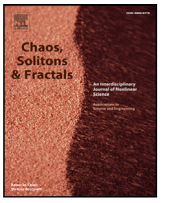
Chaos, Solitons and Fractals 208 (2026) 118128

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

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



When moving away reduces polarization: Selective depolarization by endogenous migration in attraction–repulsion opinion dynamics

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 polarization
Spatial segregation
Agent-based model
Attraction–repulsion dynamics
Echo chambers
Coevolutionary dynamics

ABSTRACT

Opinion polarization poses a fundamental challenge for contemporary societies, yet how spatial mobility coevolves with attraction–repulsion influence remains poorly understood. We develop an agent-based model in continuous two-dimensional space that couples tolerance-based attraction–repulsion opinion updates with endogenous migration: agents interact locally and then move toward ideologically similar neighbors while avoiding dissimilar ones. Extensive parameter sweeps show that mobility acts as a *selective depolarizer*. Above a critical tolerance threshold, mobility suppresses repulsion-driven extremization by enabling agents to leave antagonistic neighborhoods before repeated hostile encounters accumulate. At the same time, mobility strengthens spatial assortativity, yielding a robust *depolarized segregation* regime in which ideological polarization remains low while spatial clustering is high—demonstrating that echo-chamber-like structure need not coincide with ideological extremism. Across conditions, tolerance sets the dominant phase boundary, whereas exposure, interaction radius, and movement speed primarily modulate transition locations and time scales. As a mechanistic model, these results highlight a trade-off between reduced extremization and reduced cross-cutting contact, and suggest that interventions focusing only on mobility or exposure may have limited impact when tolerance is low; linking model parameters to concrete policy levers requires empirical calibration.

1. Introduction

Opinion polarization – the consolidation of populations into mutually hostile ideological camps – has become a defining challenge for contemporary democracies [1,2]. The consequences are far-reaching: severe polarization has been linked to legislative gridlock, erosion of democratic norms, and broader social fragmentation. Central to this phenomenon is the emergence of "echo chambers"—epistemic environments where selective exposure and homophilic ties insulate individuals from alternative perspectives [3,4]. Crucially, these chambers are not merely virtual constructs. Ideological clustering often has a *spatial* dimension: people physically congregate with those who share their beliefs, creating segregated communities of opinion analogous to Schelling's classical segregation dynamics [5]. Indeed, individuals increasingly "vote with their feet", relocating to communities of ideological peers [6]—a process of geographic self-sorting that Tiebout (1956) first theorized in the context of economic localism [7]. This spatial dimension of polarization implies that understanding the phenomenon requires models capable of capturing the *bidirectional coupling* between opinion dynamics and spatial structure: opinions shape where people go, and where people are shapes the opinions they encounter.

E-mail address: poyeker@gmail.com.

https://doi.org/10.1016/j.chaos.2026.118128
Received 25 January 2026; Received in revised form 18 February 2026; Accepted 20 February 2026

A complex-systems perspective provides the natural framework for this inquiry. Collective outcomes – consensus, fragmentation, extremization, spatial segregation – characteristically emerge from simple local rules and feedback loops rather than centralized coordination [8]. Evolutionary game theory has demonstrated that microscopic interaction rules can generate clustering, domain coarsening, and phase transitions across parameter space [9]—phenomena directly relevant to opinion fragmentation. This style of work encourages mapping out regime diagrams rather than relying on single-parameter anecdotes, and treating spatial patterns as first-class observables rather than mere visualizations.

Agent-based models of opinion formation constitute the primary theoretical toolkit for studying polarization [10–12]. These models simulate how repeated micro-level interactions might lead to macro-level outcomes such as consensus, persistent pluralism, or polarized camps. Classical frameworks – including bounded-confidence models [13] and cultural dissemination models [14] – provided foundational insights, yet they share a critical limitation: they struggle to generate the *active animosity* characteristic of modern polarization. Pure assimilation models predict global consensus, contradicting the empirical persistence of stable polarization in hyper-connected societies [2]. Bounded-confidence models address this by introducing a threshold of influence, producing fragmentation into distinct clusters—but these clusters merely *ignore* one another. Neither mechanism captures the reported (and contested) "backfire effect", whereby exposure to opposing views can push individuals toward *more* extreme positions rather than promoting convergence [15]. Recent syntheses further emphasize that modern collective behavior often depends on network architecture and higher-order interactions, while success-driven opinion formation can reshape social tensions in ways not captured by purely pairwise or mean-field assumptions [16–18].

The Attraction–Repulsion Model (ARM) [19] addresses this gap by introducing *negative ties*. In the ARM, if two agents' opinions are relatively close (within a tolerance threshold T), interaction draws them together (attraction); if their opinions exceed this threshold, interaction drives them apart (repulsion), pushing each toward more extreme positions. This bifurcation is critical: unlike bounded-confidence models where distant agents simply sever ties, ARM agents maintain a repulsive link that *actively* drives them apart—capturing the psychological phenomenon where out-group exposure reinforces in-group identity [20]. The ARM reveals what Axelrod et al. termed the "Polarization Trap": under moderate tolerance levels, a population may appear to converge toward a centrist consensus, but if the distribution tightens sufficiently – or an external shock widens it slightly beyond the tolerance threshold – the dominant force flips from attraction to repulsion, causing the system to suddenly bifurcate toward extremes. This nonlinear instability suggests that moderate polarization is inherently unstable; societies tend toward either full consensus or extreme division, with little stable ground in between. This carries a conditional implication: when interactions operate in a repulsive regime, increased exposure *can* accelerate differentiation rather than heal divides. This offers a possible mechanism consistent with findings that some online contexts exacerbate division, without claiming universality [1,15]. Subsequent work has extended repulsion-based dynamics to hypergraph structures [21], derived mean-field descriptions [22], and incorporated affective dimensions where susceptibility itself depends on opinion state [23].

However, the ARM, like most opinion models, treats interaction structure as exogenous. A robust finding from coevolutionary game theory is that strategic outcomes depend jointly on (i) how individual states change and (ii) how interaction opportunities are generated [24]. Agents reshape their social environment in response to disagreement – rewiring networks, shifting attention, or physically relocating – and these choices feed back into future opinion trajectories. In evolutionary game theory, this idea is crystallized by *coevolutionary games*, where strategy evolution couples to changes in network structure, reputation, mobility, or age. This coevolutionary viewpoint is essential for polarization models because "exposure" is rarely exogenous: individuals avoid, seek, or reweight contacts based on perceived similarity, and those choices alter the future trajectory of opinions. Mobility, in particular, is not a mere detail but a primary "control knob" of emergent collective behavior. Reviews of cyclic dominance [25] and off-lattice social dilemmas [26] demonstrate that continuous-space formulations with endogenous movement generate qualitatively distinct regimes—soft borders, metastable coexistence, cluster morphologies unreproducible on fixed grids. For polarization, this implies that understanding whether societies converge, fragment, or extremize requires coupling opinion updates to spatial dynamics.

Recent work has begun to address this coevolution of opinions and interaction geography. Starnini et al. [27] studied mobile agents with random movement and simple opinion updates, finding that even random walks combined with homophily produce emergent "echo chamber" clusters—physically separated subpopulations of like-minded individuals. Crucially, their model revealed that echo chambers can exist *within* physical groups: due to bounded confidence, different opinions can coexist in the same spatial cluster without mixing, segregated by an invisible threshold. Alraddadi et al. [28] demonstrated that *directed* mobility – actively moving toward similar and away from dissimilar neighbors – is essential for generating strong segregation. Random mobility acts as a "mean field" approximation that mixes the population; directed mobility produces what they termed the "Paradise State"—a perfectly segregated society where every agent is surrounded exclusively by agreeable neighbors and opinion dynamics effectively freeze. Pasimeni et al. [29] explicitly extended the ARM to continuous 2D space, identifying *visibility* (neighborhood range) as a critical parameter analogous to the "halving distance" in the original ARM. High visibility yields few large polarized clusters that centralize in space; low visibility fragments populations into isolated local pockets. Their spatial dynamics also reveal a distinct morphology: dominant clusters tend to centralize, pushing minority clusters to the periphery—a "centralized segregation" that creates physical barriers to diverse views.

Insights from evolutionary game theory illuminate the mechanisms driving such spatial sorting. Xiao et al. [30] demonstrated that in spatial public goods games, "leaving bads" (escaping defectors) outperforms "approaching goods" (seeking cooperators) by blocking invasion pathways. This distinction directly applies to ideological sorting: mobility driven by avoidance of dissimilarity can increase local homogeneity while reducing cross-cutting contacts, whereas mobility driven by seeking similarity accelerates sorting without necessarily stabilizing clusters. Extensions have introduced more sophisticated migration rules. Xiao et al. [31] modeled "emotional" agents who accumulate positive ("happiness") or negative ("anger") utility from neighbors; when cumulative anger

exceeds a threshold, the agent migrates. This emotional migration proves to be a double-edged sword: at low velocities, it facilitates cluster formation; at high velocities, constant movement prevents stable reciprocal bonds. Zhang et al. [32] introduced fairness-driven mobility, where agents move toward environments perceived as "fair". In polarization contexts, "fairness" often correlates with ideological validation, driving migration patterns that reinforce homogeneity under the guise of seeking equity.

Several methodological precedents from spatial evolutionary games inform our modeling choices. First, *heterogeneous responsiveness* – where agents differ in willingness to update – can induce nonlinear population-level effects and reshape phase boundaries [33]. In polarization terms, a minority of highly mobile agents may disproportionately influence segregation patterns. Second, *conformity-biased protocols* enhance reciprocity relative to random neighbor selection [34], suggesting that exposure rules should be treated as explicit mechanisms rather than implicit byproducts of geometry. Third, *asymmetric interactions* can either promote or undermine cooperation depending on baseline self-organization efficacy [35]—a "dual effect" cautioning against assuming heterogeneity will uniformly buffer or intensify polarization. Fourth, *feedback between individual decisions and environment* produces multistability and hysteresis [36]. Fifth, *aggregated neighborhood signals* (rather than single-encounter updates) induce self-organization effects that alter phase behavior [37]. Collectively, these results establish that polarization dynamics in mobile populations likely exhibit regime transitions, history dependence, and sensitivity to microscopic rules that static-network models cannot capture.

Building on this foundation, the present work proposes a coupled model of opinion and space integrating ARM-style attraction–repulsion dynamics with endogenous migration. Agents occupy a continuous 2D plane, interact within a visibility radius, and update opinions according to the tolerance-based rule: attraction when $|\Delta O| \leq T$, repulsion otherwise. Concurrently, each agent experiences a social force – attraction toward ideological allies, repulsion from antagonists – that determines spatial displacement. We implement this by calculating a weighted vector sum of influences from neighbors, where weights scale with opinion alignment. In effect, agents drift toward like-minded clusters while distancing themselves from ideological opponents, endogenously forming spatial opinion domains. A key mechanism in our model is the *orientation bias* β_i governing each agent's movement strategy, which interpolates between *approaching goods* ($\beta_i = 0$) and *leaving bads* ($\beta_i = 1$). Higher-β agents emphasize avoidance of dissimilar neighbors; lower-β agents emphasize attraction to similar ones. This parameterization allows systematic exploration of how the balance between approach and escape behaviors shapes collective outcomes—a question directly motivated by findings in evolutionary game theory that "leaving bads" can outperform "approaching goods" in sustaining cooperation [30].

Our central question is whether spatial mobility amplifies or attenuates polarization. On one hand, mobility could intensify division by enabling rapid sorting into homogeneous enclaves, potentially eliminating the cross-cutting contacts that might otherwise moderate opinions. On the other hand, by allowing agents to escape hostile encounters before cumulative repulsion drives them to extremes, mobility might suppress polarization—echoing the ARM finding that limiting exposure checks polarization. A related question is whether spatial segregation and ideological polarization necessarily go hand in hand, or whether they can be decoupled: it remains an open question whether mobile populations might achieve spatial homogeneity (echo chambers) while retaining ideological moderation, or whether clustering and extremization are inseparable. We map the resulting phase space using polarization, Moran's I, good-component count, and mixed-neighborhood fraction. Together these capture ideological spread, spatial assortativity, clustering, and echo-chamber formation. We pay particular attention to tolerance T, which governs whether interactions are attractive or repulsive, while also examining exposure, interaction radius, and orientation bias. We also attend to the multi-objective nature of intervention design: as Han et al. [38] demonstrated for cooperation, maximizing one collective indicator (e.g., reducing polarization) need not optimize others (e.g., welfare, openness). Our analysis therefore reports multiple outcome measures rather than a single polarization metric.

In what follows, we present the model architecture, derive key observables, and systematically explore the parameter space. The results illuminate how endogenous mobility interacts with ideological repulsion to shape the polarization landscape—offering insights into when "avoiding the other" mitigates extremism versus when it merely solidifies division.

2. Model

We develop an agent-based model that couples ARM-style opinion dynamics with endogenous spatial migration in continuous two-dimensional space. The architecture synthesizes three theoretical traditions: (i) the attraction–repulsion paradigm from computational social science [19], (ii) orientation-driven mobility mechanisms from evolutionary game theory [30], and (iii) coevolutionary feedback loops that allow spatial structure and opinion states to co-determine each other [24]. The resulting framework permits systematic exploration of how spatial self-sorting interacts with ideological polarization. Fig. 1 summarizes the model workflow.

2.1. Agents, state variables, and spatial embedding

Consider a population of N mobile agents embedded in a continuous two-dimensional square domain $\Omega = [0, L] \times [0, L]$ with periodic boundary conditions, where L is the linear system size. The mean agent density is $\rho = N/L^2$; together with the interaction radius R_{int} (defined below), this density determines the expected number of neighbors per agent, $\langle |\mathcal{N}_i| \rangle \approx \pi R_{\text{int}}^2 \rho$, which governs the intensity of local interactions. Each agent i is characterized by two coupled state variables:

1. A **spatial position** $\mathbf{r}_i(t) = (x_i(t), y_i(t)) \in \Omega$, governing physical location and local interaction opportunities;
2. An **ideological position** $s_i(t) \in [0, 1]$.

