When Talk is Not Enough: Social Context and the Persistence of Ideological Polarization

İrem Betül Koçak,* Isaac D. Mehlhaff,† and Başak Taraktaş‡

July 2, 2026

Abstract

Scholars have long recognized the promise of interpersonal interaction in promoting political opinion change. We argue, however, that theories of persuasion have paid insufficient attention to the broader social context in which interactions occur. The potential for interaction to produce opinion change depends not only on micro-level mechanisms, but also on macro-level features like social network structure and the distribution of opinions in the population. We build an agent-based model of repeated encounters among individuals embedded in social networks. We find that more frequent interaction increases opinion updating, but larger social networks can dampen this effect by concentrating citizens in ideological silos. Macro-level opinion distributions also condition micro-level change; highly polarized environments produce large but infrequent updates, whereas moderate environments produce smaller but more frequent changes. The potential for political talk to encourage opinion change in polarized contexts may be limited.

*PhD Student, Department of Industrial Engineering, Boğaziçi University, irembetulkocak@gmail.com

†Neubauer Family Assistant Professor, Department of Political Science, The University of Chicago, imehlhaff@uchicago.edu

‡Associate Professor, Department of Political Science and International Relations, Boğaziçi University, basak.taraktas@boun.edu.tr.

By nearly every measure—partisan hostility, ideological distance, policy disagreement—the United States is more deeply divided than it has been in decades (Abramowitz, 2010; Iyengar and Westwood, 2015). Similar stories are playing out across established democracies, with mutual distrust and ideological division becoming defining features of political life (Mehlhaff, 2025; Reiljan, 2020). Yet at the same time, Americans are more connected today than at any point in modern history. Social media platforms, round-the-clock news cycles, and high geographic mobility have multiplied opportunities for interpersonal interaction across lines of political difference.

For many years, scholars and practitioners have offered a simple prescription to stem the tide of ideological polarization: get people talking to each other. Whether in structured intergroup contact programs (Blair et al., 2025) or more casual encounters (Rossiter and Carlson, 2024), talking across party lines has been held up as a solution to political polarization (Becker, Porter, and Centola, 2019; Druckman and Nelson, 2003; Esterling, Fung, and Lee, 2021). An increasingly influential theoretical framework in cognitive psychology suggests individuals are more likely to change their opinions when they discuss issues with others, especially those with different views (Altay et al., 2023; Mercier and Sperber, 2017). To the extent that interaction promotes exposure to countervailing arguments and perspectives, this theory implies that micro-level opinion updating during interpersonal encounters—when aggregated to the macro-level—is key to reducing polarization.[1]

But if conversation is the antidote to division, why has division continued to intensify as opportunities for interaction have proliferated? We argue that the relationship between interpersonal interaction and opinion change is more complex than commonly assumed. Citizens' experiences with politics cannot be boiled down to a series of isolated conversations, like those common in controlled experiments. Rather, citizens are embedded in complex social environments that shape the interactions they can have, their openness to future interactions and, ultimately, their ideas on the issues of the day (Huckfeldt and Sprague, 1995; Mutz, 2006). Partially for these reasons, real-

[1]Existing work has thoroughly examined how meso-level institutions like political parties contribute to polarization (Akboga, Sahin, and Arik, 2025; Barber and Pope, 2019; Druckman, Peterson, and Slothuus, 2013). While we touch on aspects of party messaging below, our primary objective is to understand how citizens can produce change at the micro-level.

world interactions are not often cross-cutting. Instead, they occur in homogeneous environments where people mostly talk to friends and family who already share their viewpoints (Barberá et al., 2015; McPherson, Smith-Lovin, and Cook, 2001). In such settings, interaction may reinforce existing beliefs rather than challenge them (Bishin et al., 2016; Sunstein, 2017).

We instead propose that the relationship between interpersonal interaction and opinion dynamics is contingent on both contextual and individual-level attributes (Druckman, 2004), including the influence of messaging from political elites. The psychological mechanisms by which individuals process disagreement, the structure of individuals' social networks, and the distribution of opinions in the larger environment all work together to shape the outcomes of political talk. That is, micro-level attitude changes are partially determined by the very macro-level trends they produce.

We build an agent-based model to isolate how the structure and frequency of social encounters shape the likelihood of opinion updating and thus the potential for interaction to reduce ideological polarization. By simulating opinion dynamics under different interaction regimes, we show that opinion change is more common when interaction is frequent and one's social network is small. More frequent interaction with others suggests more opportunities to interact across lines of political difference, increasing the likelihood that individuals revise their beliefs away from the extremes. At the same time, however, larger social networks act as a countervailing force. More social ties implies more interaction among trusted friends and family, which can discourage attitude change by decreasing individuals' exposure to diverse ideas. Messaging from political elites can encourage additional persuasion among individuals, but only when the population is not already highly polarized.

Further, we show the frequency and magnitude of updating also depends on features of the initial opinion distribution. When public opinion is highly polarized, individuals experience large but infrequent shifts in opinion. When public opinion is more moderate, individual opinion is more malleable but less likely to change drastically as a result of interaction. This pattern of results suggests that, all else equal, scholars' prescription for calming ideological divisions may hold some promise, but it may also be thwarted by the high levels of polarization and connectivity that characterize twenty-first century politics.

We make two main contributions to the literature on attitude polarization. Substantively, we demonstrate how the path from interpersonal interaction to opinion change is shaped not only by micro-level psychological mechanisms, but also by the broader political environment in which citizens are embedded. Existing literature on political persuasion and partisanship has fruitfully considered the former source of variation but tends to overlook the latter. We demonstrate their joint importance for both the frequency and magnitude of opinion change. Methodologically, we show how agent-based modeling—an under-used tool in the study of political behavior—can profitably contribute to understanding complex social phenomena (e.g. Jackson et al., 2017). Our models allow us to systematically test macro-level implications of micro-level theories of opinion updating, examine the long-run impact of repeated interaction, and manipulate contextual conditions—all of which are difficult to accomplish with the observational or experimental methods most commonly used in political science and psychology.

We begin by outlining our expectations regarding the effect of interpersonal interaction on political attitudes, drawing on recent theoretical advances in cognitive psychology. We then describe our approach to agent-based modeling and outline the contributions it can help scholars make to the study of political behavior, despite being a rarely used research design in this literature. The subsequent sections articulate the specifics of our model, the variables we use to capture the model's implications for micro-level attitudes, and our strategies for statistical analysis. We show the effects of interpersonal interaction on both the rate and magnitude of attitude change under three theoretical updating mechanisms: a scenario in which agents are rational Bayesian updaters, one in which they can experience backlash to counter-attitudinal information, and one in which they receive communications from political elites. We close by discussing the implications of our findings for a collection of literatures throughout the social sciences.

Theory and Expectations

Political scientists have long recognized that ordinary citizens learn about and form opinions on political issues largely through conversation with family, friends, neighbors, and coworkers (Berelson, Lazarsfeld, and McPhee, 1954; Campbell et al., 1960). This form of talk is casual, episodic, and driven by social interests more than instrumental ones (Carlson and Settle, 2022). Yet it is precisely through these everyday exchanges that citizens encounter perspectives different from their own, receive social signals about the acceptability of various opinions, and occasionally revise their beliefs. A growing body of experimental and observational evidence suggests that cross-cutting political talk can reduce ideological polarization, increase awareness of competing rationales, and, in some cases, produce convergence in policy preferences (Becker, Porter, and Centola, 2019; Druckman and Nelson, 2003; Esterling, Fung, and Lee, 2021; Mutz, 2006).

Cognitive psychology provides a micro-level account of why such interactions can produce opinion change. The interactionist theory of reasoning treats human cognition as fundamentally social, optimized for persuasion in interpersonal exchange (Mercier and Landemore, 2012; Mercier and Sperber, 2017). In this view, opinion updating is an active, relational process: Citizens talk, question, challenge, and justify, and through these exchanges they encounter countervailing reasons, detect weaknesses in their own arguments, and occasionally revise their views (Altay et al., 2023; Trouche, Sander, and Mercier, 2014). Environments that regularly expose individuals to disagreement therefore increase exposure to opposing arguments and should provide the best opportunities for belief updating (Schneiderhan and Khan, 2008; Westwood, 2015).

