Learning from Shared News:
When Abundant Information Leads to Belief Polarization*

Renee Bowen†
Danil Dmitriev‡
Simone Galperti§

April 26, 2022

Abstract

We study learning via shared news. Each period agents receive the same quantity and quality of first-hand information and can share it with friends. Some friends (possibly few) share selectively, generating heterogeneous news diets across agents akin to echo chambers. Agents are aware of selective sharing and update beliefs by Bayes' rule. Contrary to standard learning results, we show that beliefs can diverge in this environment leading to polarization. This requires that (*i*) agents hold misperceptions (even minor) about friends' sharing and (*ii*) information *quality* is sufficiently low. Polarization can worsen when agents' social connections expand. When the *quantity* of first-hand information becomes large, agents can hold opposite extreme beliefs resulting in severe polarization. We find that news aggregators can curb polarization caused by news sharing. Our results hold without media bias or fake news, so eliminating these is not sufficient to reduce polarization. When fake news is included, it can lead to polarization but *only* through misperceived selective sharing. We apply our theory to shed light on the evolution of public opinions about climate change in the US.

JEL codes: D82, D83, D90

Keywords: polarization, echo chamber, selective sharing, selection neglect, learning, information quality, fake news, misspecification

---

*We are grateful to the editor and four anonymous referees for very helpful comments. We also thank S. Nageeb Ali, Jon Eguia, Myles Ellis, Harry Pei, Jacopo Perego, Joel Sobel, Jiemau Wu, and participants in conferences and seminars at UCSD, PhDEI, SIOE, PennState, NBER POL, UC Berkeley, UC Davis, NYU, MEDS, Harvard, Chicago, WEAI, NASMES, CETC, and Stony Brook for helpful comments and suggestions. All remaining errors are ours.

†Corresponding author: RBC1407, 9500 Gilman Dr, La Jolla, California 92093, trbowen@ucsd.edu

‡ddmitrie@ucsd.edu

§sgalperti@ucsd.edu

1 Introduction

Recent decades have witnessed rising polarization in politics, media, and public opinions—especially in the US. In their review about this phenomenon, Fiorina and Abrams (2008) stress the importance of focusing on polarization in opinions. Pew Research Center (2014) notes that both the left and right ends of the political spectrum have a greater impact on the democratic process than those with mixed views—by being the most likely to vote, donate to campaigns, and participate directly in politics. Alesina et al. (2020) find that "Americans are polarized not only in their views on policy issues and attitudes towards government and society, but also about their perceptions of the same, factual reality." Social divisions, in general, have been linked to negative economic and political outcomes, such as inequality, political gridlock, weak property rights, low investment or growth. As a consequence, economists have devoted significant attention to its causes.[1]

The causes of belief polarization may seem obvious at first glance, but reality presents several puzzles. Some have blamed misinformation or fake news. However, this explanation is at odds with recent evidence that people are reasonably good at detecting real from fake news (Angelucci and Prat (2021)). Others have blamed the ease of access to distorted information through the Internet.[2] However, the Internet has also brought an abundance of information, which should lead people to learn quickly and beliefs to converge (not diverge) according to standard economic models. The evidence on the effects of the Internet is also mixed: Some papers find that social media increases polarization (e.g., Allcott et al. (2020)),

[1] Papers about polarization in the US include Fiorina and Abrams (2008); Pew Research Center (2014, 2020b); Egan and Mullin (2017); Desmet and Wacziarg (2018); Bertrand and Kamenica (2018); Funk and Tyson (2021). Papers about the consequences of polarization include Zak and Knack (2001); Keefer and Knack (2002); Bartels (2008); Bishop (2009); McCarty et al. (2009); Gilens (2012); Barber and McCarty (2015).

[2] For various discussions on these causes of polarization, see "Facebook Throws More Money at Wiping Out Hate Speech and Bad Actors", Wall Street Journal (May 15, 2018); "Should the Government Regulate Social Media?", Wall Street Journal (June 25, 2019); Periser (2011); Flaxman et al. (2016); Sunstein (2017); Azzimonti and Fernandes (2018); Tucker et al. (2019); Zhuravskaya et al. (2020).

others find more polarization offline than online (e.g., Boxell et al. (2018); Zhuravskaya et al. (2020)).[3] Finally, some have blamed consumption of news from biased echo chambers. This explanation is also incomplete: If people get some (even small) amounts of unbiased informative news—which is often the case (Pew Research Center (2014))—they should still learn the truth according to standard economic models.

To shed light on these issues and clarify possible drivers of belief polarization, this paper offers a theoretical framework that focuses on how people learn from shared news. We ask if polarization can result merely from how people consume and share information through social connections, whether online or offline. If so, will technology-driven abundance and sharing of information reduce or increase polarization? Our answers rely on a simple, yet flexible, model that incorporates two key empirical findings. First, people often share information *selectively* (for instance, by remaining silent when information is unfavorable), which causes their listeners to consume unbalanced diets of second-hand news. Second, people tend to *neglect selection* in their second-hand information and, thus, misperceive its content. Misperception is akin to reading too little into the absence of news from friends, or more intuitively, reading too little into their silence.[4] We find that this misperception and the quality of information play critical roles in causing some people to learn incorrectly while others learn correctly. This causes their beliefs to polarize. Our theory requires neither preexisting differences in people's worldviews nor misinformation. It can explain why changes in people's information ecosystem triggered by the Internet may affect polarization.

In our model agents learn about a binary state of the world, A or B. In every period,

[3]For further discussion on the relationship between polarization and social media, see also Allcott and Gentzkow (2017); Bursztyn et al. (2019); Mosquera et al. (2019); Levy (2021). For recent reviews of this literature see Barberá (2020) and Zhuravskaya et al. (2020).

[4]For evidence of selective sharing, see Shin and Thorson (2017); Weeks et al. (2017); Shin et al. (2018); Pogorelskiy and Shum (2019); Levy (2021); Zhuravskaya et al. (2020). Unbalanced news diets appear in a wealth of evidence on echo chambers and media bubbles (Levy and Razin, 2019; Zhuravskaya et al., 2020, for a review). Bertrand and Kamenica (2018) also stress the importance of media diet driving social differences. Evidence on selection neglect appears in Esponda and Vespa (2018), Pogorelskiy and Shum (2019) Enke (2020), Esponda et al. (2021), and Jin et al. (2021), which we review in Section 3.

each agent receives first-hand information—an objective i.i.d. signal about the state—with probability γ, and no signal otherwise. We refer to the signal's informativeness—i.e., how likely it is to match the true state—as its *quality*. In every period, each agent can stay silent or share her signal with her social connections, called *friends*. We capture selective sharing by assuming that some agents, called *dogmatic*, share only signals supporting one state; the other agents, called *normal*, share every signal. For example, a dogmatic agent may be someone who holds a particularly strong view on whether to vaccinate children and share only articles in favor of that view; normal agents share any article on the topic. We refer to an agent's sources of second-hand signals as her echo chamber, because these constitute the group of people she is connected to for the purpose of news sharing. As suggested by the empirical literature, we capture selection neglect by assuming that agents have a misspecified mental model of the signal process. That is, each agent believes that each of her friends gets a signal with probability $\hat{\gamma} < \gamma$. Observationally, this is consistent with the agent excessively attributing her friends' silence to absence of news rather than suppression of news. Except for this misspecification, our agents update beliefs using standard Bayes' rule.

Our first contribution is to characterize how agents learn in this environment. Consistent with the prevailing narrative, we find that echo chambers can lead people to learn incorrectly, but this result comes with important qualifications—conceptually and practically. Conceptually, selective sharing by itself cannot lead to incorrect learning, let alone polarization. This is because if an agent fully understands how her friends select what to share, she will adjust for it and her beliefs will not be systematically distorted. People's misperception of selection in their news is necessary for echo chambers to distort beliefs. But this is still not enough. Moreover, we show that information quality also needs to be sufficiently low for distortions to occur, which has practical implications as we explain in the paper.

To be more specific, we analyze learning in the short and long run, i.e., after one round of signals (scarce information) and after infinitely many rounds (abundant information). In the short run, our agent's *expected* posterior can differ from her prior. To see why, suppose $\hat{\gamma} \approx 0$. In this case, one can show that it is as if the agent takes signals at face value and updates her belief towards the state supported by the majority of just the signals she actually gets from her friends. This creates two distorting forces, which potentially act in opposition. Suppose the agent believes ex ante that state A is (objectively) more likely. Then, her friends tend to receive more signals that support A on average, pushing the majority of shared signals towards A. Thus, selective sharing and misperception create a form of confirmation bias distorting the agent's expected posterior in the direction of her prior. Note that this force distorts beliefs even if the agent has a balanced echo chamber, namely the same number of B-leaning and A-leaning dogmatic friends. To see the second force, suppose the agent has more B-leaning than A-leaning dogmatic friends. This imbalance pushes the majority of shared signals towards B. But for this force to dominate, information quality needs to be sufficiently *low*. To see why, suppose to the contrary that information quality is very high. Given the prior, the B-leaning dogmatists most likely get signals supporting A and stay silent, so on average a majority of shared signals will still support A. In this case, if A is sufficiently more likely ex ante, the agent's posterior can be distorted towards A even if she has a majority of B-leaning dogmatic friends.

In the long run, abundant information boosts the distorting power of echo chambers, thereby exaggerating incorrect learning. But these distortions require unbalanced echo chambers, unlike in the short run—distortions due purely to the confirmation bias are wiped out as the effect of the prior goes away. For information quality below a specific threshold, the agent's posterior converges to a belief that assigns probability one to the state favored by the majority of her dogmatic friends, *irrespective* of the truth. For higher quality, her belief converges to the truth despite the effects of her echo chamber. These short- and long-run properties hold even if the agent does not take shared signals at face value (i.e., not just for $\hat{\gamma} \approx 0$). They remain true when selection neglect is minimal, i.e., when $\hat{\gamma}$ is very close to the true γ. Thus, even minor misperceptions can distort learning.

Our second contribution is to show how these distorting forces at the individual level can cause polarization at the social level. To begin, we emphasize the central role of information quality. If some agents have unbalanced echo chambers towards different states *and* information quality is sufficiently low, their beliefs will move apart on average in the short run and almost surely in the long run. However, note that in our setting polarization does not mean that all agents with echo chambers leaning towards a state polarize in the direction of that state. Indeed, for intermediate information quality, some agents with echo chambers leaning moderately against the true state can still learn it correctly. We find that raising information quality can *increase* polarization. Though perhaps unexpected, this has a simple intuition. If sufficiently many agents learn incorrectly with low information quality, increasing it causes some to start learning correctly. This shrinks the gap between the number of correct and incorrect learners and, thus, creates a more divided society. Next, we analyze how the expansion of social connections can be another driver of polarization, even though more connections may provide additional information. Fixing its quality, we obtain empirically testable conditions on how the internal structure of echo chambers has to change for polarization to fall. In short, this happens if agents' normal friends grow sufficiently faster than their dogmatic friends. Instead, a proportional growth of echo chambers that leaves their shares of dogmatic friends unchanged can exacerbate polarization.

Our theory allows us to investigate what policies may reduce the effects of news sharing on polarization. The central role of information quality suggests some solutions. An obvious one is that news outlets provide higher-quality information, but this may be hard to incentivize. Another is to exploit news aggregators. Although they may exist for other reasons, we show how aggregators can provide higher-quality information even when they *lose* some information by summarizing facts.

We illustrate our results through two quantitative exercises. The first simulates the evolution of beliefs with and without misperception of selective sharing in unbalanced echo chambers. It shows that, even when misperceptions are small (i.e., $\hat{\gamma}$ is close to γ), a large divergence in beliefs can occur, and quickly. The second exercise uses our model to replicate qualitative features of climate-change opinions in the US over the last few decades. In particular, Saad (2021) provides data between 1997 and 2021 on the views of Democrats and Republicans on whether the effects of global warming have already begun. In 1997, 46% of each group viewed those effects as already happening. By 2021, only 29% of Republican held this view, in contrast to 82% of Democrats. Our simulations suggest that modest misperception and small echo-chamber imbalance suffice to explain these patterns. Moreover, the observed polarization between Democrats and Republicans may not arise only from differences in their news diets, but also requires consuming sufficiently low quality information. This may be the result of well-documented skepticism campaigns by interest groups that began in the early 1990s (Egan and Mullin (2017)). Finally, the data shows an acceleration in polarization around 2011, which we can replicate with an expansion of echo chambers that mimics the surge in social-media use around that time. These simulations point to how the model can help disentangle multiple factors related to news sharing that may contribute to polarization on various topics.

Finally, our analysis goes to the heart of how new communication channels and formats on the Internet can affect polarization. They can lower information quality in some cases. For instance, tweets and social-media posts tend to be short and few people read the linked articles (Bakshy et al. (2015); Gabielkov et al. (2016)). People may also misperceive how news-feed algorithms work on social media, which we model with alternative misspecifications and show that they have similar implications to selection neglect (Section 7). All this can lead to polarization, even without deliberate misinformation. Yet, the Internet has arguably magnified the spread of fake news through social connections. As a byproduct of our analysis, we find that selective sharing is one (and in a sense the only) channel through which fake news can cause polarization. This could explain why fake news have become particularly concerning in recent decades. However, the Internet can also curb polarization—for instance, by helping people access high-quality information and form broader and more-balanced social connections.

Related Literature The economics literature discusses at least three possible causes of belief polarization. The one most closely related to our work is behavioral biases.[5] We highlight misperception of selective sharing, building on recent experimental evidence. Section 3 reviews this evidence in detail. We offer a tractable and flexible way to model this misperception that uncovers its consequences for polarization. A second cause of polarization is heterogeneity in preferences (Dixit and Weibull (2007); Pogorelskiy and Shum (2019)), which would exacerbate the polarization we find. A third cause is biased or multidimensional information sources.[6] We assume unbiased sources of first-hand information—which can be interpreted as media outlets. This allows us to focus on selective sharing and show that removing all media biases may not suffice to curb polarization.

Our paper fits into the growing literature on model misspecification and learning, starting from the classic work of Berk (1966). We analyze short- and long-run learning in a more specialized model and demonstrate the interaction with social information sharing to generate polarization. Bohren (2016), Bohren and Hauser (2018), and Frick et al. (2020) analyze social learning under model misspecification. In particular, Bohren and Hauser (2018) study when agents with different models of the world have no limit beliefs (i.e., beliefs cycle) or different limit beliefs (disagreement). Mailath and Samuelson (2020) also consider agents with different models of the world. Although close in spirit, our disagreement results are driven by a fundamentally different mechanism, as all our agents have the same model of

[5]See, e.g., Levy and Razin (2018); Hoffmann et al. (2019); Enke et al. (2020).

[6]See, e.g., Mullainathan and Shleifer (2005); Andreoni and Mylovanov (2012); Levendusky (2013); Conroy-Krutz and Moehler (2015); Reeves et al. (2016); Perego and Yuksel (2022).

the world. We also emphasize the role of information quality and its implications for curbing polarization. Molavi et al. (2018) study incorrect learning in social networks when non-Bayesian agents exhibit imperfect recall. By contrast, our agents are Bayesian, which allows us to leverage familiar methods for studying the effects of misperception of selective sharing across a variety of environments and applications.[7]

The idea of an echo chamber as the group of friends from whom one gets information is key in our model. This links our work to the literature on Bayesian and non-Bayesian learning in networks.[8] One closely related paper is Levy and Razin (2019), which shows that an updating heuristic called "Bayesian Peer Influence" can cause beliefs to polarize. However, their notion of polarization is that the entire society's *consensus* shifts towards a common extreme belief. By contrast, our notion is that agents' beliefs diverge to different extremes.

The verifiability of shared information and the possibility of not receiving first-hand information renders our model similar to Dye (1985). Allowing for this possibility is one often-used way to give selective sharing a chance to be effective: Otherwise, silence can be immediately interpreted as bad news (e.g., Ben-Porath et al., 2018; DeMarzo et al., 2019).

2 Model

We study a stylized model of learning from information shared through social connections. We provide the formal details here and discuss the main assumptions and supporting evidence in Section 3.

Time t is discrete, where $t = 0, 1, 2, \ldots$. A state of the world $\omega \in \{A, B\}$ realizes at $t = 0$.

[7]A nonexhaustive list of other recent work on misspecified learning includes Nyarko (1991); Esponda and Pouzo (2016); Fudenberg et al. (2017); He (2018); Heidhues et al. (2018); Jehiel (2018); Esponda et al. (2019); Ba and Gindin (2020); Dasaratha and He (2020); He and Libgober (2020); Frick et al. (2020); Fudenberg et al. (2020); Li and Pei (2020).

[8]See DeMarzo et al. (2003); Golub and Jackson (2010); Eyster and Rabin (2010); Acemoglu et al. (2010); Perego and Yuksel (2016); Azzimonti and Fernandes (2018); Pogorelskiy and Shum (2019); Spiegler (2020).

For example, ω can represent whether the effects of global warming have already begun, or whether vaccines can harm children. There is a fixed group of agents who seek to learn ω.

Information. Each agent receives first-hand information from original sources and second-hand information shared by other agents. For each $t \geq 1$, agent i receives first-hand information with probability $\gamma \in (0, 1]$ in the form of a private signal $s_{it} \in \{a, b\}$; with probability $1 - \gamma$ she receives no signal. Signals are partially informative:

$$\mathbb{P}(s_{it} = a | \omega = A) \quad = \quad \mathbb{P}(s_{it} = b | \omega = B) = q, \tag{1}$$

where $\frac{1}{2} < q < 1$. We refer to q as the information *quality*. The events of receiving a signal and its realization are i.i.d. across agents and time.[9]

Selective Sharing. Agents share their first-hand information with other agents with whom they have a social connection. We call these connections *friends*. After receiving s_{it}, agent i can share it with all her friends or stay silent. If she receives no signal, she stays silent. Thus, she can selectively suppress information, but cannot fabricate information. That is, sharing signals takes the form of verifiable information. Concretely, an agent can share a newspaper article, but cannot edit its content. We intentionally rule out tampering with or fabricating information (e.g., fake news) to highlight the role of selective sharing.

We introduce three types of information-sharing behavior. An agent is *normal* if she shares any s_{it}, *A-dogmatic* if she shares only $s_{it} = a$, and *B-dogmatic* if she shares only $s_{it} = b$. One interpretation is that some agents dogmatically believe in their conviction that only one state is true and share only information that supports it. Each agent's type is exogenous and known to her friends.

[9] In reality, people receive correlated news. However, strong evidence suggests that they often neglect correlation, especially in second-hand news (Enke and Zimmermann, 2017; Eyster et al., 2018; Pogorelskiy and Shum, 2019). Under correlation neglect, we can allow for arbitrary correlation between the agents' signals within each period and our main results are qualitatively unchanged.

