameter space.

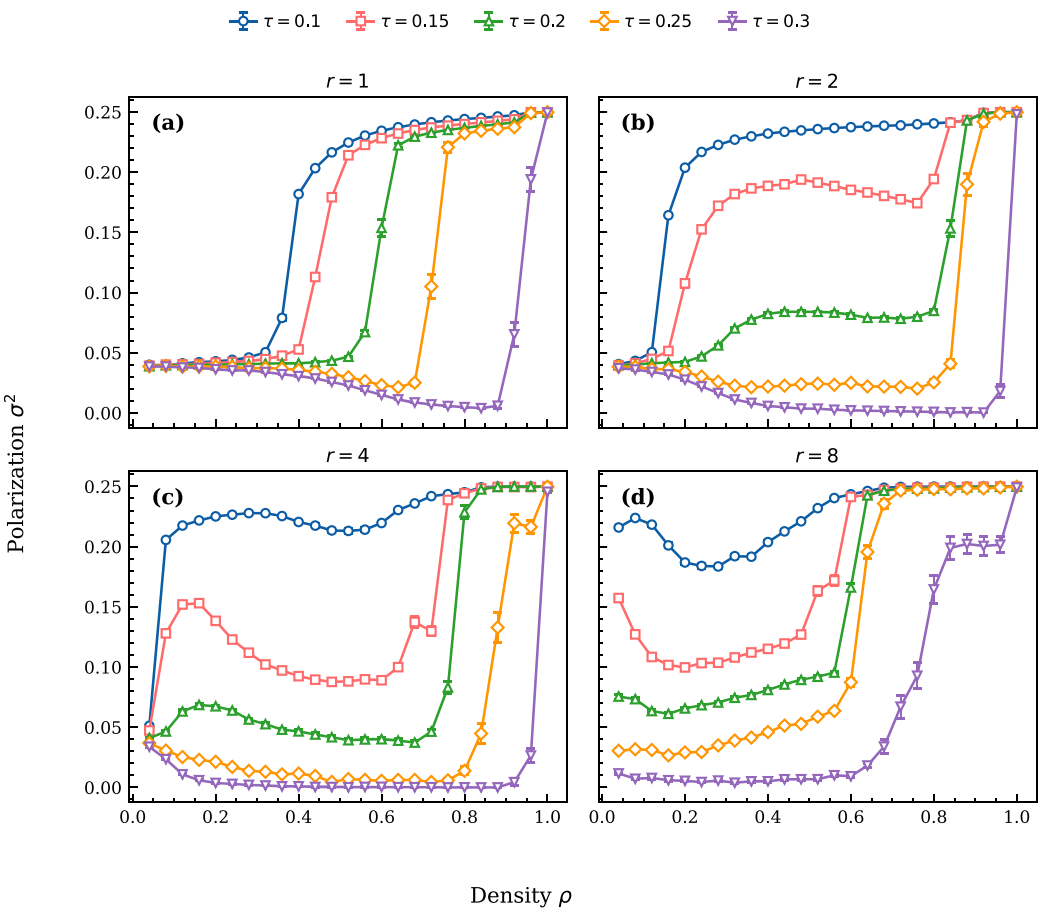

Fig. 6. Steady-state polarization Var(o) (denoted σ_o^2 in plots) versus density ρ for four neighborhood ranges: (a) $r = 1$, (b) $r = 2$, (c) $r = 4$, (d) $r = 8$. Each curve corresponds to a different tolerance τ (see legend). Error bars: standard error of the mean over 100 realizations. Points at $\rho = 1$ correspond to the eviction-disabled limit. Fixed: $p_m = 0.8$, $E = 0.1$, $R = 0.2$. Tolerance is the dominant driver of polarization; density amplifies repulsive interactions, especially in small neighborhoods where additional neighbors have the largest marginal impact.

4. Discussion

The central contribution of this work is coupling two previously separate mechanisms—ideological repulsion from the Attraction–Repulsion Model (ARM) [1] and spatial exclusion inspired by Schelling-type segregation [25]—into a bidirectional feedback between opinion dynamics and spatial structure. Opinion divergence triggers eviction when differences exceed the tolerance threshold; eviction then reshapes local interaction networks and subsequent opinion updates. Echo chambers are reinforced both by repelling discordant opinions and by physically removing dissenters. Unlike the original ARM, where the interaction network is fixed [1], here the network endogenously reconfigures as agents relocate—a feature shared with adaptive network models [26,28,29] and co-evolutionary games on structured populations [5,6], where coevolving link deletion/addition can generate distinct macroscopic regimes [42,43]. In our lattice setting, eviction acts as a structural-update analogue of coevolving link turnover, realized through physical displacement rather than explicit edge rewiring. By transplanting the "disconnect/reconnect" logic from social networks to lattice neighborhoods, our model yields a mixed regime and a p_m-driven dual effect absent from well-mixed ARM formulations.

