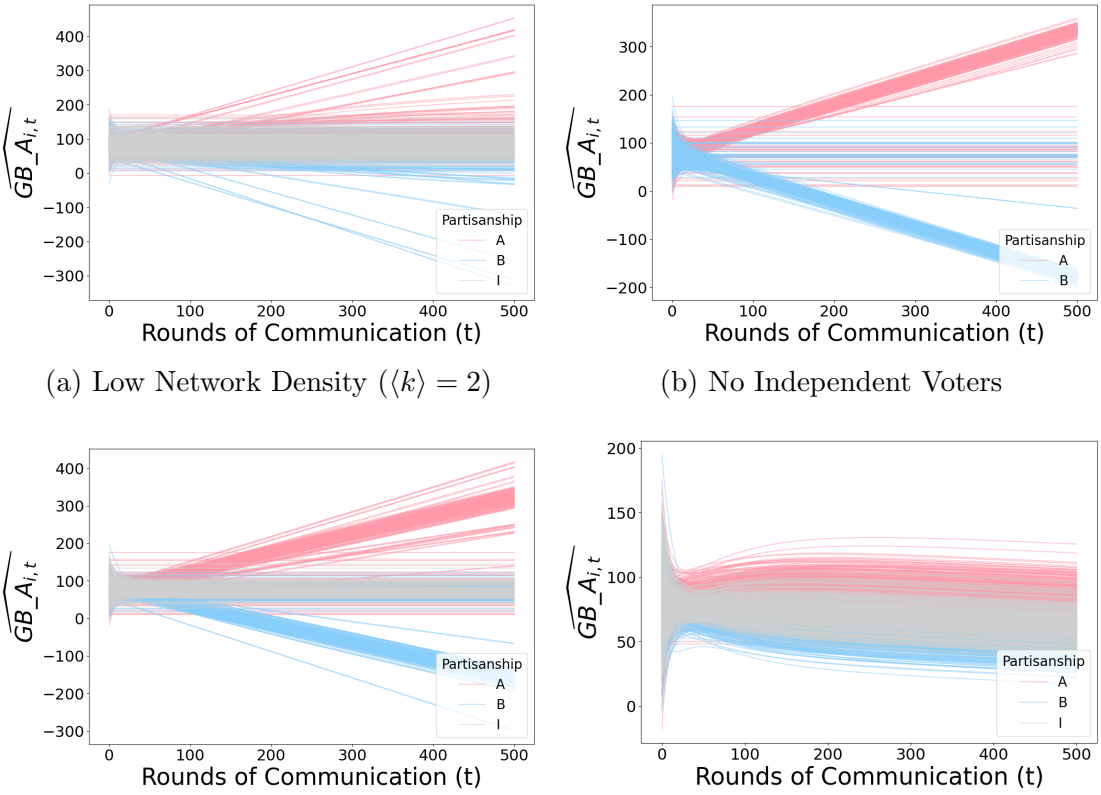
When Networks Mislead: How Partisan Communication Undermines Democratic Decision-Making

Online Appendix

1 Appendix

1.1 Conditions for Unbounded Belief Divergence in Political Networks

To further illustrate the theoretical conditions under which partisan bluffing can lead to unbounded belief divergence, we examine four alternative network configurations that systematically remove or modify the stabilizing mechanisms identified in our main analysis. These scenarios demonstrate how structural features of communication networks determine whether democratic systems converge toward bounded collective delusion or spiral toward complete detachment from reality. Figure 1 presents belief evolution patterns for candidate A quality under configurations that eliminate key circuit breakers in democratic communication.



(a) Low Network Density ($\langle k \rangle = 2$)

(b) No Independent Voters

(c) Partisans Ignore Independent Voices

(d) Independents Trust Partisan Sources

Figure 1: Conditions for Unbounded Belief Divergence in Political Networks

Figure 1(a) demonstrates how reduced network density can destabilize democratic communication by creating isolated partisan clusters with insufficient cross-cutting exposure. When average degree $\langle k \rangle$ decreases from our benchmark value of 6 to 2, the network fragments into small, homogeneous communities where partisan bluffing can operate without corrective influence from diverse information sources. The dramatic belief divergence visible in both directions—with some partisan A supporters reaching extreme positive assessments above 400 while some partisan B supporters develop extreme negative assessments below −200—illustrates how network fragmentation enables unbounded bias amplifica-

tion within isolated echo chambers. This finding reveals that democratic systems require sufficient connectivity density to maintain epistemic coherence across different partisan communities.