Whether these micro-level processes produce meaningful opinion change, however, depends on the social environments in which they unfold. Citizens tend to cluster in politically homogeneous circles with family and close friends (Carlson and Settle, 2022). In these settings, individuals seldom encounter dissent despite interacting frequently (Barberá et al., 2015; Flache and Macy, 2011; Mäs and Flache, 2013; Sunstein, 2017). By contrast, weak ties and more heterogeneous environments should expand opportunities for cross-cutting interaction, increasing the likelihood that individuals encounter arguments that challenge their views (Granovetter, 1973; Mutz, 2006).

We add yet another layer to the story: Interpersonal dynamics must be examined in light of the broader communication environment in which they occur. Much of the psychological evidence on reasoning focuses on isolated dyadic interactions, but real-world opinion formation unfolds within complex flows of information over extended periods of time (Huckfeldt and Sprague, 1995). In particular, citizens often engage in a "two-step flow" of communication, in which politically attentive individuals receive information from media or elites and transmit it through interpersonal networks (Katz and Lazarsfeld, 1955; Nisbet and Kotcher, 2009; Norris and Curtice, 2008). Because citizens rely on partisan cues to interpret political information (Lenz, 2012; Zaller, 1992), elite messaging can shape the content of interpersonal discussion, which in turn influences how opinions evolve through social interaction. Interpersonal reasoning and elite cue-taking are thus complementary processes, jointly structuring the informational inputs individuals bring to political talk.

Our theoretical expectations build on these insights in three ways. First, we treat argument exchange as the proximate micro-mechanism of opinion change; individuals update when exposed to reasons that withstand scrutiny. Second, we situate this mechanism within structured social networks, where homophily and clustering constrain exposure to disagreement while weak ties and heterogeneous environments expand it. Third, we extend these dynamics to the macro level, proposing that environments generating more frequent and diverse interaction should yield more frequent and larger opinion changes and, in the aggregate, lower polarization.

Crucially, however, the macro-level distribution of opinion is not merely an outcome of these processes——it also shapes the conditions under which they operate. When opinion is highly polarized and concentrated at the extremes, cross-cutting encounters carry substantial updating potential; the ideological distance between individuals is large, creating strong pressure for revision. Yet such encounters are least likely to occur in precisely these settings, as homophily and network clustering limit exposure to disagreement (Barberá et al., 2015; McPherson, Smith-Lovin, and Cook, 2001). The result is a pattern of infrequent but potentially large individual updates. Conversely, when the population is more moderate and evenly distributed, cross-cutting encounters are more routine and updating is more frequent. However, because the ideological distance between interlocutors is smaller to begin with, each individual update is likely to be more modest. The macro-level distribution of opinion and the micro-level process of updating therefore work together to shape the impact of interpersonal interaction on opinion change.

With these constraints in view, we expect that settings with more interpersonal interaction will produce more frequent opinion updating, as individuals face more opportunities for argument exchange. When such interaction includes exposure to substantively different viewpoints, the magnitude of aggregate change should also increase, reflecting both incremental revisions and occasional larger shifts. However, interaction can also reinforce existing beliefs when social structure suppresses cross-cutting exposure or when individuals hold strongly entrenched views. Our empirical strategy therefore evaluates how context conditions the relationship between interpersonal interaction, opinion change, and polarization. We show below that greater interaction does not uniformly produce more updating, and that both the frequency and magnitude of opinion change depend on features of the initial opinion distribution.

Studying Political Behavior with Agent-Based Models

Investigating these intricate social dynamics and their impact on opinion change in the real world presents methodological challenges for observational and experimental designs—the methods most often relied upon in public opinion and psychology research. Observational studies, while valuable for identifying general relationships and broad trends, typically only provide a snapshot of opinion at a single point in time or, at most, a handful of points in a time series. This makes it difficult to isolate the influence of particular social features—such as the frequency, content, or emotional tone of discussions—on individual opinion change and subsequent macro-level patterns. Similarly, while experiments offer greater control, they simplify social interactions to a degree that may not fully capture their multifaceted nature. Researchers' capacity to replicate the spontaneity, nuanced communication, and evolving relationships that characterize real-world opinion exchanges within a controlled laboratory setting is inherently limited.

Instead, we use agent-based modeling (ABM) to overcome these limitations. By creating computational simulations of interacting individuals governed by clearly defined rules, ABM allows us to systematically explore how micro-level interaction dynamics give rise to macro-level opinion phenomena (Epstein, 1999; Macy and Willer, 2002). This *in silico* laboratory enables the controlled manipulation of interaction parameters, such as network structure, agent characteristics (e.g. opinion strength and openness to interaction), and the rules governing opinion updating during encounters. Scholars have used ABM to study important psychological phenomena such as motivated cognition (Dandekar, Goel, and Lee, 2013), political discussion networks (Huckfeldt, Johnson, and Sprague, 2004), the formation of group identity (Gray et al., 2014), and the emergence of cooperation (Bear and Rand, 2016), but it remains severely under-utilized relative to the survey- and experiment-based methods that are common currency in political science and psychology (Jackson et al., 2017).

ABM is especially useful for showing how micro-level psychological mechanisms—such as biased information processing, emotional asymmetries, value-driven updating, or social influence—scale into macro-level patterns of polarization, stability, and rapid opinion shifts (Banisch and Shamon, 2024; Sobkowicz, 2016; Weinans et al., 2024; Zhu et al., 2022). Work on these subjects tend to emphasize emotional or informational processes (Dandekar, Goel, and Lee, 2013; Deffuant et al., 2000; Mäs and Flache, 2013). ABM allows us to explore counterfactual scenarios, test theoretical propositions, and manipulate contextual conditions independent from cognitive dynamics, making it suited to evaluating the causal role of interpersonal interaction in opinion change (Deffuant et al., 2000; Epstein, 1999; Flache and Macy, 2011).

For instance, ABM can help us understand how varying levels of homophily in social networks might either foster or hinder opinion polarization, or how different communication strategies might influence the spread and persistence of misinformation across a population—scenarios that are challenging to study comprehensively through observation or controlled experiments alone. As such, our modeling approach can provide valuable, novel insights into the complex and often non-linear relationships between interpersonal interaction and opinion dynamics.

Model Description

We develop an agent-based model to study how individuals update their opinions on a single uni-dimensional issue through repeated interpersonal encounters. Each agent i at time step t holds a continuous opinion $x_{i,t}$, inspired by the Deffuant model of social influence (Deffuant et al., 2000). These opinions evolve over time through interactions with other agents and under different opinion updating mechanisms. Below, we explain the spatial environment in which agents are embedded, rules governing agents' movement and interaction, and procedures for updating agents' opinions at each time step. We test two main updating mechanisms: a Bayesian-style rule and a rule that incorporates the possibility of a backlash effect. We also test each mechanism in an expanded, two-step communication flow where opinion leaders receive messaging from political elites.

Environment and Agent Movement

Agents are randomly distributed over a bounded space divided into a 32×32 square grid of "patches." Each patch represents a spatial unit capable of hosting multiple agents (Epstein and Axtell, 1996; Wilensky, 1999).

We provide additional structure to this environment by embedding agents within a small-world network of persistent social ties, reflecting the fact that strong ties like friends and family exert powerful influence on opinion formation and political engagement (Bakshy, Messing, and Adamic, 2015; Bond et al., 2012; Lazer et al., 2010). In a small-world network, an agent's friends are also likely to be friends with one another (Watts and Strogatz, 1998), creating tightly knit neighborhoods where information can travel quickly.[2]

[2]We vary each agent's number of network ties across model iterations.

We choose to distribute agents over a discretized, networked space for three reasons. First, it reflects our theoretical focus on local encounters and face-to-face conversation, rather than less personal interactions over digital or mass media channels. Second, it facilitates the type of interaction we seek to model, where dyadic interactions may occur when two or more agents co-locate. An unbounded environment would make interaction vanishingly rare, as agents would be distributed too sparsely across the space. Finally, a discrete spatial environment is a parsimonious way to represent a community of individuals, allowing us to model local interaction, movement constraints, and neighborhood effects (Railsback and Grimm, 2011). Our simulated environment thus represents a realistic depiction of everyday life, where individuals typically move through similar routes and encounter similar sets of people on a daily basis.

At each time step, agents move at least one patch in a random direction. The number of patches agents can move depends in part on the total number of agents, which we vary across simulations. Interactions between agents can occur if two or more agents occupy the same patch, but only one dyadic interaction is allowed per patch, per time step. When multiple agents share a patch, the model randomly selects two agents. These agents are considered to have encountered each other, but an encounter does not guarantee interaction. We model agents' likelihood of initiating interaction as a function of their past interaction experiences, described below.