Spatial eviction fundamentally alters the ARM's repertoire. In the well-mixed ARM, polarization depends primarily on tolerance and exposure: low tolerance paired with frequent cross-group encounters drives runaway extremism [1]. Our spatial model confirms that low τ yields divergent opinions, but shows that eviction reorganizes interaction patterns, either amplifying or counteracting this tendency. Eviction functions as adaptive rewiring: by removing an agent's most dissimilar neighbor with probability p_m, it limits exposure to ideologically incongruent views and mirrors coevolving-network mechanisms in which structural update frequency steers macro-level outcomes [42,43]. From a mobility perspective, p_m also plays the role of an endogenous relocation rate, analogous to migration-frequency controls in adverse-neighborhood settings [31]. When $p_m = 0$, the model reduces to the classic ARM outcome. As p_m increases, eviction disrupts sustained cross-group contact and yields a consensus steady state. Thus p_m continuously interpolates between qualitatively different attractors.

Spatial eviction can polarize locally yet depolarize globally, depending on frequency. At low p_m, occasional evictions create locally homogeneous enclaves: agents remove discordant neighbors, sharpening boundaries between ideological communities [33,44].

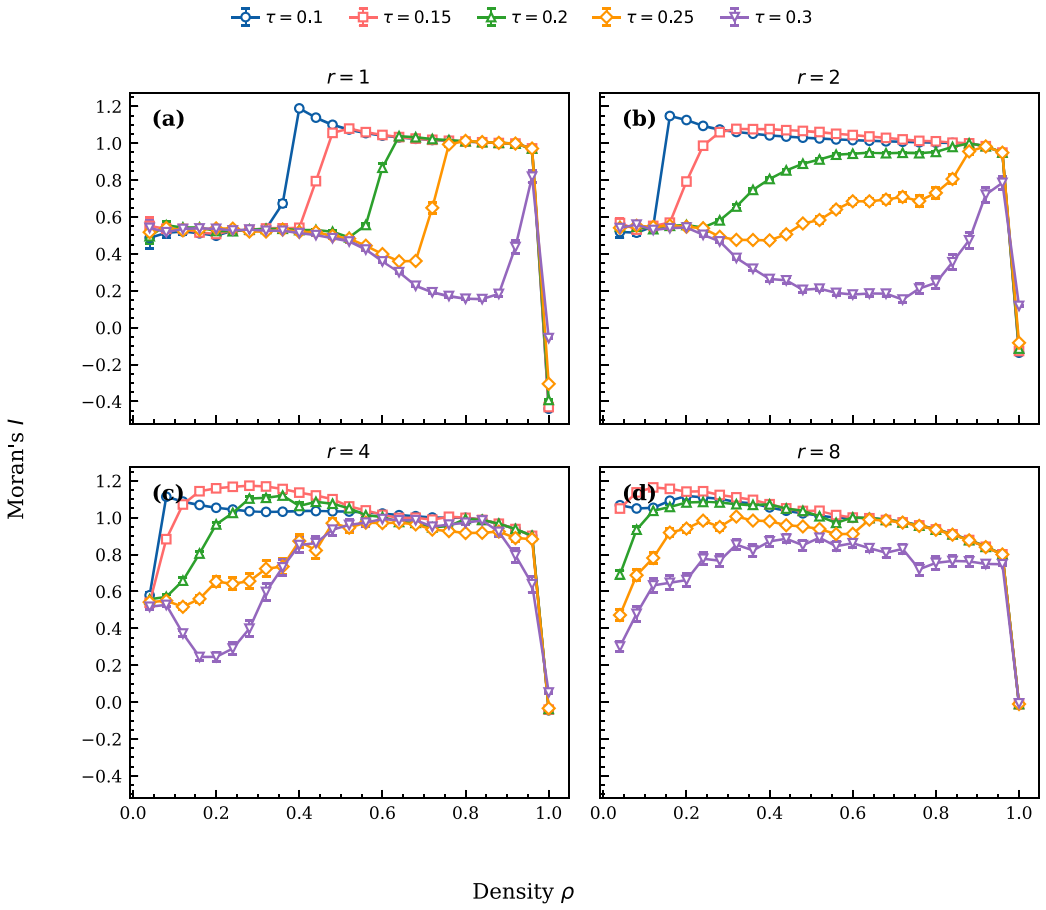

Fig. 7. Steady-state Moran's I versus density ρ for four neighborhood ranges: (a) $r = 1$, (b) $r = 2$, (c) $r = 4$, (d) $r = 8$. Moran's I is computed using Eq. (9) with $\mathcal{N}_i$ defined by the same neighborhood range r; values at different r therefore correspond to different measurement scales. Points at $\rho = 1$ correspond to the eviction-disabled limit. Curves and error bars as in Fig. 6. Fixed: $p_m = 0.8$, $E = 0.1$, $R = 0.2$. Spatial clustering mirrors polarization trends but is scale-dependent: large neighborhoods dilute apparent autocorrelation, so local opinion clusters that form at any r appear less pronounced when measured over wider regions.

Expelling outsiders insulates groups and drives opinions toward extremes through mutual reinforcement, akin to Schelling-style self-segregation. At high p_m, constant reshuffling prevents large extremist regions from consolidating; outliers are relocated before nucleating extreme clusters, and the population converges toward moderate consensus. This depolarizing effect parallels the finding of Baumgaertner et al. [30] that random relocation breaks polarization deadlocks. Our eviction mechanism is selective rather than random, producing richer dynamics, including a transitional mixed regime.