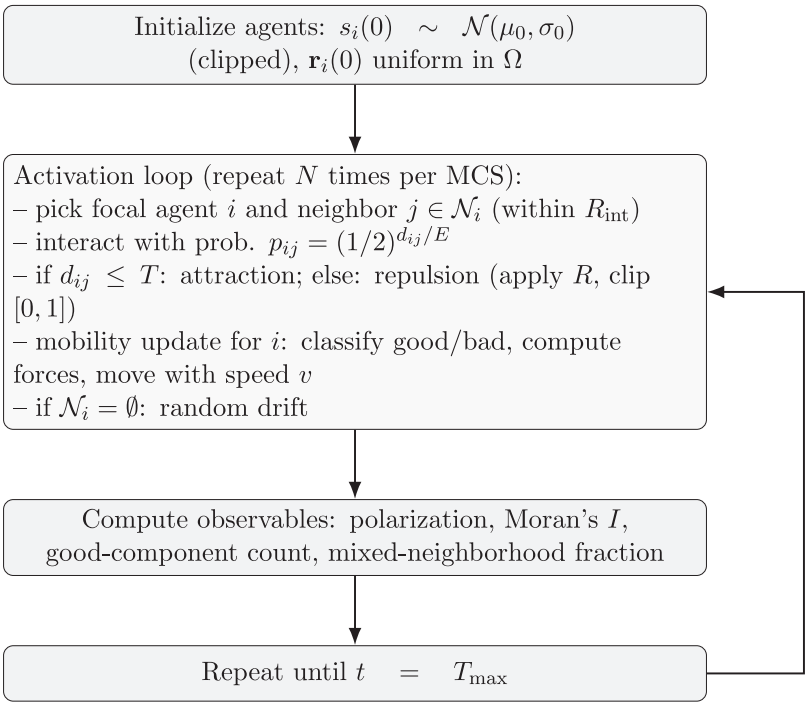


Fig. 1. Workflow of the coupled opinion–mobility model. Each activation updates opinion (ARM rule) and then applies mobility for the same agent; N activations define one MCS, after which observables are measured. The interaction probability $p_{ij} = (1/2)^{d_{ij}/E}$ follows Eq. (6), where $d_{ij} = |s_i - s_j|$ is the ideological distance and E is the exposure parameter.

We assume a one-dimensional opinion space ($D = 1$) throughout. This one-dimensional representation is a parsimonious baseline for mechanism identification and is not intended to capture multidimensional issue coupling.

The bounded ideological domain $[0, 1]$ reflects the empirical observation that opinions on most issues occupy finite, measurable scales—a foundational assumption of the ARM framework [19]. At initialization, each ideological position $s_i(0)$ is drawn from a truncated Gaussian distribution with mean μ_0 and standard deviation σ_0, clipped to $[0, 1]$. This initialization produces a population with moderate initial consensus, allowing us to observe subsequent convergence, fragmentation, or polarization trajectories. Spatial positions are initialized uniformly at random over Ω, consistent with standard protocols in orientation-driven migration studies [26,30].

We define two fundamental distance metrics that govern all subsequent dynamics:

- **Ideological distance** between agents i and j:

$$d_{ij}(t) = |s_i(t) - s_j(t)|. \tag{1}$$

  so $d_{ij} \in [0, 1]$.
- **Spatial distance** between agents i and j:

$$r_{ij}(t) = \|\Delta\mathbf{r}_{ij}(t)\|_2. \tag{2}$$

Under periodic boundary conditions, we compute the minimum-image displacement $\Delta\mathbf{r}_{ij}(t)$ component-wise:

$$\Delta x_{ij}(t) = x_i(t) - x_j(t) - L \cdot \text{round}\left(\frac{x_i(t) - x_j(t)}{L}\right), \tag{3}$$

$$\Delta y_{ij}(t) = y_i(t) - y_j(t) - L \cdot \text{round}\left(\frac{y_i(t) - y_j(t)}{L}\right), \tag{4}$$

so that $r_{ij}(t) = \sqrt{\Delta x_{ij}(t)^2 + \Delta y_{ij}(t)^2}$ and $\hat{\mathbf{r}}_{ij}(t)$ is formed from the same displacement.

Spatial interactions are inherently *local*: for each agent i, we define the spatial neighborhood as

$$\mathcal{N}_i(t) = \{j \neq i : r_{ij}(t) \le R_{\text{int}}\}, \tag{5}$$

where $R_{\text{int}} > 0$ is the *interaction radius* (or "visibility range"). This parameter captures the empirical reality that face-to-face influence and social observation operate within finite geographic ranges [27,29]. Imposing locality – rather than adopting mean-field

formulations where any agent can influence any other – allows spatial structure to emerge endogenously from initial conditions and mobility rules [24]. Crucially, the same neighborhood $\mathcal{N}_i(t)$ serves dual purposes: (i) delimiting the pool of candidates for ideological interaction, and (ii) providing the spatial signals that guide movement decisions.

2.2. Local attraction–repulsion opinion dynamics

Our ideological updating rule adapts the Attraction–Repulsion Model (ARM) [19] to spatially embedded populations. The ARM's central contribution is the introduction of *negative social influence*: rather than merely ignoring dissimilar others (as in bounded-confidence models), agents actively *repel* from those exceeding a tolerance threshold—a mechanism grounded in psychological research on group identity reinforcement and the "backfire effect" [15,20].

Partner selection. At each activation, a focal agent i is selected uniformly at random. Agent i then selects one candidate neighbor j uniformly from $\mathcal{N}_i(t)$. If $\mathcal{N}_i(t) = \emptyset$ (the agent is spatially isolated), no ideological change occurs during this activation.

Local interaction constraint (key departure from original ARM). Unlike the well-mixed ARM [19] where interaction partners are chosen uniformly from the *entire* population, we restrict partner selection to the spatial neighborhood $\mathcal{N}_i(t)$ defined by the interaction radius R_{int}. This departure has fundamental consequences: interaction opportunities now *coevolve* with mobility. Even when the exposure parameter E would permit ideological engagement between distant agents, spatial segregation can physically suppress such encounters by placing dissimilar agents beyond each other's visibility range. Conversely, when agents cluster with like-minded neighbors, their effective interaction pool shrinks to an ideological echo chamber—not by choice, but by geographic constraint. This local selection mechanism introduces spatial structure as a first-order determinant of interaction patterns, enabling the model to capture how physical geography mediates – and is mediated by – opinion dynamics. Accordingly, the present model isolates geographically mediated contact and should not be interpreted as a full model of long-range online information diffusion.

Interaction probability (exposure rule). Conditional on selecting j, an ideological interaction occurs with probability

$$p_{ij}(t) = \left(\frac{1}{2}\right)^{d_{ij}(t)/E}, \tag{6}$$

where $E > 0$ is the *exposure* parameter ("halving distance"). The exponential decay encodes the intuition that ideologically distant agents are less likely to engage meaningfully; compared to sharp cutoffs (as in bounded-confidence models), this smooth functional form avoids discontinuities and interpolates between high- and low-engagement regimes, capturing selective attention and homophilic filtering even within physical proximity [1,3].

Attraction–repulsion update (tolerance rule). When interaction occurs, the update direction is governed by a *tolerance threshold* $T \geq 0$:

- If $d_{ij}(t) \leq T$, the interaction is *attractive*: the focal agent i adjusts its position toward j in ideology space, reflecting assimilation and social influence;
- If $d_{ij}(t) > T$, the interaction is *repulsive*: the focal agent i adjusts its position away from j, capturing identity-protective cognition and reactive devaluation of out-group positions.

In the present study, T is modeled as a global fixed parameter to establish a baseline phase diagram before introducing heterogeneous or time-varying tolerance.

Let $R \in [0, 1]$ denote the *responsiveness* parameter—the fractional distance moved per interaction. The update rule is

$$s_i(t^+) = s_i(t) + \alpha_{ij}(t) \cdot R \cdot \big(s_j(t) - s_i(t)\big), \tag{7}$$

where $\alpha_{ij}(t) = +1$ if $d_{ij}(t) \leq T$ (attraction) and $\alpha_{ij}(t) = -1$ otherwise (repulsion). Post-update, $s_i(t^+)$ is clipped to $[0, 1]$ to maintain bounded opinions. The neighbor's opinion remains unchanged during this activation and updates only when selected as a focal agent.

2.3. Orientation-driven mobility with similarity-defined categories

Agent movement follows the vector-composition framework of Xiao et al. [30,39], adapted from evolutionary game theory to opinion dynamics. The key innovation is that "good" versus "bad" neighbors are defined by *ideological similarity* rather than strategy types, creating a direct coupling between opinion states and spatial forces.

Each agent moves with constant speed $v > 0$, updating its position according to

$$\mathbf{r}_i(t + 1) = \mathbf{r}_i(t) + v \cdot \hat{\mathbf{v}}_i(t), \tag{8}$$

where $\hat{\mathbf{v}}_i(t)$ is a unit direction vector constructed from local social forces.

Similarity-based partition. Within $\mathcal{N}_i(t)$, we partition neighbors into two categories using the same tolerance threshold T that governs opinion dynamics:

$$\mathcal{G}_i(t) = \{j \in \mathcal{N}_i(t) : d_{ij}(t) \leq T\} \quad \text{("good": similar, alliance-potential)}, \tag{9}$$

$$\mathcal{B}_i(t) = \{j \in \mathcal{N}_i(t) : d_{ij}(t) > T\} \quad \text{("bad": dissimilar, repulsion-inducing)}. \tag{10}$$

Reusing the same threshold T for both opinion dynamics and mobility serves three purposes. First, it enforces *theoretical parsimony*: agents who attract each other ideologically also attract each other spatially, while those who repel ideologically also induce spatial flight—reducing free parameters and avoiding arbitrary decoupling of attitude and behavior. Second, it reflects the empirical observation that affective polarization – negative evaluations of out-groups – manifests simultaneously in attitudes and behaviors, including residential choices [6,20]. Third, it creates a *self-reinforcing feedback loop*: repulsive ideological encounters drive spatial separation, which in turn reduces future cross-cutting exposure, potentially locking the system into segregated configurations.

Orientation-driven forces. Let $\hat{\mathbf{r}}_{ij}(t) = \Delta\mathbf{r}_{ij}(t)/r_{ij}(t)$ denote the unit vector pointing from neighbor j to focal agent i (using the minimum-image displacement). We define an opinion-alignment influence function

$$h(a_{ij}(t)) = a_{ij}(t)^{w_a}, \quad w_a > 0, \tag{11}$$

where $a_{ij}(t) = \max(0, 1 - d_{ij}(t))$ measures opinion alignment (1 = identical, 0 = maximally opposed) and w_a is the alignment-weight exponent used in the implementation. This replaces spatial distance weighting: more aligned neighbors exert stronger influence on movement orientation, while the attraction/repulsion sign is still determined by similarity category [26,30].

Following Xiao et al.'s construction [30], we compute two force components:

$$\mathbf{f}_i^{(G)}(t) = -\sum_{j\in\mathcal{G}_i(t)} h(a_{ij}(t)) \cdot \hat{\mathbf{r}}_{ij}(t), \tag{12}$$

$$\mathbf{f}_i^{(B)}(t) = +\sum_{j\in\mathcal{B}_i(t)} h(a_{ij}(t)) \cdot \hat{\mathbf{r}}_{ij}(t). \tag{13}$$

The force $\mathbf{f}_i^{(G)}$ points *toward* ideologically similar neighbors (attraction to "goods"), while $\mathbf{f}_i^{(B)}$ points *away* from dissimilar neighbors (repulsion from "bads"). These are direct analogues of Xiao et al.'s $\mathbf{f}_i^{(C)}$ and $\mathbf{f}_i^{(D)}$, with strategy categories replaced by ideological similarity categories.

We adopt the "social force" paradigm [26,30] rather than pairwise displacement: each agent integrates signals from *all* visible neighbors into a single direction vector before moving, yielding smoother trajectories than sequential pairwise displacements.

Each component is normalized (when nonzero) to obtain unit vectors $\hat{\mathbf{f}}_i^{(G)}$ and $\hat{\mathbf{f}}_i^{(B)}$. The combined orientation force is then

$$\mathbf{f}_i^{(GB)}(t) = (1 - \beta_i) \cdot \hat{\mathbf{f}}_i^{(G)}(t) + \beta_i \cdot \hat{\mathbf{f}}_i^{(B)}(t), \tag{14}$$

where $\beta_i \in [0, 1]$ interpolates between "approaching goods" ($\beta_i = 0$) and "leaving bads" ($\beta_i = 1$). A central finding from evolutionary game theory is that $\beta = 1$ ("leaving bads" dominance) outperforms $\beta = 0$ by blocking invasion pathways in spatial games [30]. In polarization contexts, this suggests that avoidance-driven mobility may be more potent than attraction-driven mobility in shaping emergent spatial patterns.

Steric repulsion. To prevent agent overlap and excessive local crowding, we incorporate a short-range steric repulsion term following Xiao et al. [30]:

$$\mathbf{g}_i(t) = \sum_{j\in\mathcal{N}_i(t)} \frac{\hat{\mathbf{r}}_{ij}(t)}{1 + \exp\left(\frac{r_{ij}(t) - r_f}{\sigma_s}\right)}, \tag{15}$$

where r_f sets the characteristic repulsion length scale and σ_s controls kernel steepness. Including steric forces is necessary because without short-range repulsion, ideologically similar agents would collapse onto identical positions, creating degenerate "super-agents" that distort density-dependent dynamics. The sigmoidal kernel provides a soft exclusion zone: repulsion activates strongly when agents approach within distance r_f, preventing unphysical collisions while remaining negligible at larger separations. This preserves agent individuality and ensures that local densities remain interpretable as population concentrations.