Selective news sharing contributes to creating heterogeneous information diets (Pew Research Center, 2014; Levy and Razin, 2019). Agent i's diet depends on the composition of friends she listens to, namely the number d_{Ai} of A-dogmatic friends, d_{Bi} of B-dogmatic friends, and n_i of normal friends. We refer to $e_i = (d_{Ai}, d_{Bi}, n_i)$ as i's *echo chamber*. If $d_{Ai} \neq d_{Bi}$, we say that i's echo chamber—hence, her information diet—is *unbalanced* and we refer to $d_{Ai} - d_{Bi}$ as its *imbalance*. Otherwise, we say that e_i is balanced. Hereafter, we refer to the majority (minority) of an agent's dogmatic friends as her dogmatic majority (minority). In reality people may also have heterogeneous news diets because they choose to listen to different first-hand sources of information. We abstract from this aspect to focus on the effects of news sharing.[10]

Timing. Within each period the timing is as follows: (1) signals realize; (2) each agent i receives s_{it} with probability γ; (3) each agent i shares her signal (if any) with friends as specified by her type; (4) agents update beliefs based on all received signals.

Beliefs. We are interested in the beliefs of normal agents. They have a common prior $\pi \in (0, 1)$ that $\omega = A$. Given a sequence $\mathbf{s}_i^t$ of information that agent i receives up to t (i.e., her signals, her friends' shared signals, and their silence), let $\mu(\mathbf{s}_i^t)$ be her Bayesian posterior that $\omega = A$. To examine learning in the short run, we will consider $\mu(\mathbf{s}_i^1)$; to examine learning in the long run and so the effects of abundant information, we will consider the (probability) limit of $\mu(\mathbf{s}_i^T)$ as $T \to \infty$, denoted by $\mu(\mathbf{s}_i^\infty) = \text{plim}_{T\to\infty} \mu(\mathbf{s}^T)$. We will introduce a formal measure of belief polarization in Section 5. However, intuitively, polarization requires that agents' beliefs move *systematically* apart. It is well known that $\mu(\mathbf{s}_i^1)$ and $\mu(\mathbf{s}_j^1)$ can differ in completely standard Bayesian models simply because agents i and j observe different signal realizations. Therefore, we adopt a more demanding condition for short-run polarization

[10]Section 7 discusses several richer models of news sharing. These include probabilistic selective sharing, endogenizing selective sharing, news re-sharing, and more heterogeneity across agents in addition to their echo chambers.

that looks at differences between the expectations $\mathbb{E}[\mu(\mathbf{s}_i^1)]$ and $\mathbb{E}[\mu(\mathbf{s}_j^1)]$. Another reason for considering $\mathbb{E}[\mu(\mathbf{s}_i^1)]$ is that we can view agent i as representative of a large group of individuals that are similar within their group, but differ between groups. Then, by the Law of Large Numbers $\mathbb{E}[\mu(\mathbf{s}_i^1)]$ approximates the empirical average belief of the respective group and may be used to study intra-group polarization.

One might think that selective sharing and unbalanced echo chambers should suffice to give rise to belief polarization. This is not the case. Hereafter, let $I_{\{\omega=A\}}$ equal 1 if $\omega = A$ and 0 otherwise.

Remark 1. For any echo chamber e_i and $\gamma \in (0,1]$, we have

$$\mathbb{E}\left[\mu(\mathbf{s}_i^1)\right] = \pi \qquad \text{and} \qquad \mu(\mathbf{s}_i^\infty) = I_{\{\omega=A\}}.$$

If an agent precisely understands the effects of her echo chamber on her information diet, selective sharing simply results in a specific information structure that is perhaps less informative than under full sharing. Nonetheless, her belief must satisfy standard properties of Bayesian updating.

Misperception. To overcome the challenge posed by Remark 1, we assume that agents misperceive their second-hand information by neglecting, at least partially, its selection. To generate this in a tractable way, we assume that each agent perceives that the i.i.d. probability of getting a signal is $\hat{\gamma} \in (0,1]$ where $\hat{\gamma} < \gamma$. The agent continues to use Bayes' rule to calculate $\mu(\mathbf{s}^t)$, yet applied to this slightly misspecified model of the world. As a result, she updates as if she excessively treats silence as absence of news and thus partially ignores news selection. In other words, she reads too little into friends' silence.

3 Evidence on Selected News and Misperception

We discuss the supporting evidence for our main assumptions: selective news sharing and misperception in the form of selection neglect. The analysis beginning in Section 4 does not rely on anything mentioned here, so the reader may skip this section without confusion.

3.1 Selective News Sharing

Similar to previous literature (e.g., Acemoglu et al., 2010, 2013), our baseline model takes the types of news-sharing behavior as given. One reason is that our focus is not understanding *why* people tend to share some news more than others, but understanding the *consequences* of this tendency for social learning. The types of selective sharing we assume are consistent with the types observed in the empirical literature on information disclosure (e.g., Pogorelskiy and Shum, 2019; Jin et al., 2021, and references therein). Moreover, a mounting body of evidence about news sharing online suggests that individuals often share selectively (Bakshy et al., 2015; Del Vicario et al., 2016; Flaxman et al., 2016; Halberstam and Knight, 2016; Quattrociocchi et al., 2016; Schmidt et al., 2018; Ookalkar et al., 2019). Section 7.3 discusses selective sharing in more detail, including its possible endogeneity.

3.2 Misperception and Selection Neglect

In Enke (2020) (hereafter, simply Enke), selection neglect manifests itself in beliefs that take selection too little into account and the received information too much at face value (what Kahneman (2011) calls "what you see is all there is"). Qualitatively, this is what happens in our model when $\hat{\gamma} < \gamma$, as explained above. A significant share of Enke's subjects appears to fully neglect selection, which is observationally consistent with $\hat{\gamma}$ close to zero, but other subjects seem to take selection into account, albeit partially (see Enke's Figure 2). Partial selection neglect is consistent with $\hat{\gamma}$ far from zero and illustrates that degrees of selection neglect are consistent with human behavior. Enke's findings show that selection neglect can be weakened, but not removed. Intermediate values of $\hat{\gamma}$ allow us to capture varying degrees of selection neglect, and simulations in Section 6 show that a value of $\hat{\gamma}$ very close to γ is sufficient to generate polarization seen in the data.

Evidence for partial selection neglect has been found in other papers. Jin et al. (2021) study verifiable information disclosure through an experiment. One of their main findings is that "receivers [...] are insufficiently skeptical about undisclosed information—the extent to which no news is bad news;" but they *do* respond to non-disclosure to some extent. Jin et al. (2021) also provide an extensive review of papers on voluntary disclosure with similar evidence of insufficient, yet not zero, skepticism. Esponda and Vespa (2018) study learning from evidence that involves sample selection due to unobservables. Unobservables can include other agents' private information that drives their choices (as selective sharing in our model). They find that subjects do not understand selection and respond to the observed data but not to the possibility that the data may be biased by selection. However, they also find that *quantitatively* subjects partially account for selection. Esponda et al. (2021) study base-rate neglect: the phenomenon whereby people ignore information in their prior when updating, in contrast to the Bayesian benchmark. Our model with $0 < \hat{\gamma} < \gamma$ can be interpreted as consistent with base-rate neglect when it comes to responding to silence: One can show that if an agent neglects, even partially, the prior probability that a friend is uninformed, she will overestimate the likelihood of no signal arrival from observing silence and hence excessively treat silence as absence of news. Finally, in an experiment on the effects of shared news on voting outcomes, Pogorelskiy and Shum (2019) find that subjects share news selectively and tend to be partially unresponsive to others' selective sharing.

3.3 Incorrect Mental Models

The papers above discuss psychological mechanisms that may drive selection neglect. Enke (2020) argues that neglect likely results from people having an incorrect mental model of the data-generating process, whereby some key aspects (e.g., suppressed information) do not come to mind or are only partially accounted for due to computational complexity and cognitive overload. Such incorrect models may result from intuitive system-1 reasoning (Kahneman (2011)), or from relying on a class of problems the agents know how to solve. This explanation is also related to the idea of the "naive intuitive statistician" in cognitive psychology (Fiedler and Juslin, 2006; Juslin et al., 2007). Jin et al. (2021) suggest that their subjects use an incorrect model exhibiting naivete about strategic use of non-disclosure. Esponda and Vespa (2018) view their findings as a failure by the agents to learn that their mental model is incorrect. Esponda et al. (2021) also conclude that their findings "indicate that an incorrect mental model [...] is the main driver behind [...] biased beliefs."

Our assumption of $\hat{\gamma} < \gamma$ is also consistent with a phenomenon called "illusory superiority" or "better-than-average" heuristic in psychology (Cross, 1977; Svenson, 1981; Odean, 1998; Zuckerman and Jost, 2001). People often have unjustifiably favorable views of themselves relative to others on various characteristics, which may include how well informed they are or how good they are at getting and understanding information. This can lead an agent to incorrectly think that others are systematically *less* informed than they are. It is important to note that in our model an agent's belief about the arrival rate of her own signals is irrelevant for how she learns. This is because if she does not get a signal, she knows it; if she gets a signal, she knows its meaning by knowing (1). Therefore, our analysis extends to settings where an agent is misspecified only about the news arrival rate of her sources of second-hand information. Agents may know the true arrival rate of their own information, but believe that friends get less news than they do.

These mechanisms are in line with our adoption of a misspecified model of learning. Section 7 considers other model misspecifications and shows that they all lead agents to neglect selection in ways similar to $\hat{\gamma} < \gamma$. These include misspecifications about the friends' types, their news-sharing behavior, or the information quality. Section 7 also discusses the case of $\hat{\gamma} > \gamma$. This misspecification can also lead to our main results on polarization, but implies that an agent accounts for selectivity too much—as if she treated her friends as *more* informed than they actually are, or were excessively skeptical that information is being withheld. This highlights that our main results obtain for many misspecifications that distorts how agents treat absence of news. We maintain the simplest form of this misspecification in the main text for ease of exposition.

3.4 Learning about Selection Neglect

Evidence strongly suggests that people do not learn to correctly take into account the selection in their information. Enke (2020) finds that nudging subjects to remember selection leads them to take it into account, but not entirely. This is a stark result because Enke's subjects are told from the outset exactly how their information is selected and they are reminded explicitly about it rather than having to "relearn" it from the data. Jin et al. (2021) provide their subjects with several forms of feedback, yet conclude that their under-reaction to bad news "is not easily eliminated, even if receivers are provided information about aggregate disclosure behavior and have played as senders for many rounds." Esponda et al. (2021) estimate how their analog of $\hat{\gamma}$ evolves over experiment rounds. They find that, even though subjects receive abundant and precise feedback, $\hat{\gamma}$ converges to a level significantly above 0 and below γ. Since in reality people are unlikely to receive as much feedback and face more complex learning problems, we should not expect learning about γ to occur. Moreover, several theoretical papers show how incorrect mental models can persist (e.g., Schwartzstein

(2014), Gagnon-Bartsch et al. (2018), and Fudenberg and Lanzani (2020)).

This evidence supports a key feature of our model: that the agent's view of the world rules out the true γ from the set of possibilities. Even if we allowed the agent to contemplate multiple possible $\hat{\gamma}$ and learn about them, she would not converge to the true γ if it is not among the possible options. To capture this absence of learning about the true signal generating process in the simplest way, we assume a fixed degenerate prior about $\hat{\gamma}$. Ruling out the true signal process is a defining feature of the literature on models with misspecification as listed in the Introduction. However, we should note that our results may not be robust to the agents learning the true γ.

4 Single-Agent Learning

Before examining belief polarization, we study how a generic normal agent updates her belief under the effects of selective sharing and misperception. Hence, we drop all i subscripts in this section. Note that we, as the external observer, will calculate the distributions of the agent's belief using the correct model of the world (i.e., γ not $\hat{\gamma}$).

4.1 Short Run

We begin with short-run learning. Recall that $\mu(\mathbf{s}^1)$ is the Bayesian posterior probability that the agent assigns to state A given all the information she obtains after one period.

We first show that, in the presence of misperception, selective news sharing can distort learning even if it does *not* give rise to unbalanced news diets. Specifically, $\hat{\gamma} < \gamma$ causes the agent's expected posterior to be distorted towards the state she deems more likely ex ante. This is reminiscent of updating distortions usually called confirmatory bias (Rabin (1998)).

Proposition 1. *Fix any agent with a balanced echo chamber* $e = (d_A, d_B, n)$ *where* $d_A =$

$d_B > 0$. *Then,*

$$\left(\mathbb{E}[\mu(\mathbf{s}^1)] - \pi\right)\left(\pi - \frac{1}{2}\right) > 0.$$

To give some intuition, it is useful to write the agent's posterior after one period (see Appendix A for more details). Let a_A be the number of a-signals her A-dogmatic friends received and b_B the number of b-signals her B-dogmatic friends received. The agent also receives n signals from her normal friends plus her own signal. Among these, let a_N and b_N be the number of a-signals and b-signals. By Bayes's rule her posterior belief is

$$\mu(\mathbf{s}^1) \;=\; \frac{\pi}{\pi + (1-\pi)Q^M \hat{\Gamma}^S}, \tag{2}$$

where

$$Q \equiv \frac{1-q}{q}, \qquad M \equiv a_A + a_N - (b_B + b_N),$$
$$\hat{\Gamma} \equiv \frac{\hat{\gamma}(1-q) + (1-\hat{\gamma})}{\hat{\gamma}q + (1-\hat{\gamma})}, \qquad S \equiv (d_B - b_B) - (d_A - a_A).$$

The term Q^M captures the agent's interpretation of the received signals, which is always correct: By verifiability of information, sharing a signal leaves no doubt that the signal was actually received—hence, $\hat{\gamma}$ is irrelevant. The term $\hat{\Gamma}^S$ captures how the agent misperceives the silence of her dogmatic friends, which happens for $d_A - a_A$ of A-dogmatic friends and $d_B - b_B$ of B-dogmatic friends. According to the agent's mental model, a silent friend got an unfavorable signal with probability $\hat{\gamma}$ or no signal with probability $1 - \hat{\gamma}$. Since $\hat{\Gamma}$ is decreasing in $\hat{\gamma}$, the misperception $\hat{\gamma} < \gamma$ shrinks $\hat{\Gamma}^S$ if $S < 0$ and inflates $\hat{\Gamma}^S$ if $S > 0$, thereby distorting the posterior upward or downward depending on S. It is therefore not immediate that the *average* distortion goes in any specific direction. Misperception could inflate or deflate updating, but have no effect on average.

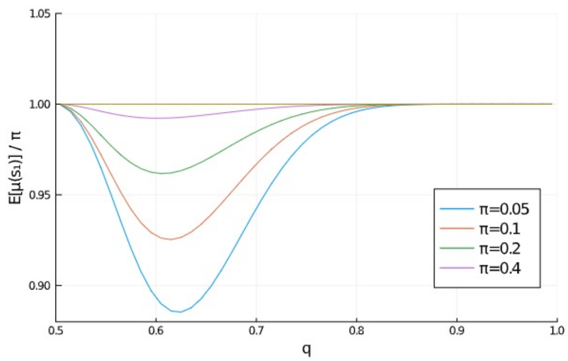
The agent's prior resolves this ambiguity. One way to see why is to consider $\hat{\gamma} \approx 0$.[11] In this case, the agent attributes silence almost entirely to lack of news and, therefore, essentially takes every shared signal at face value. Formally, $\hat{\Gamma} \approx 1$ and hence $\mu(\mathbf{s}^1)$ depends entirely on Q^M. Since $Q < 1$, the agent updates her belief in the direction of the majority of *shared* signals, namely towards A if $a_A + a_N > b_B + b_N$ and towards B otherwise. As far as normal friends are concerned, they share a *and* b-signals with frequencies that are consistent with the prior π, so they do not distort the expected posterior from π. But if state B is ex ante more likely ($\pi < \frac{1}{2}$), for instance, B-dogmatic friends are more likely to share b-signals than A-dogmatic friends are to share a-signals—for any q. Therefore, the majority of shared signals is distorted in favor of B on average, which results in $\mathbb{E}[\mu(\mathbf{s}^1)] < \pi$. The result shows that this distortion works similarly even if the agent treats silence only partially at face value (i.e., for every $\hat{\gamma} < \gamma$).

Figure 1(a) illustrates Proposition 1. The graph shows the ratio $\mathbb{E}[\mu(\mathbf{s}^1)]/\pi$ as a function of q, which would be constant at 1 without misperception by Remark 1. The figure illustrates that in spite of a balanced echo chamber ($d_A = d_B = 10$), after a single round of updating, expected beliefs are *always* distorted away from state A because the prior favors state B (i.e., $\pi < \frac{1}{2}$). The more the prior favors state B (lower π), the more distorted is the expected posterior.

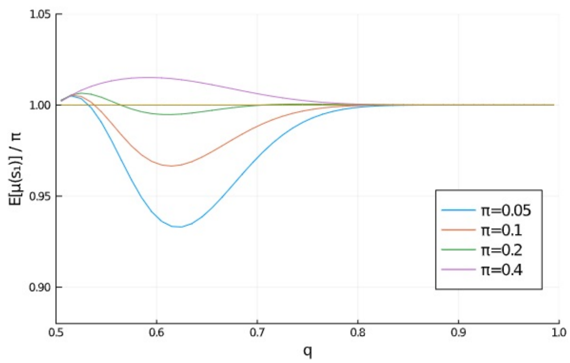
We now include the effects of echo-chamber imbalance. We find that it can distort the agent's expected posterior towards the conviction of the majority of her dogmatic friends. For this to happen, however, the information quality needs to be sufficiently low.

Proposition 2. *Fix any agent with an unbalanced echo chamber $e = (d_A, d_B, n)$. There*

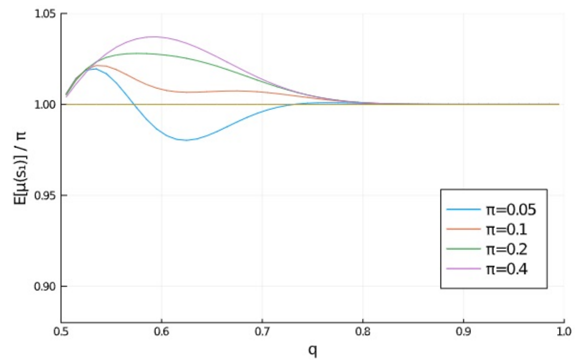
[11] Note that $\hat{\gamma} \approx 0$ is observationally consistent with full selection neglect as in Enke (2020) (see Section 3). At the same time, it guarantees that Bayesian posteriors are well defined, which would not be true for $\hat{\gamma} = 0$.



(a) $n = 1$, $d_A = 10$, $d_B = 10$



(b) $n = 1$, $d_A = 11$, $d_B = 10$



(c) $n = 1$, $d_A = 12$, $d_B = 10$

Figure 1: Distortion of expected posterior after one round of signals with $\gamma = 0.8$ and $\hat{\gamma} = 0.5$

exists $q_{SR}(e, \gamma, \hat{\gamma}) \in \left(\frac{1}{2}, 1\right]$ *such that, if* $q < q_{SR}(e, \gamma, \hat{\gamma})$*, then*

$$\left(\mathbb{E}[\mu(\mathbf{s}^1)] - \pi\right)(d_A - d_B) > 0.$$

Figures 1(b) and 1(c) illustrate this result. They also show the possibility that $q_{SR}(e, \gamma, \hat{\gamma}) < 1$ and that learning is distorted towards the agent's dogmatic minority when q is sufficiently high. Thus, it is not always true that an agent's echo chamber distorts learning towards her dogmatic majority, even if her under-reaction to silence favors that majority.

