

Opinion dynamics, weighted memes and relational structure: comparison of five experimental conditions (Exp1–Exp5)

Circulation version – NetLogo / ABM analysis

Date: 2026-01-25

Summary

We analyze a NetLogo simulator of opinion dynamics at two levels: (i) a "direct" level where opinion adoption depends on prevalence gaps, influence, polarization and network structure; (ii) a "meme" level where prevalence stems from the stock of internal representations (quantity) and the opinion of a weighted balance (weight). Using five experimental conditions (Exp1–Exp5) contrasting (a) the empirical distribution of agents and the network from a survey (2022) versus a random distribution, and (b) two meme injection schemes (center-targeted vs. global), we compare camp inversions, opinion–meme spread (Wgap), volatility, and a dynamic relational distance proxy. The results show that the Exp1 condition (empirical distribution + targeted injection center) maximizes reversals and volatility while increasing Wgap, whereas random distributions strongly reduce reversals despite a higher average relational distance. We discuss these differences as the expression of generative mechanisms (seeding, cognitive coherence, homophily) and propose an empirical anchor and a robustness strategy (Exp2–Exp5) compatible with a NetLogo/complexity science submission.

1) Introduction

Understanding the dynamics of opinion formation and transformation within polarized societies remains a central challenge for the social sciences, particularly in contexts marked by ideological fragmentation, asymmetric information flows, and uneven cognitive engagement. Traditional approaches based on static attitudinal measurements or linear causal models often fail to capture the nonlinear, path-dependent, and interaction-driven mechanisms through which opinions evolve over time. In response to these limitations, agent-based modeling (ABM) has emerged as a powerful methodological framework for exploring opinion dynamics as emergent phenomena arising from local interactions, heterogeneous cognitive states, and evolving social networks.

Within this tradition, a growing body of work emphasizes the need to move beyond purely scalar representations of opinion—typically modeled as positions along a unidimensional ideological axis—and to explicitly account for the cognitive substrates that sustain, stabilize, or destabilize those opinions. In particular, recent theoretical developments highlight the role of internal representations—arguments, narratives, or “memes”—that agents accumulate, exchange, and weight differently when forming judgments. These internal structures not only condition opinion

strength and susceptibility to change, but also mediate the effects of social influence, network exposure, and external shocks.

The present article builds on this perspective by proposing and empirically grounding a two-level agent-based model of opinion dynamics that explicitly distinguishes between (i) a **surface level** operating directly on opinion polarity and prevalence, and (ii) a **subsymbolic cognitive level** operating through meme stocks and meme-weighted opinion reconstruction. At the surface level, agents update their opinions through social influence, event-driven perturbations, and network-mediated exposure. At the cognitive level, opinions are not directly transmitted; instead, agents exchange, accumulate, and reweight memes whose internal balance determines both opinion direction and resistance to change.

A central contribution of this work lies in the formalization and empirical validation of the concept of **Wgap**—defined as the weighted imbalance between pro and contra meme stocks—and its role as a mediating variable between relational distance and opinion volatility. We hypothesize that opinion inversions (i.e., sign changes in opinion polarity) are not randomly distributed across the ideological space, but are structurally concentrated among agents characterized by (a) low prevalence, (b) high cognitive imbalance (W_{gap}), and (c) exposure to relationally distant peers. Conversely, agents located at ideological extremes are expected to exhibit higher internal saturation, greater meme coherence, and lower rates of inversion, even when subject to significant social pressure.

To test these hypotheses, we combine empirical data from a 2022 survey used to initialize agent distributions and social ties, with multiple simulation experiments conducted under five distinct experimental conditions (Exp1–Exp5). These conditions systematically vary the empirical versus random initialization of agents and networks, as well as the scope and targeting of meme injection processes. By comparing outcomes across three-group and five-group ideological segmentations, and by introducing both static and dynamic measures of relational distance, we assess the robustness and generality of the proposed mechanisms.

Beyond its empirical findings, this article aims to contribute methodologically to the literature on cognitive ABMs by demonstrating how internal representational dynamics can be formally linked to observable macro-patterns such as polarization, volatility, and ideological lock-in. The model is implemented in NetLogo and provides explicit procedural hooks—most notably through the recompute-from-memes routine—that allow researchers to trace opinion change back to its cognitive determinants. In doing so, the work bridges opinion dynamics, meme-based cognition, and complexity science, offering a reusable framework for studying belief dynamics in polarized environments.

2) Theoretical framework and hypotheses

2.1 Opinion Dynamics as a Complex Adaptive System

Opinion formation is increasingly conceptualized as a complex adaptive process resulting from the interaction of heterogeneous agents embedded in evolving social networks. Rather than converging toward stable equilibria, such systems often exhibit nonlinear trajectories, path dependence, and sensitivity to initial conditions. Small perturbations—whether endogenous or exogenous—can produce disproportionate effects, particularly when cognitive and relational structures are unevenly distributed across the population.

Within this framework, opinions are not treated as static attributes, but as dynamic states continuously reshaped by social exposure, internal cognitive processing, and feedback from the collective environment. Agents differ not only in their current opinion polarity, but also in the depth, coherence, and internal consistency of the representations supporting those opinions. These differences give rise to heterogeneous responses to influence, producing asymmetric diffusion patterns and localized zones of instability.

2.2 From Scalar Opinions to Meme-Based Cognition

Most classical models of opinion dynamics represent opinions as scalar variables evolving through averaging, bounded confidence, or threshold mechanisms. While analytically tractable, such representations conflate opinion direction with opinion strength and leave unexplained why some agents are highly resistant to change whereas others exhibit repeated reversals.