Final direction. The raw direction vector combines orientation forces, steric repulsion, and optional noise:

$$\mathbf{u}_i(t) = \hat{\mathbf{f}}_i^{(GB)}(t) + \kappa \cdot \mathbf{g}_i(t) + \boldsymbol{\xi}_i(t), \tag{16}$$

where $\kappa \geq 0$ weights steric repulsion and $\boldsymbol{\xi}_i(t)$ represents an optional noise term. In simulations we take $\boldsymbol{\xi}_i(t) = \eta \cdot \hat{\boldsymbol{\xi}}_i(t)$, where $\hat{\boldsymbol{\xi}}_i(t)$ is a random unit vector and $\eta \geq 0$ is the noise amplitude (set to zero in deterministic runs). The movement direction is the normalized vector:

$$\hat{\mathbf{v}}_i(t) = \frac{\mathbf{u}_i(t)}{\|\mathbf{u}_i(t)\|}. \tag{17}$$

If $\|\mathbf{u}_i(t)\| = 0$, we assign a random direction. Agents with empty neighborhoods ($\mathcal{N}_i(t) = \emptyset$) perform random drift: a step of length v in a uniformly random direction.

2.4. Mobility preferences and exposure–mobility coupling

Each agent possesses an individual orientation bias $\beta_i \in [0, 1]$ governing its movement trade-off between approaching similar neighbors and escaping dissimilar ones. To allow systematic exploration of how mobility preferences interact with exposure, we construct β_i as a clipped affine function of the global exposure parameter E plus an optional idiosyncratic offset:

$$\beta_i = \mathrm{clip}_{[0,1]}\Big(\underbrace{\mathrm{clip}_{[0,1]}\big(\beta_0 + k_{\beta E}(E - E_{\mathrm{ref}})\big)}_{\tilde{\beta}(E)} + \varepsilon_i\Big), \tag{18}$$

where:

- β_0 is a baseline orientation weight;
- $k_{\beta E}$ is a coupling slope linking exposure to mobility preference;
- E_{ref} is a reference exposure value (calibrated to match default ARM parameters);
- $\varepsilon_i \sim \mathcal{N}(0, \sigma_\beta)$ is an optional fixed (time-independent) heterogeneity term drawn at initialization; we set $\sigma_\beta = 0$ in the present study, reserving exploration of population-level heterogeneity for future work.

Why couple β to exposure E? By default we set $k_{\beta E} = 0$, so exposure governs only interaction probability and mobility preferences remain independent. We then treat $k_{\beta E} \neq 0$ as an explicit behavioral hypothesis: populations that are more "open" (high E) might also exhibit different spatial responses to disagreement. With $k_{\beta E} > 0$, higher exposure shifts the population toward escape-dominated movement; with $k_{\beta E} < 0$, the opposite tendency emerges. This coupling allows systematic exploration of how cognitive openness correlates with spatial behavior, bridging psychological dispositions and geographic outcomes.

2.5. Simulation protocol and observables

Update schedule. We define one *Monte Carlo step* (MCS) as N independent activations, each selecting a focal agent uniformly at random with replacement. During each activation, an agent i executes two sequential operations: (i) attempt one local ARM ideological interaction with a randomly selected neighbor j (Eq. (7)), then (ii) update its spatial position via the mobility rule (Eqs. (8)–(17)). When an interaction occurs, only the focal agent i updates its opinion; the neighbor's opinion is unchanged until it is selected as a focal agent in its own activation. Movement is applied only to the focal agent i, keeping the opinion and mobility updates distinct. This is a random-sequential asynchronous update scheme, not a synchronous all-pairs update. This ordering – *interact first, then relocate* – implements a natural causal logic: agents first experience social encounters with their current neighbors, then decide whether to stay or move based on the resulting local environment. In the baseline implementation, one movement attempt occurs per activation; we treat alternative time-scale separations as extensions.

Polarization metric. To quantify ideological polarization, we adopt the ARM measure: the variance of ideological positions [19]:

$$\mathcal{P}(t) = \mathrm{Var}\big(\{s_i(t)\}_{i=1}^{N}\big), \tag{19}$$

where high $\mathcal{P}$ indicates a population spread across the ideological spectrum (polarized) and low $\mathcal{P}$ indicates concentration around a central position (consensus).

Spatial observables. To rigorously characterize the emergent socio-spatial structure, we compute multiple complementary metrics.

First, we measure the spatial correlation of opinions using **Moran's I** [40], a standard index of spatial autocorrelation. For scalar opinions, it is calculated as:

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(s_i - \bar{s})(s_j - \bar{s})}{\sum_i (s_i - \bar{s})^2}, \tag{20}$$

where w_{ij} is the binary spatial adjacency weight (1 if $r_{ij} \le R_{\text{int}}$, 0 otherwise), $W = \sum_{i,j} w_{ij}$ is the total number of spatial links, and $\bar{s}$ is the global mean opinion. Values of I theoretically range from -1 to $+1$. In our context, a high positive I indicates the formation of spatial "echo chambers" where neighbors hold highly similar views, whereas $I \approx 0$ implies that spatial location predicts nothing about ideology (random mixing).

Second, we define the **good-component count** N_g as the number of connected components in the subgraph induced by "good" edges, where an edge connects i and j if $r_{ij} \le R_{\text{int}}$ and $d_{ij} \le T$. Formally, let $G_G = (V, E_G)$ with $V = \{1, \dots, N\}$ and $E_G = \{(i,j) : r_{ij} \le R_{\text{int}},\ d_{ij} \le T\}$; then

$$N_g = \#\text{components}(G_G), \tag{21}$$

which provides a direct proxy for the number of like-minded clusters.

Third, to quantify the persistence of mixed regions, we compute the **mixed-neighborhood fraction**

$$f_{\text{mix}}(t) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}\big[\exists j \in \mathcal{N}_i : d_{ij} \le T \ \wedge\ \exists k \in \mathcal{N}_i : d_{ik} > T\big], \tag{22}$$

the fraction of agents who simultaneously have at least one similar neighbor and at least one dissimilar neighbor within R_{int}.

2.6. Parameter summary

Table 1 summarizes the core parameters, along with default values and scan ranges used in the implementation.

Table 1
Parameter values used in simulations. "Fixed" indicates parameters held constant across all experiments; others show the values or ranges scanned.

Parameter	Meaning	Value
N	population size	1000 (fixed)
D	opinion dimension	1 (fixed)
L	domain size	20 (fixed)
R_{int}	interaction radius	1.0, 2.0, 3.0
E	exposure	0–0.5
T	tolerance	0–0.5
R	responsiveness	0.25 (fixed)
v	move speed	0, 0.1, 0.2, 0.5
β_0	baseline orientation bias	0–1
σ_β	mobility heterogeneity sd	0.0 (fixed)
$k_{\beta E}$	exposure–mobility coupling	−5, −2, 0, 2, 5
E_{ref}	exposure reference	0, 0.1, 0.2
w_a	alignment-weight exponent	1.0 (fixed)
κ	steric weight	1.0 (fixed)
r_f	steric length scale	0.2 (fixed)
σ_s	steric steepness	0.1 (fixed)
η	noise amplitude	0.0 (fixed)

3. Results

We organize the results around three central questions raised in the Introduction: (i) whether spatial mobility amplifies or attenuates ideological polarization, (ii) how tolerance and exposure jointly determine the polarization–consensus boundary, and (iii) whether spatial segregation and ideological polarization emerge in tandem or can be decoupled. Our simulations reveal that mobility acts as a *selective depolarizer*: it suppresses extremization in moderate-tolerance regimes while preserving – or even enhancing – spatial sorting. This dual role produces a counterintuitive outcome: mobile populations can be simultaneously less polarized in opinion space and more segregated in physical space than their static counterparts. Below we present systematic parameter sweeps that map out this behavior.

3.1. Polarization landscape across tolerance and exposure

Fig. 2 presents the phase diagram of mean polarization P across the tolerance–exposure (T, E) plane for four move speeds. The central finding is that mobility systematically shrinks the parameter region sustaining high polarization. In the static case ($v = 0$), elevated polarization dominates most of the low-tolerance regime and extends to moderate and high exposure values; consensus emerges only when tolerance is sufficiently large and exposure is simultaneously low. Introducing mobility ($v = 0.1$, $v = 0.2$) progressively contracts this high-P region and sharpens the phase boundary. At $v = 0.5$, strong polarization survives only in a narrow corner combining very low tolerance with moderate-to-high exposure.

The mechanism underlying this contraction relates to the ARM's repulsion dynamics: when $d_{ij} > T$, each interaction pushes agents toward opposite extremes. In a static population, agents cannot escape their local neighborhoods, so repeated repulsive encounters accumulate and drive extremization. Mobility provides an "escape valve": agents can relocate away from dissimilar neighbors *before* sufficient repulsive updates accumulate to push them toward the boundaries. This escape mechanism is most effective when tolerance is moderate—large enough that some neighbors remain attractive, yet small enough that repulsion would otherwise dominate. At very low T, even mobile agents find few attractive partners anywhere, rendering escape ineffective; polarization persists. At high T, attraction already dominates and mobility merely accelerates convergence.

3.2. Time-resolved ideology distributions

Figs. 3–5 reveal how opinion distributions evolve over time across different tolerance and mobility regimes, exhibiting two qualitatively distinct attractors that correspond to Axelrod et al.'s "polarization trap" [19]. At very low tolerance (most clearly at $T = 0.10$, and to a lesser extent at $T = 0.15$ depending on R_{int}), the population tends to accumulate near the ideological boundaries; this accumulation can become long-lived, though the boundary-pinning is most pronounced at $T = 0.10$. This boundary accumulation occurs because low T ensures that most pairwise distances exceed the tolerance threshold, triggering repulsion that drives agents toward extremes; once agents reach the boundaries, the clipping constraint $s \in [0, 1]$ prevents further movement, and subsequent repulsive encounters reinforce the extreme position. Notably, mobility up to $v = 0.2$ does not reliably eliminate boundary accumulation at very low T—agents relocate spatially but may remain trapped in the repulsion-dominated regime in opinion space. However, mobility does alter the *symmetry* of polarization: in the static case ($v = 0$), density accumulates at *both* boundaries ($s = 0$ and $s = 1$), producing symmetric bi-polarization; with mobility ($v > 0$), the distribution tends toward asymmetric uni-polarization, concentrating predominantly at one boundary. This symmetry-breaking arises because mobile agents can spatially segregate into like-minded clusters, and stochastic early-stage dynamics determine which ideological pole dominates—once a majority cluster forms, it attracts further converts while the minority pole diminishes.

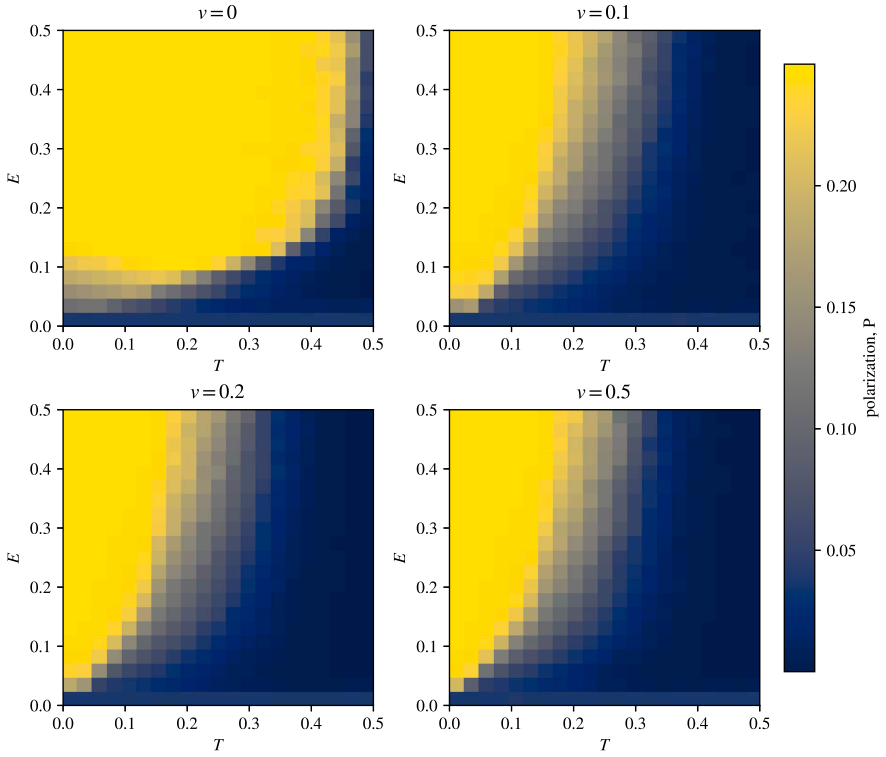


Fig. 2. Mean polarization P across tolerance T and exposure E for four move speeds $v \in \{0, 0.1, 0.2, 0.5\}$. Warmer colors indicate higher polarization. Increasing mobility shrinks the high-polarization region, especially at moderate tolerance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

At higher tolerance ($T \geq 0.20$), the system escapes this trap and converges toward a central consensus, with the density collapsing into a narrow band near $s = 0.5$ that sharpens and stabilizes more rapidly as v increases. This transition reflects the shift from repulsion-dominated to attraction-dominated dynamics: when T is large enough, most neighbors fall within the tolerance threshold, and assimilative interactions pull the population toward the mean. The interaction radius R_{int} modulates these dynamics without altering the qualitative picture—larger R_{int} produces slightly tighter, more centered consensus bands at high T, reflecting the averaging effect of larger neighborhoods, but the boundary between polarization and consensus regimes remains primarily controlled by tolerance rather than spatial range.

3.3. Coevolution of ideological and spatial structure

Figs. 6–9 track four complementary observables over time, revealing a striking decoupling between ideological polarization and spatial segregation—one of our central findings. Comparing the polarization trajectories (Fig. 6) with Moran's I (Fig. 7) exposes this decoupling most clearly. In static populations ($v = 0$), all tolerance levels eventually reach high polarization, though higher T delays the onset. Mobility fundamentally alters this picture: for $v > 0$, only low-tolerance populations ($T = 0.10$) continue to polarize, while moderate and high tolerance ($T \geq 0.15$) trajectories remain low and do not approach the high-polarization plateau characteristic of static populations. However, Moran's I tells a different story—for mobile populations, even moderate-to-high tolerance cases achieve substantial spatial correlation, with I rising rapidly before plateauing (or mildly declining) at late times. The lowest-tolerance case ($T = 0.10$) shows the strongest spatial correlation, but higher-T mobile populations exhibit *both* reduced polarization *and* moderate-to-high spatial correlation. This is what we term "depolarized segregation": agents cluster spatially with like-minded neighbors (high I) without drifting to ideological extremes (low P). The mechanism is straightforward—mobility enables agents to physically escape dissimilar neighbors, reducing cumulative repulsion and hence polarization, yet the same mobility causes like-minded agents to aggregate spatially. In effect, agents achieve local homogeneity through relocation rather than through opinion change.