To see why simply listening to more A- than B-dogmatic friends does not guarantee that the agent will systematically update her belief towards A, consider again $\hat{\gamma} \approx 0$. As before, the agent updates based on the majority of the signals she actually receives. Again, signals from normal friends do not distort her belief, so let's focus on dogmatic friends. If $\omega = A$, then on average qd_A of A-dogmatic friends share an a-signal, while $(1-q)d_B$ of B-dogmatic friends share a b-signal. Since $q > \frac{1}{2}$ and $d_A > d_B$, a-signals tend to have a majority on average. If $\omega = B$ instead, then on average $(1-q)d_A$ of A-dogmatic friends share an a-signal, while qd_B of B-dogmatic friends share a b-signal. Thus, now b-signals tend to have a majority if q is sufficiently large. In this case, if state B is sufficiently likely (i.e., π is small), the agent's belief can be distorted toward B on average despite $d_A > d_B$, because the confirmation-bias force behind Proposition 1 dominates. The result shows that this logic extends for every $\hat{\gamma} < \gamma$. This is not obvious, as a lower q weakens the belief response to both observed signals and silence ($Q \to 1$ and $\hat{\Gamma} \to 1$ as $q \to \frac{1}{2}$ in expression (2)).

Two intuitive sets of conditions lead the majority of dogmatic friends to prevail for all q (i.e., $q_{SR}(e, \gamma, \hat{\gamma}) = 1$). First, if $d_A - d_B$ is sufficiently large, the force that pushes the agent's belief in favor of A dominates for all q, even if the confirmation-bias force pushes her belief in favor of B. Second, if both forces are aligned, clearly the effects of echo-chamber imbalances prevail for all q. This second case is summarized in Corollary 1 below.

Corollary 1. Fix any agent with an unbalanced echo chamber $e = (d_A, d_B, n)$. If $(d_A - d_B)(\pi - \frac{1}{2}) \geq 0$, then $\left(\mathbb{E}[\mu(\mathbf{s}^1)] - \pi\right)(d_A - d_B) > 0$ for all $q \in \left(\frac{1}{2}, 1\right)$.

4.2 Long Run – Abundant Information

We showed that with one round of signals echo chambers can distort beliefs. One may expect these distortions to vanish when information becomes abundant (i.e., in the long run after many signals). In fact, abundant information can instead exacerbate the effect of misperceived selective sharing and cause beliefs to be almost certainly incorrect. With probability 1 and irrespective of the true state, the agent's posterior converges to a degenerate belief on the state favored by her dogmatic majority. Unlike in the short run, these distortions require information quality to be sufficiently low.

Proposition 3. *Fix any agent with an unbalanced echo chamber* $e = (d_A, d_B, n)$*. There exists* $q_{LR}(e, \gamma, \hat{\gamma}) \in \left(\frac{1}{2}, 1\right)$ *such that the following holds with probability 1:*

1. *If* $q < q_{LR}(e, \gamma, \hat{\gamma})$*, then*

   (i) *if* $d_A > d_B$ *the agent is certain the state is A in the long run (i.e.,* $\mu(\mathbf{s}^\infty) = 1$*);*

   (ii) *if* $d_B > d_A$ *the agent is certain the state is B in the long run (i.e.,* $\mu(\mathbf{s}^\infty) = 0$*);*

2. *If* $q > q_{LR}(e, \gamma, \hat{\gamma})$ *or* $d_A = d_B$*, then the agent learns the true state in the long run (i.e.,* $\mu(\mathbf{s}^\infty) = I_{\{\omega = A\}}$*).*

Balanced echo chambers always result in correct learning, so the distortion in Proposition 1 does not survive in the long run. This is because with abundant information the prior no longer matters, as the received signals dominate the agent's updating.

This result is the outcome of a non-trivial race between two kinds of information. First-hand information provides an increasingly accurate estimate of the state, which would result in perfect learning in a standard setting. Second-hand signals also provide more information, but are selected in ways the agent does not correctly take into account. It turns out that with low q the distortion in each step of updating unveiled in Proposition 2 accumulates over time, leading the posterior astray. By contrast, high-q information eventually removes the distortions caused by echo-chamber imbalance. While in the short run distortions may vanish only at $q = 1$, in the long run they vanish for a range of $q < 1$. This will be important when we consider ways to mitigate incorrect learning and the resulting polarization (Section 5.1).

For intuition, consider again $\hat{\gamma} \approx 0$ so that the agent updates based on the majority of shared signals. For $q \approx \frac{1}{2}$ and $d_A > d_B$, in every period it is more likely to have a majority of a-signals than of b-signals shared by dogmatic friends, independent of the true state. This eventually drives the agent's belief towards A. By contrast, when q is large, the majority of

A-dogmatic friends will tend to induce a majority of a shared signals in every period only if the state is actually A. If it is instead B, A-dogmatic friends will tend to get unfavorable signals and stay silent, so the minority of B-dogmatic friends tends to induce a *majority* of b-signals in every period. This accumulates over time, even if $q < 1$. Thus, sufficiently high q is enough to curtail the effect of misperception and results in correct learning.

The threshold q_{LR} that separates correct and incorrect long-run learning has intuitive comparative statics properties.

Proposition 4. *$q_{LR}(e, \gamma, \hat{\gamma})$ strictly increases as $|d_A - d_B|$ increases, or $\gamma - \hat{\gamma}$ increases, or n decreases.*

The threshold increases in the echo-chamber imbalance and misperception, as both strengthen the forces leading posteriors astray; it decreases in the number of normal friends, as they provide more unselected information. In fact, one can show that $q_{LR}(e, \gamma, \hat{\gamma}) \to \frac{1}{2}$ as $\hat{\gamma} \to \gamma$.

These results uncover some subtleties in how echo chambers can drive people's beliefs. Even if the underlying information is the same for all, people with many but moderately unbalanced dogmatic friends can learn the truth, while others with few but severely unbalanced dogmatic friends can end up believing something false.

4.3 Making and Losing Friends

Advances in technology—such as the rise of social media—have expanded the group of friends from which many agents receive second-hand information. How do these changes affect individual learning? This section addresses this, focusing on the long run.

Suppose the echo chamber of an agent changes as she makes or loses friends. When does this shrink the range of information qualities for which she learns incorrectly?

Proposition 5. *Fix any agent with echo chamber $e = (d_A, d_B, n)$ that satisfies $d_A > d_B$ and $n \geq 1$. For any other echo chamber $e' = (\lambda_A d_A, \lambda_B d_B, \lambda_N n)$ with $\lambda_N \geq 0$, $\lambda_A \geq 0$ and*

$\lambda_B \geq 0$ *that satisfy* $\lambda_A d_A > \lambda_B d_B$, *we have* $q_{LR}(e', \gamma, \hat{\gamma}) < q_{LR}(e, \gamma, \hat{\gamma})$ *if*

$$\begin{aligned} \lambda_N - 1 \quad \geq \quad & \left(\frac{\lambda_A d_A - \lambda_B d_B}{d_A - d_B} - 1 \right) \left(1 + \frac{1}{n} \right) \\ & + \frac{d_A d_B}{d_A - d_B} \cdot \frac{1}{n} \cdot \max \left\{ (\lambda_A - \lambda_B) \frac{2}{2 - \hat{\gamma}},\ (\lambda_A - \lambda_B) \right\}. \end{aligned} \tag{3}$$

To understand this condition, begin with the first term in parentheses, which is the net growth rate of the echo-chamber imbalance. The first line of (3) requires this rate to be sufficiently *smaller* than the growth rate of normal friends in order to lower q_{LR}. The second line of (3) takes into account what happens to each group—hence, the flow of selected signals—of A- and B-dogmatic friends. If the A-group grows more, lowering q_{LR} requires an even larger growth of normal friends. If the B-group grows more, this partially compensates the change in the imbalance, so it requires a smaller growth of normal friends. In short, an increase in friends involves a trade-off between access to information and the scope for echo chambers to distort beliefs. Interestingly, one can show that q_{LR} always rises if all friends grow at the same rate ($\lambda_A = \lambda_B = \lambda_N$). This provides more scope for polarization.[12]

We now ask a different question: Fixing information quality, what changes in an agent's echo chamber suffice to stop its power of distorting beliefs and re-establish correct learning? Specifically, given $\hat{q} < q_{LR}$ and $\lambda_A = \lambda_B = \lambda$, what λ_N suffices to lower q_{LR} below $\hat{q}$?

Proposition 6. *Fix any agent with echo chamber* $e = (d_A, d_B, n)$ *that satisfies* $d_A > d_B$ *and* $n \geq 1$ *and any* $\hat{q} \in \left(\frac{1}{2}, q_{LR}(e, \gamma, \hat{\gamma})\right)$. *For any other echo chamber* $e' = (\lambda d_A, \lambda d_B, \lambda_N n)$ *with* $\lambda \geq 0$ *and* $\lambda_N \geq 0$, *we have that* $q_{LR}(e', \gamma, \hat{\gamma}) < \hat{q}$ *if*

$$\lambda_N > \frac{d_A - \hat{q}(d_A + d_B)}{(2\hat{q} - 1)n} \lambda - \frac{1}{n}.$$

Propositions 5 and 6 may have several practical implications. For instance, the growth

[12] Appendix A.5 provides a more general version of Proposition 5 that also covers the case of $\lambda_A d_A < \lambda_B d_B$.

and types of an agent's friends may be estimated using data from social-media platforms about their news-sharing habits and composition. Given the desired λ_N, one can estimate how long (if ever) it will take before her echo chamber stops distorting her beliefs (i.e., q_{LR} falls below $\hat{q}$). Alternatively, social-media algorithms often control how people form new connections. Knowing the effects of echo chambers' composition on people's learning can inform how to design such algorithms so as to take into account the distortions of selective news sharing.

5 Belief Polarization in Society

We now examine belief polarization, treating the set of normal agents $\mathcal{N}$ as our society of interest. We exclude dogmatic agents based on the interpretation that they have degenerate—hence, unchangeable—beliefs.

We begin by defining a measure of polarization. Polarization does not simply mean heterogeneous beliefs but rather the existence of groups with sharply different beliefs (Esteban and Ray (1994)). Such sharp differences usually emerge over time. For this reason, we first examine polarization in long-run beliefs. Denote the vector of echo chambers in $\mathcal{N}$ by

$$\mathbf{e} = \{(d_{Ai}, d_{Bi}, n_i)\}_{i \in \mathcal{N}}.$$

By Proposition 3, $\mathbf{e}$ induces a distribution of long-run beliefs *across the agents* in $\mathcal{N}$, characterized by which agents converge to having a degenerate belief on state $\omega \in \{A, B\}$. Let

$$\mathcal{N}_A(\mathbf{e}) = \{i \in \mathcal{N} : \mu(\mathbf{s}_i^\infty) = 1\} \qquad \text{and} \qquad \mathcal{N}_B(\mathbf{e}) = \{i \in \mathcal{N} : \mu(\mathbf{s}_i^\infty) = 0\}.$$

Given echo chambers $\mathbf{e}$, define *long-run polarization* $\Pi(\mathbf{e})$ as the normalized sum of pairwise differences between long-run beliefs of agents in $\mathcal{N}$.[13] That is,

$$\Pi(\mathbf{e}) \equiv \frac{2}{|\mathcal{N}|^2} \sum_{i,j \in \mathcal{N}} \left| \mu(\mathbf{s}_i^\infty) - \mu(\mathbf{s}_j^\infty) \right| = \frac{4|\mathcal{N}_A(\mathbf{e})||\mathcal{N}_B(\mathbf{e})|}{|\mathcal{N}|^2},$$

which takes values in $[0,1]$ and attains its maximum when $|\mathcal{N}_A(\mathbf{e})| = |\mathcal{N}_B(\mathbf{e})|$. Given the true ω, we call $\mathcal{N}_\omega(\mathbf{e})$ the set of "eventually correct" agents and $\mathcal{N}_{-\omega}(\mathbf{e})$ the set of "eventually incorrect" agents. Because in our model beliefs are about some facts, in a polarized society some agents must be "right" and some "wrong." Therefore, depending on the initial situation, we can move to higher polarization not only as the result of more agents switching to the wrong side, but also as the result of more agents switching to the right side.[14]

Our previous results imply that selective information sharing and misperception can cause beliefs to polarize in the long run.

Corollary 2. Fix any society $\mathcal{N}$ with echo chambers $\mathbf{e}$ that satisfy $d_{Ai} > d_{Bi}$ and $d_{Aj} < d_{Bj}$ for some $i, j \in \mathcal{N}$. There always exists $q > \frac{1}{2}$ such that $\Pi(\mathbf{e}) > 0$ with probability 1.

This formalizes the common narrative that if echo chambers skew agents' news diets in opposite directions—where the imbalances can be arbitrarily small—then their beliefs can polarize. People on the left and right of the political spectrum tend to have more like-minded friends than not, a fact that is often cited as a possible cause of polarization (e.g., Pew Research Center (2014)). However, our results qualify this narrative: Polarization requires

[13] Note that by standard continuity arguments

$$\Pi(\mathbf{e}) = \operatorname*{plim}_{t \to \infty} \frac{2}{|\mathcal{N}|^2} \sum_{i,j \in \mathcal{N}} \left| \mu(\mathbf{s}_i^t) - \mu(\mathbf{s}_j^t) \right|.$$

[14] The results in this section hold for more general measures of polarization, such as that axiomatized by Esteban and Ray (1994). Applied to our long-run beliefs, their measure takes the form $\nu^{1+\alpha}(1-\nu) + (1-\nu)^{1+\alpha}\nu$, where $\nu = |\mathcal{N}_A(\mathbf{e})|/|\mathcal{N}|$, $\alpha \in (0, \alpha^*]$, and $\alpha^* \approx 1.6$. Other notions of polarization have been studied in the literature, including polarization along some individual characteristics (like income, wealth, education, racial segregation) and "affective polarization" (Mason, 2015; Rogowski and Sutherland, 2016; Mullinix, 2016; Iyengar and Krupenkin, 2018; Iyengar et al., 2019). These notions are beyond the scope of our theory.

low q and misperception of the effects of echo chambers (recall Remark 1), but does not require fake news nor that people look at the world through fundamentally incompatible paradigms. It does not even require that echo chambers be unbalanced in opposite directions: By Propositions 3 and 4, if $d_{Ai} \geq d_{Bi}$ for all $i \in \mathcal{N}$ and $d_{Aj} - d_{Bj} > d_{Ak} - d_{Bk}$ for some $j, k \in \mathcal{N}$, then there always exists $q > \frac{1}{2}$ such that $\Pi(\mathbf{e}) > 0$ if the true state is $\omega = B$. For such $\mathcal{N}$, if q is intermediate, some agents will be eventually correct despite their echo chamber, while others will be eventually incorrect.

These observations highlight the importance of information quality for echo chambers to give rise to belief polarization. Intuition may suggest that as people receive better information, disagreement should decline. In fact, the following result provides conditions for polarization to be non-monotonic in q. Let $\mathcal{D}_\omega$ be the set of agents who have a majority of ω-dogmatic friends—hence, they can be eventually incorrect if the true state is *not* ω.

Proposition 7. *Fix any society* $\mathcal{N}$ *that has echo chambers* $\mathbf{e}$ *which satisfy* $q_{LR}(e_i, \gamma, \hat{\gamma}) \neq q_{LR}(e_j, \gamma, \hat{\gamma})$ *for all* i, j *and fix* ω. *Then,* $\Pi(\mathbf{e})$ *decreases in* q *over* $\left(\frac{1}{2}, 1\right)$ *if and only if* $|\mathcal{D}_{-\omega}| \leq \frac{1}{2}(|\mathcal{N}| + 1)$. *Otherwise,* $\Pi(\mathbf{e})$ *is single peaked in* q.

To see the intuition, assume the true state is B. As q rises, more agents have correct long-run beliefs because for more of them q exceeds their individual q_{LR}. Thus, as q rises, one by one, agents in $\mathcal{N}_A(\mathbf{e})$ will switch to $\mathcal{N}_B(\mathbf{e})$—but not vice versa.[15] If the eventually incorrect agents outnumber the eventually correct agents initially when $q \approx \frac{1}{2}$ (i.e., $|\mathcal{D}_A| > \frac{1}{2}(|\mathcal{N}| + 1)$), this gradual migration into the set of eventually correct agents will first *increase* polarization and then decrease it towards zero. This point simply highlights that, although increasing the quality of first-hand information can counteract the power of echo chambers to distort beliefs, such increases may need to be significant to actually curb polarization.

If we now fix information quality, how do changes of the echo chambers in society affect

[15] This is where we use the assumption that $q_{LR}(e_i, \gamma, \hat{\gamma}) \neq q_{LR}(e_j, \gamma, \hat{\gamma})$ for all i, j.

long-run polarization? Consider a society $\mathcal{N}$ and values of $\lambda \geq 1$ such that each agent i's echo chamber can be expressed as $e_i^\lambda = (\lambda d_{Ai}, \lambda d_{Bi}, \lambda n_i)$. The parameter λ can be interpreted as technological advances that increase all types of social connections uniformly. The next result shows that polarization is increasing or single peaked in λ.

Proposition 8. *Fix ω. Then, $\Pi(\mathbf{e}^\lambda)$ increases in λ if $|\mathcal{D}_{-\omega}| \leq \frac{1}{2}(|\mathcal{N}|+1)$. Otherwise, $\Pi(\mathbf{e}^\lambda)$ is increasing or single peaked in λ.*

In this result, the expansion of the echo chambers affects polarization by inducing agents to switch from learning correctly to learning incorrectly (Proposition 5). However, the expansion of echo chambers can also curb polarization by promoting correct learning. By Proposition 6, this happens if technology changes cause normal friends to grow sufficiently faster than echo-chamber imbalances. These points offer a new perspective on the evidence showing that polarization seems more pronounced for demographic groups that are least likely to use the Internet and social media (Zhuravskaya et al., 2020). Their echo chambers may be smaller but also more unbalanced, while social media may give people access to a broader pool of potential normal friends.

Finally, we consider belief polarization in the short run. To this end, we now interpret each $i \in \mathcal{N}$ as a group of individuals who all have an echo chamber with the same composition e_i. Assume $e_i \neq e_j$ if $i \neq j$. We can summarize the beliefs within each group with their empirical average and use these statistics to quantify intra-group polarization. If group i is large, its empirical average belief is well approximated by $\mathbb{E}[\mu(\mathbf{s}_i^1)]$ by the Law of Large Numbers. Thus, we can define *short-run polarization* as

$$\Pi_{SR}(\mathbf{e}) = \frac{2}{|\mathcal{N}|^2} \sum_{i,j \in \mathcal{N}} \left| \mathbb{E}[\mu(\mathbf{s}_i^1)] - \mathbb{E}[\mu(\mathbf{s}_j^1)] \right|.$$

Standard Bayesian learning without misperceptions implies $\Pi_{SR}(\mathbf{e}) = 0$ (Remark 1). By contrast, selective news sharing with misperceptions can lead to $\Pi_{SR}(\mathbf{e}) > 0$. For instance, Proposition 2 implies the following.