Together, these countervailing mechanisms give rise to three dynamical regimes, each interpretable as a distinct type of attractor in the system's state space. The *extremist phase* ($p_m \approx 0$, low τ) constitutes a stable polarized attractor: the population bifurcates into two antagonistic camps at the ideological poles in opinion space ($f_{\text{ext}} > 0.80$) with vanishing moderate views. Crucially, this ideological polarization is *not* accompanied by spatial sorting: because no eviction occurs ($p_m = 0$), agents retain their initial random positions and Moran's I remains near zero (Section 3.2). Polarization in this regime is therefore purely ideological rather than spatial. Once formed, the opposing camps are self-reinforcing—repulsion drives remaining moderates toward the extremes while the frozen spatial configuration continues to expose agents to dissimilar neighbors, sustaining the repulsive feedback—making the phase resistant to perturbation, as confirmed by the robustness tests in Section 3.4. The *consensus phase* (high p_m, $f_{\text{ext}} < 0.15$) constitutes a consensus attractor: agents converge toward a unimodal consensus as eviction continuously filters out ideologically discordant elements from local neighborhoods. This convergence occurs not through increased attraction but through a filtering process that homogenizes the local environment, endogenously achieving the exposure limitation that Axelrod et al. [1] identified as necessary to prevent extreme polarization. Although Moran's I reaches high values in this phase (~0.9), this partly reflects the low-variance amplification effect discussed in Section 3.2: as opinions converge, even weak residual spatial correlations produce large I because the denominator of Eq. (9) shrinks. The consensus phase's spatial structure is therefore better characterized by direct inspection of opinion snapshots (Fig. 4c) than by I alone. The *mixed regime* occupies the intermediate region: neither polarization nor consensus dominates, and the system exhibits a patchwork of moderate and extremist pockets that coexist in delicate balance. Small perturbations—whether in parameter values or through stochastic noise, as demonstrated by the eviction noise experiments (Section 3.4)—can tip

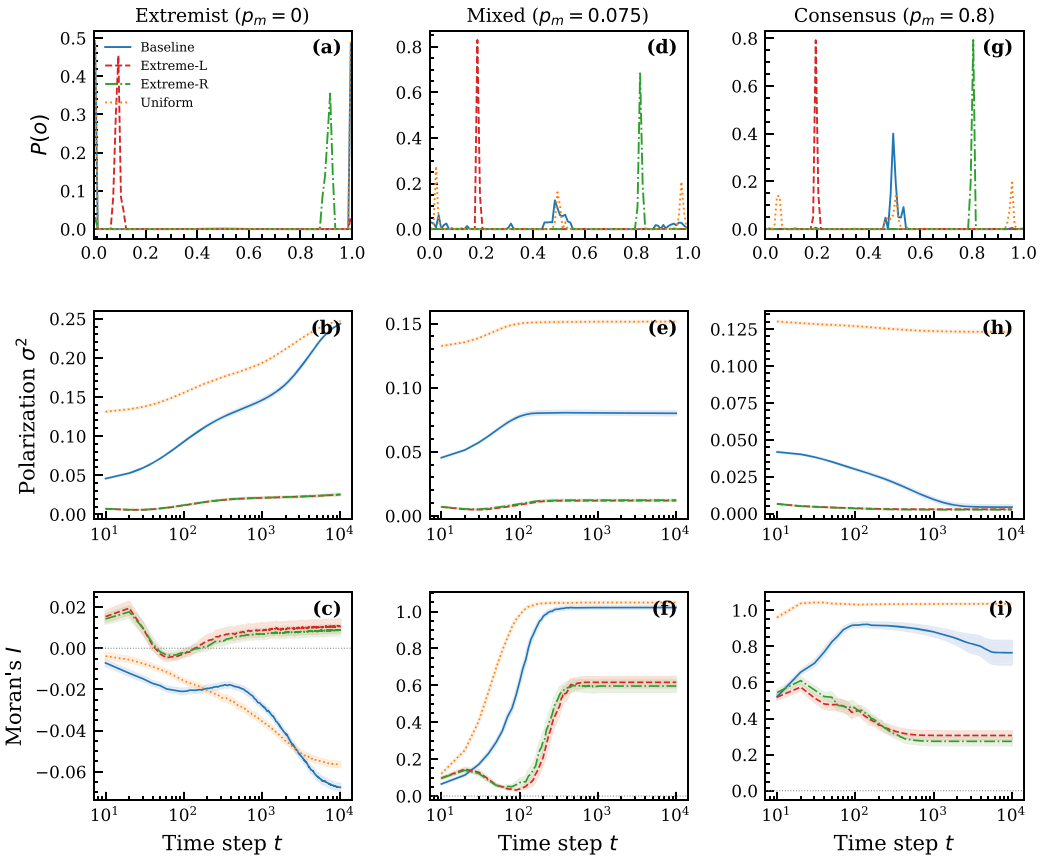

Fig. 8. Robustness to initial conditions. Columns: extremist (a–c; $p_m = 0$), mixed (d–f; $p_m = 0.075$), and consensus (g–i; $p_m = 0.8$). Rows: steady-state opinion distribution $P(o)$ (top), polarization Var(o) (middle; denoted σ_o^2 in plots), and Moran's $I(t)$ (bottom). Four initial conditions are compared: baseline ($\mu = 0.5$, $\sigma = 0.2$), left-skewed ($\mu = 0.2$, $\sigma = 0.1$), right-skewed ($\mu = 0.8$, $\sigma = 0.1$), and broad ($\mu = 0.5$, $\sigma = 0.5$). Shaded bands: ±SEM. Parameters: $\tau = 0.2$, $E = 0.1$, $R = 0.2$, $r = 4$, $\rho = 0.6$. Phase classification is robust to initial conditions: the regime type (extremist, mixed, or consensus) remains stable regardless of initialization, although the consensus location tracks the initial mean.