The remaining observables illuminate the spatial structure underlying this regime. The mixed-neighbor fraction f_{mix} (Fig. 8), which quantifies agents simultaneously having both similar and dissimilar neighbors, decays toward zero under mobility across all tolerance levels—indicating the universal formation of echo chambers in the literal, spatial sense. This echo-chamber formation occurs regardless of whether the population polarizes: even depolarized populations (high T, $v > 0$) converge to $f_{\text{mix}} \approx 0$, with larger R_{int} accelerating this transition. The good-component count N_g (Fig. 9) reveals the cluster-level dynamics: at small interaction radius ($R_{\text{int}} = 1$), N_g exhibits a characteristic "nucleation–merger" profile with an initial rise as small clusters form, followed by decay as

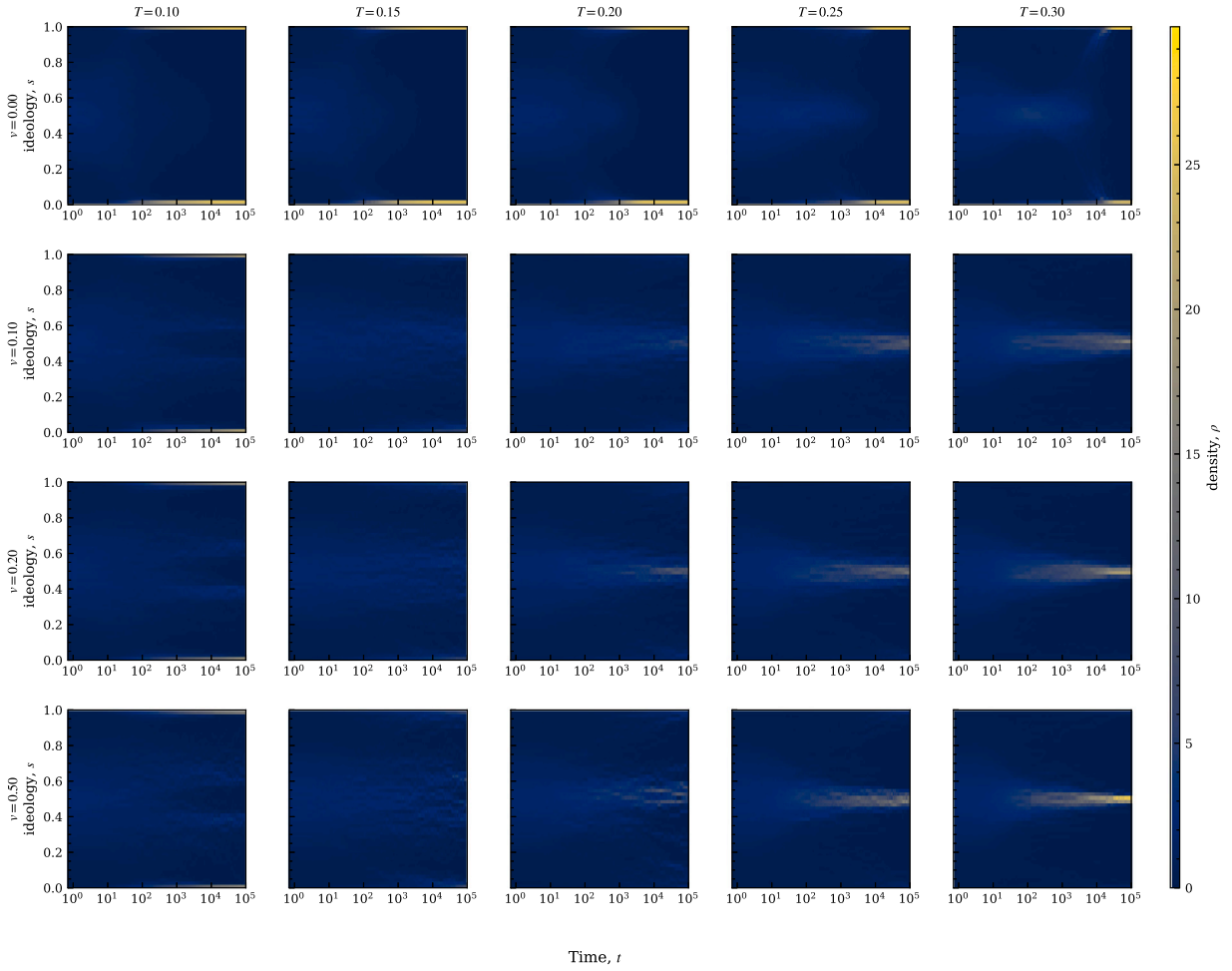


Fig. 3. Ideology density $\rho(s,t)$ over log time for $R_{\text{int}} = 1.0$ at baseline exposure $E = 0.1$. Columns vary tolerance T and rows vary move speed v. Very low tolerance favors boundary accumulation, whereas higher tolerance drives central convergence.

clusters merge or dissolve. As R_{int} increases, this hump flattens dramatically—larger visibility ranges connect nascent clusters from the outset, and by $R_{\text{int}} = 3$ the system effectively behaves as a single large domain with internal opinion gradients rather than a mosaic of isolated clusters.

Fig. 10 provides direct visual confirmation of these regimes by displaying four representative spatial configurations. Panel (a) shows the initial condition: agents are uniformly distributed in space with Gaussian-distributed opinions centered at $s = 0.5$. Panel (b) illustrates *polarized segregation* at low tolerance ($T = 0.05$, $v = 0.2$): agents have clustered into spatially separated groups, and opinions have drifted to the boundaries (blue ≈ 0, red ≈ 1), yielding both high P and high I. Panel (c) demonstrates the key phenomenon of *depolarized segregation* at high tolerance ($T = 0.50$, $v = 0.2$): spatial clustering is clearly visible, but opinions remain moderate (gray tones near $s = 0.5$), confirming that echo chambers need not produce ideological extremism. Panel (d) shows *static consensus* ($T = 0.80$, $v = 0$, $R_{\text{int}} = 2$): without mobility, agents remain spatially mixed but reach ideological consensus via repeated local averaging. The contrast between panels (b) and (c) visually captures our central finding: mobility enables spatial sorting without necessarily driving polarization.

3.4. Orientation bias: approaching goods versus leaving bads

Fig. 11 examines how the baseline orientation bias β_0—which interpolates between pure "approaching goods" ($\beta_0 = 0$) and pure "leaving bads" ($\beta_0 = 1$)—affects steady-state outcomes across different tolerance levels. This analysis directly tests Xiao et al.'s [30] finding from evolutionary game theory that escape-dominated movement outperforms approach-dominated movement in sustaining cooperation.

Polarization exhibits a modest but consistent decline as β_0 increases from 0 toward 1, with the effect most pronounced at low tolerance ($T = 0.1, 0.15$). This confirms that "leaving bads" — prioritizing escape from dissimilar neighbors — is more effective at

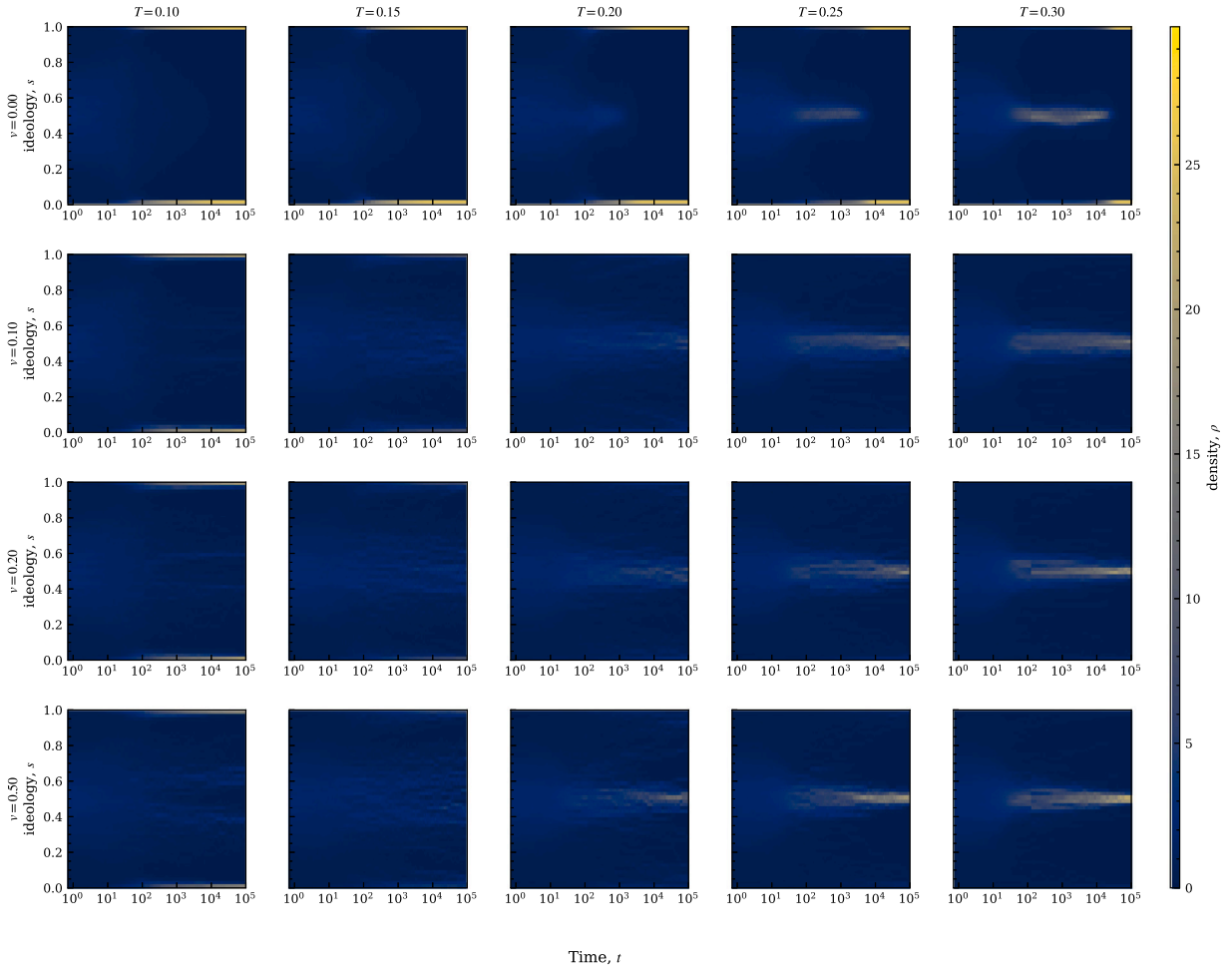


Fig. 4. Ideology density $\rho(s,t)$ over log time for $R_{\text{int}} = 2.0$ at baseline exposure $E = 0.1$. Columns vary tolerance T and rows vary move speed v. The same polarization-to-consensus transition appears, with boundary pinning concentrated at low T.

suppressing polarization than "approaching goods" — prioritizing attraction to similar neighbors. The mechanism is straightforward: escape-dominated movement reduces cumulative exposure to repulsive interactions, whereas approach-dominated movement merely accelerates clustering without preventing the repulsive encounters that drive extremization. However, at the pure escape limit ($\beta_0 = 1$), an interesting bifurcation emerges: low-tolerance populations ($T \leq 0.15$) exhibit sharp upward jumps in polarization, while higher-tolerance populations ($T \geq 0.2$) show slight further decreases. For low T, pure escape movement eliminates all attractive forces toward like-minded neighbors, causing agents to disperse rather than aggregate; the resulting spatial mixing paradoxically increases exposure to dissimilar neighbors and reignites repulsion-driven polarization. For high T, the same dispersion is benign because most interactions are attractive anyway, and the slight additional mixing marginally reduces polarization. This finding suggests that the optimal mobility strategy depends on the tolerance regime: balanced approach-escape mixtures work best at low tolerance, while pure escape can be effective at high tolerance. Notably, the overall effect size remains modest compared to tolerance: increasing T from 0.1 to 0.15 reduces polarization far more than shifting β_0 across its entire range.

Moran's I shows weak sensitivity to β_0 across most of the parameter range. However, at $\beta_0 = 1$, Moran's I jumps upward for all tolerance levels, with the largest increases at intermediate T. This spike reflects a qualitatively different spatial structure at the pure-escape limit, where the lack of attractive forces fundamentally alters clustering dynamics.

The mixed-neighborhood fraction f_{mix} exhibits the most dramatic boundary effect. For $\beta_0 < 1$, f_{mix} remains low across all tolerance levels, indicating successful sorting into homogeneous neighborhoods. At $\beta_0 = 1$, however, f_{mix} spikes sharply for *all* T, revealing that pure escape movement prevents the formation of stable like-minded clusters and leaves agents persistently exposed to mixed ideological environments. This reinforces the interpretation that pure "leaving bads" is counterproductive: agents flee dissimilar neighbors but, lacking attractive forces toward similar ones, fail to find ideological refuge.