To address this limitation, we adopt a meme-based cognitive representation in which opinions emerge from the balance of internal stocks of pro and contra representations. Memes are defined here as minimal cognitive units—arguments, narratives, frames, or justifications—that agents can acquire, transmit, forget, or reweight through interaction. Importantly, memes do not merely accumulate; they interact asymmetrically, eroding or reinforcing one another depending on their polarity and weight.

In this formulation, opinion polarity is not directly transmitted between agents. Instead, social influence operates by modifying meme stocks, which are subsequently recomposed into an expressed opinion. This distinction introduces a temporal and cognitive lag between exposure and observable opinion change, allowing for partial influence, latent instability, and delayed inversions.

2.3 Prevalence, Saturation, and Cognitive Resistance

Prevalence is defined as the total quantity of memes held by an agent, irrespective of polarity. It functions as a proxy for cognitive saturation and ideological entrenchment. Agents with low prevalence possess relatively few internal representations and are therefore weakly anchored; their opinions tend to be volatile and highly sensitive to local influence. By contrast, agents with high prevalence exhibit dense internal meme ecologies, conferring resistance to change even when exposed to opposing viewpoints.

This distinction is central to understanding why opinion change is unevenly distributed across the ideological spectrum. Extremist agents are not merely extreme because of opinion polarity; they are also typically characterized by high prevalence and strong internal coherence. Centrist agents, by contrast, often occupy regions of low saturation, making them more susceptible to perturbations and cross-camp influence.

2.4 The Wgap as a Cognitive Imbalance Indicator

To operationalize the internal structure of meme-based cognition, we introduce the concept of **Wgap**, defined as the weighted difference between positive and negative meme stocks. Unlike raw meme counts, Wgap captures both directionality and intensity, reflecting the degree of internal cognitive imbalance favoring one side of the opinion spectrum.

Formally, Wgap increases when weighted pro memes dominate contra memes, and decreases in the opposite case. Values near zero indicate either balanced internal representations or low overall meme content. Crucially, Wgap is hypothesized to mediate the relationship between social exposure and observable opinion change: relational distance influences meme acquisition, which in turn modifies Wgap, eventually leading—under certain conditions—to opinion inversion.

2.5 Relational Distance: Static and Dynamic Measures

Social influence is not uniform across network ties. The impact of a neighbor depends not only on tie existence, but also on ideological distance. We distinguish between two complementary notions of relational distance.

Static relational distance is defined as the ideological distance between agents as initialized from empirical or random distributions. It reflects long-term structural polarization embedded in the network.

Dynamic relational distance, by contrast, evolves over time as agents update their opinions. It captures moment-to-moment exposure to dissimilar views and allows for the identification of transient zones of tension where agents are repeatedly confronted with discordant signals.

We hypothesize that dynamic relational distance is a stronger predictor of cognitive imbalance (Wgap) and opinion volatility than static distance alone.

2.6 Hypotheses

Based on the preceding framework, we formulate the following hypotheses:

H1 — Concentration of inversions among weakly anchored agents

Opinion inversions are disproportionately concentrated among agents with low prevalence and low internal saturation.

H2 — Mediating role of W_{gap}

The effect of relational distance on opinion inversion is mediated by W_{gap} : greater exposure to ideologically distant neighbors increases cognitive imbalance, which in turn raises the probability of inversion.

H3 — Central instability hypothesis

Agents located in centrist regions of the opinion space exhibit higher volatility, higher inversion rates, and higher sensitivity to meme injection than agents at ideological extremes.

H4 — Extremist stability hypothesis

Agents at ideological extremes show lower inversion rates despite experiencing opinion oscillations within their own camp, reflecting high prevalence and internal meme coherence.

H5 — Robustness across initialization regimes

The qualitative relationships between relational distance, W_{gap} , and volatility persist across empirical and random initializations, as well as across different meme injection scopes.

Together, these hypotheses frame opinion dynamics as a cognitively mediated process in which internal representational structure, rather than opinion polarity alone, governs stability, change, and polarization trajectories.

3) Empirical data and simulation design

3.1 Empirical Basis: Survey-Derived Opinion Structure

The empirical foundation of the model relies on data drawn from a large-scale survey conducted in Québec in 2022, designed to capture opinion polarization, issue salience, and relational proximity among respondents on a central political cleavage. The survey provided not only distributions of opinion polarity, but also indirect indicators of conviction strength and relational alignment, allowing for the construction of empirically grounded agent constellations.

From this dataset, agents were initialized with heterogeneous opinion values spanning the interval $([-1, +1])$, corresponding to strong opposition and strong support respectively. These values were not sampled uniformly, but calibrated to reproduce the observed empirical density across ideological positions. In addition, prevalence values were assigned so as to reflect differential anchoring across the opinion spectrum, with higher prevalence concentrated toward the extremes and lower prevalence around the center.

Crucially, the survey also informed the initial structure of the social network. Each agent was connected to a fixed number of neighbors selected according to ideological proximity rules derived from the empirical distribution, producing a network in which homophily is present but incomplete. This structure ensures that agents are predominantly exposed to similar opinions while still encountering cross-cutting ties, a configuration consistent with empirical findings on political discussion networks.

3.2 Agent Attributes and State Variables

Each agent in the simulation is characterized by a set of state variables evolving over time:

- **Opinion:** a continuous scalar in $([-1, +1])$, representing expressed stance.
- **Prevalence:** an integer-valued measure representing the total number of internal representations held by the agent.
- **Influence:** a bounded variable capturing the agent's capacity to transmit memes to others.
- **Meme stocks:** two internal reservoirs (meme-plus and meme-minus) corresponding to representations supporting positive or negative polarity.
- **Weighted meme stocks:** (meme-plus_w, meme-minus_w) allowing differential impact of memes on opinion recomposition.