the system toward either the extremist or consensus attractor. This balance parallels the rewiring-versus-imitation interplay in co-evolutionary network models, where the interplay of topological adaptation and opinion dynamics controls the transition between consensus and fragmented states [26,45]. Durrett et al. [46] further showed that even minor differences in rewiring rules—rewire-to-same versus rewire-to-random—produce qualitatively different (discontinuous versus continuous) phase transitions; our model's random relocation is structurally closer to the rewire-to-random variant, consistent with the smooth crossover we observe through the intermediate mixed regime. When one process dominates, the system settles into a clear attractor (extremist or consensus). In the mixed regime, the competition between polarizing and homogenizing forces leaves the system sensitive to stochastic perturbations—consistent with our robustness results showing that eviction noise most strongly affects the mixed regime (Section 3.4). This multi-stability, with a fragile intermediate state separating two robust attractors, is reminiscent of tipping-point dynamics in complex systems [23] and suggests that pluralistic states in which diverse viewpoints coexist are inherently precarious under repulsive social forces unless actively sustained by sufficient spatial mobility.

Having characterized these regimes, we now situate them within the broader landscape of opinion dynamics and spatial segregation models. Compared with the well-mixed ARM of Axelrod et al. [1], our spatial model produces a richer phase diagram by making exposure an emergent property of agent movements rather than an exogenous parameter. The consensus phase demonstrates that self-organized segregation can endogenously replicate the polarization-preventing effect of reduced exposure, while the mixed regime—where subcommunities with different opinions coexist—has no analog in a fully connected population but emerges naturally on a lattice with partial mobility. Furthermore, the exposure parameter E and responsiveness R modulate the phase boundaries in intuitive directions: higher E widens the extremist region by increasing the frequency of cross-group encounters that trigger repulsion, while higher R sharpens and accelerates the transitions between phases (Section 3.1).

Our findings extend the voter-model mixing results of Baumgaertner et al. [30] to a setting with explicit repulsive interactions. Prior work showed that random relocation hastens consensus by disrupting entrenched opinion clusters; our model confirms that mobility promotes consensus, but only when sufficiently frequent to overcome the polarizing force of negative influence. Moreover, because our eviction rule selectively targets the most dissimilar neighbor rather than relocating agents at random, the transition

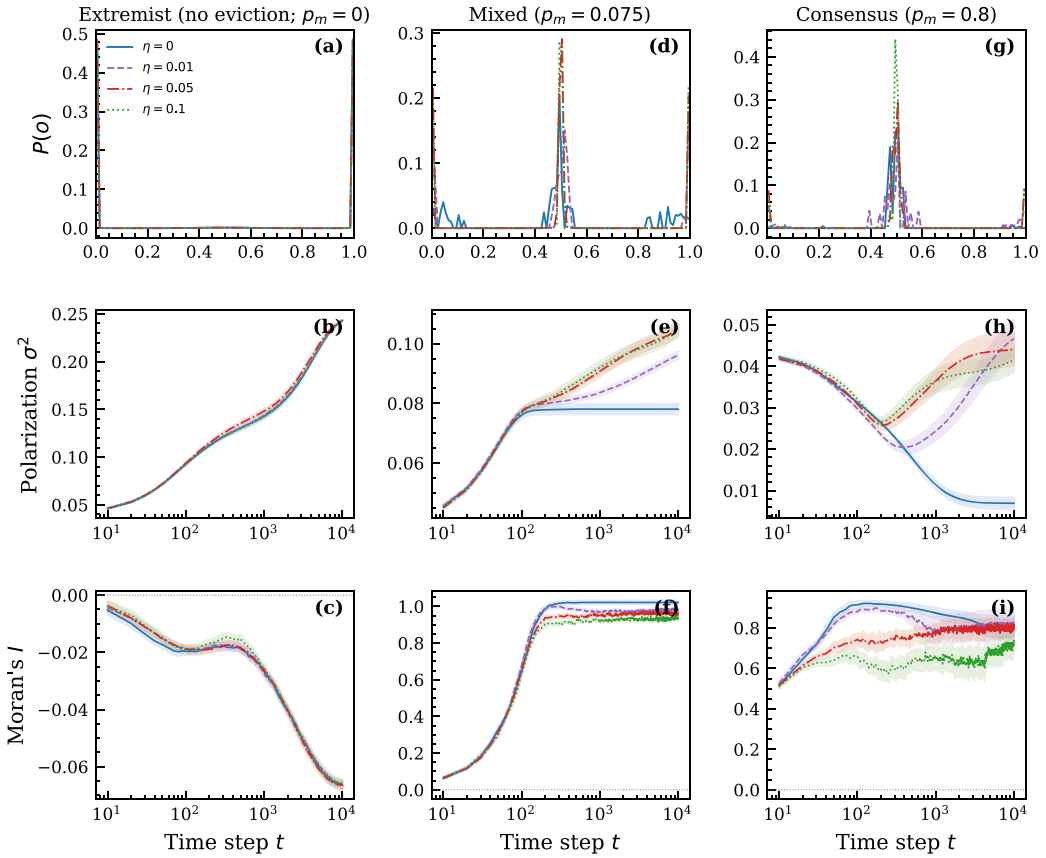

Fig. 9. Robustness to eviction noise η. Layout as in Fig. 8, but comparing four noise levels: $\eta = 0$ (baseline), 0.01, 0.05, and 0.1. Because eviction attempts occur with probability p_m, η has no effect at the extremist reference point ($p_m = 0$). In the mixed regime, increasing η strongly perturbs the steady state, while the consensus phase remains nearly unchanged. Parameters as in Fig. 8. The mixed regime's high sensitivity to noise is consistent with its interpretation as a fragile transitional state between two robust attractors.