Figure 1 displays a stylized example of how the simulation proceeds. In panel (a), agents are randomly assigned opinion positions and distributed across the grid, with a small number of network ties connecting them. These ties remain constant throughout the simulation. Panel (b) depicts the moving and updating process occurring at each time step: Agents move to adjacent patches, some interact with each other, and they update their opinions according to rules we detail below. Finally, panel (c) shows a hypothetical configuration at the end of the simulation. Network ties remain consistent, but agents in this example have converged on the "negative" opinion.

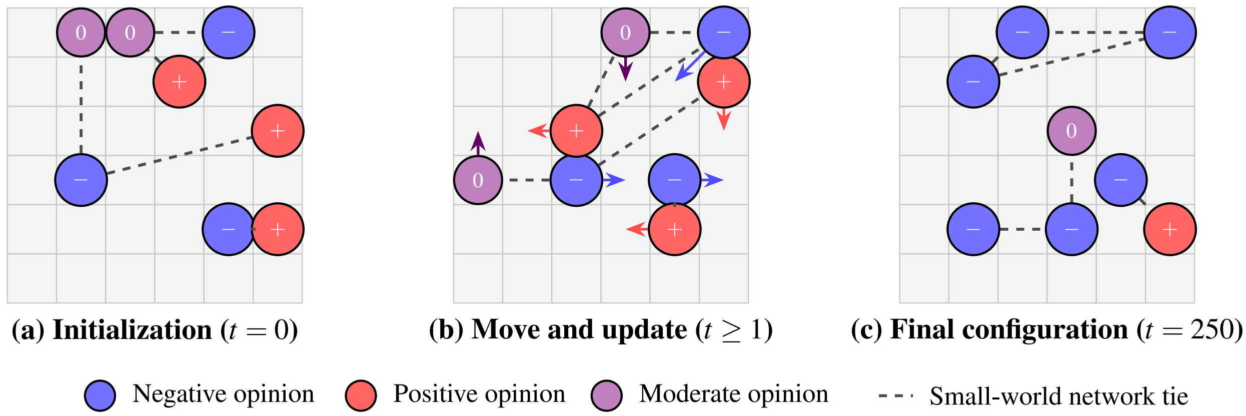

Figure 1: Stylized Example of Agent-Based Simulation. Agents (circles) are randomly distributed across a grid of patches and connected by a small-world network of ties (dashed lines). At each of 250 time steps, agents move to adjacent patches, interact with co-located agents, and update their opinions.

Opinions and Updating Mechanisms

Agents hold continuous opinions on the interval $[-1, 1]$, where 0 represents a moderate position and values closer to -1 and 1 indicate more extreme ones. Agents' initial opinions are randomly drawn from bimodal distributions designed to reflect different degrees of polarization.[3]

Suppose agents i and j encounter one another. If they have similar opinions (i.e. $x_{i,t} \cdot x_{j,t} > 0$), they always interact. This rule reflects well-known homophily dynamics in political communication (Carlson and Settle, 2022; Mutz, 2006), which refers to the tendency for individuals to associate and communicate more frequently with others who share similar characteristics or viewpoints (McPherson, Smith-Lovin, and Cook, 2001). For encounters involving differing opinions ($x_{i,t} \cdot x_{j,t} < 0$), agents interact with probability $P_{ij,t} \leq 1$, reflecting their combined willingness to engage in cross-cutting conversation. We model $P_{ij,t}$ as the geometric mean of both agents' openness to interaction, $\ell_{i,t}$ and $\ell_{j,t}$.

We represent each agent i's openness to initiating conversation at time t—$\ell_{i,t}$—as a function of their experiences in previous conversations. Agents begin the simulation with $\ell_{i,t}$ drawn from

[3] To ensure our model reflects empirical reality, we derive our initial opinion distributions from survey data. The Appendix contains further details on model initialization.

a uniform distribution over the unit interval. They then update this value at each subsequent time step, depending on the type of agents they interact with.

Suppose agent i interacts with agent j. If agent j is moderate ($-0.5 \leq x_{j,t} \leq 0.5$), we increase agent i's likelihood of initiating conversation: $\ell_{i,t} = 1.1 \cdot \ell_{i,t-1}$, reflecting the notion that exposure to reasonable viewpoints can encourage further discussion. Conversely, if agent j is extreme ($x_{j,t} < -0.5$ or $x_{j,t} > 0.5$), we decrease agent i's likelihood by an identical proportion: $\ell_{i,t} = 0.9 \cdot \ell_{i,t-1}$. This update captures the intuition that when someone interacts with an extreme, like-minded compatriot, they can become further entrenched in their own views and less willing to entertain others. Conversely, it also captures the intuition that encountering someone with an extreme opinion in the *opposite* direction of one's own can be a "turn-off" to further conversation on that topic, as individuals attempt to insulate themselves from cognitive discomfort.

If an interaction occurs, agents update their opinions based on a function of each other's opinions, their own openness, and a personal learning rate α_i, which depends on the number of network connections for agent i. When two agents are friends, α_i is doubled to account for the increased trust that strong ties place in each other.

We explore the dynamics of opinion change under two cognitive mechanisms: Bayesian-style updating and a backlash effect. We further examine the implications of elite messaging—another common source of political information for citizens—for opinion dynamics governed by each mechanism. Under the Bayesian updating mechanism, agents proportionally adjust their opinions toward the views of their discussion partners:

$$x_{i,t+1} = x_{i,t} + x_{j,t} \cdot \alpha_i \cdot \ell_{i,t}. \tag{1}$$

This formulation parallels classic models of continuous opinion adjustment (Deffuant et al., 2000).

Although research is mixed on this question (Coppock, 2022; Lodge and Taber, 2013; Wood and Porter, 2019), it is possible that citizens could exhibit a backlash effect to encountering counter-attitudinal opinions, wherein cross-cutting information leads them to become further entrenched in their original positions. Under this "backlash" updating mechanism, when an agent's interlocutor is on the opposite side of the opinion midpoint, the update rule becomes:

$$x_{i,t+1} = x_{i,t} - \frac{x_{j,t} \cdot \alpha_i \cdot \ell_{i,t}}{10}, \tag{2}$$

where we divide by 10 to prevent opinions from becoming resistant to any change.

Finally, we design a mechanism to emulate the effects of mass media or messaging from preferred political elites, which are important sources of citizens' attitudes (Lenz, 2012; Zaller, 1992). To simulate a shock to the information environment coming from elite messaging, we first allow the model to run for 50 time steps. At $t = 50$, we introduce a one-time exogenous shock to agents whose opinions $x_{i,50}$ fall within the top tercile of the opinion distribution.[4] By targeting the shock to the most extreme agents and observing downstream effects on opinion dynamics in the full population, we simulate the two-step flow of communication, wherein opinion leaders receive a message from an elite source and then pass it on to their social contacts (Druckman and Nelson, 2003; Katz and Lazarsfeld, 1955; Norris and Curtice, 2008). We employ this elite messaging shock under both Bayesian and backlash-based updating mechanisms, enabling us to evaluate whether sudden elite interventions can redirect trajectories of opinion change that interpersonal influence alone would not produce.

Variables and Analysis

Each simulation runs for 250 time steps. This duration was selected based on exploratory trials showing that the system consistently reaches a stable state within 200 time steps across all parameter combinations. Setting the limit to 250 steps provides an adequate safety margin to guarantee full convergence for every run. For every combination of population size, neighborhood size, and initial opinion distribution, we conduct 50 simulation runs, allowing us to incorporate variation in simulation outcomes into error estimates in downstream analyses.

[4] This setup would be analogous to a politician making a public statement that is targeted at their co-partisans, who are already primed to take such political messages into account.

To evaluate the dynamics of opinion updates, we specify two complementary dependent variables to measure both the qualitative reversal of ideological alignments and the quantitative magnitude of opinion displacement. First, we define the "camp-switching rate" (CSR) to capture discrete ideological realignments. CSR measures the proportion of agents whose ideological polarity has changed since the beginning of the simulation, indicating that their opinion has crossed the neutral threshold. This metric provides an aggregate measure of how frequently opinions shift over the course of a simulation.