Corollary 3. Fix any society $\mathcal{N}$ with echo chambers $\mathbf{e}$ that satisfies $d_{Ai} > d_{Bi}$ and $d_{Aj} < d_{Bj}$ for some groups i, j. There always exists $q > \frac{1}{2}$ such that $\Pi_{SR}(\mathbf{e}) > 0$.

Thus, as long as some groups of people have echo chambers with opposite imbalances, our model can also account for polarization in the short run. In contrast to the long run, where this requires low q, short-run polarization can arise even for high q (Propositions 1 and 2 and Corollary 1). This could cause temporary polarization: Even if all agents eventually learn correctly, their beliefs may polarize in the short run.

The consequences of polarization is beyond the scope of our analysis, but several strands of research suggest that polarization leads to negative political and economic outcomes.[16] One consequence that is often mentioned and extensively studied is gridlock. According to Brady et al. (2008), political dissensus in the US —what they predict polarization leads to— renders parties "less likely to be able to make policy coalitions with the requisite numbers to beat filibusters or to override presidential vetoes." They find evidence of this, especially for energy and environmental policy (see Section 6.2 for an application to climate change). They argue that polarization may make it harder to tackle other long-term reforms to entitlement programs, Social Security, and health care. It is worth noting that gridlock refers to group decision problems. For such problems, even if more people learn the truth, it may lead to a situation of "narrow majorities" and hence gridlock. Online Appendix D.1 develops an extension of our model where Pareto-inefficient gridlock can arise as polarization occurs.

5.1 Mitigating Polarization

Ferejohn et al. (2020) note that challenges to shaping the character of democratic institutions

[16]In addition to the references in the Introduction, see, e.g., Layman et al. (2006); Epstein and Graham (2007); Nivola and Brady (2007, 2008); McCarty (2019). Beinart (2008) also argues that polarization may damage US foreign policy (e.g., its credibility, effectiveness, and responsiveness).

include "managing the development of media and information technologies to ensure they enhance, rather than degrade, robust pluralism and civil political engagement." This section takes a step in that direction.

How could a social planner address polarization caused by shared news? Selective sharing, misperceptions, and echo chambers of friends seem difficult to influence. It may be more straightforward to influence the quality of their first-hand information. For instance, this may involve incentivizing newspapers to spend more on reporters, data gathering, and fact checking. But this direct way may be infeasible for technological or economic reasons. This section suggests a feasible way to increase the quality of information that people ultimately receive without changing q of the original sources.

The last decades have witnessed the rise of *news aggregators*, online platforms that summarize news for their users (e.g., *The Drudge Report*, *Apple News*, *Yahoo! News*). The reason may be that aggregators help people handle the overload of daily news, or pool news from different sources into one convenient access point. By summarizing news, aggregators throw away some information relative to the totality of the aggregated signals. Yet, this summary can have higher quality than each aggregated signal *individually*, which is the key observation for our purposes. Through the lens of our theory news aggregators can also serve another function: undermine the distortions of selective news sharing and thus curb polarization.[17]

There are many ways to aggregate signals. To make our point we consider the following simple form. Divide time into blocks of M periods, where M is an odd number. For every $t = 1, 2, \ldots$, define $\hat{s}^i_{Mt}$ as a new signal that is sent to agent i at the end of each time block

[17] Other papers studying news aggregators in an economic context include Athey et al. (2017) and Hu et al. (2019). Athey et al. (2017) explore experimentally the impact of news aggregators on the consumption of news from other outlets, while the focus of Hu et al. (2019) is differentiation between personalized news-aggregation providers.

and reports whether more a- or b-signals realized in that block:

$$\hat{s}^i_{Mt} = \begin{cases} 0 & \text{if } \sum_{k=(t-1)M+1}^{tM} I_{\{s_{ik}=a\}} < \frac{M}{2} \\ 1 & \text{if } \sum_{k=(t-1)M+1}^{tM} I_{\{s_{ik}=a\}} > \frac{M}{2}. \end{cases}$$

Clearly, $\hat{s}^i_{Mt}$ conveys less information than do the aggregated M signals together. However, $\hat{s}^i_{Mt}$ has higher quality than each s_{it}. To see this, suppose $M = 3$ and $\omega = A$:

$$\begin{aligned} \mathbb{P}(\hat{s}^i_3 = 1|\omega = A) &= \mathbb{P}\left(\sum_{k=1}^{3} I_{\{s_{ik}=a\}} \geq 2 \middle| \omega = A\right) \\ &= q^3 + 3q^2(1-q) > q = \mathbb{P}(s_{it} = a|\omega = A). \end{aligned}$$

Thus, substituting s_{it} with $\hat{s}^i_{Mt}$ worsens the quantity of information for the agents, but improves its quality. Note that in standard models this substitution would be irrelevant for long-run learning.

How much aggregation is enough to curb polarization? Let $\widehat{\Pi}$ be the long-run polarization when signals s_{it} are replaced with $\hat{s}^i_{Mt}$.[18]

Proposition 9. *Fix any society $\mathcal{N}$ with echo chambers $\mathbf{e}$ and information quality q such that $\Pi(\mathbf{e}) > 0$. Let $\bar{q}_{LR} = \max_{i \in \mathcal{N}} q_{LR}(e_i, \gamma, \hat{\gamma})$. Then, $\widehat{\Pi}(\mathbf{e})$ equals zero if*

$$M > -\frac{2\ln(1-\bar{q}_{LR})}{(2q-1)^2}.$$

Partial news aggregation can suffice to curb polarization, because Proposition 3 showed that undoing the effects of misperceived selective sharing in the long run does not require perfect information quality. Note that each agent's aggregator summarizes only her primitive signals, but the level of aggregation required needs to take into account the agents with

[18]The threshold in Proposition 9 is a conservative condition based on tail bounds for Binomial cumulative distributions, which do not have a closed form. Numerical methods may provide tighter conditions.

whom she shares. This is because, through news sharing, how much agent i aggregates her signals has learning externalities on her friends. Our common threshold for M internalizes such externalities by taking into account those agents for whom the effects of misperceived selective sharing are the hardest to overcome. This may call for institutional intermediaries or platforms that aggregate news taking into account these externalities.

6 Quantitative Analysis and Applications

We now apply our theory in two directions. First, we develop simulations showing that, even when misperceptions are small (i.e., γ and $\hat{\gamma}$ are close), large divergence in beliefs can occur and quickly. Thus, even small misperceptions can be of empirical importance. Second, we show that the mechanisms we identified can explain the specific way in which opinions about climate change have polarized in the US over the last few decades.

6.1 Simulations on the Speed of Belief Divergence

To better understand how the evolution of beliefs depends on misperception, we compare it with how learning would occur if we turned misperception off. We do so through numerical simulations. We set the true state to be A and consider two agents, Alice and Bob, whose echo chambers have opposite imbalance. We use the parameter values in Table 1.

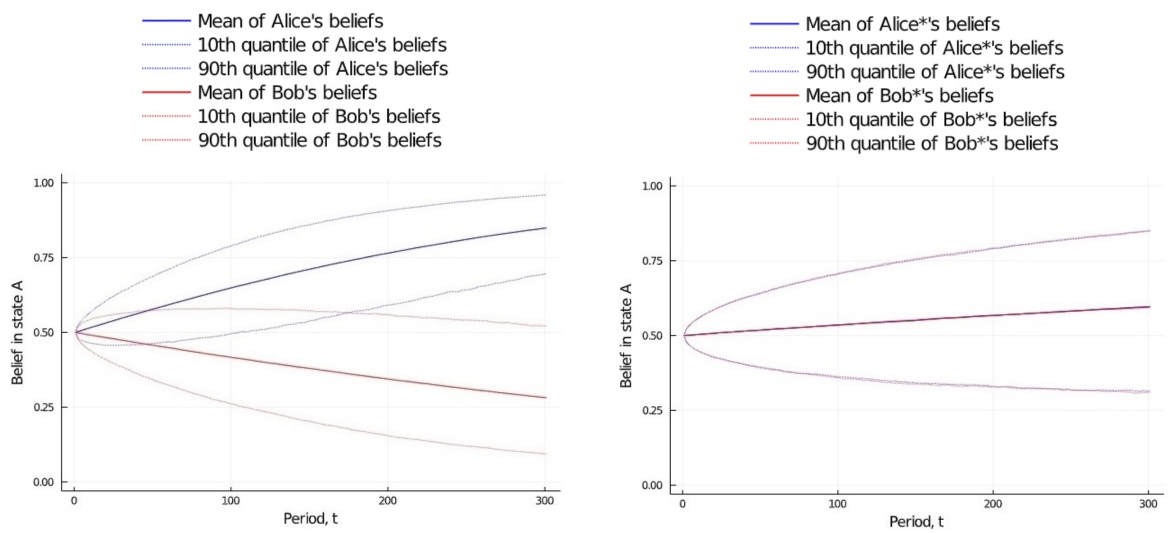
Setting a high news arrival rate $\gamma = 0.95$ is without loss of generality, as it simply involves interpreting a period as a sufficiently long length of time. We set $\hat{\gamma} = 0.9\gamma$ to consider small misperceptions. Since we set $\omega = A$, Alice always learns correctly, but Bob can learn incorrectly. The number of friends in Table 1 represents a relatively large echo chamber, with a relatively small imbalance given other parameters. In reality, there is wide variation in echo chambers in society (see, e.g., Pew Research Center (2014); Bakshy et al. (2015); Cinelli

News Arrival Rate	γ	0.95
Misperception	$\hat{\gamma}$	0.9γ
Echo Chambers	$n^{\text{Alice}} = n^{\text{Bob}}$	16
	$d_A^{\text{Alice}} = d_B^{\text{Bob}}$	24
	$d_B^{\text{Alice}} = d_A^{\text{Bob}}$	16
Information Quality	q_ℓ	$0.7 \times \frac{1}{2} + 0.3 \times q_{LR}$
	q_h	$0.3 \times \frac{1}{2} + 0.7 \times q_{LR}$

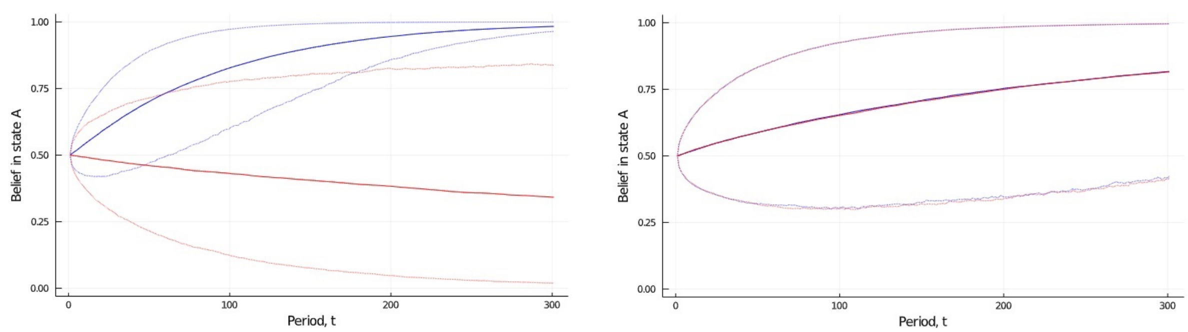
Table 1: Simulation parameters

et al. (2021)). For this reason, Appendix B presents further simulations varying the echo chambers' size and imbalance for comparison and shows different patterns of polarization with misperception. We vary the speed of learning by varying q. For our mechanism to be active and generate incorrect learning, we need $q < q_{LR} \equiv q_{LR}(e, \gamma, \hat{\gamma})$. We simulate two "treatments" with $q_\ell < q_h$, where $q_\ell = 0.7 \times 0.5 + 0.3 \times q_{LR}$ and $q_h = 0.3 \times 0.5 + 0.7 \times q_{LR}$. Finally, we construct benchmark agents, Alice* and Bob*, who are identical to Alice and Bob respectively, except that they have no misperception: They use γ to update beliefs.

We simulate the evolution of beliefs as follows. We generate 10,000 histories of signals that last for 300 periods. We then calculate four distributions of 10,000 belief trajectories starting at $\pi = 0.5$, one for each of Alice, Bob, Alice*, and Bob*. We can interpret each distribution as what can happen to the corresponding agent's beliefs from an ex-ante viewpoint, or as the cross-section of a specific group in society, of which one of our agents is a representative. We plot these distributions in Figure 2. The blue lines capture the distributions of beliefs of Alice and Alice*, while the red lines represent Bob and Bob*. The dark lines are the means of the distributions and the light lines are the 10% and 90% quantiles.



(a) Low q with misperception (b) Low q with *no* misperception



(c) High q with misperception (d) High q with *no* misperception

Figure 2: Belief paths for Alice and Bob (left), Alice* and Bob* (right)

Figures 2(a) and 2(c) illustrate that our mechanism need not take a very long time to distort Bob's learning relative to Alice's and, consequently, generate polarization—even for a minimal level of misperception. For $q = q_\ell$, Alice's and Bob's beliefs diverge at comparable speeds and learning is *faster* than it would be without misperceptions (Figures 2(b) and 2(d)). For $q = q_h$, signals are more informative in each period; therefore, those who learn correctly (Alice, Alice*, Bob*) will do so faster. For Bob, the higher q_h weakens the distorting power of his echo chamber, which causes his belief to move more slowly in the wrong direction (Figure 2(c)). This introduces an asymmetry in the speed at which Alice's and Bob's beliefs change. Alice becomes quickly convinced that the state is A, while Bob drifts slowly towards B. The net effect is rapid polarization—faster than for q_ℓ.

Changing echo chambers has the following qualitative effects. Enlarging the echo-chamber imbalance or its size increases the speed of divergence. For small echo chambers with small imbalances belief divergence continues to occur, but is slower for all agents—as expected. (See Figures 6 and 7 in Appendix B.)

6.2 Application: Opinions about Climate Change in the US

Climate change has become one of the most divisive issues of our time. Egan and Mullin (2017) (henceforth EM) document that US public opinions on climate change showed little to no polarization until the early 1990s and, "over the course of a little more than 20 years, the environment was transformed from the least to the most polarized issue" (EM, p. 219). Similarly, Pew Research Center (2020a) finds that addressing climate change has become one of the most polarizing issues in the US: 21% of Republicans consider it a top priority as opposed to 78% of Democrats. Some argue that this polarization affects the US ability to act on climate change, by miring policies in gridlock.[19] In addition, some authors suggest a connection between this gridlock and information acquisition on the Internet (e.g., see Helmuth et al., 2016).

In this section we suggest how our theory can be used to shed light on polarization of climate-change opinions and its relationship to the Internet as a source of information. Our model can be used to disentangle the role of information quality and quantity, echo-chamber size and composition, and the extent of misperception. Future research can fine-tune model

[19]See, for example, "Political Tribalism and Climate Policy Gridlock" (https://www.greentechmedia.com/articles/read/political-tribalism-and-climate-policy-gridlock) and Helmuth et al. (2016).

parameters through detailed simulations and empirical estimation, and such fine-tuning can determine which drivers are most important to guide business practices and policy-making. In reality, a combination of many forces—some outside of our model—are likely at work to determine polarization, but we provide a place to begin such exploration.

Climate change is a multifaceted issue. In this application, we consider Gallup survey data between 1997 and 2021 that reports Americans' opinions about whether the effects of global warming have already begun (Saad, 2021). We can think of this as a binary objective state of the world as in our model, since the effects (sea level rise, ice caps loss, heat waves, etc.) have either already begun (state A) or they have not (state B). Figure 3 illustrates the shares of Republicans, Democrats, and Independents who agree with the statement that "the effects of global warming have already begun." Focusing on Democrats and Republicans, initially about half of each political group agrees with the statement. For Republicans, this share declines slightly at first, hovering in the mid 40's for some time, and then accelerates to finish at 29%. By contrast, the share of Democrats climbs strongly to finish at 82%.

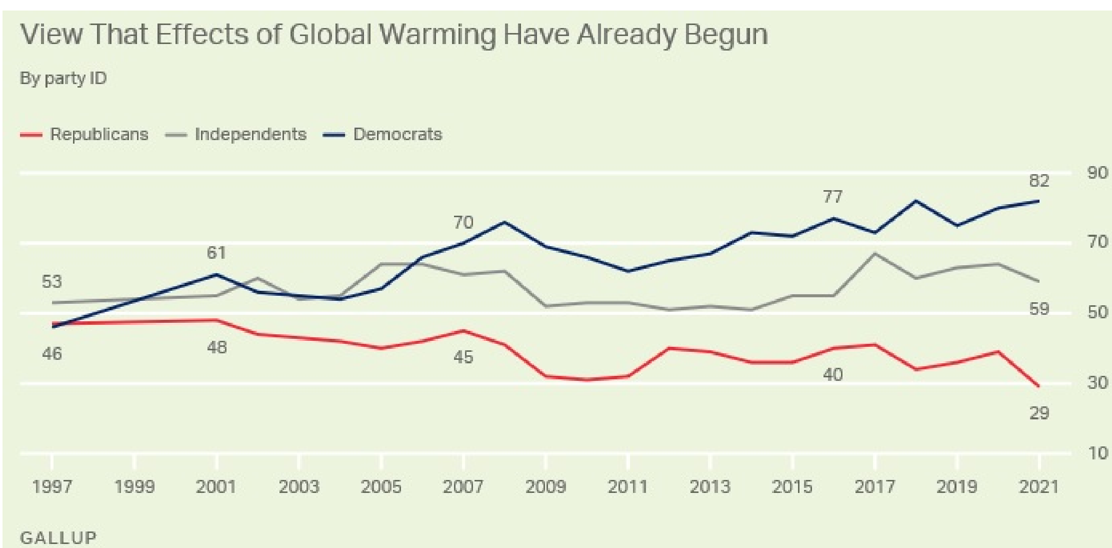


Figure 3: US public opinion on climate change, 1997–2021 (source: Saad (2021)).

We argue that these patterns can be explained by selective news sharing in unbalanced echo chambers combined with misperception and low quality of information. It is well-known that Democrats and Republicans tend to have different echo chambers and news diets (Pew Research Center (2014, 2020b)). In addition, EM note that people often do not experience climate change directly and have difficulties handling its complex scientific content, so they have to rely on information shared by others. EM also argue that people often rely on low-quality information to form opinions about climate change (see also Boykoff (2008), Mayer (2012), Moser (2014)).

It is important to note that Figure 3 does not report respondents' beliefs directly, but their agreement with options from a coarse set. To deal with this, we assume that there is a threshold belief that the effects of global warming have already begun above which a person states to agree with the corresponding option. Assuming this threshold is fairly stable, we can use our model to simulate the cross section of Democrats' and Republicans' beliefs over time and see whether the share above that threshold evolves similarly to Figure 3. To be fully impartial, we assume that the relevant threshold belief is constant and equals 0.5 for all agents.