The most extreme breakdown occurs in Figure 1(b), which examines a purely partisan society with no independent voters. Under this configuration, partisan bluffing operates without any honest communication to provide corrective signals, leading to explosive belief divergence that accelerates throughout the communication process. Partisan A supporters develop increasingly inflated assessments that approach 500 by round 500, while partisan B supporters spiral toward assessments approaching -300, representing complete collective delusion on both sides. The symmetric but opposite trajectories demonstrate how the absence of independent "circuit breakers" allows strategic bias to compound indefinitely through network transmission, confirming our theoretical prediction that democratic systems require some proportion of honest actors to maintain stability.

Figure 1(c) reveals how trust structures determine democratic stability by examining scenarios where partisans completely ignore independent voices (setting $w_{p,I} = 0$ instead of our benchmark $w_{p,I} = 0.5$). Even though independents maintain accurate beliefs throughout the process (visible as stable gray lines near the true value), their inability to influence partisan belief formation eliminates their circuit-breaking function. The result resembles the purely partisan scenario, with explosive belief divergence occurring because partisans effectively isolate themselves from corrective information despite its availability in the network. This demonstrates that the mere presence of independent voters is insufficient for democratic stability—their voices must carry sufficient credibility with partisan actors

to moderate extreme biases.

Finally, Figure 1(d) examines the consequences of corrupting independent judgment by allowing independents to partially trust partisan sources (setting $w_{I,p} = 0.3$ instead of our benchmark $w_{I,p} = 0$). This configuration destroys the epistemic anchor function that independents serve in our main analysis, as their beliefs become contaminated by strategic bias transmitted through partisan networks. The result shows independents losing their stabilizing role and developing biased assessments that drift substantially from truth, while partisan groups experience amplified divergence because their traditional source of corrective information has been compromised. This scenario illustrates how democratic stability depends critically on preserving institutions and actor types that remain insulated from partisan strategic incentives.

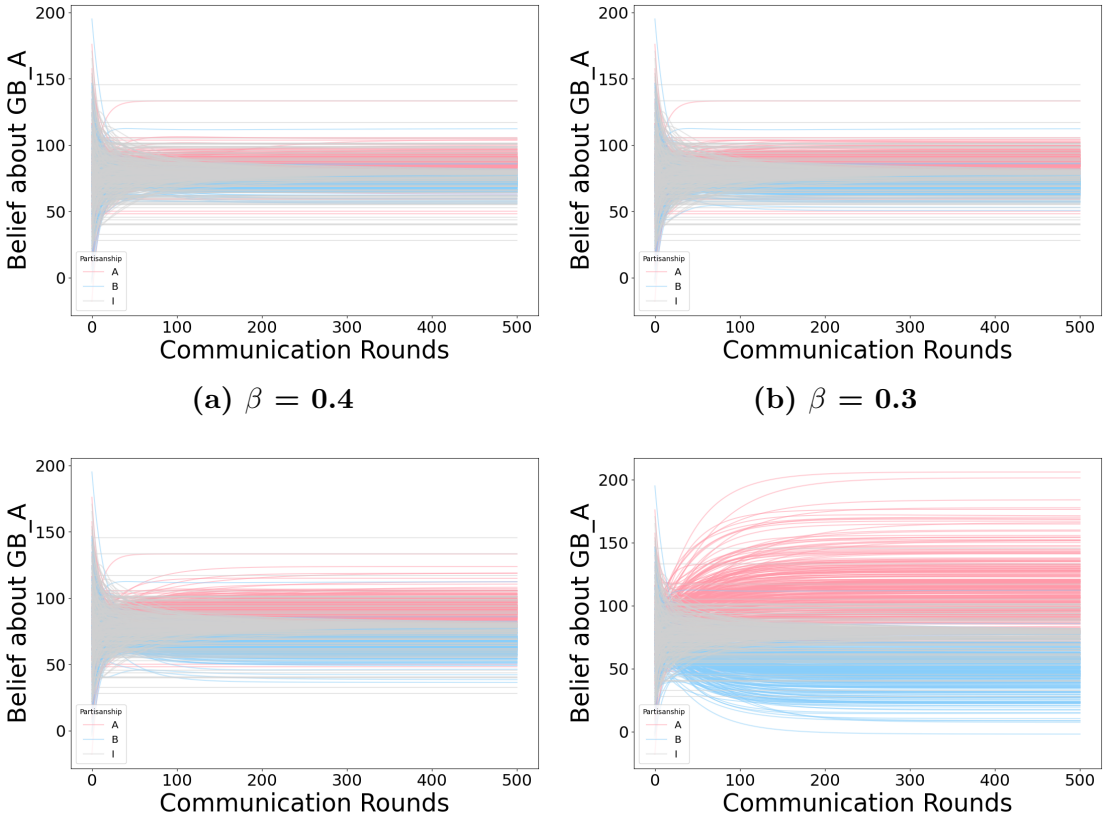
These results provide crucial insights into the structural prerequisites for democratic stability in networked communication environments. They demonstrate that bounded collective delusion—while problematic—represents a preferable outcome to the unbounded belief divergence that occurs when key stabilizing mechanisms are removed. The findings suggest that institutional designers should focus on preserving sufficient network connectivity, maintaining populations of independent actors, ensuring that independent voices carry credibility with partisan audiences, and protecting the epistemic independence of honest communicators from strategic contamination. Without these structural safeguards, democratic communication systems can spiral toward complete detachment from reality, making accurate collective choice impossible regardless of the objective quality differences between electoral alternatives.