However, relying solely on aggregate net change can obscure important micro-level dynamics. For example, similar proportions of a population radicalizing in opposing directions would yield a net change close to zero, erroneously suggesting that the system is static. Moreover, agents may undergo ideological intensification or moderation without crossing the neutral threshold, a meaningful source of opinion change that would not be captured by CSR. To measure these dynamics, we calculate the mean absolute change (MAC). MAC quantifies the average absolute difference between each agent's initial and updated opinions, thereby measuring the overall magnitude of opinion displacement independent of direction. Together, CSR and MAC allow us to distinguish between environments in which opinions change rarely but substantially and those in which opinions change frequently but only incrementally.

Our two main independent variables describe the structural properties of the interaction environment. First, the interaction ratio captures the mean number of encounters per agent. This measure scales encounter frequency by population size and reflects the effective intensity of social contact. Second, neighborhood size represents the average number of network neighbors assigned to each agent in the small-world network. This variable captures the size of agents' immediate social environment and the number of potential communication partners they may encounter at each time step.

Finally, each simulation is initialized with opinions randomly drawn from bimodal distributions, with either moderate or extreme opinions and low or high variation. An extreme, low-variation distribution represents a high-polarization environment, while a moderate, high-variation distribution represents low polarization (Mehlhaff, 2024). Because we expect the effect of interpersonal interaction to depend on the surrounding opinion environment, our regression models include interaction terms between interaction ratio and initial opinion distribution. We estimate ordinary least squares (OLS) regressions where each observation represents a full simulation run.

Results

To examine how initial opinion distributions and individual-level interactions drive opinion change, we analyze the simulated data under several parameter configurations. We examine results under each updating mechanism in turn, followed by simulations that incorporate elite messaging.

Bayesian Updating

To visually show the evolution of opinion dynamics over the course of the simulation, Figure 2 plots CSR and MAC at each time step.[5] Thin lines show results for each of the 50 model iterations, while thick lines give locally smoothed averages. While opinion change tends to increase over time, there is a stark difference between extreme- and moderate-opinion contexts. Extreme opinion scenarios see limited opinion change; CSR in these contexts remains barely above zero for the duration of the simulation. MAC is slightly higher and increasing over time, suggesting that opinion dynamics in polarized contexts are characterized by rare but substantial shifts.

In contrast, moderate opinion contexts generate more frequent opinion updates; agents in these scenarios are much more likely to change their opinion at each time step. In the moderate-opinion, low-variation context, about 40 percent of agents change their opinions over the course of the simulation, representing a substantial amount of persuasion. Since they start out closer to the midpoint, however, even small shifts can lead to opinion change, so MAC remains relatively low.

[5] We specifically examine the version of the model where population and network size parameters are set to their maximum values. We use the regression models below to give a broader view across parameter values.

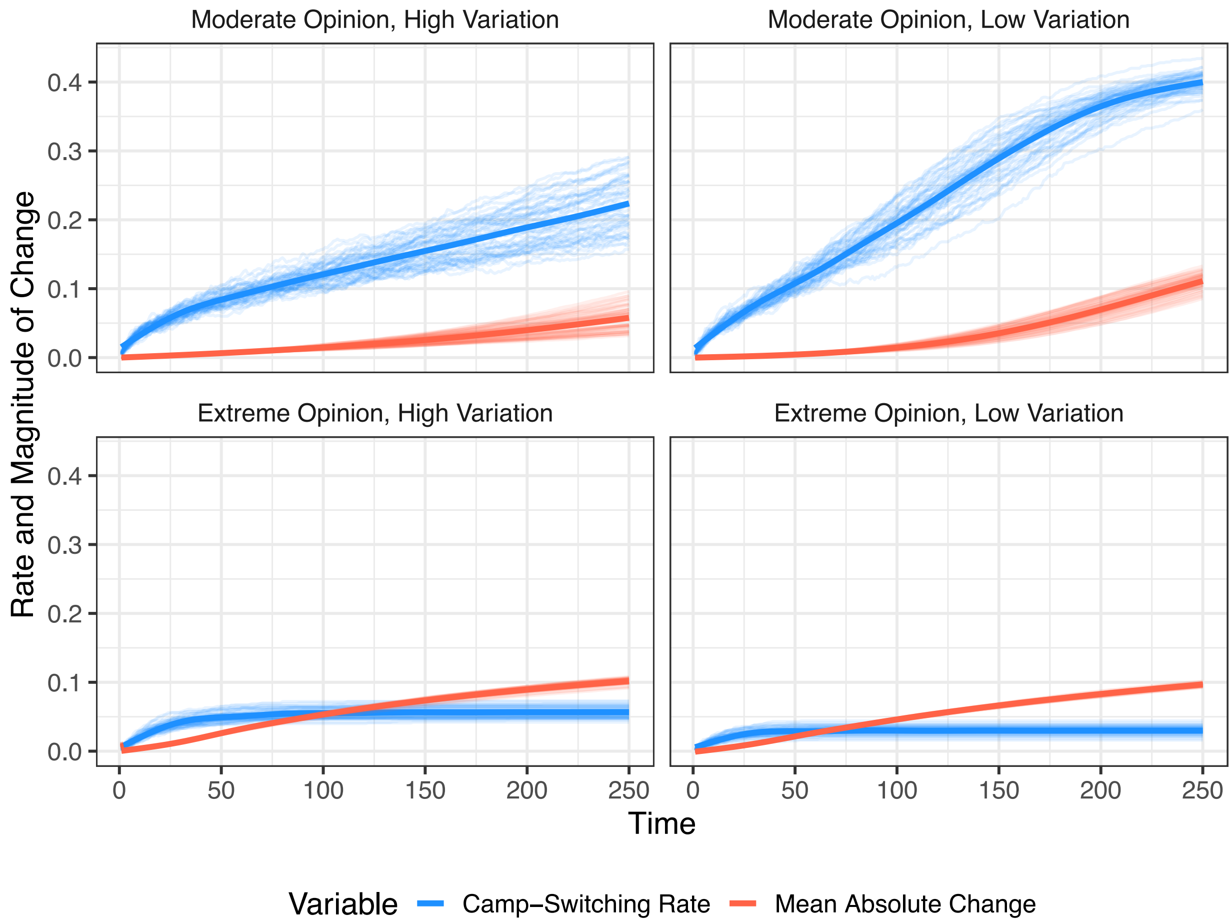

Figure 2: Camp-Switching Rate and Mean Absolute Change over Time (Bayesian Updating). Thin lines represent individual simulation runs, thick lines represent locally smoothed averages.

To investigate these dynamics more formally, Table 1 reports OLS regression results for CSR and MAC under Bayesian updating.[6] Across both models, interaction is a strong and statistically significant predictor of opinion dynamics. Higher levels of interaction are associated with more frequent ideological realignment and larger magnitude of opinion adjustment. As observed in Figure 2, the initial distribution of opinion in the population also affects the dynamics of opinion change. Relative to the baseline context of moderate opinions with high variation (that is, the least polarized scenario), extreme contexts are associated with lower CSR but higher MAC. When

[6]We do not control for population size, since it is already incorporated into the interaction ratio. The Appendix reports model diagnostics.

the environment is sharply polarized, agents shift their opinions by large amounts, but they do so infrequently. This likely reflects the fact that in extreme contexts, agents are more likely to start out far from the ideological midpoint, so they must make larger shifts to cross it. In moderate contexts, agents are more likely to start out closer to the midpoint, so crossing it requires less drastic adjustments.

More critically, the structure of the opinion environment conditions the effect of interaction. Interaction leads to less frequent camp-switching in extreme environments than in moderate ones. At the same time, however, interaction also leads to greater magnitude of opinion change in those same contexts. This suggests that while interaction does not make opinion change more likely in polarized contexts, it does make those rare crossing events more dramatic when they do occur. In moderate contexts, by contrast, interaction results in smaller opinion updates, but those small changes can still lead to camp-switching due to agents being more tightly clustered around the midpoint.

Finally, neighborhood size exerts a consistent stabilizing influence. Larger neighborhood sizes reduce both camp-switching and overall opinion movement, suggesting that broader local interaction environments dampen the volatility of opinion updates. When agents have larger social circles, they are less likely to participate in cross-cutting interaction, which has the effect of isolating them in ideological silos and reducing the likelihood of opinion change. This finding is consistent with the notion that social networks can serve as echo chambers, reinforcing existing beliefs and limiting exposure to diverse viewpoints (Barberá, 2020). It also aligns with previous research showing that cross-cutting exposure is better facilitated by weak ties rather than strong ones (Mutz, 2006).