We begin with a basic simulation that uses the following parameters. The common prior is $\pi = 0.5$, which removes any initial bias and allows us to focus on the effects of echo chambers and information quality. The arrival rate of information is $\gamma = 0.95$. As noted in Section 6.1, this high γ simply involves adjusting the interpretation of a period accordingly. We let $\hat{\gamma} = 0.9\gamma$, to test whether minor misperceptions can explain the evidence. The echo chambers of Democrats and Republicans are symmetric: All have the same proportion of dogmatic and normal friends, but the Democrats' imbalance favors state A while the Republicans' favors B. Examining various simulations, we conclude that the general trends in Figure 3 appear most consistent with small echo chambers, small imbalances, and $q = q_h$ (see Figure 7(g) in Appendix B). Using this, we generate belief trajectories for 10,000 Democrats and 10,000 Republicans conditional on A being the true state, as in the simulations of

Section 6.1. This conditioning is consistent with the scientific consensus, as recently reported by the Intergovernmental Panel on Climate Change (Masson-Delmotte et al., 2021a,b, page 10). The simulation lasts 50 periods, so each corresponds to roughly half a year in Figure 3.

Figure 4 reports the share of Democrats and Republicans whose simulated beliefs are above 0.5.[20] Consistent with Figure 3, Democrats update beliefs towards state A and faster than Republicans, who update towards state B. As time goes by, Republicans appear to learn less and incorrectly. This is because the distorting force of their echo chambers and the informative force of their signals almost balance each other when $q = q_h$.

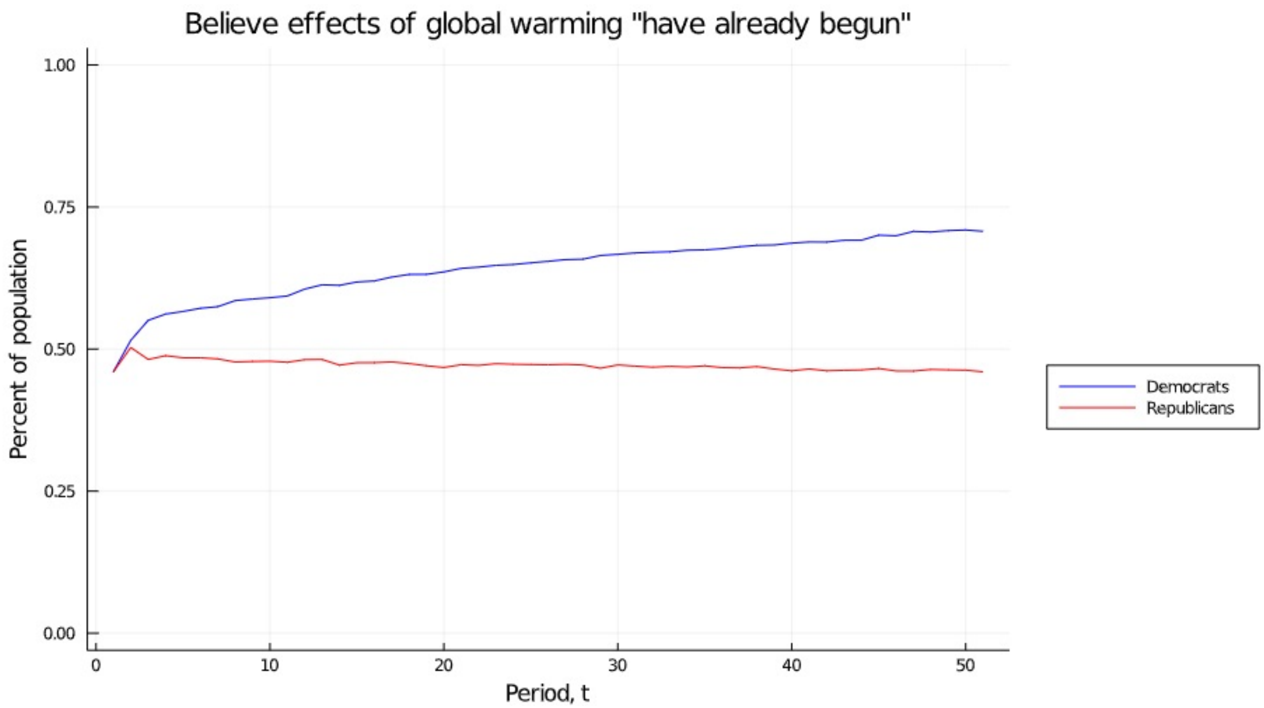


Figure 4: Basic simulation – small echo chambers.

This simulation and our model with misperception offers a first insight into the possible drivers of polarization in US opinions on climate change. This polarization cannot arise only

[20]Given the common prior 0.5, the share of Republicans and Democrats with beliefs above 0.5 would be mechanically 0% or 100% in the first period of the simulation, depending on the tie-breaking rule. Since this is a technical artifact, we initiate the time series at 46% for each political group as in the data. Alternatively, we could assume that, for each group, 46% (54%) of agents has a prior slightly above (below) 0.5.

from different news diets between Democrats and Republicans (Remark 1). It also requires that the quality of information about climate change be sufficiently low, otherwise both Democrats and Republicans would learn correctly despite having opposite echo chambers. Such a low q may be, in part, the result of skepticism campaigns started in the early 1990s. As discussed in EM and Mann (2021), groups representing fossil-fuel interests organized well-funded campaigns to undermine the credibility and informativeness of evidence about climate change, arguably lowering the quality of the available information.

More fine-tuned analysis outside the scope of this paper can be used to pin down model parameters more precisely, but we demonstrate that even a minor (reasonable) modification can improve the match substantially. One feature that emerges from a closer look at Figure 3 is that Democrats' and Republicans' opinions roughly track each other until about 2011, but seem to diverge more steadily thereafter. One thing that happened in the years around 2011 is the surge in the use of social media, such as Facebook, Youtube, and others (Samur, 2018; Ortiz-Ospina, 2019). Presumably, this expanded people's echo chambers, which can be easily replicated in the simulation of our model. To do this as transparently as possible, we permanently increase the echo-chamber size from small to large at period $t = 25$ (where large is obtained by scaling all types of friends proportionally to the values used in Section 6.1).

Figure 5 reports this modified simulation. As expected, the expansion of echo chambers speeds up learning for both Republicans and Democrats, but more so for Democrats. The share of Democrats with belief above 0.5 reaches 80% at $t = 50$, which is closer to the actual data. The share of Republicans declines somewhat more quickly after $t = 25$ to end around 40%. While it does not fall as far as 29% as in the data, this drop in 2021 seems significantly larger than in previous years, indicating that additional forces may have played a role.

The modified simulation serves as an example of how the data can guide parameter choices, but also provides a second insight. Growth in online networks and social media are two often blamed culprits for polarization. Indeed, the growth of the Internet in the early

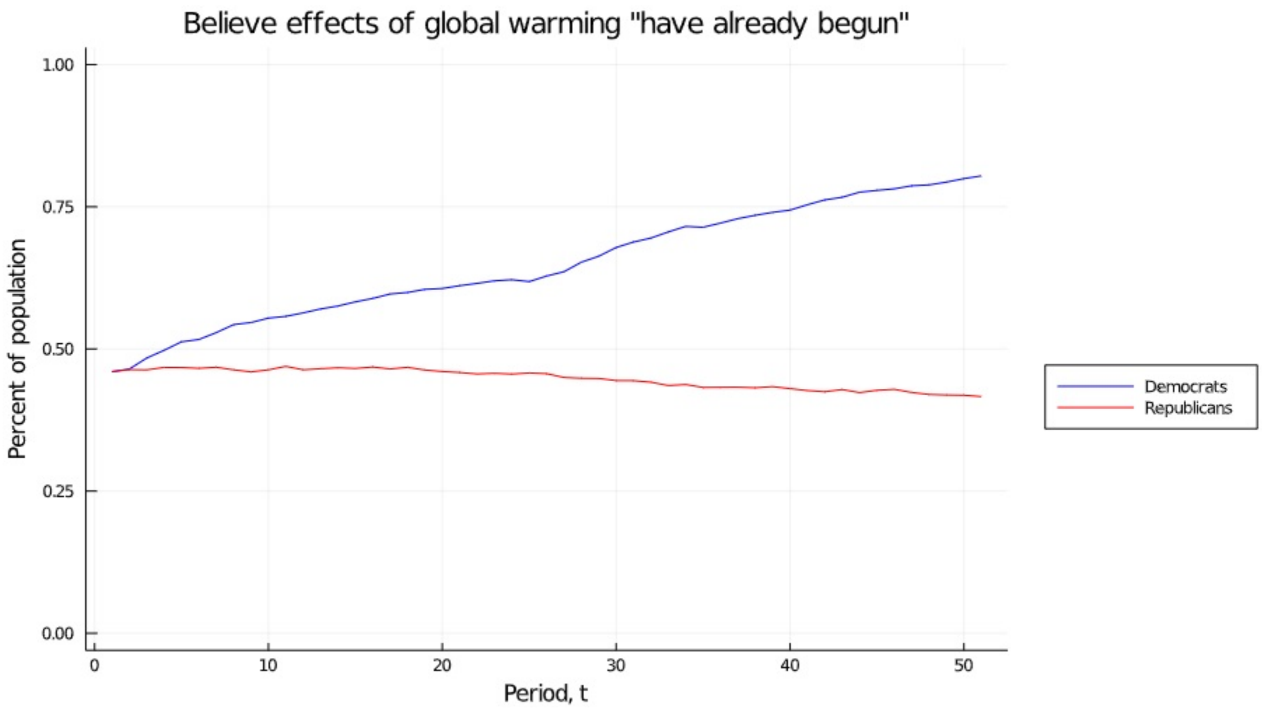


Figure 5: Modified simulation – echo chambers expand at $t = 25$.

1990s expanded online connections (for example, more information sharing through email), and this may have increased the quality threshold q_{LR} for correct learning, thereby triggering the polarizing mechanism (Propositions 5 and 8). This would explain little polarization up to the 1990s and increasing polarization thereafter. The major growth in social-media use around 2011 likely compounded the problem, accelerating polarization. Thus, the model helps identify pivotal moments in the use of the Internet and social media that could have driven polarization in views about climate change. These appear to be the late 1990s and the years around 2011.

Once again, these simulations highlight the role of information quality. Note that, while we refer to "high" information quality ($q = q_h$) as better explaining the data, this q is still below the threshold q_{LR} to generate correct learning. Thus, efforts to increase information quality about climate change can be part of the solution to alleviate polarization and gridlock. However, Proposition 7 tells us that this increase in q needs to be sufficiently large.

Proposition 9 suggests that a way to do this is by aggregating various information sources.

As previously mentioned, in reality, some combination of the aforementioned forces have likely been at work. We leave it to future research for a careful analysis of which are most important.

7 Discussion and Extensions

This section discusses other aspects of our baseline model in Section 2 as well as possible extensions.

7.1 Other Forms of Model Misspecification

We consider other ways in which agents may have misspecified mental models of the world that can lead them to misperceive information. This clarifies the main mechanism through which selective news sharing can lead to polarization: the combination of unbalanced echo chambers and incorrect response to absence of news. Throughout this section, the true properties of first-hand information (i.e., (1)) as well as timing remain as in the baseline model.

First, we explore what happens if the agents' misspecified model involves *over*-estimating the news arrival rate for friends: $\hat{\gamma} > \gamma$. This may capture an agent who treats her friends as more informed than they actually are, or is excessively skeptical that information is being withheld. In short, she reads too *much* into silence. This case may be an instance of the so-called "below-average effect" in psychology (Erev et al. (1994); Kruger (1999)). Online Appendix C.1 shows how our results change if $\hat{\gamma} > \gamma$. In a nutshell, now the agent updates as if she excessively treats silence of a dogmatic friend as suppression of news and therefore as bad news for the state that friend prefers. As a result, while for $\hat{\gamma} < \gamma$ the distorting power of echo chambers pushes the agent's belief in the direction of the state preferred by the majority of her dogmatic friends, for $\hat{\gamma} > \gamma$ it pushes her belief in the *opposite* direction. The case of $\hat{\gamma} < \gamma$ seems more consistent with both the evidence on selection neglect in Section 3 and the common understanding of the effects of echo chambers on beliefs.

Next, we consider misspecifications that do not involve the rate of news arrival. That is, assume that all agents correctly assign probability γ to the arrival of first-hand information in each period (i.e., $\hat{\gamma} = \gamma$). Instead, we analyze misspecifications of the following variables:

(I) *Probabilities with which friends share signals.* Suppose normal agents share any first-hand signal s_{it} with probability $\nu \in (0, 1]$ and stay silent with probability $1 - \nu$. Each A-dogmatic agent shares $s_{it} = b$ with probability $f \in [0, 1]$ and $s_{it} = a$ with probability $g \in [0, 1]$, where $0 \leq f < g \leq 1$; with the remaining probabilities, the agent stays silent. Each B-dogmatic agent is like an A-dogmatic agent, except for swapping the probabilities of sharing a- and b-signals. Our baseline model corresponds to $\nu = g = 1$ and $f = 0$. Misspecification (I) means that each agent knows all her friends' types, but her mental model replaces the true sharing probabilities f, g, and ν with incorrect ones $\hat{f}$, $\hat{g}$, and $\hat{\nu}$ where $\hat{f} < \hat{g}$.

(II) *Friends' types.* Given three types as in the baseline model, there are many possible ways in which an agent can misclassify her friends. For conciseness, we consider the case where some dogmatic friends are misclassified as normal. The sharing behavior is deterministic ($\nu = g = 1$ and $f = 0$). Let $\hat{n}_A$ and $\hat{n}_B$ be the number of A- and B-dogmatic friends that an agent misclassifies as normal. That is, she treats these friends as sharing all their signals, while in reality they share only signals favorable to one state.

(III) *Quality of first-hand information.* In each agent's mental model, the probability with which a signal matches the state is $\hat{q} \in \left(\frac{1}{2}, 1\right)$ instead of the true q. Note that this misspecification differs conceptually from all others considered in this paper, which are about how friends share news.

Proposition 10. *Each of misspecifications (I), (II), and (III) alone can cause belief polarization as the result of incorrect learning. This happens if and only if the true information quality q is sufficiently low and there are appropriate, real or perceived, imbalances in echo chambers.*

An echo-chamber imbalance means slightly different things depending on the misspecification (see Online Appendix C for the details). For (I), it means a different number of A- and B-dogmatic friends as well as a different gap in the probabilities of sharing signals ($f - g \neq \hat{f} - \hat{g}$). For (II), it means a disagreement between the real and perceived difference in the number of dogmatic friends ($d_A - d_B \neq \hat{d}_A - \hat{d}_B$). For (III), it means a different number of A- and B-dogmatic friends.

Despite these differences, all these misspecifications cause incorrect learning and polarization through the same fundamental mechanism as in the baseline model. That is, the agents respond to silence incorrectly by misperceiving how much of it depends on lack rather than suppression of information. Moreover, information quality plays the same role in enabling and preventing polarization. This further supports our insights about mitigating polarization by aggregating news.

Misperceiving silence is the only mechanism through which (III) causes polarization in our model. Indeed, if $\gamma = 1$ and hence silence must mean unfavorable first-hand information, the agents always learn correctly in the long run despite $\hat{q} \neq q$. Moreover, for (III) to cause polarization the agents must over-estimate the information quality, that is, $\hat{q} > q$. This case is related to the idea of "fake news:" Such news are false or very uninformative (low q), yet people mistakenly treat them as reliable and informative (high $\hat{q}$). Our results then suggest that fake news can cause incorrect learning and polarization, but *only indirectly through selective news sharing.* This may explain why, even though fake news have always existed, they have become especially dangerous in the age of social media. This also provides a rationale for fact-checking as a way to realign $\hat{q}$ with q.

7.2 Applicability to News Media Outlets and Social Media

The previous discussion of other misspecifications allows us to clarify how our insights relate to settings such as news media outlets and social-media platforms like Facebook or Twitter. First, our analysis applies unchanged if we interpret a friend as a news media outlet. In this case, normal friends correspond to outlets that report news objectively; dogmatic friends instead correspond to outlets that report news selectively with some slant, which can be modeled as in (I).

Second, unlike peer-to-peer news sharing that occurs offline at family gatherings, work, or parties, online sharing on social media is often mediated by algorithms (for example, on Facebook). These algorithms select what we see in our news feed, possibly using different probabilities depending on the kind of news as in (I). As such, they play a similar role of our dogmatic agents. Moreover, news-feed algorithms tend to be very complex and lack transparency. It is then likely that most people may have misspecified models about how these algorithms work as in (I), possibly adding to the causes of selection neglect discussed in Section 3. In this case, news-feed algorithms can also trigger our main mechanism of incorrect learning and lead to polarization.

Finally, our theory can also apply in settings like Twitter, where there are a small number of "serial tweeters" and a mass of followers. In this context, examples of our dogmatic agents are plentiful, and information suppression on an issue can take the form of tweeting about other issues. Followers may misperceive the tweet selection of their followee, either by under-estimating his news arrival rate as in the baseline model, or in one of the forms discussed above ((I), (II), or (III)).

7.3 Selective News Sharing and Endogeneity

The selective sharing of our dogmatic agents can be interpreted in at least two ways. The first is that dogmatic agents care about the state (i.e., it affects their underlying preferences), but have extreme beliefs that are very hard to change—perhaps because they are stubborn, narrow minded, or blindly follow and promote their conviction. Formally, they have a degenerate prior belief that places probability 1 on either A or B, which does not change with new information. Alternatively, dogmatic agents can change their views, yet much more slowly than non-dogmatic agents.[21] As a result, how they selectively share information is very persistent. A second interpretation is that dogmatic agents do not care about the state (their underlying preferences are state independent) and simply share information to promote their agenda. This is related to the framework in Jin et al.'s (2021) experiment—and, of course, more general models of selective sharing of verifiable information starting in the 80s. In that experiment, "senders" face the same sharing decisions as in our model: disclose the signal they have or stay silent, but nothing else. They also have monotonic preferences in receivers' belief about the state, while receivers want to learn correctly.

Future research may endogenize selective sharing in settings similar to ours. In Online Appendix D, we take a first step in this direction by exploring an extension in which the type of selective sharing we assume in the paper emerges endogenously. This extension uses several standard assumptions to keep things simple. Generalizing these assumptions points to future work.

7.4 News Re-Sharing

In reality, people can re-share news received from friends, and thus any agent may receive third-, fourth-, or nth-hand information. Our model assumed no signal re-sharing, but we

[21] For studies on people's reluctance to change worldview see, e.g., Edwards (1968), Nisbett and Ross (1980), Evans (1989), Nickerson (1998), and Galperti (2019).

can easily extend our analysis to allow this. For instance, Jane may receive a signal shared by Tom through a friend of hers. The simplest case is when Jane knows the paths that a signal travelled to reach her. This is similar to knowing the origin of a re-tweet on Twitter. In this case, re-sharing is equivalent to adding Tom as a direct friend in Jane's echo chamber. If the paths allow all signals to reach Jane, it is equivalent to adding Tom as a friend of Tom's type; if the paths allow only signals favorable to state ω to reach Jane, it is equivalent to adding Tom as an ω-dogmatic friend.