The good-component count N_g completes the picture. For $\beta_0 < 1$, N_g is relatively insensitive to β_0, remaining roughly constant (or slightly increasing) until the discontinuity at $\beta_0 = 1$. At $\beta_0 = 1$, N_g drops sharply for all tolerance levels, with the most dramatic

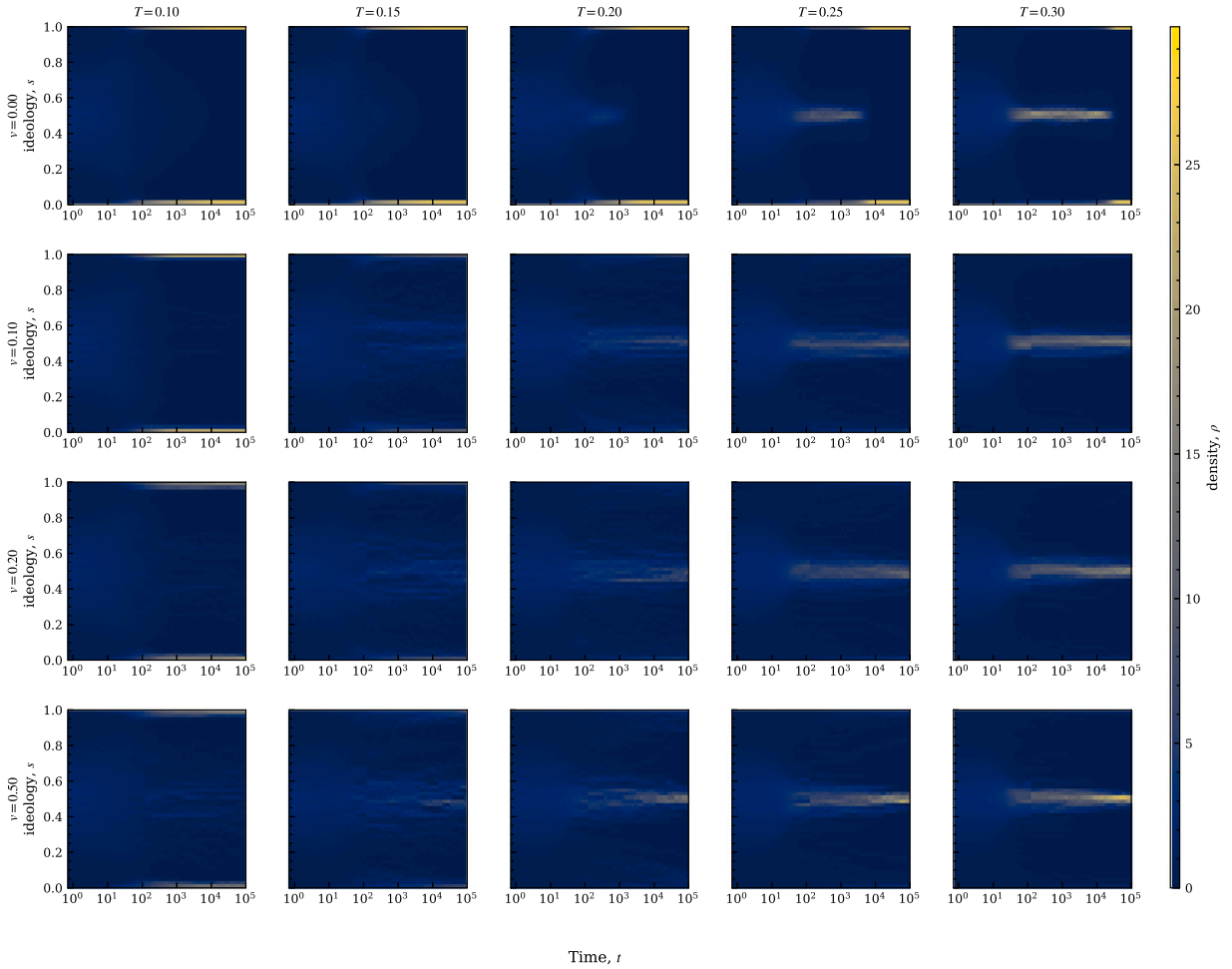


Fig. 5. Ideology density $\rho(s,t)$ over log time for $R_{\text{int}} = 3.0$ at baseline exposure $E = 0.1$. Columns vary tolerance T and rows vary move speed v. Tolerance remains the dominant control parameter even under larger interaction range.

collapse occurring at $T = 0.10$ (to approximately 3 components), while higher tolerance values settle at single-digit component counts (approximately 5–10). This universal collapse reflects the dissolution of coherent like-minded clusters when attractive forces vanish entirely; the especially severe drop at $T = 0.1$ occurs because strong repulsion combined with pure escape movement scatters agents without allowing stable cluster formation.

3.5. Exposure–mobility coupling effects

Fig. 12 examines how exposure–mobility coupling affects outcomes across different baseline exposure levels. The coupling parameter $k_{\beta E}$ links an agent's mobility bias β_i to global exposure E: positive $k_{\beta E}$ shifts agents toward escape-dominated movement ("leaving bads") as exposure increases, while negative $k_{\beta E}$ favors approach-dominated movement ("approaching goods"). Across all parameter combinations, the tolerance T exerts the strongest influence on outcomes – polarization P declines monotonically with T, Moran's I exhibits a characteristic U-shaped dependence, and the good-component count N_g decreases steadily – confirming that tolerance remains the master parameter of the system. However, the effectiveness of mobility-preference coupling is *critically dependent* on the baseline exposure reference E_{ref}. At $E_{\text{ref}} = 0$ (i.e., a reference point far below the baseline exposure $E = 0.1$), varying $k_{\beta E}$ produces substantial curve separation, with negative coupling sustaining higher polarization and higher N_g (more fragmented clustering) at intermediate T. Strikingly, at $E_{\text{ref}} = 0.1$ coupling effects nearly vanish—all curves collapse regardless of $k_{\beta E}$. This collapse is a direct consequence of the parameterization in Eq. (18): when the actual exposure E equals the reference value E_{ref}, the coupling term $k_{\beta E}(E - E_{\text{ref}})$ vanishes identically, rendering β_i independent of $k_{\beta E}$. The $E_{\text{ref}} = 0.1$ column thus represents a *parametric zero-point* of the coupling mechanism rather than an emergent dynamical property. At $E_{\text{ref}} = 0.2$, coupling sensitivity re-emerges with different characteristics.

These findings carry practical implications for interventions targeting mobility preferences, such as policies encouraging or discouraging residential sorting. When $E = E_{\text{ref}}$, altering individuals' approach/avoidance tendencies will have minimal impact on

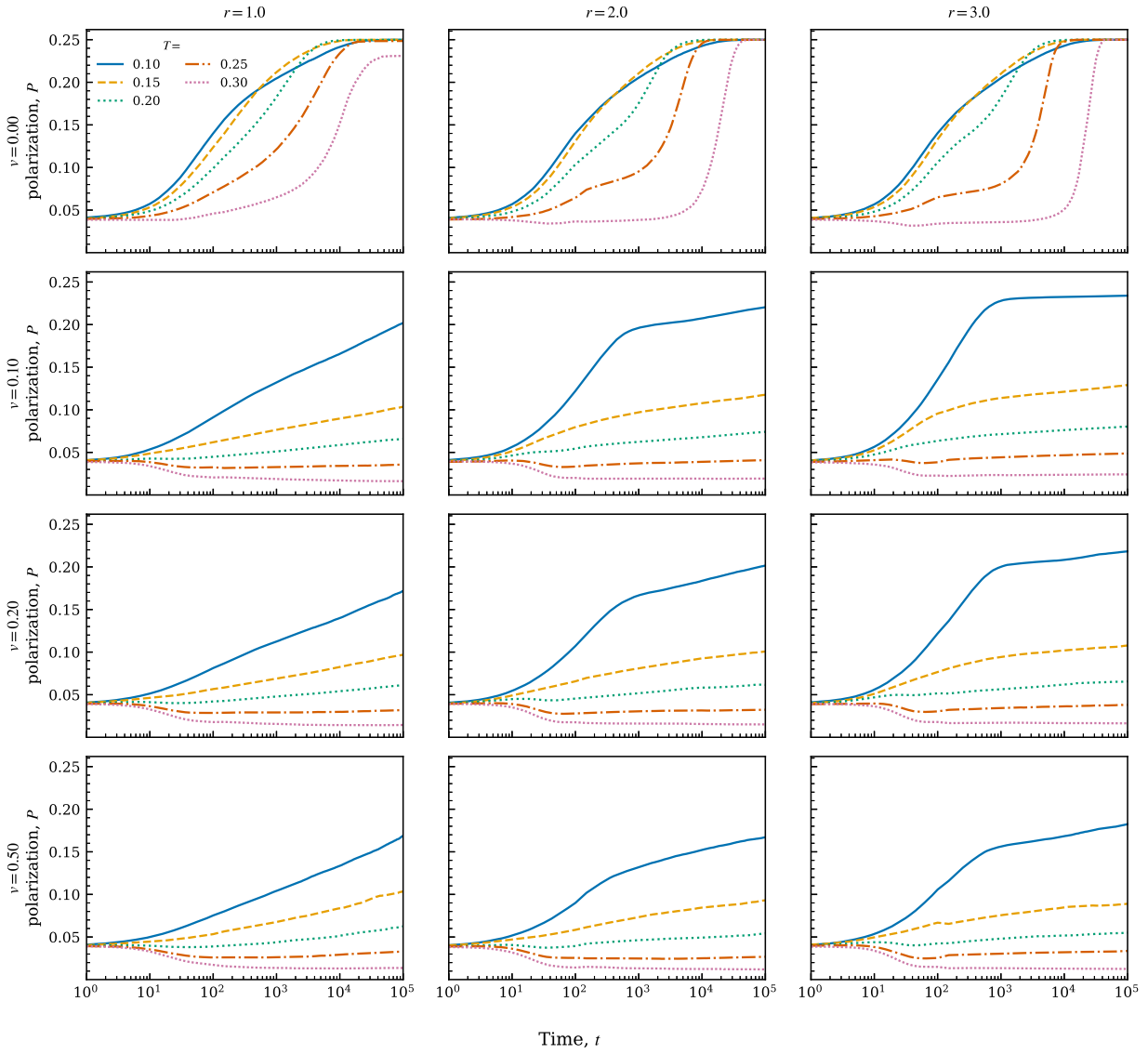


Fig. 6. Time series of polarization P on a log-time axis. Columns vary interaction radius R_{int} and rows vary move speed v; colors denote tolerance T. Mobility keeps polarization low for moderate/high tolerance, while very low tolerance remains polarization-prone. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

polarization or segregation—this is a built-in consequence of the coupling parameterization rather than a substantive finding about moderate-exposure societies. Conversely, in low-exposure environments where selective attention is strong, mobility-preference interventions could substantially reshape collective outcomes—but the direction of effect depends on the sign of the coupling.

4. Discussion

The central question motivating this work was whether spatial mobility amplifies or attenuates ideological polarization. Our results provide a nuanced answer: mobility acts as a *selective depolarizer*, suppressing extremization when tolerance exceeds a critical threshold while remaining ineffective – or even counterproductive – in low-tolerance regimes. This selectivity resolves an apparent tension in the literature. On one hand, mobility enables agents to escape hostile encounters before cumulative repulsion drives them to ideological extremes—an "escape valve" mechanism analogous to findings in evolutionary games that allowing partners to avoid bad interactions can improve collective outcomes. On the other hand, mobility accelerates spatial sorting into homogeneous enclaves, potentially eliminating the cross-cutting contacts that moderate opinions. Our simulations reveal that these two effects can be decoupled: mobile populations can simultaneously exhibit *lower* ideological polarization and *higher* spatial segregation than their static counterparts—a regime we term "depolarized segregation". In this state, echo chambers form (high segregation) without

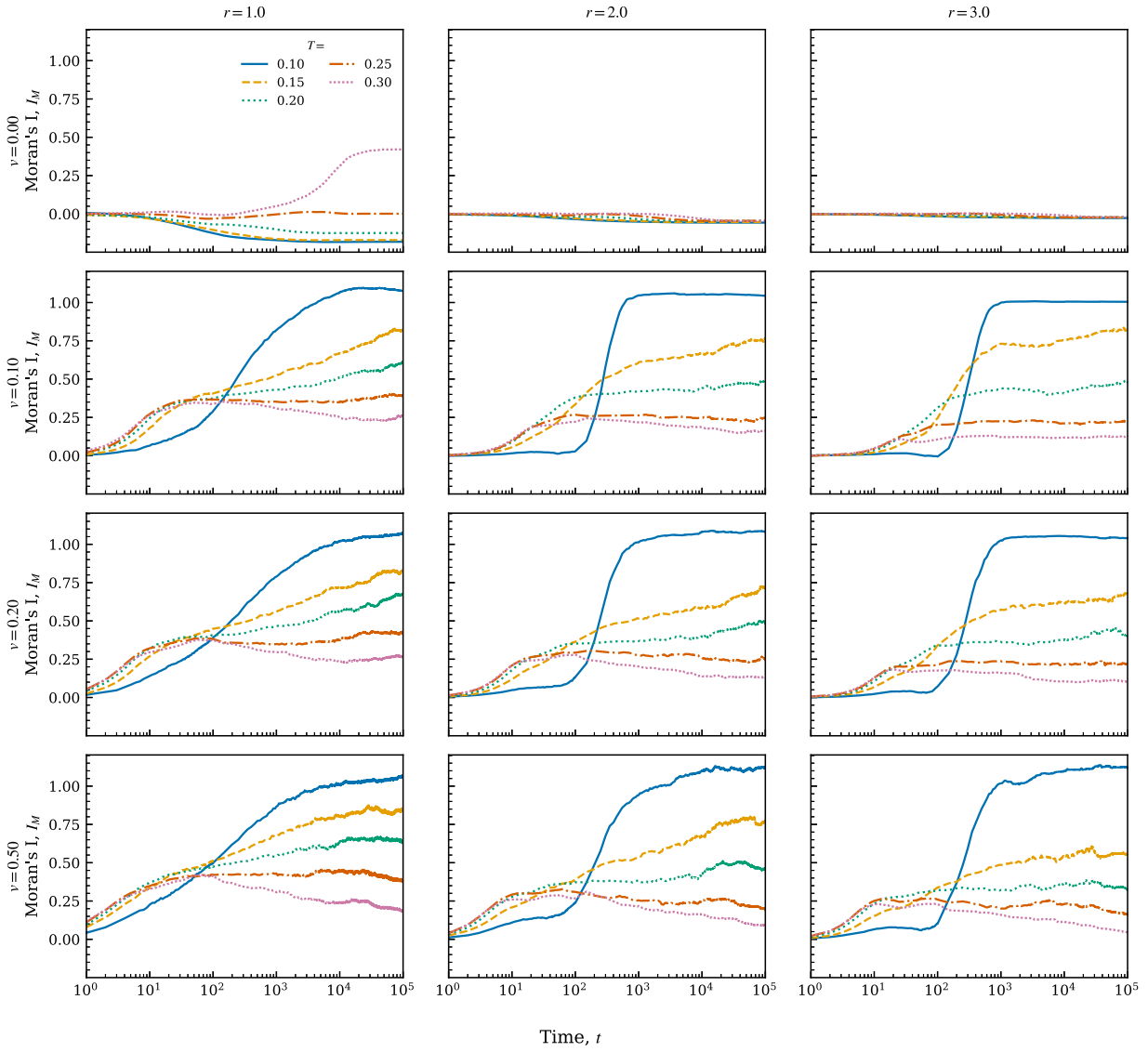


Fig. 7. Time series of Moran's I on a log-time axis. Columns vary interaction radius R_{int} and rows vary move speed v; colors denote tolerance T. Spatial assortativity can increase under mobility even when ideological polarization stays low. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ideological extremism (low polarization). Physical isolation from opposing views reduces the repulsive interactions that drive opinion divergence, even as it eliminates opportunities for persuasion. This mechanism aligns with findings from spatial models incorporating negative influence, where segregation consistently reduces bi-polarization by minimizing cross-group repulsive encounters [12,41]. This finding challenges the widespread assumption that echo chambers inevitably produce extremism, showing instead that under the right conditions, like-minded clustering can coexist with moderate opinions.