between polarized and consensus states passes through the intermediate mixed regime—a feature absent from models with purely random mixing and consistent with targeted partner-focused coevolutionary updates that can redirect invasion dynamics [47].

In the context of negative influence and segregation, Feliciani et al. [3] demonstrated that when repulsive interactions drive opinion updates, pre-segregated groups actually hinder polarization by limiting cross-group contact. Our high-p_m consensus phase echoes this insight: frequent eviction creates a de facto segregated state in which intolerant agents rarely encounter adversaries, curtailing the repulsive interactions that would otherwise drive opinions to extremes. Crucially, whereas Feliciani et al. examined static segregation imposed as an exogenous initial condition, our model produces and sustains segregation dynamically—segregation emerges from opinion differences and simultaneously feeds back to influence those opinions, making it both cause and consequence of the polarization trajectory.

Our model also generalizes Schelling-type segregation [25] by endowing agents with mutable opinions rather than fixed traits. Arcón et al. [33] showed that combining Schelling mobility with positive influence (opinion imitation) creates and sustains echo chambers; in contrast, our coupling of mobility with repulsive dynamics reveals that sufficient spatial mobility can actually *erode* polarization rather than merely freezing group differences, since eviction disperses extremists before stable clusters can form. In a complementary direction, Pasimeni et al. [48] extended the ARM into two-dimensional space and found that geographic constraints produce large consensus clusters that centralize to avoid conflicts—a result consistent with our consensus-phase phenomenology. Additionally, Okada [32] explored positive and negative spatial influences on a static network and found that repulsive interactions can maintain opinion diversity; our model complements this by demonstrating that adding physical relocation to spatial negative influence yields a qualitatively different phase structure, including the mixed regime that static topologies do not exhibit.

These results carry implications beyond the specific model studied here. The mixed regime suggests that sustained ideological diversity may require active maintenance through mechanisms analogous to spatial mobility: open dialogue, cross-cutting ties, or policies that promote exposure to diverse perspectives. Without such "mixing" efforts, systems with repulsive social forces tend toward polarized echo chambers or enforced consensus. The eviction mechanism abstracts real-world phenomena from neighborhood ostracism to social media deplatforming [37–39], suggesting qualitative relevance for exclusion in both physical and digital spaces. In particular, the dual role of eviction—polarizing at low frequency yet depolarizing at high frequency—echoes findings that platform

removal can reduce harmful content while pushing displaced users toward more extreme alternatives [39]. Our model offers a mechanistic account: infrequent exclusion creates insulated enclaves that radicalize, whereas pervasive reshuffling dissolves them. Thus the *rate* of exclusion, not merely its presence, shapes whether outcomes are polarizing or moderating—though we caution against direct policy extrapolation from an idealized lattice model. The psychological mechanism—social exclusion heightening susceptibility to extremist ideologies [36]—aligns with our prediction that infrequent eviction pushes displaced agents into isolated enclaves where radicalization intensifies.

Several limitations suggest productive extensions. First, our model relocates evicted agents to uniformly random empty sites; real mobility is geographically constrained and often involves self-selection. Distance-limited or preference-based moves would test whether the consensus phase reflects genuine integration or whether localized relocation strengthens echo-chamber enclaves. A particularly informative variant would replace random relocation with a "move-to-similar" rule, whereby evicted agents seek neighborhoods populated by like-minded individuals. Under such a rule, displaced agents would self-sort into ideologically compatible enclaves rather than landing at arbitrary locations, likely accelerating echo-chamber formation and increasing spatial clustering (Moran's I). Because the depolarizing effect of eviction at high p_m relies on the disruption of nascent extremist clusters (Section 4), preference-based relocation would weaken this mechanism: the consensus phase would likely shrink, while extremist enclaves would become more spatially entrenched. The mixed regime—already fragile under random relocation—might narrow or vanish altogether. In essence, the current random-relocation assumption represents a *best case* for depolarization; realistic preference-based mobility would shift the balance toward polarization, consistent with Schelling-type self-segregation dynamics [25,33]. Second, opinions are one-dimensional. Pasimeni et al. [48] found that additional opinion dimensions reduce the number of stable clusters, and Bramson et al. [13] argued that multi-dimensional spaces better capture real political attitudes; extending our framework would show whether the three-phase structure persists as cross-dimensional tolerance interactions alter the repulsion landscape. Third, all agents share uniform τ and p_m. Introducing a minority of "firebrand" agents with very low tolerance and high eviction propensity would test whether such subpopulations expand the mixed regime by nucleating persistent extremist cores or instead create a regime where a small radical faction coexists with a moderate majority. Fourth, a mean-field or bifurcation analysis could yield semi-analytic expressions for the critical phase boundaries as functions of $(p_m, \tau, E, R, \rho, r)$, moving beyond numerical isocontours toward scaling relations that expose the essential control parameters. Fifth, our model is implemented on a regular two-dimensional lattice with local Manhattan neighborhoods, yet many real-world social systems—especially online platforms—exhibit scale-free, small-world, or higher-order interaction structures [7]. On scale-free networks, high-degree hub agents would wield disproportionate eviction power, potentially concentrating exclusionary influence in a few nodes and reshaping the phase boundaries. Moreover, recent work has shown that higher-order (group-level) interactions can fundamentally alter collective dynamics in spatial networks [49–51]; extending the eviction mechanism to simplicial complexes or hypergraphs—where group consensus, rather than pairwise disagreement, triggers expulsion—could reveal qualitatively new regimes. This limitation is particularly relevant given that we motivate our eviction mechanism partly through social media deplatforming, which occurs on platforms with decidedly non-lattice topologies. Finally, external perturbations—media shocks, institutional interventions, or periodic opinion injection—could probe whether the extremist and consensus attractors can be destabilized and identify strategies that promote cross-group exposure without triggering repulsive encounters.