As it is well known from the literature on rational learning in networks (see, e.g., Mueller-Frank (2013), Golub and Sadler (2016), Mueller-Frank and Neri (2021)), complications can arise if Jane does not know the origin of a signal or the paths it travelled. Since our signals are verifiable, if there is no more than one path, there is no confounding or correlation issue. But if there are multiple paths, two re-shared signals may come from the same origin, which introduces correlation. In this case, however, consistent evidence shows that people tend to neglect correlation (Enke and Zimmermann (2017)), which has been previously assumed in social-learning models (Eyster and Rabin (2010))). Under this assumption, each path from a source to Jane acts as a separate friend in her echo chamber. As long as correlation is sufficiently weak, one can show that it does not matter for long-run beliefs.

Finally, we can capture re-sharing in a reduced form as increasing the news-arrival rate for an agent's friends (i.e., γ). This is because they can now share both a signal they got directly from its source (e.g., a newspaper) and a signal they got from one of their friends. The consequences can be easily seen through our results. If re-sharing increases γ but agents do not take this into account fully, it effectively becomes another mechanism that leads to $\hat{\gamma} < \gamma$.

7.5 Agents' Heterogeneity

A final comment is in order on the heterogeneity between normal agents that we allow. We assumed that only the composition of echo chambers can differ between them. Otherwise, they are identical and have the same model of the world: The prior π, the signal distribution (γ and (1)), and the form of information misperception ($\hat{\gamma}$, (I), (II), or (III)) are the same for all. This setting helps to focus on the role of different information diets due to echo chambers as a driver of belief polarization. It is intuitive that adding differences between agents can introduce other drivers of polarization, which can be easily inferred from our results.

8 Concluding Remarks

We studied if and when learning from shared news can lead to belief polarization. Our answer is consistent with some common narratives about news sharing, yet highlights several qualifications. Selective sharing alone cannot lead to polarization, even if it gives rise to unbalanced news diets. It has to be combined with some misperception that causes people to incorrectly respond to others' selection of news, which has been documented in the literature on selection neglect. Moreover, this key mechanism leads to polarization if (and only if) the quality of first-hand information is sufficiently low. These results clarify the relationship between information quality, selective sharing, and polarization.

Our analysis goes to the heart of why new communication channels and formats on the Internet may contribute to, worsen, or curb polarization. First, the dramatic expansion of communication between people may have increased the consumption of selected second-hand news (e.g., on social media). Second, the quality of consumed information may have worsened: Tweets and social-media posts tend to be short, imprecise, and absorbed only superficially. Third, the Internet has offered bad actors a megaphone to spread fake news, and we found that it is selective sharing—not fake news per se—that can distort beliefs. Bad actors may also try to expand echo-chamber imbalances or, more subtly, release bits of true but low-quality news with high frequency (like Tweets), which amplifies the power of misperceived selective sharing. However, the Internet may also help people access higher-quality news and form more balanced social connections, which can curb polarization.

Several directions remain for future research. For instance, the role of unbalanced echo chambers in our analysis begs the question of what happens when social links form endogenously. In this process, people may follow their demand for information or other socio-economic forces (identity, class, race, ideology, work career, etc.). Some people may tend to link with like-minded friends, creating a vicious cycle between belief polarization and echo chambers' imbalance. Others may tend to link with reliable sources of information, possibly with opposite effects. Which tendency prevails is ultimately an empirical question. Recent studies on homophily in social networks, which include Golub and Jackson (2012), Baccara and Yariv (2013), and Halberstam and Knight (2016), can guide analysis in this direction. We hope our framework can prompt further theoretical investigations that expand on our first step in studying endogenous selective sharing, belief polarization, and its consequences such as gridlock. On the empirical side, future research could consider more detailed simulations of our theory that examine the evolution of opinions about other social topics (such as the use of vaccines or other public-health issues). Our approach to polarization and its drivers may guide data collection efforts, for instance, by highlighting the usefulness of more survey questions recording opinions about objective facts and people's diets of second-hand news.

Renee Bowen, UC San Diego and NBER

Danil Dmitriev, UC San Diego

Simone Galperti, UC San Diego

Appendix

A Proofs

A.1 Proof of Proposition 1

Consider an agent with n normal friends and $d > 0$ dogmatic friends of each type.

Without loss of generality, we can ignore the normal friends and assume that $n = 0$. The reason is that, using the Law of Total Expectation, we can rewrite $\mathbb{E}[\mu]$ as a sum over all possible signal realizations of dogmatic friends, where in each term we have the expected posterior conditional on a given signal realization. The remaining uncertainty in this conditional posterior are signal realizations of normal friends. Since normal friends stay silent if and only if they truly receive no signal, the agent's misspecification plays no role and the expectation of that conditional posterior must be equal to the "prior." That is, it equals the posterior updated only on the signals of dogmatic friends. Hence, we can focus on the dogmatic friends.

Let a_A be the number of signals $s = a$ that the A-dogmatic friends receive, and b_B be the number of $s = b$ that the B-dogmatic friends receive. Denote $\mathbf{s} = \{a_A, b_B\}$. Given the correct γ, the posterior that $\omega = A$ is

$$\mu^*(\mathbf{s}) = \frac{\pi \mathbb{P}^*(\mathbf{s}|A)}{\pi \mathbb{P}^*(\mathbf{s}|A) + (1-\pi)\mathbb{P}^*(\mathbf{s}|B)},$$

where

$$\mathbb{P}^*(\mathbf{s}|A) = \frac{d!d!}{a_A!(d-a_A)!b_B!(d-b_B)!}\gamma^{a_A+b_B}q^{a_A}(1-q)^{b_B}(\gamma(1-q)+(1-\gamma))^{d-a_A}(\gamma q+(1-\gamma))^{d-b_B},$$
$$\mathbb{P}^*(\mathbf{s}|B) = \frac{d!d!}{a_A!(d-a_A)!b_B!(d-b_B)!}\gamma^{a_A+b_B}(1-q)^{a_A}q^{b_B}(\gamma q+(1-\gamma))^{d-a_A}(\gamma(1-q)+(1-\gamma))^{d-b_B}.$$

Given the incorrect $\hat{\gamma}$, the agent's posterior belief given $\mathbf{s}$ will be

$$\mu(\mathbf{s}) = \frac{\pi \mathbb{P}(\mathbf{s}|A)}{\pi \mathbb{P}(\mathbf{s}|A) + (1-\pi)\mathbb{P}(\mathbf{s}|B)}, \tag{4}$$

where $\mathbb{P}(\mathbf{s}|A)$ and $\mathbb{P}(\mathbf{s}|B)$ are calculated replacing γ with $\hat{\gamma}$. To understand each term consider $\mathbb{P}^*(\mathbf{s}|A)$, which is the conditional probability of observing $\mathbf{s}$ given $\omega = A$. Then, $(\gamma q)^{a_A}$ is the probability of getting a_A signals $s = a$ from A-dogmatic friends; $(\gamma(1-q))^{b_B}$ is the probability of getting b_B signals $s = b$ from B-dogmatic friends; $(\gamma q + (1-\gamma))^{d_B - b_B}$ is the probability of observing $d_B - b_B$ B-dogmatic friends staying silent, as it is either a genuine silence (with prob. $1-\gamma$) or a suppressed signal $s = a$ (with prob. γq); $(\gamma(1-q) + (1-\gamma))^{d_A - a_A}$ is the probability of observing $d_A - a_A$ A-dogmatic friends staying silent, as it is either a genuine silence (with prob. $1-\gamma$) or a suppressed signal $s = b$ (with prob. $\gamma(1-q)$). For $\mathbb{P}^*(\mathbf{s}|B)$, the probabilities q and $1-q$ are reversed because the true state is B.

Consider the expectation of the difference between μ^* and μ:

$$\begin{aligned}
\mathbb{E}[\mu - \mu^*] &= \sum_{\mathbf{s}} \left(\pi \mathbb{P}^*(\mathbf{s}|A) + (1-\pi)\mathbb{P}^*(\mathbf{s}|B)\right)\left(\mu(\mathbf{s}) - \mu^*(\mathbf{s})\right) \\
&= \sum_{\mathbf{s}} \pi \mathbb{P}^*(\mathbf{s}|A) \left(\frac{\mathbb{P}(\mathbf{s}|A)}{\mathbb{P}^*(\mathbf{s}|A)} \cdot \frac{\pi \mathbb{P}^*(\mathbf{s}|A) + (1-\pi)\mathbb{P}^*(\mathbf{s}|B)}{\pi \mathbb{P}(\mathbf{s}|A) + (1-\pi)\mathbb{P}(\mathbf{s}|B)} - 1 \right) \\
&= \sum_{\mathbf{s}} \pi \mathbb{P}^*(\mathbf{s}|A) \left(\frac{1 + \rho Q^{a_A - b_B} \Gamma^{a_A - b_B}}{1 + \rho Q^{a_A - b_B} \hat{\Gamma}^{a_A - b_B}} - 1 \right),
\end{aligned}$$

where

$$Q = \frac{1-q}{q}, \quad \Gamma = \frac{\gamma(1-q) + (1-\gamma)}{\gamma q + (1-\gamma)}, \quad \hat{\Gamma} = \frac{\hat{\gamma}(1-q) + (1-\hat{\gamma})}{\hat{\gamma} q + (1-\hat{\gamma})}, \quad \rho = \frac{1-\pi}{\pi}. \tag{5}$$

Using the expression of $\mathbb{P}^*(\mathbf{s}|A)$, we can write

$$\begin{aligned}
\mathbb{E}[\mu - \mu^*] \quad &= \pi \textstyle\sum_{a_A, b_B} \left(\frac{d!}{a_A!(d-a_A)!} \cdot \frac{d!}{b_B!(d-b_B)!} \gamma^{a_A + b_B} q^{a_A + b_B} (\gamma(1-q) + (1-\gamma))^{2d - a_A - b_B} \right) \\
&\qquad \times \left(\frac{1-q}{q}\right)^{b_B} \left(\frac{\gamma(1-q)+(1-\gamma)}{\gamma q + (1-\gamma)}\right)^{b_B - d} \left(\frac{1+\rho Q^{a_A - b_B}\Gamma^{a_A - b_B}}{1+\rho Q^{a_A - b_B}\hat{\Gamma}^{a_A - b_B}} - 1 \right) \\
&= \pi \textstyle\sum_{a_A, b_B} \left(\frac{d!}{a_A!(d-a_A)!} \cdot \frac{d!}{b_B!(d-b_B)!} \gamma^{a_A + b_B} q^{a_A + b_B} (\gamma(1-q) + (1-\gamma))^{2d - a_A - b_B} \right) \\
&\qquad \times \Gamma^{-d} \left(\frac{Q^{b_B}\Gamma^{b_B} + \rho Q^{a_A}\Gamma^{a_A}}{Q^{b_B}\hat{\Gamma}^{b_B} + \rho Q^{a_A}\hat{\Gamma}^{a_A}} Q^{b_B}\hat{\Gamma}^{b_B} - Q^{b_B}\Gamma^{b_B} \right)
\end{aligned}$$

$$= \pi\Gamma^{-d} \sum_{0\le x\le y\le d} \left(\tfrac{d!d!}{x!(d-x)!y!(d-y)!}\gamma^{x+y}q^{x+y}(\gamma(1-q)+(1-\gamma))^{2d-x-y}\right) \qquad (6)$$
$$\times \left(\tfrac{Q^y\Gamma^y+\rho Q^x\Gamma^x}{Q^y\hat{\Gamma}^y+\rho Q^x\hat{\Gamma}^x}Q^y\hat{\Gamma}^y - Q^y\Gamma^y + \tfrac{Q^x\Gamma^x+\rho Q^y\Gamma^y}{Q^x\hat{\Gamma}^x+\rho Q^y\hat{\Gamma}^y}Q^x\hat{\Gamma}^x - Q^x\Gamma^x\right).$$

The key is that while the original distribution $\mathbb{P}^*(\mathbf{s}|A)$ is not symmetric between a_A and b_B, the last line involves a symmetric distribution between x and y. We want to prove that the sum in (6) is negative for $\rho > 1$, which will imply $\mathbb{E}[\mu - \mu^*] < 0$ for $\pi < \frac{1}{2}$.

Consider the derivative with respect to ρ of the term in the second line of (6), denoted by Δ_{xy}:

$$\frac{\partial \Delta_{xy}}{\partial \rho} = \frac{Q^{x+y}(\Gamma^x\hat{\Gamma}^y - \Gamma^y\hat{\Gamma}^x)}{(Q^y\hat{\Gamma}^y+\rho Q^x\hat{\Gamma}^x)^2}Q^y\hat{\Gamma}^y + \frac{Q^{x+y}(\Gamma^y\hat{\Gamma}^x - \Gamma^x\hat{\Gamma}^y)}{(Q^x\hat{\Gamma}^x+\rho Q^y\hat{\Gamma}^y)^2}Q^x\hat{\Gamma}^x,$$

which is negative if and only if

$$\frac{\Gamma^x\hat{\Gamma}^y - \Gamma^y\hat{\Gamma}^x}{(Q^y\hat{\Gamma}^y+\rho Q^x\hat{\Gamma}^x)^2}Q^y\hat{\Gamma}^y < \frac{\Gamma^x\hat{\Gamma}^y - \Gamma^y\hat{\Gamma}^x}{(Q^x\hat{\Gamma}^x+\rho Q^y\hat{\Gamma}^y)^2}Q^x\hat{\Gamma}^x.$$

Recall that $y \ge x$. Note that $\Gamma^x\hat{\Gamma}^y - \Gamma^y\hat{\Gamma}^x > 0$ if and only if $\hat{\Gamma}^{y-x} > \Gamma^{y-x}$. If $y = x$, this holds with equality and the derivative above is 0. If $y > x$, $\hat{\Gamma}^{y-x} > \Gamma^{y-x}$ is equivalent to $\hat{\Gamma} > \Gamma$, which in turn is equivalent to $\hat{\gamma} < \gamma$. From here on, we assume $y > x$.

Then, the derivative of Δ_{xy} is negative if and only if

$$\frac{Q^y\hat{\Gamma}^y}{(Q^y\hat{\Gamma}^y+\rho Q^x\hat{\Gamma}^x)^2} < \frac{Q^x\hat{\Gamma}^x}{(Q^x\hat{\Gamma}^x+\rho Q^y\hat{\Gamma}^y)^2}.$$

Note that $Q\hat{\Gamma} < 1$, which implies $(Q\hat{\Gamma})^y < (Q\hat{\Gamma})^x$. Using this, we can obtain the equivalent inequality

$$(2-(1+\rho)^2)(Q\hat{\Gamma})^{x+y} < \rho^2((Q\hat{\Gamma})^{2x} + (Q\hat{\Gamma})^{2y})$$

For $\rho > 1$, this inequality holds, as the left side is negative and the right side is positive. Given that this holds for any $x < y$, it follows that the derivative of the entire sum in (6) is negative for $\rho > 1$. Note that this sum is equal to 0 (term by term) at $\rho = 1$. This implies that the sum becomes negative for all $\rho > 1$ as desired. In other words, given $\hat{\gamma} < \gamma$, moving the prior from 50-50 towards a state will make the unconditional expected posterior of that state *higher* than the prior.

A.2 Proof of Proposition 2

The proof strategy is to first find the derivative of $\mathbb{E}[\mu|\omega]$ with respect to q at $q = \frac{1}{2}$ and then show how its sign depends on $d_A - d_B$. Using continuity of $\mathbb{E}[\mu|\omega]$ in q and the fact that $\mathbb{E}[\mu|\omega] = \pi$ at $q = \frac{1}{2}$, we will obtain the desired conclusion.

Using (4) and (5), for a given realization $\mathbf{s} = (a_A, b_B, a_N, b_N)$, an agent's posterior that $\omega = A$ can be written as

$$\mu(\mathbf{s}) = \frac{\pi}{\pi + (1-\pi) Q^M \hat{\Gamma}^S}.$$

To compute $\mathbb{E}[\mu|\omega]$, it is useful to use iterated expectations and condition on the set of friends who receive a signal. Let $\mathbb{E}[\mu|\omega, x_A, x_B, x_N]$ be the expected posterior conditional on the event that the state is ω and that x_A A-dogmatic friends, x_B B-dogmatic friends, and x_N normal friends received a signal. For simplicity, x_N includes the agent's own signal. Abusing notation a bit, let $N = n+1$. We can then write

$$\mathbb{E}[\mu|\omega] = \sum_{x_A=0}^{d_A} \sum_{x_B=0}^{d_B} \sum_{x_N=0}^{N} \frac{d_A! d_B! N!}{x_A!(d_A - x_A)! x_B!(d_B - x_B)! x_N!(N - x_N)!} \cdot \gamma^{x_A + x_B + x_N} (1-\gamma)^{d_A + d_B + N - x_A - x_B - x_N} \mathbb{E}[\mu|\omega, x_A, x_B, x_N].$$

The derivative of $\mathbb{E}[\mu|\omega]$ with respect to q is

$$\frac{\partial}{\partial q} \mathbb{E}[\mu|\omega] = \sum_{x_A=0}^{d_A} \sum_{x_B=0}^{d_B} \sum_{x_N=0}^{N} \frac{d_A! d_B! N!}{x_A!(d_A - x_A)! x_B!(d_B - x_B)! x_N!(N - x_N)!} \cdot \gamma^{x_A + x_B + x_N} (1-\gamma)^{d_A + d_B + N - x_A - x_B - x_N} \frac{\partial}{\partial q} \mathbb{E}[\mu|\omega, x_A, x_B, x_N]. \tag{7}$$

We now find $\frac{\partial}{\partial q} \mathbb{E}[\mu|\omega, x_A, x_B, x_N]$ and evaluate it at $q = \frac{1}{2}$.