1.2 Individual-Level Belief Trajectories Across Distributional Parameters

This appendix provides detailed visualization of how individual voter beliefs evolve over time under different configurations of the partisan benefit distribution. Using our benchmark scenario where $GB_A = 77.9$ and $GB_B = 46.4$, we examine belief trajectories when partisan benefits follow a Gamma(1, β) distribution for varying values of the shape parameter β. These individual-level dynamics illuminate the precise mechanisms through which extreme partisan preferences contaminate collective belief formation and drive supporters of inferior candidates toward systematic collective delusion.

Figure 2 presents belief trajectories for all 1,000 voters across four different distributional configurations, with each line representing an individual voter's assessment of candidate A quality over 500 communication rounds. The color coding distinguishes partisan A supporters (red), partisan B supporters (blue), and independents (gray), revealing how different distributional assumptions generate dramatically different patterns of belief convergence and divergence.

The progression across panels reveals the profound influence of distributional tail behavior on democratic communication dynamics, demonstrating how increasing prevalence of extreme partisans systematically degrades collective belief formation. Figure 2 (a) presents the baseline case where $\beta = 0.4$ creates a highly concentrated distribution with minimal extreme preferences. Under this configuration, belief trajectories for all voter types converge toward truth with minimal systematic bias, demonstrating that democratic communication can achieve its theoretical potential when strategic incentives remain moderate. The tight clus-



(a) $\beta = \mathbf{0.4}$ **(b)** $\beta = \mathbf{0.3}$

(c) $\beta = \mathbf{0.2}$ **(d)** $\beta = \mathbf{0.1}$

Figure 2: Individual Belief Evolution Under Alternative Partisan Benefit Distributions

tering of individual trajectories around true values illustrates how concentrating partisan preferences enables beneficial information aggregation to dominate harmful bias amplification.

Figure 2 (b) reveals the initial signs of democratic degradation as $\beta = 0.3$ introduces a longer distributional tail with more extreme partisan preferences. While most belief trajectories remain relatively stable, we begin to observe systematic bias among partisan B supporters with increased variance across individual voters. The systematic bias among partisan B supporters becomes visible

but remains bounded, suggesting that moderate increases in extreme preferences can limit the scope of accurate collective judgment without creating complete collective delusion. Independent voters (gray lines) maintain consistent trajectories across all distributional configurations, confirming their role as epistemic anchors that provide stable reference points regardless of partisan dynamics.

The deterioration accelerates in Figure 2 (c), where $\beta = 0.2$ creates a moderate tail distribution that substantially amplifies bias effects without completely destabilizing the system. Partisan B supporters exhibit systematic overestimation of candidate A quality, with the range of individual trajectories widening considerably compared to the concentrated distribution scenario, and the rate of belief divergence increasing noticeably. This intermediate case reveals the nonlinear relationship between distributional parameters and democratic performance: modest increases in the prevalence of extreme partisans can generate large deteriorations in collective accuracy through cascading effects in network communication.

Finally, Figure 2 (d) demonstrates the complete breakdown of democratic communication when $\beta = 0.1$ creates a heavy-tailed distribution with numerous extreme partisans. We observe dramatic belief polarization that persists throughout the entire communication process, with partisan B supporters (blue lines) exhibiting severe upward drift in their assessments of candidate A, many individual trajectories reaching assessments above 150 despite the true value of 77.9. This systematic overestimation reflects the amplified influence of extreme partisan B voters who transmit heavily biased signals that contaminate belief formation throughout their network neighborhoods. The wide dispersion of belief trajectories among partisan B supporters indicates that extreme preferences cre-

ate heterogeneous bias amplification, with voters embedded in different network configurations experiencing varying degrees of collective delusion.

These individual-level dynamics provide crucial evidence for our theoretical framework by demonstrating how certain voters with extreme partisan benefits function as "bias super-spreaders" who contaminate collective belief formation far beyond their numerical representation. The visualization reveals that extreme partisans do not simply hold biased beliefs in isolation—they actively transmit these biases through network communication, creating cascading effects that shift entire partisan communities toward collective delusion. The progressive amplification of bias effects as distributional tails expand confirms that democratic vulnerability depends critically on the prevalence of extreme preferences rather than just the mean level of partisan attachment.