In summary, coupling ideological repulsion with spatial eviction produces a three-phase structure—extremist, mixed, and consensus—governed primarily by the eviction probability p_m and tolerance τ. The key new insight relative to the existing literature is the dual, frequency-dependent role of exclusion: the same mechanism that entrenches polarization at low rates dissolves it at high rates, with a fragile pluralistic regime in between. This finding extends the Attraction–Repulsion Model into spatially structured settings and provides a mechanistic framework for understanding how exclusionary social processes shape collective opinion.

5. Conclusion

We introduce an agent-based model that couples Attraction–Repulsion opinion dynamics with spatial eviction on a two-dimensional lattice. Parameter sweeps, temporal analysis, and robustness tests yield three findings. First, the model exhibits three phases controlled primarily by p_m and τ: an *extremist* phase with polarization at the ideological poles in opinion space, a *consensus* phase with a unimodal consensus, and a fragile *mixed* regime with coexisting moderate and extreme subpopulations. Phase classification is robust to initial conditions and persists across exposure and responsiveness values, indicating genuine dynamical attractors. Second, spatial eviction plays a dual role: at low rates it sharpens local homogeneity and amplifies polarization; at high rates it prevents stable extremist clusters, endogenously limiting cross-group exposure. Third, higher density and smaller neighborhoods intensify polarization and spatial clustering; the $\rho = 1$ limit disables eviction, reducing the model to pure opinion dynamics and eliminating the consensus attractor.

Overall, incorporating spatial exclusion enriches collective outcomes by coupling ideological repulsion with physical relocation, producing phase behavior and spatial patterning absent from well-mixed formulations. Future extensions could examine multi-dimensional opinions, heterogeneous tolerance, realistic mobility constraints, and external perturbations to clarify when exclusion mitigates or exacerbates polarization.

Code Availability

The NetLogo source code implementing the model described in this paper is publicly available at COMSES Net: https://www.comses.net/codebases/b8547f51-20a7-49db-b57b-f22274a3bf0b/releases/1.0.0/.

Data availability

No data was used for the research described in the article.