These results both confirm and extend prior work on spatial opinion dynamics. Consistent with Starnini et al. [27], we find that mobility combined with homophily produces emergent echo chambers—physically separated clusters of like-minded agents. However, our ARM-based framework goes further by revealing that such chambers need not coincide with extreme opinions, a distinction invisible in earlier bounded-confidence models where agents simply ignore those beyond their tolerance range rather than actively repelling each other. The importance of this distinction is underscored by network studies showing that even sparse cross-community ties can dramatically increase polarization when negative influence operates [42]—a finding that helps explain why our mobile agents, by reducing such bridging contacts, can achieve depolarization. In Starnini et al.'s model, heterogeneous groups were fragile and tended to fragment, but opinion clusters formed by random movement could still assume only a limited range of internal diversity. By contrast, our model's repulsive interaction mechanism shows that even as groups segregate, opinions can remain centrist—a novel insight into echo chamber morphology. Alraddadi et al. [28] demonstrated that directed mobility

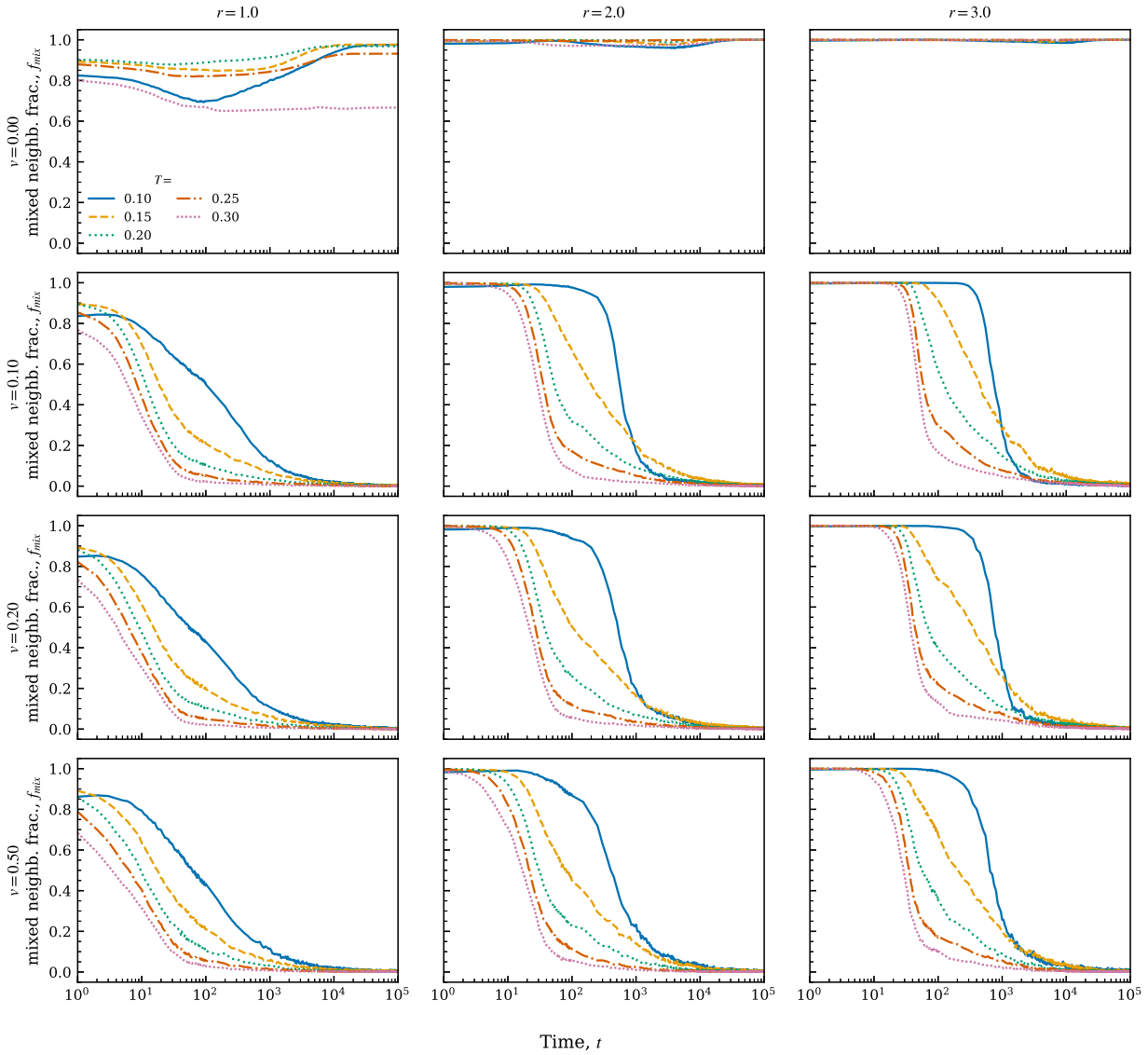


Fig. 8. Time series of mixed-neighbor fraction f_{mix} on a log-time axis. Columns vary interaction radius R_{int} and rows vary move speed v; colors denote tolerance T. Mobility drives mixed neighborhoods toward extinction, indicating robust echo-chamber formation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(toward similar, away from dissimilar) generates stronger segregation than random movement; our results confirm this pattern while adding a crucial nuance: the ideological consequences of directed mobility depend critically on tolerance. In low-tolerance settings, directed movement leads to polarized enclaves (repulsion dominates), whereas at higher tolerance, the same movement rule yields depolarized segregation. The "Paradise State" – perfect segregation with frozen opinions, akin to a stable pluralism – appears in our model at high tolerance and high mobility, where agents achieve local consensus through relocation rather than opinion conversion. Pasimeni et al. [29] identified agents' visibility range as a key parameter governing whether spatial clustering leads to polarization; we complement this by showing that mobility speed interacts with tolerance to reshape the phase diagram of outcomes. Faster mobility effectively broadens the parameter region leading to consensus by giving agents more opportunity to re-sort before opinions can polarize, thus contracting the high-polarization phase. Notably, our phase boundary analysis reveals a critical tolerance threshold below which even unlimited mobility cannot prevent polarization—implying that a society's willingness to physically re-assort can compensate, to a point, for lower open-mindedness. Our analysis of orientation bias provides a nuanced extension of Xiao et al.'s [30] finding that "leaving bads" outperforms "approaching goods" in sustaining cooperation. In our context, escape-dominated movement does reduce polarization—but only up to a point. At the pure escape limit ($\beta_0 = 1$), a bifurcation emerges: low-tolerance populations experience *increased* polarization because agents disperse without aggregating with like-minded

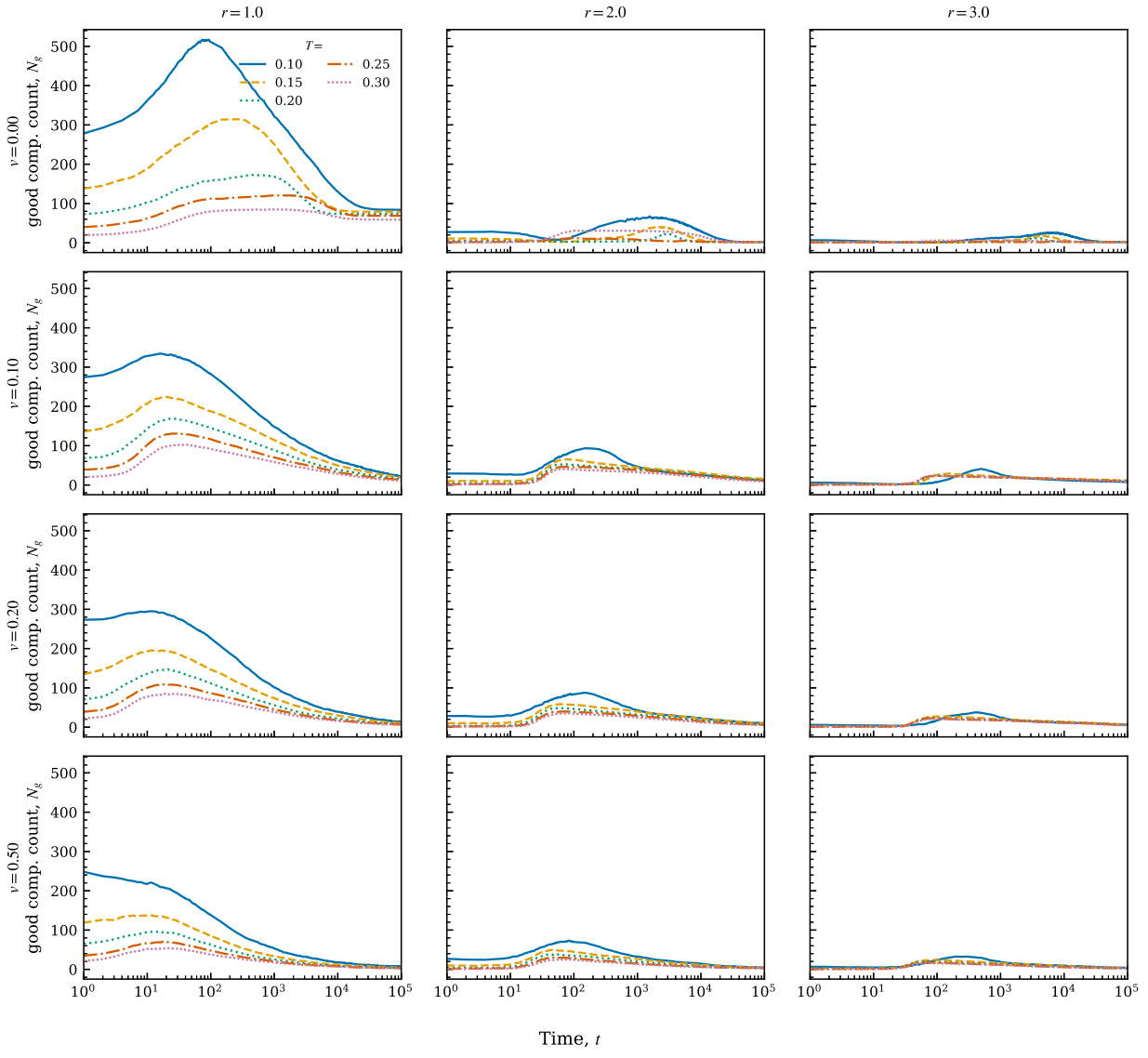


Fig. 9. Time series of good-component count N_g on a log-time axis. Columns vary interaction radius R_{int} and rows vary move speed v; colors denote tolerance T. Nucleation–merger dynamics are strongest at small interaction radius and weaken as R_{int} increases. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

peers, while high-tolerance populations benefit from slight additional mixing. This suggests that while moderate avoidance of out-group contact can depolarize, indiscriminate avoidance ($\beta \to 1$ with very low tolerance) may prevent any stable clustering, leading to persistent scattering—a society with neither polarization nor cohesion. Thus, the optimal mobility strategy is regime-dependent rather than universal.