The overarching takeaway from these Bayesian updating simulations is that the nature of opinion change depends on micro-level interactions—as existing theory would suggest—but it also depends on the social environment and opinion distribution in which agents are embedded. These macro-level features further condition the effects of interaction, indicating that political discussion has real limits when it comes to its ability to drive opinion change.

Table 1: Effect of Interaction under Bayesian Updating

	Dependent variable:	
	CSR	MAC
Interaction ratio	0.036^{***}	0.022^{***}
	(0.002)	(0.002)
Extreme opinion, high variation	-0.116^{***}	0.161^{***}
	(0.008)	(0.007)
Extreme opinion, low variation	-0.158^{***}	0.172^{***}
	(0.009)	(0.008)
Moderate opinion, low variation	0.170^{***}	0.066^{***}
	(0.003)	(0.003)
Neighborhood size	-0.009^{***}	-0.029^{***}
	(0.0004)	(0.0004)
Interaction × Extreme, high var.	-0.019^{*}	0.077^{***}
	(0.009)	(0.007)
Interaction × Extreme, low var.	-0.033^{***}	0.088^{***}
	(0.009)	(0.008)
Interaction × Moderate, low var.	-0.003	-0.006^{*}
	(0.003)	(0.003)
Constant	0.190^{***}	0.048^{***}
	(0.002)	(0.002)
Observations	4,800	4,800
R^2	0.963	0.677
Adjusted R^2	0.962	0.676

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$
Reference category: Moderate opinion, high variation

Backlash Effect

We now consider a setting in which we relax the assumption that agents always move toward their interlocutors and may instead react defensively to counterattitudinal information (Lau and

Redlawsk, 2006; Lodge and Taber, 2013).[7] Figure 3 traces CSR and MAC over the course of the simulation under this backlash rule, with thin lines again representing each model iteration and thick lines giving locally smoothed averages. The contrast between extreme and moderate contexts is still present as it was under Bayesian updating, but here it is even more stark. In extreme opinion environments, CSR stays very small, while MAC is large and rises steadily over time. Moderate settings look different. There, both outcomes are relatively large and increasing over time, with MAC being notably larger than under the Bayesian updating mechanism. This pattern of results hints at bifurcated opinion dynamics; agents close to the ideological midpoint switch across it, while those farther away from the midpoint become more entrenched in their positions.

Table 2 sharpens this picture by reporting OLS estimates for CSR and MAC under the backlash rule. More interaction is still associated with more camp-switching and larger absolute opinion movement. The surrounding opinion environment again matters just as much. Consistent with Figure 3, extreme contexts yield much lower CSR than the baseline moderate setting, yet they do not always generate the compensating increase in MAC observed under Bayesian updating. In fact, the degree of variation in opinions appears to matter more than extremity; high-variation scenarios exhibit significantly smaller opinion shifts than low-variation ones do. One possible explanation for this pattern could be that when opinion variation is low—holding group extremity constant—agents are clustered more tightly in ideological communities. They are thus more likely to come into contact with someone from the opposite camp, as there are fewer agents close to the midpoint. Under a backlash updating mechanism, such interactions would push agents further away from the midpoint, resulting in larger shifts overall.

More important for our argument is how the effects of interaction vary across opinion contexts. In extreme environments, additional interaction reduces both CSR and MAC relative to moderate settings. This is a key departure from the results in Table 1. Under Bayesian updating, polarized contexts still allowed rare but dramatic shifts when agents crossed the midpoint. Under backlash, those same contexts dampen both outcomes at once. Polarized settings therefore do not merely

[7] Cf. Coppock (2022), Guess and Coppock (2020), and Wood and Porter (2019).

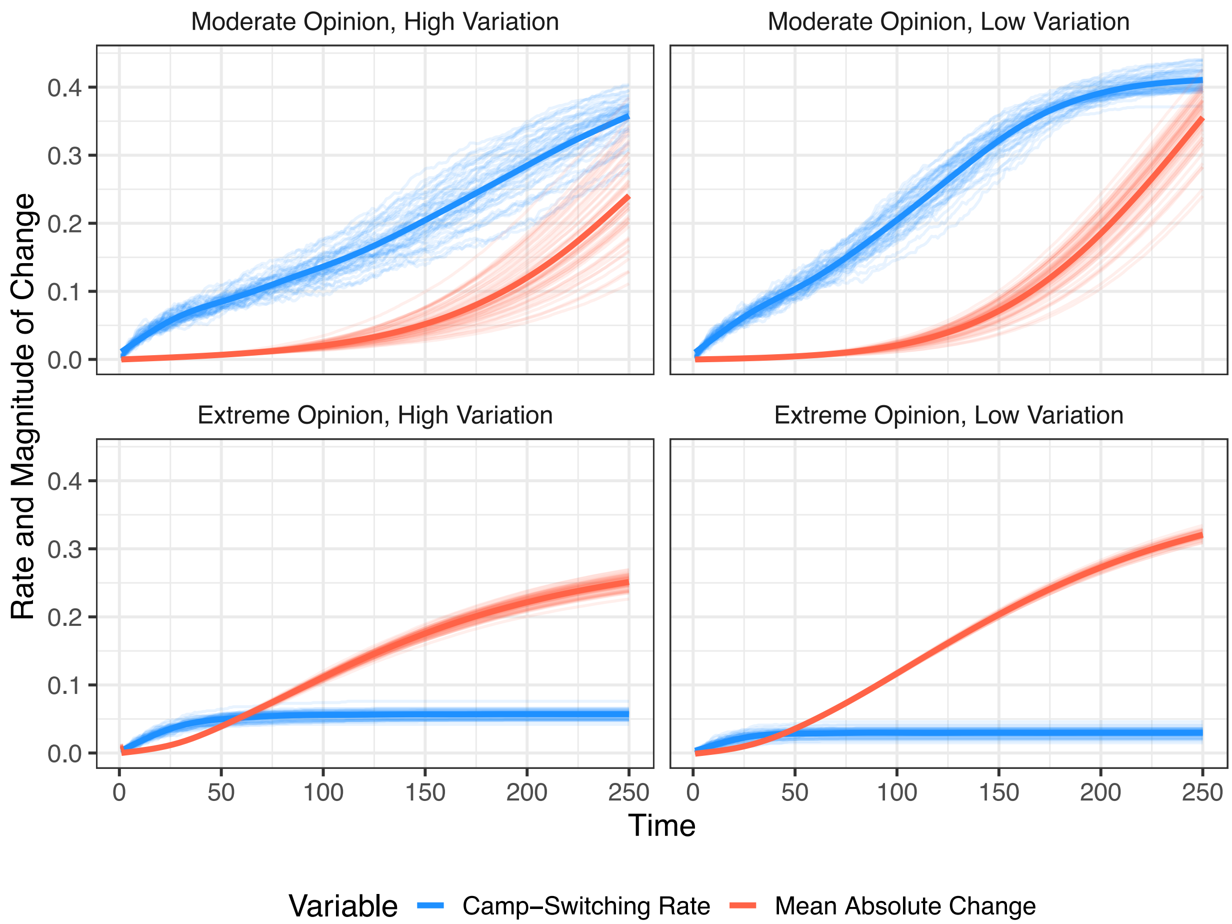

Figure 3: Camp-Switching Rate and Mean Absolute Change over Time (Backlash). Thin lines represent individual simulation runs, thick lines represent locally smoothed averages.

make opinion change uncommon, they make repeated interaction less likely to promote meaningful updating in any direction.

Neighborhood size also continues to constrain opinion change, pushing both CSR and MAC downward. Taken together, simulations under an attitudinal backlash mechanism point to a narrower pathway for opinion change. Discussion can still generate updating in moderate settings, but once agents are embedded in ideologically polarized environments, defensive reactions to disagreement suppress both the frequency and magnitude of movement.