References

[1] R. Axelrod, J.J. Daymude, S. Forrest, Preventing extreme polarization of political attitudes, Proc. Natl. Acad. Sci. 118 (50) (2021) e2102139118. https://doi.org/10.1073/pnas.2102139118
[2] A. Flache, M. Mäs, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet, J. Lorenz, Models of social influence: towards the next frontiers, J. Artif. Soci. Soc. Simul. 20 (4) (2017) 2. https://doi.org/10.18564/jasss.3521
[3] T. Feliciani, A. Flache, J. Tolsma, How, when and where can spatial segregation induce opinion polarization? two competing models, J. Artif. Soc. Soc. Simul. 20 (2) (2017) 6. https://doi.org/10.18564/jasss.3419
[4] C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. 81 (2) (2009) 591–646. https://doi.org/10.1103/RevModPhys.81.591
[5] M. Perc, A. Szolnoki, Coevolutionary games – A mini review, BioSystems 99 (2) (2010) 109–125. https://doi.org/10.1016/j.biosystems.2009.10.003
[6] M. Perc, J. Gómez-Gardeñes, A. Szolnoki, L.M. Floría, Y. Moreno, Evolutionary dynamics of group interactions on structured populations: a review, J. R. Soc. Interface 10 (80) (2013) 20120997. https://doi.org/10.1098/rsif.2012.0997
[7] M. Jusup, P. Holme, K. Kanazawa, M. Takayasu, I. Romić, Z. Wang, S. Geček, T. Lipić, B. Podobnik, L. Wang, W. Luo, T. Klanjšček, J. Fan, S. Boccaletti, M. Perc, Social physics, Phys. Rep. 948 (2022) 1–148. https://doi.org/10.1016/j.physrep.2021.10.005
[8] Z. Li, X. Chen, H.-X. Yang, A. Szolnoki, Game-theoretical approach for opinion dynamics on social networks, Chaos 32 (7) (2022) 073117. https://doi.org/10.1063/5.0084178
[9] M. Chica, M. Perc, F.C. Santos, Success-driven opinion formation determines social tensions, iScience 27 (3) (2024) 109254. https://doi.org/10.1016/j.isci.2024.109254
[10] M. Sherif, C.I. Hovland, Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change, Yale University Press, New Haven, CT, 1961.
[11] W. Jager, F. Amblard, Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change, Comput. Math. Org. Theory 10 (4) (2005) 295–303. https://doi.org/10.1007/s10588-005-6282-2
[12] S. Huet, G. Deffuant, W. Jager, A rejection mechanism in 2D bounded confidence provides more conformity, Adv. Complex Syst. 11 (4) (2008) 529–549. https://doi.org/10.1142/S0219525908001799
[13] A. Bramson, P. Grim, D.J. Singer, S. Fisher, W. Berger, G. Sack, C. Flocken, Disambiguation of social polarization concepts and measures, J. Math. Sociol. 40 (2) (2016) 80–111. https://doi.org/10.1080/0022250X.2016.1147443
[14] D. Baldassarri, P. Bearman, Dynamics of political polarization, Am. Sociol. Rev. 72 (5) (2007) 784–811. https://doi.org/10.1177/000312240707200507
[15] G. Deffuant, D. Neau, F. Amblard, G. Weisbuch, Mixing beliefs among interacting agents, Adv. Complex Syst. 3 (1–4) (2000) 87–98. https://doi.org/10.1142/S0219525900000078
[16] R. Hegselmann, U. Krause, Opinion dynamics and bounded confidence: models, analysis and simulation, J. Artif. Soc. Soc. Simul. 5 (3) (2002) 2. , https://www.jasss.org/5/3/2.html.
[17] M.W. Macy, J.A. Kitts, A. Flache, S. Benard, Polarization in dynamic networks: a hopfield model of emergent structure, in: R. Breiger, K. Carley, P. Pattison (Eds.), Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, National Academies Press, Washington, DC, 2003, pp. 162–173. https://doi.org/10.17226/10735
[18] G. Deffuant, F. Amblard, G. Weisbuch, T. Faure, How can extremism prevail? a study based on the relative agreement interaction model, J. Artif. Societ. Soc. Simul. 5 (4) (2002) 1. , https://www.jasss.org/5/4/1.html.
[19] S. de Marchi, The complexity of polarization, Proc. Natl. Acad. Sci. 119 (17) (2022) e2115019119. https://doi.org/10.1073/pnas.2115019119
[20] R. Axelrod, S. Forrest, J.J. Daymude, Reply to de marchi: modeling polarization of political attitudes, Proc. Natl. Acad. Sci. 119 (17) (2022) e2202863119. https://doi.org/10.1073/pnas.2202863119
[21] P. Dandekar, A. Goel, D.T. Lee, Biased assimilation, homophily, and the dynamics of polarization, Proc. Natl. Acad. Sci. 110 (15) (2013) 5791–5796. https://doi.org/10.1073/pnas.1217220110
[22] A. Flache, M.W. Macy, Small worlds and cultural polarization, J. Math. Sociol. 35 (1–3) (2011) 146–176. https://doi.org/10.1080/0022250X.2010.532261
[23] M.W. Macy, M. Ma, D.R. Tabin, J. Gao, B.K. Szymanski, Polarization and tipping points, Proc. Natl. Acad. Sci. 118 (50) (2021) e2102144118. https://doi.org/10.1073/pnas.2102144118
[24] S. Banisch, E. Olbrich, Opinion polarization by learning from social feedback, J. Math. Sociol. 43 (2) (2019) 76–103. https://doi.org/10.1080/0022250X.2018.1517761
[25] T.C. Schelling, Dynamic models of segregation, J. Math. Sociol. 1 (2) (1971) 143–186. https://doi.org/10.1080/0022250X.1971.9989794
[26] T. Gross, B. Blasius, Adaptive coevolutionary networks: a review, J. R. Soc. Interfac. 5 (20) (2008) 259–271. https://doi.org/10.1098/rsif.2007.1229
[27] O.J. Chu, J.F. Donges, G.B. Robertson, G. Pop-Eleches, The microdynamics of spatial polarization: a model and an application to survey data from ukraine, Proc. Natl. Acad. Sci. 118 (50) (2021) e2104194118. https://doi.org/10.1073/pnas.2104194118
[28] D. Geschke, J. Lorenz, P. Holtz, The triple-filter bubble: using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers, Brit. J. Soc. Psychol. 58 (1) (2019) 129–149. https://doi.org/10.1111/bjso.12286
[29] D. Centola, J.C. González-Avella, V.M. Eguíluz, M. San Miguel, Homophily, cultural drift, and the co-Evolution of cultural groups, J. Conflict Resolut. 51 (6) (2007) 905–929. https://doi.org/10.1177/0022002707307632
[30] B.O. Baumgaertner, P.A. Fetros, S.M. Krone, R.C. Tyson, Spatial opinion dynamics and the effects of two types of mixing, Phys. Rev. E 98 (2) (2018) 022310. https://doi.org/10.1103/PhysRevE.98.022310
[31] H.-W. Lee, C. Cleveland, A. Szolnoki, When costly migration helps to improve cooperation, Chaos 32 (9) (2022) 093103. https://doi.org/10.1063/5.0100772
[32] I. Okada, N. Okano, A. Ishii, Spatial opinion dynamics incorporating both positive and negative influence, Front. Phys. 10 (2022) 953184. https://doi.org/10.3389/fphy.2022.953184
[33] V. Arcón, J.P. Pinasco, I. Caridi, A schelling-Opinion model based on integration of opinion formation with residential segregation, in: J. Borondo, A.J. Morales, J.C. Losada, R.M. Benito (Eds.), Causes and Symptoms of Socio-Cultural Polarization: Role of Information, Communication and New Technologies, Springer, Singapore, 2022, pp. 27–50. https://doi.org/10.1007/978-981-16-5268-4_2
[34] C. Gracia-Lázaro, L.F. Lafuerza, L.M. Floría, Y. Moreno, Residential segregation and cultural dissemination: an axelrod–schelling model, Phys. Rev. E 80 (4) (2009) 046123. https://doi.org/10.1103/PhysRevE.80.046123
[35] C. Gracia-Lázaro, L.M. Floría, Y. Moreno, Selective advantage of tolerant cultural traits in the axelrod–schelling model, Phys. Rev. E 83 (5) (2011) 056103. https://doi.org/10.1103/PhysRevE.83.056103
[36] M. Pfundmair, N.R. Wood, A. Hales, E.D. Wesselmann, How social exclusion makes radicalism flourish: a review of empirical evidence, J. Soc. Iss. 80 (1) (2024) 341–359. https://doi.org/10.1111/josi.12520
[37] A. Mekacher, M. Falkenberg, A. Baronchelli, The systemic impact of deplatforming on social media, PNAS Nexus 2 (11) (2023) pgad346. https://doi.org/10.1093/pnasnexus/pgad346
[38] C. Buntain, M. Snegovaya, Post-january 6 deplatforming shows long-term effects on ideological polarization among twitter users, PNAS Nexus 4 (11) (2025) pgaf333. https://doi.org/10.1093/pnasnexus/pgaf333