Opinion is not updated directly through social interaction. Instead, interactions modify meme stocks, which are subsequently recomposed into an expressed opinion via a deterministic aggregation rule. This separation between cognitive state and expressed state is a defining feature of the model.

3.3 Dual-Level Dynamics: Direct and Meme-Based Processes

The simulator operates on two coupled but analytically distinct levels.

At the **direct level**, opinion change is governed by probabilistic adoption mechanisms modulated by opinion distance, group alignment, and influence. This level corresponds to classical opinion dynamics models and serves as a baseline for comparison.

At the **meme-based level**, which is activated when use-memes? is enabled, interactions affect internal meme stocks rather than opinions directly. Meme acquisition, decay, and cross-erosion are controlled by parameters governing gain, forgetting, and anti-leakage between opposing representations. Opinion is recomputed from meme stocks at each tick using the recompute-from-memes procedure, which enforces internal consistency between cognition and expression.

These two levels can operate independently or jointly, allowing systematic exploration of how cognitive mediation alters macroscopic outcomes.

3.4 Segmentation of the Opinion Space

To analyze heterogeneity in dynamics, agents are segmented according to their initial opinion into two alternative classification schemes.

The **three-group segmentation** distinguishes:

- left-oriented agents $((\text{opinion} < -0.3))$,
- centrist agents $((-0.3 \leq \text{opinion} \leq 0.3))$,
- right-oriented agents $((\text{opinion} > 0.3))$.

The **five-group segmentation** refines this structure by distinguishing soft and hard positions on each side:

- hard left ((opinion < -0.7)),
- soft left ((-0.7 <= opinion <= -0.3)),
- center,
- soft right ((0.3 < opinion <= 0.7)),
- hard right ((opinion > 0.7)).

These segmentations are applied consistently across empirical and synthetic conditions, enabling direct comparison of volatility, inversion rates, and cognitive imbalance across ideological zones.

3.5 Experimental Conditions (Exp1–Exp5)

Five experimental conditions were constructed to disentangle the respective roles of empirical structure, randomness, and exogenous meme injection:

- **Exp1:** Empirical distribution and network (survey-based), with targeted meme injection affecting agents with $(-0.3 \leq \text{opinion} \leq 0.3)$.
- **Exp2:** Randomized distribution and random network, without meme injection.
- **Exp3:** Randomized distribution and network, with the same targeted meme injection as Exp1.
- **Exp4:** Empirical distribution and network, with broad meme injection affecting the full opinion range.
- **Exp5:** Randomized distribution and network, with broad meme injection.

Each condition was replicated across multiple runs to assess robustness and variability, producing ensembles of trajectories for statistical analysis.

3.6 Outputs and Data Collection

At each tick, the simulator records both aggregate indicators and agent-level states. Aggregate outputs include opinion distributions, prevalence statistics, inversion counts, network dynamics, and meme-based indicators such as Wgap and saturation. Agent-level exports record opinion, prevalence, influence, meme stocks, and weighted meme values for each agent at selected iterations.

These data form the basis for subsequent statistical analyses, including mediation models, logistic regressions of inversion probability, and comparative evaluations across segmentation schemes and experimental conditions.

4) Methods

This section describes segmentation (3 and 5 groups), metrics (Wgap, volatility, static/dynamic relational distance), and statistical models (mediation and logit).

4.1 Segmentation Schemes of the Opinion Space

Analyses are conducted using two complementary segmentation schemes of the opinion space, applied consistently across all experimental conditions and simulation runs.

The **three-group segmentation** distinguishes agents according to their expressed opinion at initialization:

- **Left-oriented agents:** ($\text{opinion} < -0.3$)
- **Centrist agents:** ($-0.3 \leq \text{opinion} \leq 0.3$)
- **Right-oriented agents:** ($\text{opinion} > 0.3$)

This coarse partition isolates the ideological center as a distinct analytical category and serves as the primary lens for testing hypotheses about volatility and inversion dynamics.

The **five-group segmentation** refines this structure by distinguishing soft and hard positions on each side of the spectrum:

- **Hard left:** ($\text{opinion} < -0.7$)
- **Soft left:** ($-0.7 \leq \text{opinion} < -0.3$)
- **Center:** ($-0.3 \leq \text{opinion} \leq 0.3$)
- **Soft right:** ($0.3 < \text{opinion} \leq 0.7$)
- **Hard right:** ($\text{opinion} > 0.7$)

This finer segmentation allows for a more granular assessment of how ideological anchoring modulates internal cognitive dynamics and behavioral stability.

4.2 Opinion Inversions and Volatility

An **opinion inversion** is defined as a sign change in an agent's expressed opinion over time, i.e., a transition from positive to negative polarity or vice versa. For each agent (i), the total number of inversions across a simulation run is recorded.

Volatility is measured as the **temporal variability of opinion**, operationalized as the standard deviation of the agent's opinion trajectory across ticks. While inversions capture discrete allegiance switches, volatility captures continuous oscillations that may occur within the same ideological camp.

These two indicators are treated as analytically distinct but complementary dimensions of instability.

4.3 Meme-Based Cognitive Imbalance (Wgap)

The **Wgap** indicator measures the imbalance between weighted meme stocks within an agent's cognitive state. For agent (i) at time (t), Wgap is defined as:

$$Wgap_{i,t} = \frac{|meme-plus_{w_{i,t}} - meme-minus_{w_{i,t}}|}{meme-plus_{w_{i,t}} + meme-minus_{w_{i,t}} + \epsilon}$$

where (ϵ) is a small constant preventing division by zero.

Wgap ranges from 0 (perfect balance between opposing representations) to 1 (complete dominance of one side). It captures the internal asymmetry of the agent's representational system and is interpreted as a proxy for cognitive polarization at the individual level.