Lemma 1.

$$\frac{\partial}{\partial q}\mathbb{E}\left[\mu|\omega,x_A,x_B,x_N\right]\Big|_{q=\frac{1}{2}}=\sum_{a_N=0}^{x_N}\frac{x_N!}{a_N!(x_N-a_N)!}\left(\frac{1}{2}\right)^{x_N}\frac{\partial}{\partial q}\mathbb{E}\left[\mu|a_N,\omega,x_A,x_B,x_N\right]\Big|_{q=\frac{1}{2}}.$$

Proof. Letting $H(q;A)=q$ and $H(q;B)=1-q$, we can write

$$\mathbb{E}\left[\mu|\omega,x_A,x_B,x_N\right]=\sum_{a_N=0}^{x_N}\frac{x_N!}{a_N!(x_N-a_N)!}H(q;\omega)^{a_N}(1-H(q;\omega))^{x_N-a_N}\mathbb{E}\left[\mu|a_N,\omega,x_A,x_B,x_N\right].$$

The derivative of $\mathbb{E}\left[\mu|\omega,x_A,x_B,x_N\right]$ can thus be represented as

$$\frac{\partial}{\partial q}\mathbb{E}\left[\mu|\omega,x_A,x_B,x_N\right]=\textstyle\sum_{a_N=0}^{x_N}\frac{x_N!}{a_N!(x_N-a_N)!}\Big[a_N H(q;\omega)^{a_N-1}(1-H(q;,\omega))^{x_N-a_N}-$$

$$-(x_N-a_N)H(q;\omega)^{a_N}(1-H(q;\omega))^{x_N-a_N-1}\Big]H_q(q;\omega)\mathbb{E}\left[\mu|a_N,\omega,x_A,x_B,x_N\right]+$$

$$+\textstyle\sum_{a_N=0}^{x_N}\frac{x_N!}{a_N!(x_N-a_N)!}H(q;\omega)^{a_N}(1-H(q;\omega))^{x_N-a_N}\frac{\partial}{\partial q}\mathbb{E}\left[\mu|a_N,\omega,x_A,x_B,x_N\right].$$

If $q=\frac{1}{2}$, then $H(q;\omega)=\frac{1}{2}$ for each ω. Also, the agent will not update her prior based on any signals: $\mathbb{E}\left[\mu|a_N,\omega,x_A,x_B,x_N\right]=\pi$. The above expression thus simplifies to

$$\frac{\partial}{\partial q}\mathbb{E}\left[\mu|\omega,x_A,x_B,x_N\right]\Big|_{q=\frac{1}{2}}=\textstyle\sum_{a_N=0}^{x_N}\frac{x_N!}{a_N!(x_N-a_N)!}\left(\frac{1}{2}\right)^{x_N-1}(2a_N-x_N)\,H_q\left(\frac{1}{2};\omega\right)\pi+$$

$$+\textstyle\sum_{a_N=0}^{x_N}\frac{x_N!}{a_N!(x_N-a_N)!}\left(\frac{1}{2}\right)^{x_N}\frac{\partial}{\partial q}\mathbb{E}\left[\mu|a_N,\omega,x_A,x_B,x_N\right]\Big|_{q=\frac{1}{2}}.$$

Note that

$$\sum_{a_N=0}^{x_N}\frac{x_N!}{a_N!(x_N-a_N)!}\left(\frac{1}{2}\right)^{x_N-1}(2a_N-x_N)=0,$$

because for each positive term in the sum there is an identical term with a negative sign. We can then write

$$\frac{\partial}{\partial q}\mathbb{E}\left[\mu|\omega,x_A,x_B,x_N\right]\Big|_{q=\frac{1}{2}}=\sum_{a_N=0}^{x_N}\frac{x_N!}{a_N!(x_N-a_N)!}\left(\frac{1}{2}\right)^{x_N}\frac{\partial}{\partial q}\mathbb{E}\left[\mu|a_N,\omega,x_A,x_B,x_N\right]\Big|_{q=\frac{1}{2}}.$$

■

It remains to evaluate $\frac{\partial}{\partial q}\mathbb{E}\left[\mu|a_N,\omega,x_A,x_B,x_N\right]\Big|_{q=\frac{1}{2}}$. The following is a first intermediate step.[22]

Lemma 2.

$$\frac{\partial}{\partial q}\mathbb{E}\left[\mu|a_N,\omega,x_A,x_B,x_N\right]=\sum_{a_D=0}^{\left\lfloor\frac{x_A+x_B-1}{2}\right\rfloor}\frac{(x_A+x_B)!}{a_D!(x_A+x_B-a_D)!}\frac{\partial}{\partial q}\Big(f(a_D,q,a_N)+f(x_A+x_B-a_D,q,a_N)\Big),$$

where

$$f(k,q,a_N)=\frac{\pi H(q;\omega)^k(1-H(q;\omega))^{x_A+x_B-k}}{\pi+(1-\pi)Q^{k-x_B+2a_N-x_N}\hat{\Gamma}^{k-x_B-(d_A-d_B)}}.$$

Proof. Let $a_D\leq x_A+x_B$ be the total number of $s=a$ that A- and B-dogmatic friends have received. Using $a_B=x_B-b_B$ and $b_N=x_N-a_N$, we can write

$$\mu=\frac{\pi}{\pi+(1-\pi)Q^{a_D-x_B+2a_N-x_N}\hat{\Gamma}^{a_D-x_A-(d_A-x_A)+(d_B-x_B)}}$$

Note that μ includes the dogmatic friends who have not received a signal (d_A-x_A A-dogmatic and d_B-x_B B-dogmatic), as the agent does not know whether they did not get a signal or they suppressed it. Using this, we can obtain

$$\mathbb{E}\left[\mu|a_N,\omega,x_A,x_B,x_N\right]=\sum_{a_D=0}^{x_A+x_B}\Bigg[\frac{(x_A+x_B)!}{a_D!(x_A+x_B-a_D)!}H(q;\omega)^{a_D}(1-H(q;\omega))^{x_A+x_B-a_D}\cdot\frac{\pi}{\pi+(1-\pi)Q^{a_D-x_B+2a_N-x_N}\hat{\Gamma}^{a_D-x_B-(d_A-d_B)}}\Bigg].$$

Using binomial symmetry, we get

$$\mathbb{E}\left[\mu|a_N,\omega,x_A,x_B,x_N\right]=\sum_{a_D=0}^{\left\lfloor\frac{x_A+x_B-1}{2}\right\rfloor}\frac{(x_A+x_B)!}{a_D!(x_A+x_B-a_D)!}\Big(f(a_D,q,a_N)+f(x_A+x_B-a_D,q,a_N)\Big),$$

where $f(k,q,a_N)$ is as defined in the lemma. Taking the derivative with respect to q gives the result. ■

[22]The symbol $\lfloor x\rfloor$ denotes the largest integer smaller than x.

The next is a second intermediate step to evaluate $\frac{\partial}{\partial q}\mathbb{E}\left[\mu | a_N, \omega, x_A, x_B, x_N\right]\Big|_{q=\frac{1}{2}}$.

Lemma 3. *At* $q = \frac{1}{2}$,

$$\frac{\partial}{\partial q}\Big(f(a_D, q, a_N) + f(x_A + x_B - a_D, q, a_N)\Big)$$
$$= \left(\frac{1}{2}\right)^{x_A+x_B-1} 2\pi(1-\pi)\left[2(2a_N - x_N) + \frac{2}{2-\hat{\gamma}}(x_A - x_B) - \frac{2\hat{\gamma}}{2-\hat{\gamma}}(d_A - d_B)\right].$$

Proof. To simplify subsequent algebra, define $z(q,\hat{\gamma}) = \ln(\hat{\Gamma})\left[\ln(Q)\right]^{-1}$. Taking the derivative of $f(k, q, a_N)$ with respect to q gives

$$\frac{\partial}{\partial q} f(k,q,a_N) = \frac{\pi}{\pi + (1-\pi)Q^{k-x_B+2a_N-x_N+(k-x_B-(d_A-d_B))z(q,\hat{\gamma})}} \cdot$$
$$\cdot\Big((x_A + x_B - k)H(q;\omega)^k(1-H(q;\omega))^{x_A+x_B-k-1}\left(-H_q(q;\omega)\right) + kH(q;\omega)^{k-1}(1-H(q;\omega))^{x_A+x_B-k}H_q(q;\omega)\Big)$$
$$+H(q;\omega)^k(1-H(q;\omega))^{x_A+x_B-k}\pi(1-\pi)\cdot$$
$$\cdot\left[\frac{(k-d_B+2a_N-x_N)Q^{k-x_B+2a_N-x_N-1+(k-x_B-(d_A-d_B))z(q,\hat{\gamma})}\frac{1}{q^2}}{\left(\pi+(1-\pi)Q^{k-x_B+2a_N-x_N}\Gamma^{k-x_B-(d_A-d-B)}\right)^2} + \right.$$
$$\left. + \frac{(k - x_B - (d_A - d - B))Q^{k-x_B+2a_N-x_N+(k-x_B-1-(d_A-d_B))z(q,\hat{\gamma})}\frac{(2-\hat{\gamma})\hat{\gamma}}{(\hat{\gamma}q+(1-\hat{\gamma}))^2}}{\left(\pi + (1-\pi)Q^{k-x_B+2a_N-x_N}\Gamma^{k-x_B-(d_A-d_B)}\right)^2}\right],$$

which evaluated at $q = \frac{1}{2}$ equals

$$\left(\frac{1}{2}\right)^{x_A+x_B-1}(2k - x_A - x_B)H_q(q;\omega)\frac{\pi}{\pi+(1-\pi)}$$
$$+\left(\frac{1}{2}\right)^{x_A+x_B} 4\pi(1-\pi)\cdot\frac{(k - x_B + 2a_N - x_N) + (k - x_B - (d_A - d_B))\frac{\hat{\gamma}}{2-\hat{\gamma}}}{(\pi+(1-\pi))^2}$$
$$= \left(\tfrac{1}{2}\right)^{x_A+x_B-1}\cdot\Big[(2k - x_A - x_B)H_q(q;\omega)\pi$$
$$+2\pi(1-\pi)\Big((k - x_B + 2a_N - x_N) + (k - x_B - (d_A - d_B))\tfrac{\hat{\gamma}}{2-\hat{\gamma}}\Big)\Big].$$

Therefore, at $q = \frac{1}{2}$ we have

$$\tfrac{\partial}{\partial q}\left(f(a_D,q,a_N) + f(x_A + x_B - a_D, q, a_N)\right) = \left(\tfrac{1}{2}\right)^{x_A+x_B-1}2\pi(1-\pi)\Big(2(2a_N - x_N) + \tfrac{2}{2-\hat{\gamma}}(x_A - x_B) - \tfrac{2\hat{\gamma}}{2-\hat{\gamma}}(d_A - d_B)\Big).$$

■

We now further simplify $\frac{\partial}{\partial q}\mathbb{E}\left[\mu|\omega, x_A, x_B, x_N\right]\Big|_{q=\frac{1}{2}}$.

Lemma 4.

$$\frac{\partial}{\partial q}\mathbb{E}\left[\mu|\omega, x_A, x_B, x_N\right]\Big|_{q=\frac{1}{2}} = \frac{4\pi(1-\pi)}{2-\hat{\gamma}}\left((x_A - x_B) - \hat{\gamma}(d_A - d_B)\right).$$

Proof. From Lemma 2 and 3 we have

$$\begin{aligned}\frac{\partial}{\partial q}\mathbb{E}\left[\mu|a_N, \omega, x_A, x_B, x_N\right]\Big|_{q=\frac{1}{2}} &= \textstyle\sum_{a_D=0}^{\left\lfloor \frac{x_A+x_B-1}{2}\right\rfloor} \frac{(x_A+x_B)!}{a_D!(x_A+x_B-a_D)!}\left(\frac{1}{2}\right)^{x_A+x_B-1} 2\pi(1-\pi)\cdot \\ &\quad \cdot\left(2(2a_N - x_N) + (x_A - x_B) + (x_A - x_B - 2(d_A - d_B))\tfrac{\hat{\gamma}}{2-\hat{\gamma}}\right) \\ &= 4\pi(1-\pi)(2a_N - x_N) + \frac{4\pi(1-\pi)}{2-\hat{\gamma}}\left((x_A - x_B) - \hat{\gamma}(d_A - d_B)\right).\end{aligned}$$

The second equality follows from observing that the sum

$$\sum_{a_D=0}^{\left\lfloor \frac{x_A+x_B-1}{2}\right\rfloor} \frac{(x_A+x_B)!}{a_D!(x_A+x_B-a_D)!}\left(\frac{1}{2}\right)^{x_A+x_B-1}$$

is a binomial expansion of $\left(\frac{1}{2}+\frac{1}{2}\right)^{x_A+x_B} = 1$.

Using Lemma 2, we then have

$$\begin{aligned}\frac{\partial}{\partial q}\mathbb{E}\left[\mu|\omega, x_A, x_B, x_N\right]\Big|_{q=\frac{1}{2}} &= \textstyle\sum_{a_N=0}^{x_N} \frac{x_N!}{a_N!(x_N-a_N)!}\left(\frac{1}{2}\right)^{x_N}\Big[4\pi(1-\pi)(2a_N - x_N) \\ &\quad + \tfrac{4\pi(1-\pi)}{2-\hat{\gamma}}\left((x_A - x_B) - \hat{\gamma}(d_A - d_B)\right)\Big] \\ &= \frac{4\pi(1-\pi)}{2-\hat{\gamma}}\left((x_A - x_B) - \hat{\gamma}(d_A - d_B)\right),\end{aligned}$$

where the equality follows from the symmetry of $\sum_{a_N=0}^{x_N} \frac{x_N!}{a_N!(x_N-a_N)!}(2a_N - x_N)$ we used before. ■

Finally, we return to the derivative of $\mathbb{E}[\mu|\omega]$. Using Lemma 4, equation (7) simplifies to

$$
\begin{aligned}
\frac{\partial}{\partial q}\mathbb{E}\left[\mu|\omega\right]\Big|_{q=\frac{1}{2}} &= \sum_{x_A=0}^{d_A}\sum_{x_B=0}^{d_B}\sum_{x_N=0}^{N}\frac{d_A!d_B!N!}{x_A!(d_A-x_A)!x_B!(d_B-x_B)!x_N!(N-x_N)!}\gamma^{x_A+x_B+x_N}\cdot \\
&\quad \cdot(1-\gamma)^{d_A+d_B+N-x_A-x_B-x_N}\left(\frac{4\pi(1-\pi)}{2-\hat{\gamma}}\left((x_A-x_B)-\hat{\gamma}(d_A-d_B)\right)\right) \\
&= \frac{4\pi(1-\pi)}{2-\hat{\gamma}}\left[\sum_{x_A=0}^{d_A}\frac{d_A!}{x_A!(d_A-x_A)!}\gamma^{x_A}(1-\gamma)^{d_A-x_A}x_A\right. \\
&\quad \left. -\sum_{x_B=0}^{d_B}\frac{d_B!}{x_B!(d_B-x_B)!}\gamma^{x_B}(1-\gamma)^{d_B-x_B}x_B-\hat{\gamma}(d_A-d_B)\right] \\
&= \frac{4\pi(1-\pi)}{2-\hat{\gamma}}\left[\mathbb{E}[x_A]-\mathbb{E}[x_B]-\hat{\gamma}(d_A-d_B)\right] \\
&= \frac{4\pi(1-\pi)}{2-\hat{\gamma}}(d_A-d_B)(\gamma-\hat{\gamma}).
\end{aligned}
\tag{8}
$$

Given $\gamma > \hat{\gamma}$, if $d_A > d_B$, then the derivative is positive, which means that $\mathbb{E}[\mu|\omega]$ is distorted towards A for low q. The opposite is true if $d_A < d_B$.

A.3 Proof of Proposition 3

Recall that agent i has d_A A-dogmatic, d_B B-dogmatic, and n normal friends. Without loss of generality, assume $d_A > d_B$.[23]

Given T periods, denote the number of signals $s = a$ received by

- agent i as a_i,
- A-dogmatic friend j of agent i as a_j^A, $j \in \{1, 2, \ldots, d_A\}$,
- B-dogmatic friend j of agent i as a_j^B, $j \in \{1, 2, \ldots, d_B\}$,
- normal friend j of agent i as a_j^N, $j \in \{1, 2, \ldots, n\}$.

Denote the number of signals $s = b$ received by

[23] Our argument relates to that in Berk's (1966) main characterization result. We provide a direct proof, as this helps us show the dependence of the limit beliefs on the parameters of interest in this paper.

- agent i as b_i,
- A-dogmatic friend j of agent i as b_j^A, $j \in \{1, 2, \ldots, d_A\}$,
- B-dogmatic friend j of agent i as b_j^B, $j \in \{1, 2, \ldots, d_B\}$,
- normal friend j of agent i as b_j^N, $j \in \{1, 2, \ldots, n\}$.

Then, the number of no-signal arrivals for the same agents is given by

- $(T - a_i - b_i)$ for agent i,
- $(T - a_j^A - b_j^A)$ for A-dogmatic friend j of agent i, $j \in \{1, 2, \ldots, d_A\}$,
- $(T - a_j^B - b_j^B)$ for B-dogmatic friend j of agent i, $j \in \{1, 2, \ldots, d_B\}$,
- $(T - a_j^N - b_j^N)$ for normal friend j of agent i, $j \in \{1, 2, \ldots, n\}$.

Over the T periods, i's A-dogmatic friend j stayed silent b_j^A times, whereas her B-dogmatic friend k stayed silent a_k^B times.