[39] M.H. Ribeiro, H. Hosseinmardi, R. West, D.J. Watts, Deplatforming did not decrease parler users' activity on fringe social media, PNAS Nexus 2 (3) (2023) pgad035. https://doi.org/10.1093/pnasnexus/pgad035
[40] C. Zhang, J. Zhang, F.J. Weissing, M. Perc, G. Xie, L. Wang, Different reactions to adverse neighborhoods in games of cooperation, PLoS ONE 7 (4) (2012) e35183. https://doi.org/10.1371/journal.pone.0035183
[41] P.A.P. Moran, Notes on continuous stochastic phenomena, Biometrika 37 (1–2) (1950) 17–23. https://doi.org/10.1093/biomet/37.1-2.17
[42] A. Szolnoki, M. Perc, Emergence of multilevel selection in the prisoner's dilemma game on coevolving random networks, New J. Phys. 11 (2009) 093033. https://doi.org/10.1088/1367-2630/11/9/093033
[43] A. Szolnoki, M. Perc, Resolving social dilemmas on evolving random networks, Europhys. Lett. (EPL) 86 (3) (2009) 30007. https://doi.org/10.1209/0295-5075/86/30007
[44] A. Szolnoki, M. Perc, Conformity enhances network reciprocity in evolutionary social dilemmas, J. R. Soc. Interfac. 12 (103) (2015) 20141299. https://doi.org/10.1098/rsif.2014.1299
[45] P. Holme, M.E.J. Newman, Nonequilibrium phase transition in the coevolution of networks and opinions, Phys. Rev. E 74 (5) (2006) 056108. https://doi.org/10.1103/PhysRevE.74.056108
[46] R. Durrett, J.P. Gleeson, A.L. Lloyd, P.J. Mucha, F. Shi, D. Sivakoff, J.E.S. Socolar, C. Varghese, Graph fission in an evolving voter model, Proc. Natl. Acad. Sci. 109 (10) (2012) 3682–3687. https://doi.org/10.1073/pnas.1200709109
[47] A. Szolnoki, X. Chen, Blocking defector invasion by focusing on the most successful partner, Appl. Math. Comput. 385 (2020) 125430. https://doi.org/10.1016/j.amc.2020.125430
[48] F. Pasimeni, R. Wade, F. Alkemade, Opinion dynamic and social clustering in a 2D space: an agent based experiment, Comput. Econ. (2025). https://doi.org/10.1007/s10614-025-10961-w
[49] F. Battiston, V. Capraro, F. Karimi, S. Lehmann, A.B. Migliano, O. Sadekar, A. Sánchez, M. Perc, Higher-order interactions shape collective human behaviour, Nat. Hum. Behav. 9 (12) (2025) 2441–2457. https://doi.org/10.1038/s41562-025-02373-5
[50] A. Liu, D. Xiao, W. Li, T. Yang, C. Yang, J. Fan, W. Wang, Localized seeding for triggering the global social contagion in higher-order spatial networks, Chaos Soliton Fractal. 200 (2025) 117141. https://doi.org/10.1016/j.chaos.2025.117141
[51] Y. Gao, J. Li, F. Gao, W. Wang, Coevolution of multipathogens on higher-order networks, Chaos Soliton. Fractal. 202 (2026) 117588. https://doi.org/10.1016/j.chaos.2025.117588