****Mediation****: Formal test "distance \rightarrow Wgap \rightarrow inversions". We estimate (a) the effect of distance on Wgap (path a), (b) the effect of Wgap on reversals by controlling distance (path b), and (c) the residual direct effect of distance on reversals (c'). The intervals are obtained by bootstrapping.

Logit: logistic model of the probability of at least one inversion (0/1) from Wgap, dynamic distance, volatility, prevalence, and control terms (experiment, group). Marginal effects are reported in the revised format.

4.4 Relational Distance: Static and Dynamic Proxies

To operationalize **relational distance**, two complementary proxies are employed.

The **static relational distance** is computed from the initial network structure. Each agent is connected to a fixed number of neighbors; the absolute difference between the agent's opinion and the mean opinion of its neighbors provides a static measure of ideological distance embedded in the network.

The **dynamic relational distance** extends this concept by recomputing, at each tick, the distance between an agent's current opinion and the contemporaneous opinions of its neighbors. This dynamic proxy captures the co-evolution of opinions and network exposure and reflects how agents drift away from—or converge toward—their local relational environment over time.

Both measures are normalized to ensure comparability across runs and experimental conditions.

4.5 Mediation Analysis: Distance \rightarrow Wgap \rightarrow Inversions

A central hypothesis of the study is that relational distance affects opinion inversions **indirectly**, through its impact on internal cognitive imbalance.

This hypothesis is tested using a **formal mediation framework** comprising three equations:

1. **Mediator model:**

$$Wgap_i = \alpha_0 + \alpha_1 Distance_i + \mathbf{X}_i\alpha + \varepsilon_i \quad [$$

2. **Outcome model:**

$$Inversions_i = \beta_0 + \beta_1 Wgap_i + \beta_2 Distance_i + \mathbf{X}_i\beta + \eta_i$$

3. **Total effect model:**

$$Inversions_i = \gamma_0 + \gamma_1 Distance_i + \mathbf{X}_i\gamma + \nu_i$$

where \mathbf{X}_i includes control variables such as prevalence and influence.

Indirect effects are estimated using nonparametric bootstrap procedures to assess statistical significance.

4.6 Logistic Models of Inversion Probability

To complement count-based analyses, inversion dynamics are also examined **using logistic regression models**, where the dependent variable indicates whether an agent experiences at least one inversion during a run.

Predictors include Wgap, relational distance (static or dynamic), prevalence, and segmentation group. Marginal effects are computed to facilitate interpretation and comparison across experimental conditions.

4.7 Monte Carlo Replication and Robustness

All analyses are conducted over ensembles of simulation runs. For each experimental condition (Exp1–Exp5), multiple independent runs are generated using identical parameter settings but different random seeds.

Results are summarized using means, confidence intervals, and distributional comparisons to assess robustness. Monte Carlo aggregation allows distinguishing structural effects from stochastic fluctuations and supports comparative inference across experimental designs.

4.8 Link to NetLogo Procedures

All indicators and transformations described above are computed directly from the simulator's state variables. In particular, the recomposition of opinion from meme stocks relies on the NetLogo procedure `recompute-from-memes`, which enforces the mapping between internal representations and expressed stance.

This explicit linkage ensures that statistical analyses remain grounded in the generative mechanisms of the agent-based model.

5) Comparative results (Exp1–Exp5)

This section presents a systematic comparison of the five experimental conditions (Exp1–Exp5), focusing on three key outcome variables: **opinion inversions**, **cognitive imbalance (Wgap)**, and **relational distance (static and dynamic)**. Results are reported using both the three-group and five-group segmentations of the opinion space.

Table 1. Summary of the overall indicators by experience (averages).

experiment	mean_inversions	pct_any_inversion	mean_volatility	mean_wgap	mean_dyn_dist	mean_prevalence
Exp1-1	1.52	31.4	0.1542	0.1226	0.2029	53.57
Exp2-1	0.272	16.2	0.08205	0.09798	0.301	24.31
Exp3-1	0.23	14.2	0.08302	0.1004	0.2907	24.77
Exp4-1	0.238	12.6	0.08799	0.09328	0.166	49.4
Exp5-1	0.192	12.4	0.08483	0.09557	0.2873	24.01

5.1 Overview of Experimental Conditions

The five experimental conditions differ along two main dimensions:

(i) the initial distribution of agents (empirical vs. random), and

(ii) the scope of meme injection.

- **Exp1:** Empirical distribution (Survey 2022), meme injection targeting centrist agents ($-0.3 \leq \text{opinion} \leq 0.3$).
- **Exp2:** Random distribution, no meme injection.
- **Exp3:** Random distribution, meme injection targeting centrist agents.
- **Exp4:** Empirical distribution, meme injection applied across the full opinion range (-1 to $+1$).
- **Exp5:** Random distribution, meme injection applied across the full opinion range.

This design allows disentangling the effects of empirical anchoring, targeting strategy, and stochastic structure.

5.2 Opinion Inversions Across Segmentations

Figure 1 reports the mean number of opinion inversions per agent, computed under both the three-group and five-group segmentations.

Across all experimental conditions, **centrist agents exhibit the highest inversion rates**. This pattern is particularly pronounced in Exp1 and Exp3, where meme injection is explicitly targeted at the center of the opinion space.

In contrast, agents located at the ideological extremes (hard left and hard right) display very low inversion frequencies. While these agents may experience opinion fluctuations, such oscillations remain confined within their initial polarity and do not result in sign reversals.

Synthesis figures (Exp1–Exp5).