The theoretical significance of our findings lies in the hierarchy of parameter importance they reveal. Tolerance emerges as the master parameter: it determines whether the system resides in an attraction-dominated or repulsion-dominated regime, and this distinction overrides the effects of mobility, exposure, and interaction radius. The exposure parameter E, which governs interaction probability, proves less decisive than one might expect—even high exposure cannot rescue a population from polarization when tolerance is low, because every interaction then triggers repulsion. Mobility, despite its dramatic effects on spatial structure, cannot overcome the polarization trap at low T. The orientation bias β_0 similarly plays a subordinate role: while escape-dominated movement generally reduces polarization, the effect size is modest compared to tolerance, and the pure escape limit ($\beta_0 = 1$) can be counterproductive in low-tolerance regimes. This hierarchy carries a sobering implication: interventions targeting exposure, mobility speed, or mobility strategy may be ineffective unless the underlying tolerance landscape is favorable. The parametric zero-point at $E = E_{\text{ref}}$, where the coupling term $k_{\beta E}(E - E_{\text{ref}})$ vanishes by construction, further illustrates this hierarchy: when the actual exposure

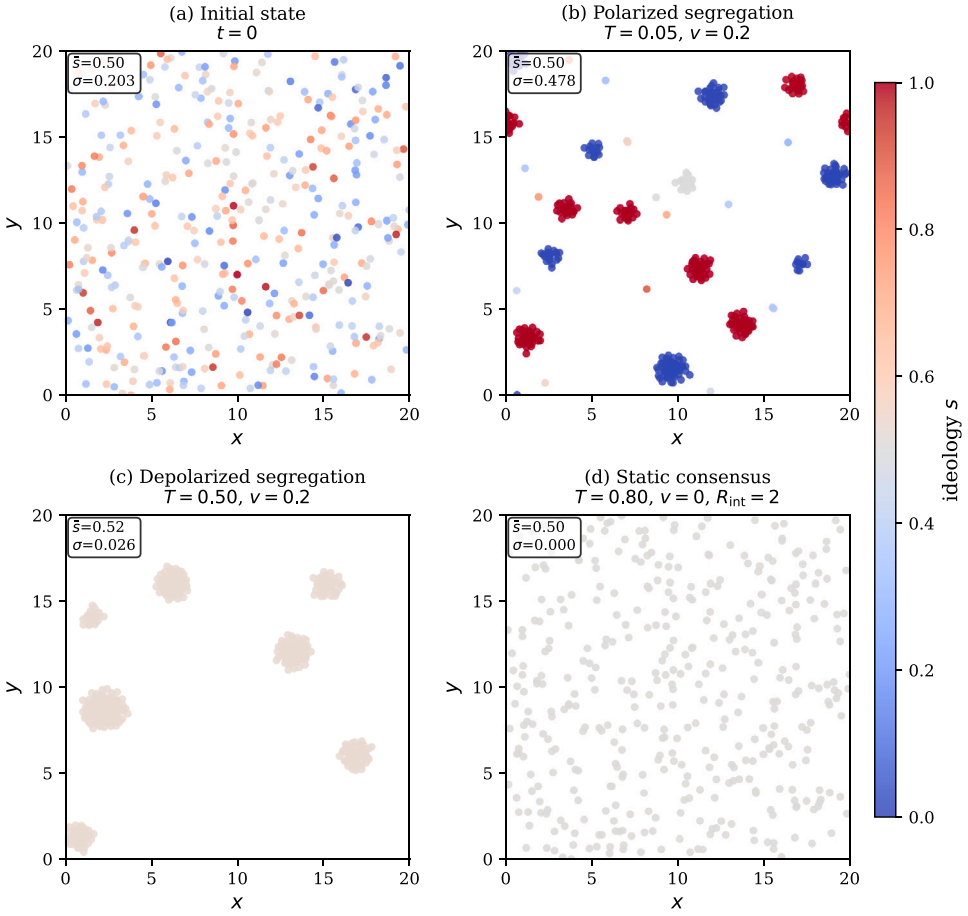


Fig. 10. Spatial snapshots illustrating the initial condition and three distinct dynamical regimes. Each panel displays agent positions colored by ideology s (blue ≈ 0, red ≈ 1, gray ≈ 0.5), with inset statistics showing mean opinion $\bar{s}$ and standard deviation σ. (a) Initial configuration ($t = 0$): uniformly random positions with Gaussian-distributed opinions centered at $\bar{s} = 0.50$ ($\sigma = 0.203$). (b) Polarized segregation ($T = 0.05$, $v = 0.2$): agents cluster spatially *and* hold extreme opinions, with $\sigma = 0.478$ indicating high polarization—distinct red and blue clusters form at the ideological boundaries. (c) Depolarized segregation ($T = 0.50$, $v = 0.2$): agents cluster spatially but opinions converge toward the center ($\bar{s} = 0.52$, $\sigma = 0.026$), demonstrating that echo chambers need not produce ideological extremism. (d) Static consensus ($T = 0.80$, $v = 0$, $R_{\text{int}} = 2$): without mobility, agents remain spatially mixed but reach near-perfect ideological consensus ($\bar{s} = 0.50$, $\sigma = 0.000$) via repeated local averaging. The contrast between panels (b) and (c) visually captures the central finding: mobility enables spatial sorting without necessarily driving polarization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

coincides with the reference value, variations in mobility-preference coupling $k_{\beta E}$ have no effect on β_i, and thus cannot influence outcomes. This is a mathematical consequence of the parameterization (Eq. (18)) rather than an emergent property of the dynamics.

As a mechanistic model, these findings offer conceptual insights rather than direct policy prescriptions—translating model parameters to real-world interventions requires identifying empirical proxies and validating the assumed mechanisms. With this caveat, the results suggest several considerations for understanding polarization dynamics. First, they caution against simple narratives equating echo chambers with extremism: physical segregation can coexist with moderate opinions if agents achieve homogeneity through relocation rather than through repulsion-driven radicalization. Second, they highlight a multi-objective trade-off in intervention design [38]: policies that reduce polarization by facilitating escape from hostile encounters may simultaneously reduce cross-cutting contact ($f_{\text{mix}} \to 0$), eliminating the bridging ties that could, under different circumstances, promote understanding. Third, our results show that neither increased exposure to opposing views nor engineered mobility (e.g., mixing people from different groups) will succeed in reducing polarization if the population's tolerance remains below the critical threshold. This aligns with empirical observations that simply exposing people to oppositional content can backfire when animosity is high—for example, Bail et al. [1] found that exposing Twitter users to accounts from the opposing political camp actually increased their polarization. This backfire effect is consistent with decades of research on motivated reasoning [43] and biased assimilation [44], which show that exposure to counter-attitudinal information often strengthens rather than weakens prior beliefs. Moreover, meta-analytic evidence indicates that negative intergroup contact exerts stronger effects on attitudes than positive contact [45], suggesting that in hostile environments the risks of cross-cutting exposure may outweigh its benefits. Such failures of well-intentioned interventions make sense in light of

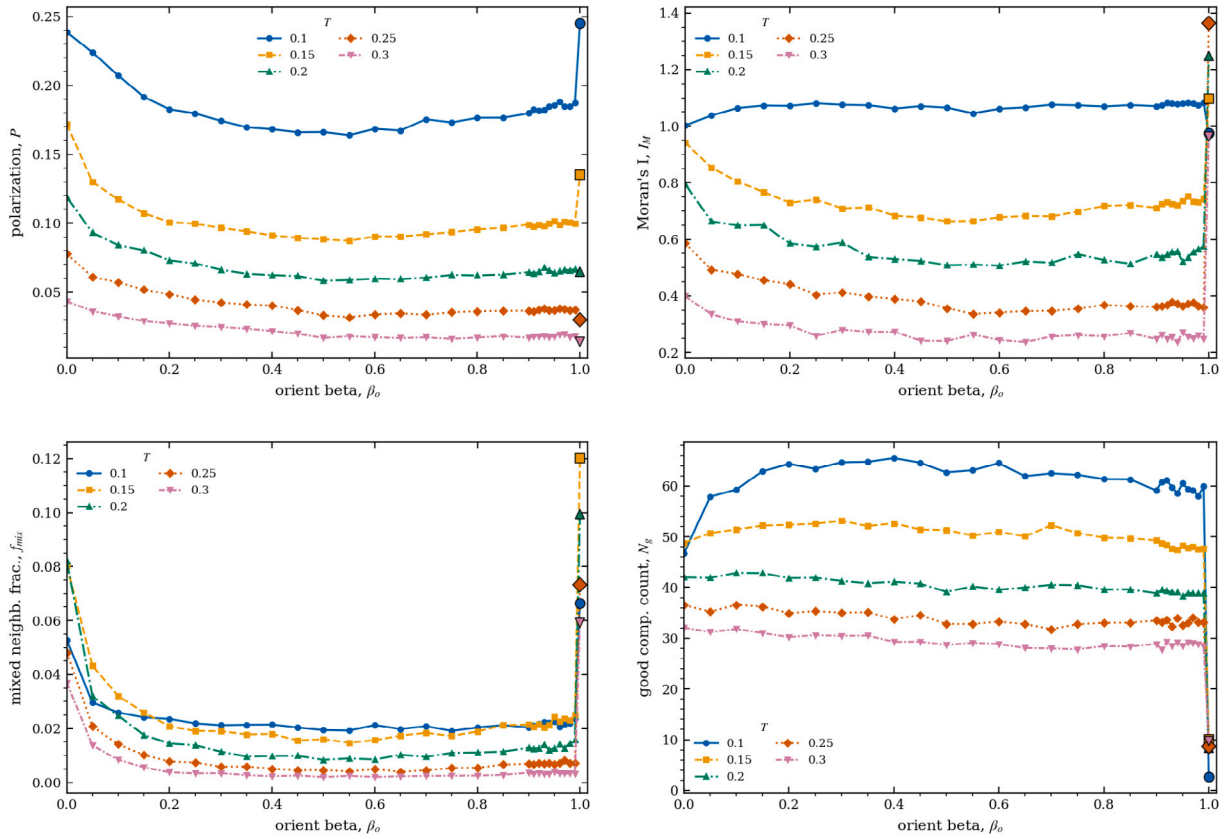


Fig. 11. Steady-state observables as functions of orientation bias β_0. Colors denote tolerance $T \in \{0.1, 0.15, 0.2, 0.25, 0.3\}$. Panels show (clockwise from top-left): polarization P, Moran's I, good-component count N_g, and mixed-neighbor fraction f_{mix}. β_0 has modest effects except at the pure-escape limit $\beta_0 = 1$, where all metrics show boundary discontinuities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

our model: if baseline tolerance is too low, additional cross-cutting exposure only triggers more repulsive interactions, entrenching hostility rather than alleviating it. Fourth, the E_{ref}-dependence of coupling effects suggests that the efficacy of mobility-preference interventions – such as policies encouraging residential integration or discouraging sorting – depends on baseline societal conditions. In low-exposure environments where selective attention is strong, such interventions may reshape collective outcomes; in moderate-exposure environments, they may have negligible impact. Fifth, our analysis of orientation bias reveals that promoting pure avoidance behavior ("leaving bads" without "approaching goods") can backfire in low-tolerance contexts: agents who only flee dissimilar neighbors without seeking similar ones fail to form stable like-minded clusters, leaving them persistently exposed to mixed environments that sustain polarization. Effective mobility-based interventions should therefore encourage balanced approach-escape strategies rather than pure avoidance.

Several limitations constrain the generalizability of these results and clarify the scope of our claims. Our model assumes a one-dimensional opinion space, so the findings should be interpreted as mechanism-level results under a single-axis abstraction rather than as a full representation of multidimensional ideological coupling. Tolerance is modeled as a global constant, and we therefore do not capture heterogeneous or time-evolving tolerance profiles across individuals and groups. The interaction layer is purely spatial: without a social network overlay, agents interact only with nearby neighbors, so long-range online ties and algorithmically mediated exposure are outside the model's scope. We also do not model external information sources, media influence, or elite cues—factors known to shape opinion dynamics independently of peer interaction. The initial conditions assume a homogeneous population with moderate opinions; dynamics from polarized initial states or with pre-existing subgroup structure may differ. Finally, our focus on deterministic dynamics (apart from partner selection) excludes the noise and bounded rationality that characterize real decision-making. We also note that the negative influence mechanism central to our model, while supported by psychological evidence on reactance and motivated reasoning, is not universally observed in empirical studies of opinion dynamics [46]. Our results are therefore most applicable to contexts characterized by affective polarization or intergroup animosity, where encounters with dissimilar others plausibly trigger distancing rather than persuasion.

Future work should address these limitations while extending the framework in several directions. Multidimensional opinion spaces would permit exploration of "issue alignment" – the tendency for opinions on disparate topics to correlate – and its interaction

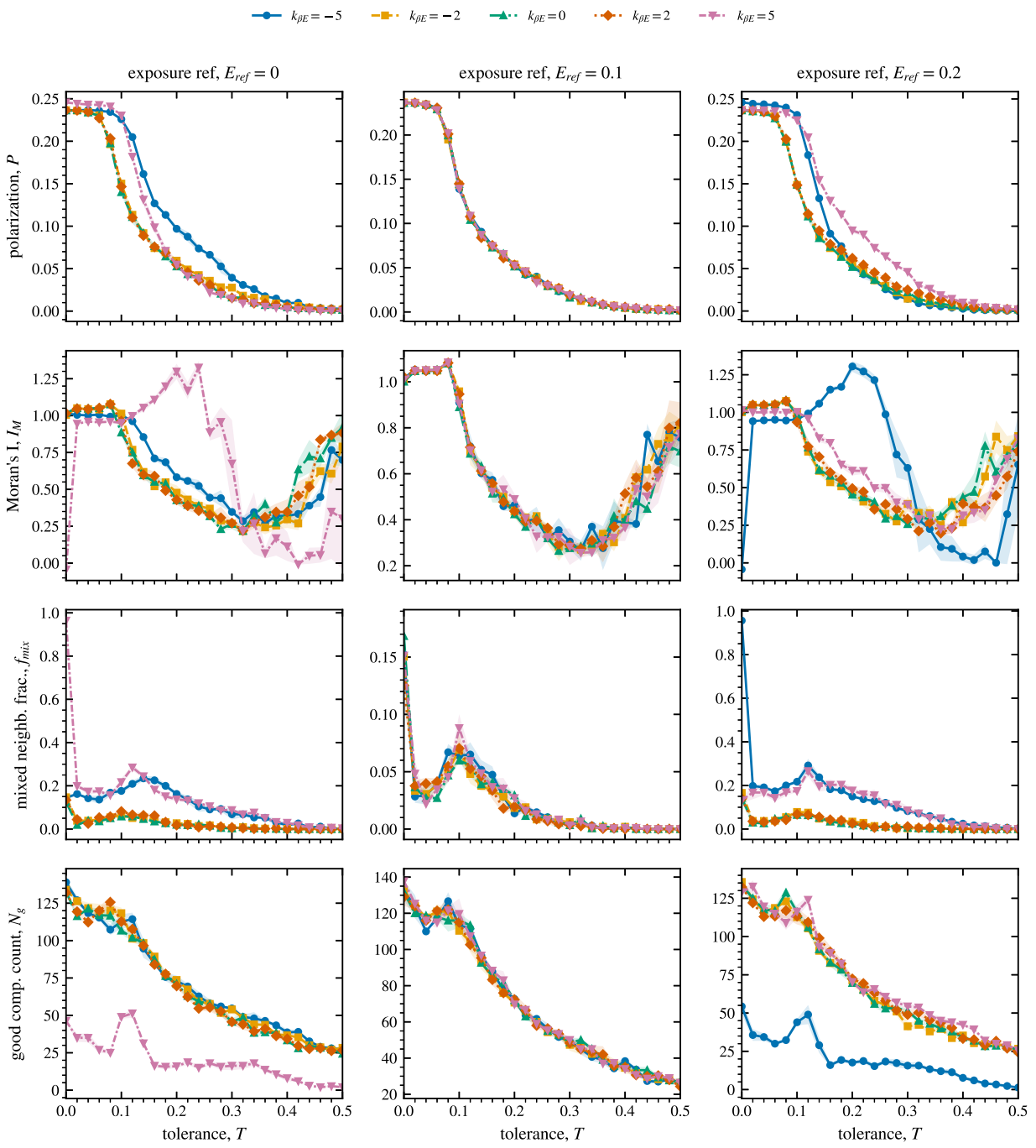


Fig. 12. Tolerance-dependent observables under exposure–mobility coupling. Columns vary exposure reference $E_{\text{ref}} \in \{0, 0.1, 0.2\}$ and colors denote coupling $k_{\beta E}$. Rows (top to bottom): polarization P, Moran's I, mixed-neighbor fraction f_{mix}, and good-component count N_g. Coupling effects are strongest away from the parametric zero-point $E_{\text{ref}} = 0.1$, where curves collapse by construction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with spatial sorting. Endogenizing tolerance as a heterogeneous and dynamic variable $T_i(t)$ would allow direct analysis of coevolution between social interaction and open-mindedness. Coupling the spatial model with an evolving social network would capture the interplay between physical and virtual exposure. Incorporating external shocks or media influence would test the robustness of the depolarized segregation regime to perturbations. Introducing heterogeneous mobility preferences ($\sigma_\beta > 0$) would allow exploration of scenarios where subpopulations differ in their approach/avoidance tendencies—a situation empirically relevant given differential

willingness to relocate by age or socioeconomic status. Empirical calibration using mobility data (e.g., residential sorting patterns) and opinion surveys could ground the model's parameter ranges. Finally, the framework could be extended to explore intervention strategies: under what conditions can targeted mobility subsidies, exposure mandates, or tolerance-enhancing education programs shift the system from polarization to depolarized coexistence?