Moreover, the persistence of biased trajectories across hundreds of communication rounds demonstrates that strategic communication creates stable equilibria that resist correction through continued social interaction. Unlike random errors that cancel through information aggregation, the systematic biases generated by extreme partisans compound through network transmission, leading to permanent departures from truth that persist indefinitely. This finding challenges optimistic assumptions about the self-correcting nature of democratic deliberation and highlights the importance of institutional mechanisms that can moderate extreme preferences or limit their influence on collective decision-making.

1.3 Interactive Effects of Candidate Quality and Partisan Benefit Distributions

The relationship between candidate quality differences and partisan benefit distributions creates complex interaction effects that determine when democratic systems remain robust versus when they become vulnerable to systematic failure. Figure 3 provides a comprehensive three-dimensional analysis of these interactions based on 1,000 simulations across systematically varied combinations of global benefit differences and distributional parameters.

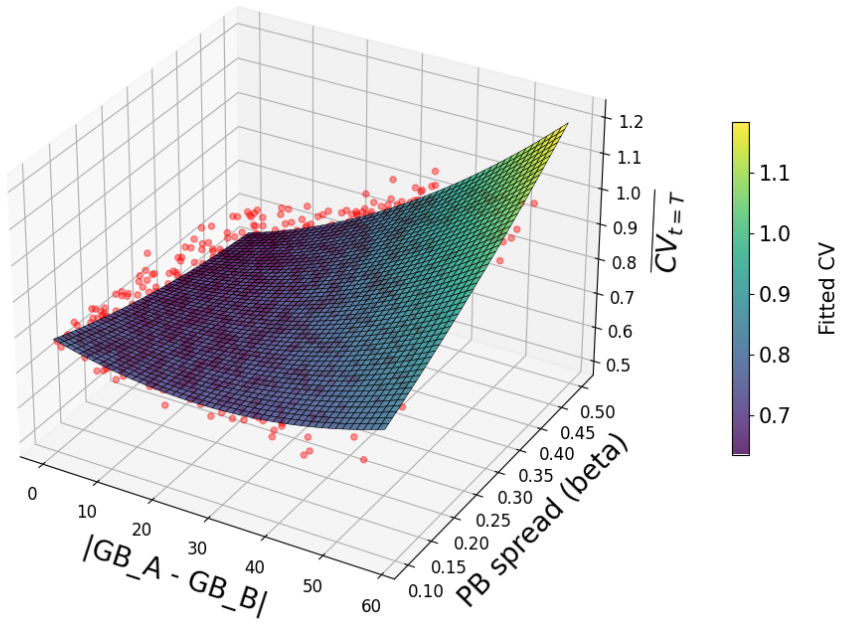


Figure 3: Interactive Effects of Candidate Quality Differences and Partisan Benefit Distributions

The surface plot reveals several critical patterns that illuminate the conditions under which democratic performance transitions between different regimes.

When candidate quality differences are large (right region of the surface), democratic performance remains remarkably robust even when extreme partisans are numerous, as overwhelming objective evidence can overcome even heavily biased communication. The mathematical intuition is straightforward: when true signal strength greatly exceeds the magnitude of strategic distortion, honest information dominates partisan bluffing regardless of network composition or distributional characteristics.

However, when quality differences are moderate—precisely the range we identified as most problematic in our main analysis—the presence of extreme partisans can completely destabilize democratic decision-making, creating sharp performance cliffs where small increases in extreme preferences generate dramatic deteriorations in collective accuracy. These cliff effects occur because moderate candidate differences create optimal conditions for strategic bias to overcome truthful signals: when true quality differences become comparable to the magnitude of bias that extreme partisans introduce, network amplification can systematically distort collective beliefs away from objective reality.

The surface exhibits smooth, continuous transitions rather than abrupt discontinuities, indicating that democratic performance degrades gradually as distributional parameters shift rather than collapsing suddenly at critical thresholds. This gradual transition pattern has profound implications for understanding democratic vulnerability in contemporary societies. Rather than facing sudden systemic failures that would trigger immediate institutional responses, democratic systems may experience slowly mounting vulnerability as partisan sorting increases extreme preferences, with degradation remaining largely invisible until competitive

elections reveal the extent of collective delusion among supporters of inferior candidates. This finding suggests that monitoring the distributional characteristics of partisan preferences may be as important as tracking mean levels of polarization for assessing democratic health and designing early warning systems for institutional breakdown.