Fig. 1 — Opinion sign inversions across experimental conditions (Exp1–Exp5)

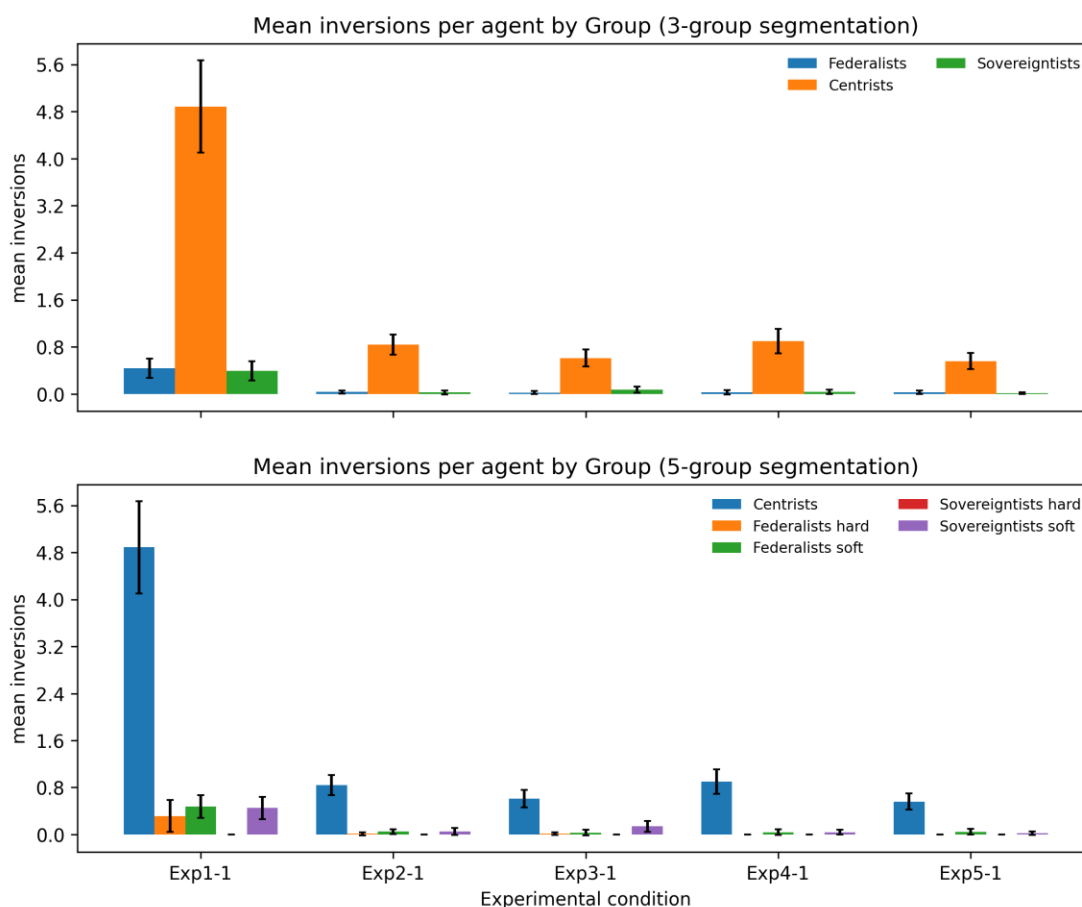


Fig. 1. Mean inversions by group (segmentation 3 and 5), Exp1–Exp5 comparison.

Fig. 2 — Mean Wgap ($|\text{opinion} - \text{meme-derived opinion}|$) across conditions (Exp1–Exp5)

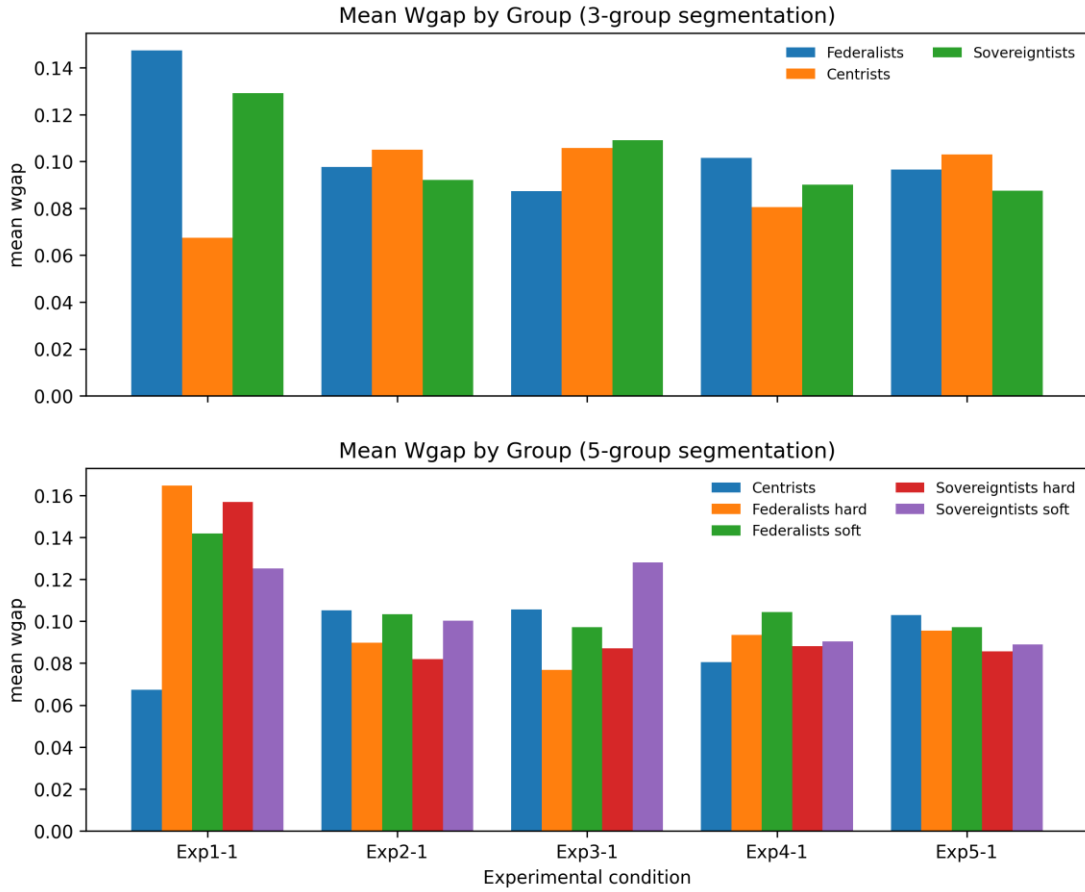


Fig. 2. Average Wgap per group (segmentation 3 and 5), Exp1–Exp5 comparison.