Table 2: Effect of Interaction under Backlash

	Dependent variable:	
	CSR	MAC
Interaction ratio	0.041^{***}	0.084^{***}
	(0.001)	(0.002)
Extreme opinion, high variation	-0.254^{***}	-0.051^{***}
	(0.001)	(0.003)
Extreme opinion, low variation	-0.282^{***}	0.123^{***}
	(0.001)	(0.003)
Moderate opinion, low variation	0.084^{***}	0.117^{***}
	(0.001)	(0.003)
Neighborhood size	-0.012^{***}	-0.065^{***}
	(0.0004)	(0.001)
Interaction × Extreme, high var.	-0.041^{***}	-0.065^{***}
	(0.001)	(0.003)
Interaction × Extreme, low var.	-0.041^{***}	-0.053^{***}
	(0.001)	(0.003)
Interaction × Moderate, low var.	-0.027^{***}	0.030^{***}
	(0.001)	(0.002)
Constant	0.312^{***}	0.203^{***}
	(0.001)	(0.002)
Observations	4,800	4,800
R^2	0.971	0.683
Adjusted R^2	0.971	0.683

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$
Reference category: Moderate opinion, high variation

Elite Messaging

We finally consider whether elite messaging changes the trajectories generated by peer-to-peer interaction alone by explicitly modeling a two-step communication flow. Figure 4 plots CSR and MAC over time across contexts, similar to Figures 2 and 3. Since we examine the impact of the two-step flow under both Bayesian and backlash updating mechanisms, we present the former in the top row of plots and the latter in the bottom row. Dashed vertical lines indicate the point at which the exogenous information shock occurs.

The most obvious difference in opinion dynamics is across moderate and extreme opinion scenarios. When opinion is more moderate, the elite intervention generates, at most, a small increase in CSR and MAC, and only in the high-variation context—the least polarized of the four opinion environments. Because we designate the top third of agents as "opinion leaders" who receive the elite cue (Katz and Lazarsfeld, 1955), there is simply not much difference between agents receiving and not receiving the shock in less polarized contexts, limiting the possibility for any trickle-down effect.

In extreme opinion environments, by contrast, the elite shock produces a much more pronounced increase in both CSR and MAC. Both CSR and MAC jump dramatically immediately after the shock, though the increase is less pronounced in low-variation contexts, which are the most polarized of the four opinion environments. Interestingly, both metrics slowly decrease following the initial jump, except for MAC under the backlash mechanism and high polarization. Here, MAC continues increasing after the shock while CSR drops nearly to zero. Overall, these patterns suggest that while elite messaging can have a significant impact on opinion dynamics in polarized contexts, the effects are not necessarily enduring; effects decay as agents return to their regular peer-to-peer communication habits, a finding echoed by Druckman and Nelson (2003).

Table 3 reports the four regression models jointly (Bayesian and backlash mechanisms, each for CSR and MAC). The "elite cue" variable captures the effect of the elite messaging shock itself. As expected, the elite cue does encourage opinion change as well as larger magnitudes of opinion shifts, regardless of the updating mechanism. Controlling for this elite cue, interaction ratio and neighborhood size again operate as they did in the absence of the shock, with more interaction typically portending more opinion change and larger social networks leading to less change. Context again matters, with extreme opinion environments associated with lower CSR and higher MAC. In the case of backlash, however, the degree of variation in the initial bimodal opinion

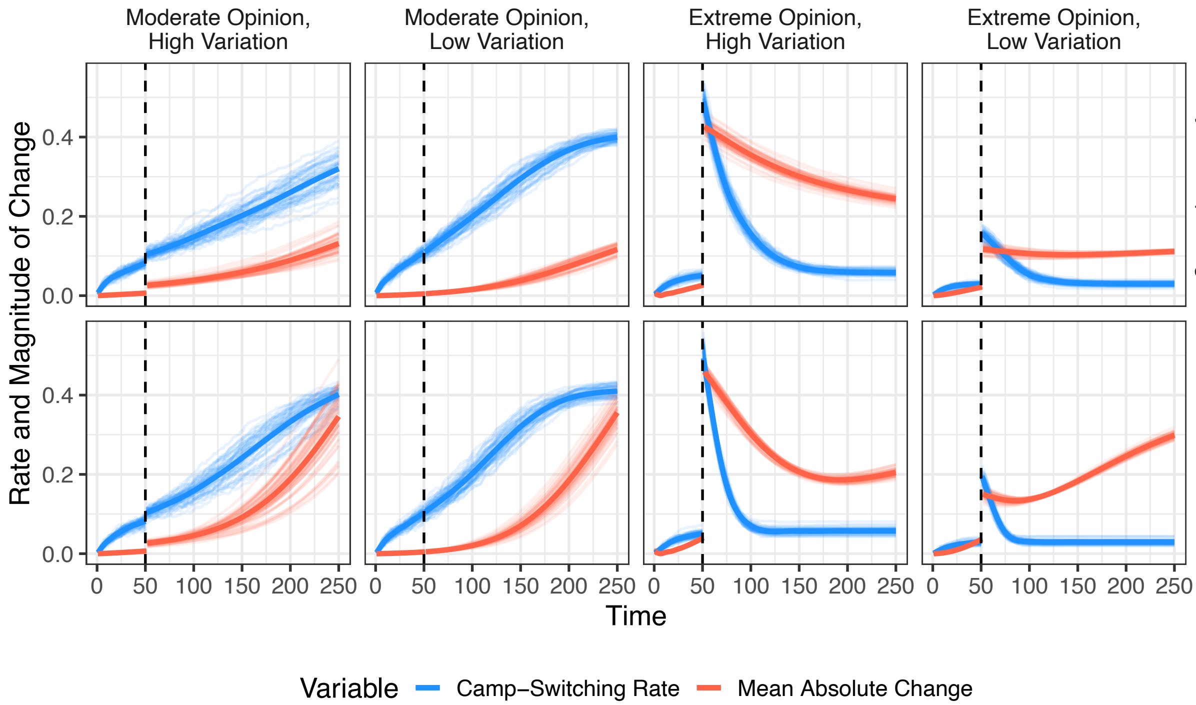

Figure 4: Camp-Switching Rate and Mean Absolute Change with Elite Messaging. Thin lines represent individual simulation runs, thick lines represent locally smoothed averages. Dashed vertical lines indicate exogenous information shock.

distributions appears to matter more than extremity does, with low-variation contexts generating more opinion change than high-variation ones.

The moderating effect of context again demonstrates how the effect of interaction is conditional on the opinion landscape. Under Bayesian updating, interaction in extreme settings decreases switching while increasing the magnitude of change, reinforcing the pattern of infrequent but large shifts. The baseline context—moderate opinion with high variation, the least polarized scenario—is the environment in which interaction drives the least amount of absolute opinion change. The story is very similar in the case of backlash effects: as polarization increases, interaction produces less outright opinion change. In extreme scenarios, it also produces less ideological movement overall. The effect of neighborhood size remains negative in all four models, aligning with the broader pattern that denser local ties reduce responsiveness to new information, even when elites inject an exogenous directional signal.

Table 3: Effect of Interaction under Elite Messaging Across Updating Mechanisms

	Bayesian Updating		*Backlash Effect*	
	CSR	MAC	CSR	MAC
Interaction ratio	0.004	0.014***	0.038***	0.090***
	(0.003)	(0.003)	(0.001)	(0.003)
Extreme opinion, high variation	−0.254***	0.223***	−0.255***	0.002
	(0.008)	(0.008)	(0.002)	(0.005)
Extreme opinion, low variation	−0.262***	0.063***	−0.284***	0.084*
	(0.011)	(0.011)	(0.002)	(0.005)
Moderate opinion, low variation	0.097***	0.002	0.081***	0.094***
	(0.004)	(0.004)	(0.002)	(0.004)
Elite cue	0.036***	0.042***	0.029***	0.038***
	(0.001)	(0.001)	(0.001)	(0.003)
Neighborhood size	−0.015***	−0.029***	−0.012***	−0.066***
	(0.001)	(0.001)	(0.001)	(0.001)
Interaction × Extreme, high var.	−0.063***	0.158***	−0.038***	−0.071***
	(0.009)	(0.009)	(0.002)	(0.004)
Interaction × Extreme, low var.	−0.036***	0.034**	−0.038***	−0.055***
	(0.011)	(0.011)	(0.002)	(0.004)
Interaction × Moderate, low var.	0.032***	0.018***	−0.023***	0.049***
	(0.004)	(0.004)	(0.002)	(0.004)
Constant	0.242***	0.085***	0.298***	0.213***
	(0.003)	(0.003)	(0.002)	(0.004)
Observations	3,600	3,600	3,600	3,600
R^2	0.931	0.563	0.948	0.629
Adjusted R^2	0.931	0.562	0.948	0.628

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$
Reference category: Moderate opinion, high variation

Taken together, we see the two-step information flow as something of a conditional accelerator. Elite cues can reliably increase opinion updating in downstream interpersonal interactions,but their effects are limited when population-level opinion is moderate. Even in highly polarized environments, elite messaging can magnify asymmetries already brought about by the underlying cognitive mechanism. Under Bayesian updating, that means larger movement without frequent crossing; with backlash, it means further consolidation into ideological camps. These conditional outcomes contribute to a consistent finding that we highlight throughout: Effects of elite communication and interpersonal interaction are not always additive, but rather they are dependent on social context.