5. Conclusion

We introduced a coupled model of opinion dynamics and spatial mobility that integrates ARM-style attraction–repulsion interactions with endogenous migration in continuous two-dimensional space. Our systematic exploration of the parameter space yielded four principal findings. First, spatial mobility acts as a selective depolarizer: it suppresses ideological extremization in moderate-to-high tolerance regimes by providing an "escape valve" through which agents relocate away from hostile neighbors before cumulative repulsion drives them to the boundaries of opinion space. Second, ideological polarization and spatial segregation can be decoupled—mobile populations may simultaneously exhibit reduced polarization and enhanced spatial clustering, a regime we term "depolarized segregation". This finding challenges the widespread assumption that echo chambers inevitably produce extremism. Third, tolerance emerges as the master parameter governing system behavior; mobility speed, exposure, interaction radius, and orientation bias all modulate outcomes but cannot overcome the polarization trap when tolerance falls below a critical threshold. Fourth, the optimal mobility strategy is regime-dependent: balanced approach-escape mixtures work best in low-tolerance populations, while pure escape movement can benefit high-tolerance populations but backfires in low-tolerance contexts by preventing stable cluster formation.

These results carry implications for both theory and practice. Theoretically, they demonstrate that the relationship between spatial structure and opinion dynamics is more nuanced than prior work has suggested: physical isolation from opposing views can reduce polarization by eliminating repulsive encounters, even as it creates echo chambers that preclude persuasion. Practically, they caution that interventions targeting exposure or mobility may prove ineffective unless the underlying tolerance landscape is favorable, and that policies promoting pure avoidance behavior can be counterproductive—effective mobility-based interventions should encourage balanced strategies that combine escape from dissimilar neighbors with attraction toward similar ones.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

I thank the anonymous reviewers for their careful reading and constructive comments, which helped improve the clarity and rigor of this manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.chaos.2026.118128.

Data availability

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

References

[1] Bail CA, Argyle LP, Brown TW, Bumpus JP, Chen H, Hunzaker MBF, Lee J, Mann M, Merhout F, Volfovsky A. Exposure to opposing views on social media can increase political polarization. Proc Natl Acad Sci 2018;115(37):9216–21. http://dx.doi.org/10.1073/pnas.1804840115.
[2] DiMaggio P, Evans J, Bryson B. Have Americans' social attitudes become more polarized? Am J Sociol 1996;102(3):690–755. http://dx.doi.org/10.1086/230995.
[3] Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci 2021;118(9):e2023301118. http://dx.doi.org/10.1073/pnas.2023301118.
[4] Pérez-Martínez H, Bauzá Mingueza F, Soriano-Paños D, Gómez-Gardeñes J, Floría LM. Polarized opinion states in static networks driven by limited information horizons. Chaos Solitons Fractals 2023;176:113917. http://dx.doi.org/10.1016/j.chaos.2023.113917.
[5] Schelling TC. Dynamic models of segregation. J Math Sociol 1971;1(2):143–86. http://dx.doi.org/10.1080/0022250X.1971.9989794.
[6] Bishop B, Cushing RG. The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart. Houghton Mifflin Harcourt; 2009.
[7] Tiebout CM. A pure theory of local expenditures. J Political Econ 1956;64(5):416–24. http://dx.doi.org/10.1086/257839.
[8] Perc M, Grigolini P. Collective behavior and evolutionary games – an introduction. Chaos Solitons Fractals 2013;56:1–5. http://dx.doi.org/10.1016/j.chaos.2013.06.002.
[9] Perc M, Jordan JJ, Rand DG, Wang Z, Boccaletti S, Szolnoki A. Statistical physics of human cooperation. Phys Rep 2017;687:1–51. http://dx.doi.org/10.1016/j.physrep.2017.05.004.
[10] DeGroot MH. Reaching a consensus. J Amer Statist Assoc 1974;69:118–21. http://dx.doi.org/10.1080/01621459.1974.10480137.
[11] Hegselmann R, Krause U. Opinion dynamics and bounded confidence: Models, analysis and simulation. J Artif Soc Soc Simul 2002;5(3):2, URL https://www.jasss.org/5/3/2.html.

[12] Flache A, Mäs M, Feliciani T, Chattoe-Brown E, Defuant G, Huet S, Lorenz J. Models of social influence: Towards the next frontiers. J Artif Soc Soc Simul 2017;20(4):2. http://dx.doi.org/10.18564/jasss.3521.

[13] Deffuant G, Neau D, Amblard F, Weisbuch G. Mixing beliefs among interacting agents. Adv Complex Syst 2000;3:87–98. http://dx.doi.org/10.1142/S0219525900000078.

[14] Axelrod R. The dissemination of culture: A model with local convergence and global polarization. J Confl Resolut 1997;41(2):203–26. http://dx.doi.org/10.1177/0022002797041002001.

[15] Nyhan B, Reifler J. When corrections fail: The persistence of political misperceptions. Political Behav 2010;32(2):303–30. http://dx.doi.org/10.1007/s11109-010-9112-2.

[16] Jusup M, Holme P, Kanazawa K, Takayasu M, Romić I, Wang Z, Geček S, Lipić T, Podobnik B, Wang L, Luo W, Klanjš̆ cek T, Fan J, Boccaletti S, Perc M. Social physics. Phys Rep 2022;948:1–148. http://dx.doi.org/10.1016/j.physrep.2021.10.005.

[17] Battiston F, Capraro V, Karimi F, Lehmann S, Migliano AB, Sadekar O, Sánchez A, Perc M. Higher-order interactions shape collective human behaviour. Nat Hum Behav 2025;9(12):2441–57. http://dx.doi.org/10.1038/s41562-025-02373-5.

[18] Chica M, Perc M, Santos FC. Success-driven opinion formation determines social tensions. IScience 2024;27(3):109254. http://dx.doi.org/10.1016/j.isci.2024.109254.

[19] Axelrod R, Daymude JJ, Forrest S. Preventing extreme polarization of political attitudes. Proc Natl Acad Sci 2021;118(50):e2102139118. http://dx.doi.org/10.1073/pnas.2102139118.

[20] Iyengar S, Sood G, Lelkes Y. Affect, not ideology: A social identity perspective on polarization. Public Opin Q 2012;76(3):405–31. http://dx.doi.org/10.1093/poq/nfs038.

[21] Huang C, Bian H, Han W. Breaking the symmetry neutralizes the extremization under the repulsion and higher order interactions. Chaos Solitons Fractals 2024;180:114544. http://dx.doi.org/10.1016/j.chaos.2024.114544.

[22] Cao F, Reed S. The iterative persuasion-polarization opinion dynamics and its mean-field analysis. SIAM J Appl Math 2025;85(4):1596–620. http://dx.doi.org/10.1137/24M1700661.

[23] Dandekar P, Goel A, Lee DT. Biased assimilation, homophily, and the dynamics of polarization. Proc Natl Acad Sci 2013;110(15):5791–6. http://dx.doi.org/10.1073/pnas.1217220110.

[24] Perc M, Szolnoki A. Coevolutionary games – a mini review. BioSystems 2010;99:109–25. http://dx.doi.org/10.1016/j.biosystems.2009.10.003.

[25] Szolnoki A, Mobilia M, Jiang L-L, Szczesny B, Rucklidge AM, Perc M. Cyclic dominance in evolutionary games: a review. J R Soc Interface 2014;11:20140735. http://dx.doi.org/10.1098/rsif.2014.0735.

[26] de Oliveira BF, Szolnoki A. Social dilemmas in off-lattice populations. Chaos Solitons Fractals 2021;144:110743. http://dx.doi.org/10.1016/j.chaos.2021.110743.

[27] Starnini M, Frasca M, Baronchelli A. Emergence of metapopulations and echo chambers in mobile agents. Sci Rep 2016;6:31834. http://dx.doi.org/10.1038/srep31834.

[28] Alraddadi EE, Allen SM, Sherborne N, Sherborne R, Whitaker RM. The role of homophily in opinion formation among mobile agents. J Inf Telecommun 2020;4(4):504–23. http://dx.doi.org/10.1080/24751839.2020.1772614.

[29] Pasimeni F, Wade R, Alkemade F. Opinion dynamic and social clustering in a 2D space: An agent based experiment. Comput Econ 2025. http://dx.doi.org/10.1007/s10614-025-10961-w.

[30] Xiao Z, Chen X, Szolnoki A. Leaving bads provides better outcome than approaching goods in a social dilemma. New J Phys 2020;22:023012. http://dx.doi.org/10.1088/1367-2630/ab6a3b.

[31] Xiao S, Zhang L, Li H, Dai Q, Yang J. Environment-driven migration enhances cooperation in evolutionary public goods games. Eur Phys J B 2022;95(4):67. http://dx.doi.org/10.1140/epjb/s10051-022-00327-8.

[32] Zhang H, Chen X, Szolnoki A. Evolution of cooperation among fairness-seeking agents in spatial public goods game. Appl Math Comput 2025;489:129183. http://dx.doi.org/10.1016/j.amc.2024.129183.

[33] Szolnoki A, Chen X. Strategy dependent learning activity in cyclic dominant systems. Chaos Solitons Fractals 2020;138:109935. http://dx.doi.org/10.1016/j.chaos.2020.109935.

[34] Habib MA, Tanaka M, Tanimoto J. How does conformity promote the enhancement of cooperation in the network reciprocity in spatial prisoner's dilemma games? Chaos Solitons Fractals 2020;138:109997. http://dx.doi.org/10.1016/j.chaos.2020.109997.

[35] Sharma G, He Z, Shen C, Tanimoto J. Asymmetric population promotes and jeopardizes cooperation in spatial prisoner's dilemma game. Chaos Solitons Fractals 2024;187:115399. http://dx.doi.org/10.1016/j.chaos.2024.115399.

[36] Wang Q, Chen X, Szolnoki A. Coevolutionary dynamics of feedback-evolving games in structured populations. Chaos Solitons Fractals 2025;193:116070. http://dx.doi.org/10.1016/j.chaos.2025.116070.

[37] Szolnoki A, Perc M. The self-organizing impact of averaged payoffs on the evolution of cooperation. New J Phys 2021;23:063068. http://dx.doi.org/10.1088/1367-2630/ac0aa6.

[38] Han TA, Duong MH, Perc M. Evolutionary mechanisms that promote cooperation may not promote social welfare. J R Soc Interface 2024;21:20240547. http://dx.doi.org/10.1098/rsif.2024.0547.

[39] Angelani L. Collective predation and escape strategies. Phys Rev Lett 2012;109(11):118104. http://dx.doi.org/10.1103/PhysRevLett.109.118104.

[40] Moran PAP. Notes on continuous stochastic phenomena. Biometrika 1950;37(1–2):17–23. http://dx.doi.org/10.1093/biomet/37.1-2.17.

[41] Feliciani T, Flache A, Tolsma J. How, when and where can spatial segregation induce opinion polarization? Two competing models. J Artif Soc Soc Simul 2017;20(2):6. http://dx.doi.org/10.18564/jasss.3419.

[42] Flache A, Macy MW. Small worlds and cultural polarization. J Math Sociol 2011;35(1–3):146–76. http://dx.doi.org/10.1080/0022250X.2010.532261.

[43] Taber CS, Lodge M. Motivated skepticism in the evaluation of political beliefs. Am J Political Sci 2006;50(3):755–69. http://dx.doi.org/10.1111/j.1540-5907.2006.00214.x.

[44] Lord CG, Ross L, Lepper MR. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. J Pers Soc Psychol 1979;37(11):2098–109. http://dx.doi.org/10.1037/0022-3514.37.11.2098.

[45] Paolini S, Gibbs M, Sales B, Anderson D, McIntyre K. Negativity bias in intergroup contact: Meta-analytical evidence that bad is stronger than good, especially when people have the opportunity and motivation to opt out of contact. Psychol Bull 2024;150(8):921–64. http://dx.doi.org/10.1037/bul0000439.

[46] Takács K, Flache A, Mäs M. Discrepancy and disliking do not induce negative opinion shifts. PLoS ONE 2016;11(6):e0157948. http://dx.doi.org/10.1371/journal.pone.0157948.