5.3 Cognitive Imbalance (Wgap) by Group

Figure 2 presents average Wgap values by segmentation group for each experimental condition.

Wgap values are consistently lowest among centrist agents, indicating a more balanced internal composition of opposing meme representations. This balance coexists with high behavioral instability, as evidenced by elevated inversion and volatility measures.

Conversely, hard ideological groups exhibit high Wgap values, reflecting strong dominance of one meme polarity over the other. This internal asymmetry corresponds to greater cognitive anchoring and reduced likelihood of opinion reversal.

Comparisons across experiments show that meme injection increases Wgap primarily in the targeted population. In Exp1 and Exp3, this effect is concentrated in the center, whereas in Exp4 and Exp5, Wgap increases across all groups.

Fig. 3 — Dynamic relational distance vs opinion volatility (agent-level, Exp1-Exp5)

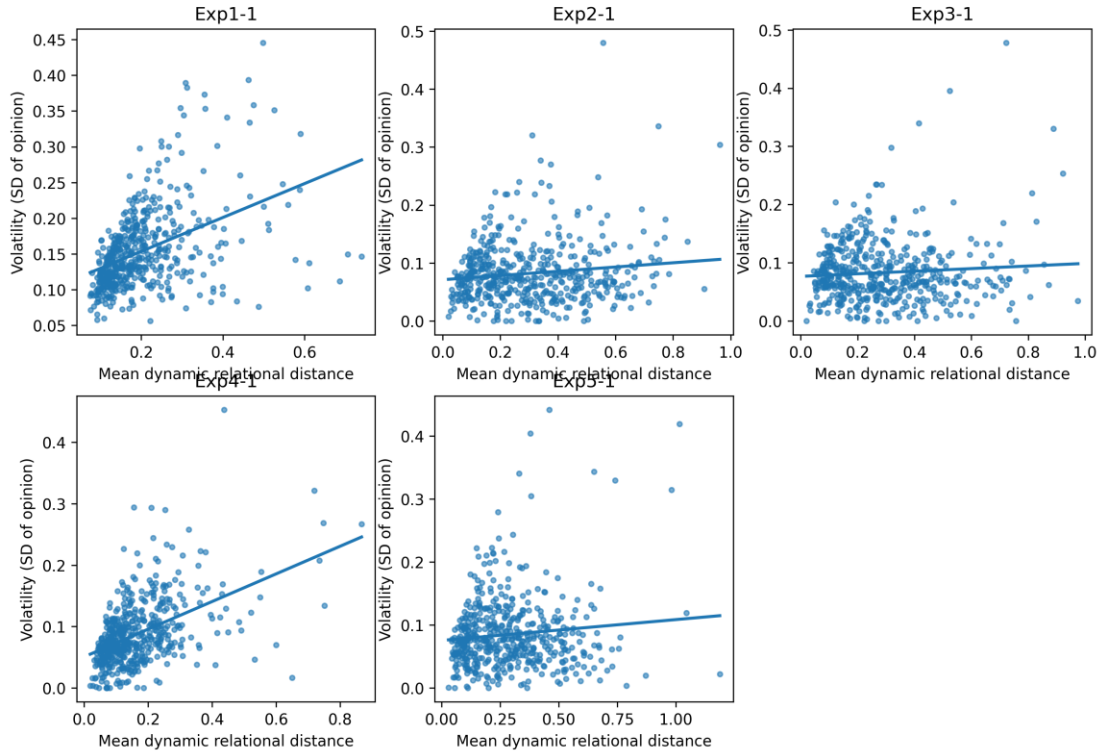


Fig. 3. Dynamic relational distance vs volatility (Exp1–Exp5 panels).

5.4 Static vs. Dynamic Relational Distance

Figure 3 plots the relationship between **dynamic relational distance** and **opinion volatility** across agents and experimental conditions.

Dynamic distance is strongly correlated with volatility, particularly in Exp1 and Exp3. Agents whose relational environment diverges over time from their own opinion trajectory exhibit larger oscillations and higher probabilities of inversion.

Static relational distance, by contrast, shows a weaker association with behavioral outcomes, suggesting that **relational misalignment is not a fixed property of network topology**, but an emergent and time-dependent process.

5.5 Mediation Results: Distance → Wgap → Inversions

Formal mediation analyses confirm that the effect of relational distance on opinion inversions is **partially mediated by Wgap**.

In all experimental conditions involving meme injection (Exp1, Exp3, Exp4, Exp5), the indirect path through Wgap is statistically significant. The direct effect of distance on inversions is substantially reduced when Wgap is included in the model, indicating that **internal cognitive imbalance constitutes a key transmission channel** between social exposure and behavioral change.

In Exp2 (random distribution, no injection), mediation effects are weaker and less consistent, underscoring the role of structured informational inputs in activating this mechanism.

5.6 Cross-Experimental Robustness

Despite differences in initialization and intervention scope, several robust patterns emerge:

1. Centrist agents consistently exhibit the highest instability (inversions and volatility).
2. High Wgap is associated with ideological rigidity and reduced inversion probability.
3. Dynamic relational distance outperforms static distance as a predictor of instability.
4. Meme injection amplifies existing structural tendencies rather than creating them de novo.