Discussion

Our central finding is that the potential for political discussion to promote attitude change and thus dampen ideological polarization is contingent on social context and the broader opinion environment. When social networks channel people into homogeneous ideological camps or when opinion distributions are concentrated at the extremes, more interaction does not reliably produce belief revision. This suggests an important update to a literature that has focused largely on micro-level mechanisms: interpersonal influence cannot be understood apart from the environments in which it takes place.

We see at least four theoretical implications of our results. First, they speak to longstanding debates on cross-cutting exposure. A range of influential theories from both political science and psychology argue that communicating across lines of disagreement should weaken opinion extremity and prompt citizens to revise their opinions (e.g. Mercier and Sperber, 2017; Mutz, 2006; Trouche, Sander, and Mercier, 2014). However, these benefits are unlikely to accrue without exposure to genuinely diverse perspectives, and such exposure depends on the structure of the broader social environment. We show that when the opinion landscape is polarized, interaction produces large but infrequent opinion updates, consistent with homophily-driven environments where individuals communicate predominantly with like-minded others (Barberá et al., 2015; McPherson, Smith-Lovin, and Cook, 2001). In such settings, interaction reproduces the limited informational range of the network rather than expanding it. Conversely, moderate opinion contexts respond more readily to interaction. Here, the effect of interaction manifests as many small incremental shifts rather than large breaks——an outcome consistent with theories of cumulative influence in networked settings (Centola and Macy, 2007; Flache and Macy, 2011).

Second, the consistent negative effect of neighborhood size across all mechanisms runs counter to widely held expectations about network structure. Small-world networks, with clustered neighborhoods and short path lengths (Watts and Strogatz, 1998), are often thought to facilitate both reinforcement within close circles and exposure through occasional weak ties (Granovetter, 1973). However, we show that larger neighborhoods depress opinion updating, suggesting that denser local networks multiply redundant connections rather than increasing exposure to genuinely new information. This finding echoes recent concerns that high-connectivity environments, such as algorithmically curated social media feeds, may saturate individuals with confirmatory signals, reducing the marginal impact of any single encounter (Mäs and Flache, 2013; Sunstein, 2017).

Third, foundational theories of public opinion emphasize that elite cues shape attitudes by providing interpretive frames or salient directional signals (e.g. Zaller, 1992). By modeling the two-step flow of communication, we show that elite interventions directly increase opinion updating, but even this result is contingent on social context: Elite cues only have a meaningful impact when the population is already sufficiently polarized. These effects trickle down to some degree from "opinion leaders" to their interlocutors, but they quickly decay once individuals resume their regular regimen of interpersonal interaction (Druckman and Nelson, 2003).

Finally, backlash theories—where opinion extremity reduces receptiveness or triggers entrenchment—are often invoked to explain ideological polarization (Bishin et al., 2016; Lau and Redlawsk, 2006; Lodge and Taber, 2013). We show that even if backlash effects exist (Coppock, 2022; Guess and Coppock, 2020; Wood and Porter, 2019), their influence on opinion dynamics is still context-dependent. In extreme environments, backlash leaves opinions effectively frozen despite interaction, while in more moderate landscapes, interaction can still encourage opinion change. This pattern further reinforces the view that macro-level opinion structure constrains micro-level cognitive processes.

Methodologically, we illustrate the value of ABM for research in political behavior. By specifying psychological mechanisms and examining repeated encounters over time, ABM makes it possible to evaluate macro-level implications of micro-level theories under controlled counterfactual conditions. This generative approach provides insights that are difficult to obtain using purely observational or experimental methods, where the underlying mechanisms of opinion change are often difficult to identify.

Several limitations should temper interpretation. Our agents follow stylized updating rules and operate on a single issue dimension, whereas real citizens combine argument evaluation, identity defense, and strategic signaling in more complex ways. This arrangement limits our ability to draw conclusions about other important aspects of polarization, such as the degree to which it is rooted in moral psychology (e.g. Haerter, Filsinger, and Freitag, 2025). Moral commitments may be held more tightly than attitudes concerning policy specifics, for instance, leaving open questions as to how morality shapes interpersonal interaction and its effects on political attitudes.

At a more mechanical level, network structure is fixed within each model iteration, even though real social ties evolve with mobility, sorting, and platform dynamics. Elite messaging is modeled as an exogenous one-time shock directed at opinion leaders, while actual elite communication is strategic, repeated, and responsive to public feedback. Finally, we abstract away from individual-level heterogeneity in political knowledge, identity strength, and susceptibility to persuasion, all of which likely condition opinion updating in practice.

These limitations point to productive next steps. Future work could incorporate additional complexity by making elite strategies endogenous to agent behavior, allow social networks to evolve over time, introduce heterogeneous agent characteristics, and extend the framework to multidimensional attitudes or affective polarization (Baldassarri and Bearman, 2007; Mason, 2018). Using empirically observed communication and network data to motivate parameter selection could also improve external validity and sharpen substantive interpretation (e.g. Mäs and Flache, 2013).

Taken together, our findings refine predominant theories of interpersonal influence. They demonstrate that interaction is not an inherently depolarizing force; its efficacy depends critically on contextual properties of the population. Where diversity already exists, interaction can nudge opinions. Where diversity is absent, it merely reproduces stability. The same cognitive or communicative mechanisms yield different system-level outcomes depending on initial opinion distributions and network structure. Efforts to reduce polarization by increasing discussion volume alone may therefore disappoint unless they also alter the contexts that determine who talks to whom and how disagreement is processed.

References

Abramowitz, Alan I. (2010). *The Disappearing Center: Engaged Citizens, Polarization, and American Democracy*. New Haven, CT: Yale University Press.

Akboga, Sema, Osman Sahin, and Engin Arik (Aug. 2025). "Polarisation over the Meaning of Democracy: The Case of Political Parties in Turkey." *Politics* 45.3, pp. 333–354.

Altay, Sacha et al. (Mar. 2023). "Information Delivered by a Chatbot Has a Positive Impact on COVID-19 Vaccines Attitudes and Intentions." *Journal of Experimental Psychology: Applied* 29.1, pp. 52–62.

American National Election Studies (2021). *ANES 2020 Time Series Study Full Release [Dataset and Documentation].*

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic (2015). "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348.6239, pp. 1130–1132.

Baldassarri, Delia and Peter Bearman (2007). "Dynamics of Political Polarization." *American Journal of Sociology* 113.4, pp. 1145–1184.

Banisch, Sven and Hawal Shamon (2024). "Validating Argument-Based Opinion Dynamics with Survey Experiments." *Journal of Artificial Societies and Social Simulation* 27.1.

Barber, Michael and Jeremy C. Pope (Feb. 2019). "Does Party Trump Ideology? Disentangling Party and Ideology in America." *American Political Science Review* 113.1, pp. 38–54.

Barberá, Pablo (2020). "Social Media, Echo Chambers, and Political Polarization." *Social Media and Democracy*. Ed. by Nathaniel Persily and Joshua A. Tucker. New York: Cambridge University Press, pp. 34–55.

Barberá, Pablo et al. (Oct. 2015). "Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?" *Psychological Science* 26.10, pp. 1531–1542.

Bear, Adam and David G. Rand (Jan. 2016). "Intuition, Deliberation, and the Evolution of Cooperation." *Proceedings of the National Academy of Sciences* 113.4, pp. 936–941.

Becker, Joshua, Ethan Porter, and Damon Centola (May 2019). "The Wisdom of Partisan Crowds." *Proceedings of the National Academy of Sciences* 116.22, pp. 10717–10722.

Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee (1954). *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: The University of Chicago Press.

Bishin, Benjamin G. et al. (July 2016). "Opinion Backlash and Public Attitudes: Are Political Advances in Gay Rights Counterproductive?" *American Journal of Political Science* 60.3, pp. 625–648.

Blair, Robert A. et al. (Nov. 2025). "Depolarizing within the Comfort of Your Party: Experimental Evidence from Online Workshops." *Political Communication* 42.6, pp. 925–950.

Bond, Robert M. et al. (2012). "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489.7415, pp. 295–298.

Campbell, Angus et al. (1960). *The American Voter*. Chicago: The University of Chicago Press.