These regularities persist across Monte Carlo replications and segmentation schemes, supporting their interpretation as structural properties of the modeled system.

6) Discussion

This section interprets the comparative results (Exp1–Exp5) in terms of **generative mechanisms, robustness, limitations, and theoretical implications** for agent-based modeling, NetLogo simulations, and meme-based cognition.

6.1 Generative Mechanisms of Opinion Change

The results consistently indicate that opinion change is not driven directly by exposure to opposing opinions, but by the **interaction between relational distance and internal cognitive structure**.

Relational distance—particularly in its dynamic formulation—acts as a destabilizing force only insofar as it produces or amplifies **internal imbalance between competing meme representations**, as captured by Wgap. Opinion inversions occur when agents are simultaneously exposed to heterogeneous social signals and maintain a relatively balanced internal meme ecology.

In this sense, **Wgap functions as a cognitive gate**:

- low Wgap enables opinion reversals,
- high Wgap inhibits them.

This mechanism explains why centrist agents, who tend to accumulate both pro and contra memes in comparable proportions, exhibit high volatility and frequent inversions, while ideologically extreme agents remain stable despite ongoing interactions.

6.2 Centrist Instability and Ideological Rigidity

Across all experimental conditions, centrist agents occupy a structurally unstable position. Their internal meme composition remains pluralistic, preventing strong anchoring, while their relational environment exposes them to competing influences.

In contrast, agents at ideological extremes accumulate highly asymmetric meme stocks. This asymmetry translates into high Wgap values, reinforcing cognitive rigidity. Even when these agents experience opinion fluctuations, such variations occur **within the same polarity** and rarely result in sign inversions.

These findings suggest that polarization in the model is not merely a consequence of homophily or selective exposure, but emerges from **self-reinforcing cognitive asymmetries**.

6.3 Robustness Across Experimental Conditions

The comparison of Exp1–Exp5 demonstrates that the observed mechanisms are robust to variations in initial conditions and intervention strategies.

- Empirical vs. random initialization affects the distribution of outcomes but not their qualitative structure.
- Meme injection intensifies existing dynamics without fundamentally altering them.
- Dynamic relational distance consistently outperforms static distance as a predictor of volatility and inversion.

Monte Carlo replications confirm that these patterns are not artifacts of specific realizations but reflect stable properties of the modeled system.

6.4 Limits of the Model

Several limitations must be acknowledged.

First, relational distance is proxied using agent indices, which encode ideological ordering by construction. While this proxy is defensible given the empirical calibration, it abstracts away from richer semantic or multidimensional opinion spaces.

Second, meme representations are modeled as scalar quantities and weights, whereas real cognitive representations are likely structured, contextual, and hierarchically organized.

Third, the model does not explicitly represent institutional media, strategic actors, or endogenous agenda-setting mechanisms, which may interact with meme dynamics in real-world contexts.

6.5 Implications for ABM, NetLogo, and Meme-Based Cognition

From an agent-based modeling perspective, these results highlight the importance of **internal state variables** that mediate between social interaction and observable behavior. Opinion change cannot be adequately modeled as a direct function of neighbor influence alone.

For NetLogo practitioners, the explicit separation between **interaction mechanisms** and **cognitive update procedures**—as implemented through recompute-from-memes—offers a modular architecture that facilitates theoretical transparency and empirical calibration.

More broadly, the findings support a conception of memes not merely as transmissible units, but as internal cognitive resources whose balance and accumulation condition behavioral flexibility. Opinion dynamics thus emerge from the coupling of network processes and internal representational structures.

7. Conclusion

This article examined opinion dynamics through a systematic comparison of five experimental conditions (Exp1–Exp5), combining empirically grounded initialization, agent-based simulation, and a meme-based cognitive architecture.

Across all conditions, the results show that **opinion change, volatility, and ideological realignment are primarily driven by internal cognitive structure rather than by social exposure alone**. Relational distance—whether static or dynamically updated—constitutes a necessary condition for opinion change, but it is not sufficient by itself. Its effects are consistently mediated by the internal balance between competing meme representations, captured by the Wgap indicator.

Agents located in the central region of the opinion space emerge as the most volatile. Their position combines heterogeneous social exposure with a relatively balanced internal meme ecology, producing maximal susceptibility to opinion inversions. By contrast, agents at ideological extremes exhibit strong internal asymmetries in meme stocks, which stabilize their opinions

despite ongoing interactions and occasional within-camp fluctuations. These agents may oscillate in intensity, but they rarely switch ideological camp.

The comparison between empirically initialized populations (based on the 2022 survey) and randomly initialized populations shows that **initial distributions modulate the magnitude of observed effects without altering the underlying generative mechanisms**. Meme injection amplifies polarization or destabilization depending on its targeting, yet the same cognitive and relational processes operate across all experimental conditions. Monte Carlo replications further confirm the robustness of these results.

Methodologically, this work underscores the importance of modeling **opinions as emergent outcomes of internal cognitive representations**, rather than as directly updated state variables. Treating opinions as the result of weighted meme accumulation provides a more realistic account of belief formation, persistence, and change, while also enabling explicit links between micro-level cognition and macro-level collective patterns.

More broadly, these results contribute to complexity science by illustrating how macro-level polarization patterns can arise from micro-level interactions between network structure and internal cognitive states. They also suggest that political instability and realignment are not anomalies of the system, but intrinsic properties of populations characterized by heterogeneous exposure and cognitively pluralistic agents.

From a complexity science perspective, the findings illustrate how large-scale polarization and instability can arise from simple local interactions when internal cognitive states are taken into account. Opinion volatility and realignment are not anomalies of the system, but intrinsic properties of populations characterized by heterogeneous exposure and cognitively pluralistic agents.

Future research may extend this framework by incorporating multidimensional opinion spaces, institutional actors, and adaptive media environments. Nonetheless, the present results already demonstrate that **meme-based cognition offers a powerful lens** for understanding the dynamics of opinion formation, stability, and transformation in complex social systems.

Finally, this study demonstrates the analytical value of integrating meme-based cognition into agent-based models implemented in NetLogo. The framework provides a flexible and theoretically grounded platform for studying political polarization, belief dynamics, and ideological change in complex social systems.

8) Methodological appendix

This appendix details the formal definitions, equations, segmentation rules, and simulation procedures underlying the analyses presented in the main text. Its objective is to ensure full transparency and reproducibility, and to clarify the explicit links between the empirical indicators, the statistical models, and the NetLogo implementation.

8.1 Opinion Segmentation Schemes

Two segmentation schemes are used throughout the analyses.

Three-group segmentation

Agents are classified according to their opinion value $o_i \in [-1, 1]$:

- **Left (Federalist):** $o_i < -0.3$
- **Center (Centrist):** $-0.3 \leq o_i \leq 0.3$
- **Right (Sovereigntist):** $o_i > 0.3$

This segmentation isolates the central ideological region where opinion reversals are most likely to occur.

Five-group segmentation

A finer ideological partition is also applied:

- **Soft Left:** $-0.7 \leq o_i < -0.3$
- **Hard Left:** $o_i < -0.7$
- **Center:** $-0.3 \leq o_i \leq 0.3$
- **Soft Right:** $0.3 < o_i \leq 0.7$
- **Hard Right:** $o_i > 0.7$

This segmentation allows for the identification of asymmetric polarization and differential stability within ideological camps.

8.2 Meme-Based Opinion Reconstruction

Opinions are not updated directly. Instead, they are derived from internal meme stocks.

Each agent i holds:

- M_i^+ : cumulative weighted pro-memes (meme-plus-w)
- M_i^- : cumulative weighted contra-memes (meme-minus-w)

The meme-derived opinion is defined as:

$$\hat{o}_i = \begin{cases} \frac{M_i^+ - M_i^-}{M_i^+ + M_i^-} & \text{if } M_i^+ + M_i^- > 0 \\ 0 & \text{otherwise} \end{cases}$$

This formulation ensures that:

- opinion polarity reflects the balance of representations,
- opinion magnitude reflects internal asymmetry,
- opinions are undefined (neutral) when no memes are present.

8.3 Wgap: Internal Cognitive Imbalance

The **Wgap** indicator captures the internal asymmetry between competing meme representations:

$$\text{Wgap}_i = \frac{|M_i^+ - M_i^-|}{M_i^+ + M_i^-}$$

Wgap ranges from 0 (perfectly balanced internal representations) to 1 (complete dominance of one side).

Conceptually:

- low Wgap \rightarrow cognitive pluralism and openness,
- high Wgap \rightarrow ideological entrenchment and resistance to change.

Wgap acts as a mediator between social exposure and opinion stability.

8.4 Relational Distance

Two measures of relational distance are employed.

Static relational distance

For agent i with neighbors $j \in \mathcal{N}_i$:

$$D_i^{\text{static}} = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} |o_i - o_j|$$

This measure captures ideological heterogeneity in the local network.

Dynamic relational distance

Dynamic distance incorporates temporal evolution:

$$D_i^{\text{dynamic}}(t) = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} |o_i(t) - o_j(t)|$$

This allows distance to co-evolve with opinion change and network dynamics, capturing feedback loops between exposure and belief updating.

8.5 Volatility and Inversions

Opinion volatility

Volatility is defined as the mean absolute change in opinion between successive ticks:

$$V_i = \frac{1}{T-1} \sum_{t=2}^T |o_i(t) - o_i(t-1)|$$

Opinion inversions

An inversion occurs when the sign of an agent's opinion changes:

$$\text{Inversion}_i(t) = \begin{cases} 1 & \text{if } \text{sign}(\downarrow(o_i(t))) \neq \text{sign}(o_i(t-1)) \\ 0 & \text{otherwise} \end{cases}$$

The total number of inversions per agent is used as a primary indicator of ideological realignment.

8.6 Statistical Models

Mediation analysis

Mediation tests follow the structure:

$$\text{Distance} \rightarrow \text{Wgap} \rightarrow \text{Inversions}$$

We estimate:

1. the effect of distance on Wgap,
2. the effect of Wgap on inversions,
3. the direct effect of distance on inversions controlling for Wgap.

Significant reduction of the direct effect supports a mediation mechanism.

Logistic regression

Inversion probability is modeled as:

$$\Pr(\text{Inversion}_i = 1) = \text{logit}^{-1}(\beta_0 + \beta_1 \text{Wgap}_i + \beta_2 D_i + \beta_3 \text{Prevalence}_i)$$

Marginal effects are computed to assess relative contributions.

8.7 NetLogo Implementation

All meme-based updates are centralized in the NetLogo procedure:

recompute-from-memes

This procedure:

1. updates prevalence from total meme stock,
2. recomputes opinion from weighted meme balance,
3. enforces bounds on opinion and prevalence,
4. synchronizes internal cognition with observable states.

Crucially, no procedure directly sets opinions; all opinion change is emergent from meme dynamics.

8.8 Reproducibility and Data Sources

Analyses rely on:

- survey-based initialization (2022 empirical distribution),
- randomized initialization (Monte Carlo baselines),
- CSV exports of agent states (Values mode),
- repeated runs with identical parameter settings across experimental conditions.

All indicators are computed post hoc from exported agent-level data, ensuring full separation between simulation dynamics and statistical analysis.