Carlson, Taylor N. and Jaime E. Settle (2022). *What Goes Without Saying: Navigating Political Discussion in America*. New York: Cambridge University Press.

Centola, Damon and Michael Macy (Nov. 2007). "Complex Contagions and the Weakness of Long Ties." *American Journal of Sociology* 113.3, pp. 702–734.

Coppock, Alexander (2022). *Persuasion in Parallel: How Information Changes Minds about Politics*. Chicago: The University of Chicago Press.

Dandekar, Pranav, Ashish Goel, and David T. Lee (2013). “Biased Assimilation, Homophily, and the Dynamics of Polarization.” *Proceedings of the National Academy of Sciences* 110.15, pp. 5791–5796.

Deffuant, Guillaume et al. (Jan. 2000). “Mixing Beliefs among Interacting Agents.” *Advances in Complex Systems* 03.01n04, pp. 87–98.

Druckman, James N. (Nov. 2004). “Political Preference Formation: Competition, Deliberation, and the (Ir)Relevance of Framing Effects.” *American Political Science Review* 98.4, pp. 671–686.

Druckman, James N. and Kjersten R. Nelson (Oct. 2003). “Framing and Deliberation: How Citizens’ Conversations Limit Elite Influence.” *American Journal of Political Science* 37.3, pp. 729–745.

Druckman, James N., Erik Peterson, and Rune Slothuus (Feb. 2013). “How Elite Partisan Polarization Affects Public Opinion Formation.” *American Political Science Review* 107.1, pp. 57–79.

Epstein, Joshua M. (May 1999). “Agent-Based Computational Models and Generative Social Science.” *Complexity* 4.5, pp. 41–60.

Epstein, Joshua M. and Robert Axtell (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Cambridge, MA: MIT Press.

Esterling, Kevin M., Archon Fung, and Taeku Lee (Apr. 2021). “When Deliberation Produces Persuasion Rather than Polarization: Measuring and Modeling Small Group Dynamics in a Field Experiment.” *British Journal of Political Science* 51.2, pp. 666–684.

Flache, Andreas and Michael W. Macy (2011). “Small Worlds and Cultural Polarization.” *Journal of Mathematical Sociology* 35.1-3, pp. 146–176.

Granovetter, Mark S (May 1973). “The Strength of Weak Ties.” *American Journal of Sociology* 78.6, pp. 1360–1380.

Gray, Kurt et al. (Apr. 2014). “The Emergence of “Us and Them” in 80 Lines of Code: Modeling Group Genesis in Homogeneous Populations.” *Psychological Science* 25.4, pp. 982–990.

Guess, Andrew and Alexander Coppock (Oct. 2020). “Does Counter-Attitudinal Information Cause Backlash? Results from Three Large Survey Experiments.” *British Journal of Political Science* 50.4, pp. 1497–1515.

Haerter, Victoria, Maximilian Filsinger, and Markus Freitag (Nov. 2025). “Divided by Morality? Moral Foundations of Affective Polarisation during Hard Times.” *Politics* 45.4, pp. 594–615.

Huckfeldt, Robert, Paul E. Johnson, and John Sprague (2004). *Political Disagreement: The Survival of Diverse Opinions within Communication Networks*. New York: Cambridge University Press.

Huckfeldt, Robert and John Sprague (1995). *Citizens, Politics, and Social Communication: Information and Influence in an Election Campaign*. New York: Cambridge University Press.

Iyengar, Shanto and Sean J. Westwood (July 2015). “Fear and Loathing across Party Lines: New Evidence on Group Polarization.” *American Journal of Political Science* 59.3, pp. 690–707.

Jackson, Joshua Conrad et al. (May 2017). “Agent-Based Modeling: A Guide for Social Psychologists.” *Social Psychological and Personality Science* 8.4, pp. 387–395.

Katz, Elihu and Paul F. Lazarsfeld (1955). *Personal Influence: The Part Played by People in the Flow of Mass Communications*. New York: The Free Press.

Lau, Richard R. and David P. Redlawsk (2006). *How Voters Decide: Information Processing in Election Campaigns*. New York: Cambridge University Press.

Lazer, David et al. (2010). “The Coevolution of Networks and Political Attitudes.” *Proceedings of the National Academy of Sciences* 107.50, pp. 214–219.

Lenz, Gabriel S. (2012). *Follow the Leader? How Voters Respond to Politicians’ Policies and Performance*. Chicago: The University of Chicago Press.

Lodge, Milton and Charles S. Taber (2013). *The Rationalizing Voter*. New York: Cambridge University Press.

Macy, Michael W. and Robert Willer (Aug. 2002). “From Factors to Actors: Computational Sociology and Agent-Based Modeling.” *Annual Review of Sociology* 28.1, pp. 143–166.

Mäs, Michael and Andreas Flache (2013). “Differentiation Without Distancing: Explaining Bi-Polarization Without Negative Influence.” *PLoS ONE* 8.11, e74516.

Mason, Lilliana (2018). *Uncivil Agreement: How Politics Became Our Identity*. Chicago: The University of Chicago Press.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (Aug. 2001). “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27.1, pp. 415–444.

Mehlhaff, Isaac D. (Aug. 2024). “A Group-Based Approach to Measuring Polarization.” *American Political Science Review* 118.3, pp. 1518–1526.

— (2025). *Mass Polarization across Time and Space*. New York: Cambridge University Press.

Mercier, Hugo and Hélène Landemore (Apr. 2012). “Reasoning Is for Arguing: Understanding the Successes and Failures of Deliberation.” *Political Psychology* 33.2, pp. 243–258.

Mercier, Hugo and Dan Sperber (2017). *The Enigma of Reason*. Cambridge, MA: Harvard University Press.

Mutz, Diana C. (2006). *Hearing the Other Side: Deliberative versus Participatory Democracy*. New York: Cambridge University Press.

Nisbet, Matthew C. and John E. Kotcher (Mar. 2009). “A Two-Step Flow of Influence? Opinion-Leader Campaigns on Climate Change.” *Science Communication* 30.3, pp. 328–354.

Norris, Pippa and John Curtice (Oct. 2008). “Getting the Message Out: A Two-Step Model of the Role of the Internet in Campaign Communication Flows During the 2005 British General Election.” *Journal of Information Technology & Politics* 4.4.

Railsback, Steven F. and Volker Grimm (2011). *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton, NJ: Princeton University Press.

Reiljan, Andres (May 2020). “”Fear and Loathing across Party Lines” (Also) in Europe: Affective Polarisation in European Party Systems.” *European Journal of Political Research* 59.2, pp. 376–396.

Rossiter, Erin L. and Taylor N. Carlson (Oct. 2024). “Cross-Partisan Conversation Reduced Affective Polarization for Republicans and Democrats Even After the Contentious 2020 Election.” *The Journal of Politics* 86.4, pp. 1608–1612.

Schneiderhan, Erik and Shamus Khan (Mar. 2008). “Reasons and Inclusion: The Foundation of Deliberation.” *Sociological Theory* 26.1, pp. 1–24.

Sobkowicz, Pawel (May 2016). “Quantitative Agent Based Model of Opinion Dynamics: Polish Elections of 2015.” *PLoS One* 11.5, e0155098.

Sunstein, Cass R. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ: Princeton University Press.

Trouche, Emmanuel, Emmanuel Sander, and Hugo Mercier (Oct. 2014). “Arguments, More than Confidence, Explain the Good Performance of Reasoning Groups.” *Journal of Experimental Psychology: General* 143.5, pp. 1958–1971.

Watts, Duncan J. and Steven H. Strogatz (1998). “Collective Dynamics of “Small-World” Networks.” *Nature* 393.6684, pp. 440–442.

Weinans, Els et al. (2024). “An Exploration of Drivers of Opinion Dynamics.” *Journal of Artificial Societies and Social Simulation* 27.1.

Westwood, Sean J. (Oct. 2015). “The Role of Persuasion in Deliberative Opinion Change.” *Political Communication* 32.4, pp. 509–528.

Wilensky, Uri (1999). *NetLogo*. http://ccl.northwestern.edu/netlogo/.

Wood, Thomas and Ethan Porter (Mar. 2019). “The Elusive Backfire Effect: Mass Attitudes’ Steadfast Factual Adherence.” *Political Behavior* 41.1, pp. 135–163.

Zaller, John R. (1992). *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.

Zhu, Jiefan et al. (2022). “An agent-based model of opinion dynamics with attitude-hiding behaviors.” *Physica A* 603, p. 127662.