Agent i's posterior satisfies

$$\mu(\mathbf{s}^T) = \frac{\pi}{\pi + (1 - \pi) Q^M \hat{\Gamma}^S},$$

where

$$\begin{aligned} M &= (a_i - b_i) + \sum_{j=1}^{n} (a_j^N - b_j^N) + \sum_{j=1}^{d_A} a_j^A - \sum_{j=1}^{d_B} b_j^B, \\ S &= \sum_{j=1}^{d_B} (T - b_j^B) - \sum_{j=1}^{d_A} (T - a_j^A). \end{aligned}$$

Thus, $\text{plim}_{T \to \infty} \mu(\mathbf{s}^T) = 1$ (resp. $\text{plim}_{T \to \infty} \mu(\mathbf{s}^T) = 0$) if and only if $Q^M \hat{\Gamma}^S$ converges to zero (resp. $+\infty$) with probability 1 as $T \to \infty$ or, equivalently, $\ln\left(Q^M \hat{\Gamma}^S\right)$ converges to $-\infty$ (resp. $+\infty$) with probability 1 as $T \to \infty$. Using $z(q, \hat{\gamma}) = \ln(\hat{\Gamma})[\ln(Q)]^{-1}$, we can write $\ln\left(Q^M \hat{\Gamma}^S\right)$ as

$\ln(Q)K(\mathbf{x}, T; q, \hat{\gamma})$, where

$$K(\mathbf{x}, T; q, \hat{\gamma}) = (a_i - b_i) + \sum_{j=1}^{n}(a_j^N - b_j^N) + \sum_{j=1}^{d_A} a_j^A - \sum_{j=1}^{d_B} b_j^B + \left(\sum_{j=1}^{d_B}(T - b_j^B) - \sum_{j=1}^{d_A}(T - a_j^A)\right) z(q, \hat{\gamma}),$$

and

$$\mathbf{x} = (a_i, b_i, (a_j^N, b_j^N)_{j=1}^{n}, (a_j^A, b_j^A)_{j=1}^{d_A}, (a_j^B, b_j^B)_{j=1}^{d_B}).$$

Given $\ln(Q) < 0$, we require that $K(\mathbf{x}, T; q, \hat{\gamma})$ converge to $+\infty$ (resp. $-\infty$) with probability 1 as $T \to \infty$. Note that

$$\lim_{T\to\infty} K(\mathbf{x}, T; q, \hat{\gamma}) = \lim_{T\to\infty} T \left(\frac{K(\mathbf{x}, T; q, \gamma)}{T}\right).$$

Using $H(q; A) = q$ and $H(q; B) = 1 - q$, by the Law of Large Numbers we have

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} \frac{K(\mathbf{x}, T; q, \hat{\gamma})}{T} &= (\gamma H(q;\omega) - \gamma(1 - H(q;\omega))) + \sum_{j=1}^{n}(\gamma H(q;\omega) - \gamma(1 - H(q;\omega))) \\
&\quad + \sum_{j=1}^{d_A} \gamma H(q;\omega) - \sum_{j=1}^{d_B} \gamma(1 - H(q;\omega)) + \\
&\quad + \left(\sum_{j=1}^{d_B}(1 - \gamma(1 - H(q;\omega))) - \sum_{j=1}^{d_A}(1 - \gamma H(q;\omega))\right) z(q, \hat{\gamma}) \\
&= -\gamma(1 + n + (1 + z(q, \hat{\gamma}))d_B) - (d_A - d_B)z(q, \hat{\gamma}) + \\
&\quad + \gamma\Big(2(1 + n) + (d_A + d_B)(1 + z(q, \hat{\gamma}))\Big) H(q;\omega).
\end{aligned}$$

Given this, $\operatorname{plim}_{T\to\infty} K(\mathbf{x}, T; q, \hat{\gamma}) = +\infty$ (resp. $-\infty$) if and only if this last expression is positive (resp. negative), which is equivalent to

$$H(q;\omega) > \text{(resp. } <\text{)} \;\; \tau(q) = \frac{1}{2} + \frac{((2 - \gamma)z(q, \hat{\gamma}) - \gamma)\,(d_A - d_B)}{2\gamma\,(2(1 + n) + (d_A + d_B)(1 + z(q, \hat{\gamma})))}. \tag{9}$$

Note that

$$\lim_{q \to \frac{1}{2}} \tau(q) = \frac{1}{2} + \frac{(\hat{\gamma} - \gamma)(d_A - d_B)}{\gamma(2-\hat{\gamma})(2(2-\hat{\gamma})(1+n) + 2(d_A + d_B))},$$

$$\lim_{q \to 1} \tau(q) = \frac{1}{2} + \frac{-\gamma(d_A - d_B)}{2\gamma(2(1+n) + (d_A + d_B))} \in (0,1).$$

In Online Appendix E, we show that $\tau(q)$ is decreasing and concave for $q \in \left(\frac{1}{2}, 1\right)$ and $\tau'\left(\frac{1}{2}\right) = 0$.

There are two cases to consider. First, suppose $\omega = B$ and hence $H(q;B) = 1 - q$. Given $\hat{\gamma} < \gamma$, condition (9) holds with "$>$" at $q = \frac{1}{2}$ and with "$<$" at $q = 1$. Given the aforementioned properties of $\tau(q)$, there exists a unique $q_{LR} \in \left(\frac{1}{2}, 1\right)$ such that $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 1$ if $q < q_{LR}$ and $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 0$ if $q > q_{LR}$.[24] Second, suppose $\omega = A$ and hence $H(q;A) = q$. Given $\hat{\gamma} < \gamma$, condition (9) holds with "$>$" at $q = \frac{1}{2}$ and hence at all $q \in \left(\frac{1}{2}, 1\right)$ by the properties of $\tau(q)$. It follows that $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 1$ for all $q \in \left(\frac{1}{2}, 1\right)$.

A.4 Proof of Proposition 4

Assuming $d_A > d_B$, we prove that q_{LR} is increasing in d_A and γ and decreasing in d_B, n and $\hat{\gamma}$. The case of $d_A < d_B$ follows similarly.

Consider the case of $\omega = B$, which is the case that can result in incorrect learning. The value of q_{LR} is the unique fixed point that satisfies

$$1 - q = \frac{1 + n + (1 + z(q,\hat{\gamma}))d_B + \frac{z(q,\hat{\gamma})}{\gamma}(d_A - d_B)}{2(1+n) + (d_A + d_B)(1 + z(q,\hat{\gamma}))},$$

or equivalently

$$q - \frac{1}{2} = \frac{(\gamma - (2-\gamma)z(q,\hat{\gamma}))(d_A - d_B)}{2\gamma(2(1+n) + (d_A + d_B)(1 + z(q,\hat{\gamma})))} \tag{10}$$

The right-hand side of the latter condition is strictly decreasing in d_B, n and $\hat{\gamma}$, which implies that q_{LR} is also decreasing in these variables. Since the right-hand side is increasing in γ, so is q_{LR}.

[24] For $q = q_{LR}$, $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T)$ may not be unique, consistent with Berk's (1966) discussion of his asymptotic carrier set $\mathcal{A}_0$ when this contains more than one point.

Finally, one can show that the derivative of the right-hand side of the former condition with respect to d_A equals

$$\frac{\left((2-\gamma)z(q,\hat{\gamma})-\gamma(1+z^2(q,\hat{\gamma}))\right)d_B+(1+n)\left((1-\gamma)z(q,\hat{\gamma})-\gamma\right)}{\gamma(2+2n+(d_A+d_B)(1+z(q,\hat{\gamma})))^2}<0,$$

where the inequality follows from $(2-\gamma)z(q,\hat{\gamma})<\gamma<\gamma(1+z^2(q,\hat{\gamma}))$ and $(1-\gamma)z(q,\hat{\gamma})<\gamma$. This implies that q_{LR} is increasing in d_A.

A.5 General Statement of Proposition 5 and its Proof

While Proposition 5 focused on the case $\lambda_A d_A > \lambda_B d_B$, we state and prove a more general result that also covers the case $\lambda_A d_A < \lambda_B d_B$.

Proposition 11 (General Statement of Proposition 5)**.** *Fix any agent with echo chamber $e = (d_A, d_B, n)$ that satisfies $d_A > d_B$ and $n \geq 1$. For any other echo chamber $e' = (\lambda_A d_A, \lambda_B d_B, \lambda_N n)$ with $\lambda_N \geq 0$, $\lambda_A \geq 0$ and $\lambda_B \geq 0$, we have $q_{LR}(e,\gamma,\hat{\gamma}) < q_{LR}(e',\gamma,\hat{\gamma})$ if*

$$\lambda_N \geq 1+\left(\frac{|\lambda_A d_A-\lambda_B d_B|}{d_A-d_B}-1\right)\left(1+\frac{1}{n}\right)+\frac{d_A d_B}{d_A-d_B}\cdot\frac{1}{n}\cdot\mathbf{J}(d_A,d_B,\hat{\gamma},\lambda_A,\lambda_B), \tag{11}$$

where

$$\mathbf{J}(d_A,d_B,\hat{\gamma},\lambda_A,\lambda_B)=\begin{cases}\max\left\{(\lambda_A-\lambda_B)\frac{2}{2-\hat{\gamma}},\ (\lambda_A-\lambda_B)\right\}, & \textit{if } \lambda_A d_A > \lambda_B d_B\\ \max\left\{\left(\lambda_B\frac{d_B}{d_A}-\lambda_A\frac{d_A}{d_B}\right)\frac{2}{2-\hat{\gamma}},\ \left(\lambda_B\frac{d_B}{d_A}-\lambda_A\frac{d_A}{d_B}\right)\right\}, & \textit{otherwise.}\end{cases}$$

Proof. We need to consider the fixed-point condition that defines q_{LR}, which depends on which state results in incorrect learning.

Case 1: Suppose $\lambda_A d_A - \lambda_B d_B > 0$. Then incorrect learning can occur in state B under both the original and the new echo chamber. A sufficient condition for $q_{LR}(e,\gamma,\hat{\gamma}) < q_{LR}(e',\gamma,\hat{\gamma})$ is the following:[25]

$$\frac{(\gamma-(2-\gamma)z(q,\hat{\gamma}))\,(\lambda_A d_A-\lambda_B d_B)}{2(1+\lambda_N n)+(\lambda_A d_A+\lambda_B d_B)(1+z(q,\hat{\gamma}))} \quad < \quad \frac{(\gamma-(2-\gamma)z(q,\hat{\gamma}))\,(d_A-d_B)}{2(1+n)+(d_A+d_B)(1+z(q,\hat{\gamma}))}, \quad \text{for all } q.$$

Given $\hat{\gamma}<\gamma$, one can show that $\gamma-(2-\gamma)z(q,\hat{\gamma})>0$ for all q. Using this and rearranging, the previous condition becomes

$$\lambda_N > 1+\left(\frac{\lambda_A d_A-\lambda_B d_B}{d_A-d_B}-1\right)\left(1+\frac{1}{n}\right)+\frac{(\lambda_A-\lambda_B)d_A d_B(1+z(q,\hat{\gamma}))}{(d_A-d_B)}\cdot\frac{1}{n}.$$

Since $z(q,\hat{\gamma})$ takes values between 0 and $\frac{\hat{\gamma}}{2-\hat{\gamma}}$, we obtain the sufficient condition

$$\lambda_N > 1+\left(\frac{\lambda_A d_A-\lambda_B d_B}{d_A-d_B}-1\right)\left(1+\frac{1}{n}\right)+\frac{d_A d_B}{d_A-d_B}\cdot\frac{1}{n}\cdot\max\left\{(\lambda_A-\lambda_B)\frac{2}{2-\hat{\gamma}},\ (\lambda_A-\lambda_B)\right\}.$$

Case 2: Suppose $\lambda_A d_A-\lambda_B d_B<0$. In this case, incorrect learning occurs in state B for the original echo chamber and state A for the new echo chamber. Then, $q_{LR}(e,\gamma,\hat{\gamma})<q_{LR}(e',\gamma,\hat{\gamma})$ if the following holds:

$$\frac{((2-\gamma)z(q,\hat{\gamma})-\gamma)\,(\lambda_A d_A-\lambda_B d_B)}{2(1+\lambda_N n)+(\lambda_A d_A+\lambda_B d_B)(1+z(q,\hat{\gamma}))} \quad < \quad \frac{(\gamma-(2-\gamma)z(q,\hat{\gamma}))\,(d_A-d_B)}{2(1+n)+(d_A+d_B)(1+z(q,\hat{\gamma}))}, \quad \text{for all } q.$$

Dividing by $(\gamma-(2-\gamma)z(q,\hat{\gamma}))$ and simplifying as before we obtain the sufficient condition

$$\begin{aligned}\lambda_N > 1&+\left(\frac{\lambda_B d_B-\lambda_A d_A}{d_A-d_B}-1\right)\left(1+\frac{1}{n}\right)\\&+\frac{d_A d_B}{d_A-d_B}\cdot\frac{1}{n}\cdot\max\left\{\left(\lambda_B\frac{d_B}{d_A}-\lambda_A\frac{d_A}{d_B}\right)\frac{2}{2-\hat{\gamma}},\ \left(\lambda_B\frac{d_B}{d_A}-\lambda_A\frac{d_A}{d_B}\right)\right\}.\end{aligned}$$

■

[25] Note that $q_{LR}(e,\gamma,\hat{\gamma})>q_{LR}(e',\gamma,\hat{\gamma})$ if the opposite inequality holds, which happens if $\lambda_A=\lambda_B=\lambda_N$ for instance.

A.6 Proof of Proposition 6

We only have to consider the case of $\omega = B$. In this case, q_{LR} is defined by condition (10). Fix d_A, d_B, n, λ and $\hat{q}$, we need to find λ_N such that

$$\hat{q} > \frac{1}{2} + \frac{(\gamma - (2-\gamma)z(q_{LR}, \hat{\gamma}))\,(\lambda d_A - \lambda d_B)}{2\gamma\,(2(1+\lambda_N n) + (\lambda d_A + \lambda d_B)(1 + z(q_{LR}, \hat{\gamma})))}.$$

Since the right-hand side is decreasing in $z(q, \hat{\gamma})$, we obtain a sufficient condition by imposing the inequality for the lowest value of $z(q, \hat{\gamma})$, which is 0. Rearranging yields the following condition:

$$\lambda_N > \frac{d_A - \hat{q}(d_A + d_B)}{(2\hat{q} - 1)n}\lambda - \frac{1}{n}.$$

A.7 Proof of Proposition 7

For this proof, we will use the notation $\Pi(q; \mathbf{e})$, $\mathcal{N}_\omega(q; \mathbf{e})$, and $\mathcal{N}_{-\omega}(q; \mathbf{e})$ to explicitly account for the dependence of Π and these sets on q. We start with the following Lemma 5.

Lemma 5. *As q increases, the set $\mathcal{N}_\omega(q; \mathbf{e})$ weakly expands and the set $\mathcal{N}_{-\omega}(q; \mathbf{e})$ weakly shrinks, both in the sense of set inclusion.*

Proof. Fix any $\hat{q} > \frac{1}{2}$. Suppose $i \in \mathcal{N}_\omega(\hat{q}; \mathbf{e})$. There are two possibilities. If $q_{LR}(e_i, \gamma, \hat{\gamma}) > \hat{q}$, then i's dogmatic majority must be towards the correct state ω. Increasing q beyond $q_{LR}(e_i, \gamma, \hat{\gamma})$ will lead the agent to learn correctly that the state is ω. Hence, $i \in \mathcal{N}_\omega(q; \mathbf{e})$ for all $q > \hat{q}$. If $q_{LR}(e_i, \gamma, \hat{\gamma}) < \hat{q}$ (for simplicity we omit the knife-edge case of equality), then i is already learning correctly and increasing q will not change her asymptotic beliefs. Thus, $i \in \mathcal{N}_\omega(q; \mathbf{e})$ for all $q > \hat{q}$. We conclude that $\mathcal{N}_\omega(q; \mathbf{e})$ does not shrink as q increases.

Now consider $j \in \mathcal{N}_{-\omega}(\hat{q}; \mathbf{e})$. This means that $q_{LR}(e_i, \gamma, \hat{\gamma}) > \hat{q}$. Increasing q beyond $q_{LR}(e_i, \gamma, \hat{\gamma})$ will lead j to learn correctly, which means she will leave $\mathcal{N}_{-\omega}(q; \mathbf{e})$.

■

Without loss of generality, label the agents so that $q_{LR}(e_i, \gamma, \hat{\gamma}) < q_{LR}(e_j, \gamma, \hat{\gamma})$ if and only if

$i < j$. Suppose $\omega = A$. Fix any $\hat{q} > \frac{1}{2}$ and consider sets $\mathcal{N}_A(\hat{q}; \mathbf{e})$ and $\mathcal{N}_B(\hat{q}; \mathbf{e})$. Let $i(\hat{q})$ be the lowest i such that $q_{LR}(e_i, \gamma, \hat{\gamma}) > \hat{q}$. As q increases to any q' that satisfy $q_{LR}(e_{i(\hat{q})}, \gamma, \hat{\gamma}) < q' < q_{LR}(e_{i(\hat{q})+1}, \gamma, \hat{\gamma})$, agent $i(\hat{q})$ will flip from $\mathcal{N}_B(q; \mathbf{e})$ to $\mathcal{N}_A(q; \mathbf{e})$. This implies

$$|\mathcal{N}_A(q'; \mathbf{e})| = |\mathcal{N}_A(\hat{q}; \mathbf{e})| + 1 \text{ and } |\mathcal{N}_B(q'; \mathbf{e})| = |\mathcal{N}_B(\hat{q}; \mathbf{e})| - 1.$$

Consider the long-run polarization:

$$\Pi(\hat{q}; \mathbf{e}) = \frac{4}{|\mathcal{N}|} \cdot |\mathcal{N}_A(\hat{q}; \mathbf{e})||\mathcal{N}_B(\hat{q}; \mathbf{e})|$$
$$\Pi(q'; \mathbf{e}) = \frac{4}{|\mathcal{N}|} \cdot (|\mathcal{N}_A(\hat{q}; \mathbf{e})| + 1) (\mathcal{N}_B(\hat{q}; \mathbf{e}) - 1)$$

Note that $\Pi(\hat{q}; \mathbf{e}) \geq \Pi(q'; \mathbf{e})$ if and only if

$$|\mathcal{N}_B(\hat{q})| \leq |\mathcal{N}_A(\hat{q})| + 1.$$

Hence, $\Pi(q; \mathbf{e})$ weakly decreases as q increases if and only if initially (i.e., at $q = \hat{q}$) the set of eventually incorrect agents is smaller than the set of eventually correct agents plus one. Since $\mathcal{N}_B(q; \mathbf{e})$ weakly shrinks in q, a necessary and sufficient condition for $\Pi(q; \mathbf{e})$ to be weakly decreasing in q is that $|\mathcal{N}_B(\frac{1}{2}; \mathbf{e})| = |\mathcal{D}_B|$ is weakly smaller than $|\mathcal{N}| - |\mathcal{D}_B| + 1$, that is, $|\mathcal{D}_B| \leq \frac{1}{2}(|\mathcal{N}| + 1)$.

A.8 Proof of Proposition 8

This result builds on Proposition 5, which implies that scaling all friends by the same factor λ increases the threshold q_{LR} below which incorrect learning occurs (see the proof).

Suppose that a minority of normal agents have an echo-chamber imbalance against ω (i.e., $|\mathcal{D}_{-\omega}| \leq \frac{1}{2}(|\mathcal{N}| + 1)$). Note that, for any q, the incorrect learners are always a minority, as they start in the minority for q close to $\frac{1}{2}$ and decrease in numbers as q increases. Now fix any q. If the agents' thresholds q_{LR} increase, weakly more agents with echo-chamber imbalances against ω will learn incorrectly. This expands the set of incorrect learners closer towards 50% of society, which implies an increase in polarization.

Suppose instead that a majority of normal agents have an echo-chamber imbalance against ω (i.e., $|\mathcal{D}_{-\omega}| > \frac{1}{2}(|\mathcal{N}|+1)$). If we start with q large enough so that a minority of agents learn incorrectly (which is always possible), then increasing the thresholds q_{LR} will cause more agents to learn incorrectly. This will expand the set of incorrect learners closer towards 50% of society, and, hence, increase polarization. Eventually, the set of incorrect learners may become the majority. At that point, further increases in q_{LR} will lead to the set becoming larger and farther from 50% of society, thereby decreasing polarization. In this case, polarization will be single-peaked.

A.9 Proof of Proposition 9

Consider $\omega = A$—the argument is the same for $\omega = B$. We want to find M such that $\mathbb{P}(\hat{s}^i_M = 1|\omega = A) > \bar{q}_{LR}$. This ensures by Proposition 3 that all agents in $\mathcal{N}$ learn correctly and hence $\hat{\Pi}(\mathbf{e}) = 0$. Now, note that

$$\begin{aligned}\mathbb{P}(\hat{s}^i_M = 0|\omega = A) &= \mathbb{P}\left(\sum_{k=0}^{M} I_{\{s_{ik}=a\}} < \frac{M}{2}|\omega = A\right) \\ &= \sum_{k=0}^{\lfloor \frac{M}{2} \rfloor} \frac{M!}{(M-k)!k!} q^k(1-q)^{M-k} \\ &\leq \exp\left(-2M\left(q - \frac{\lfloor \frac{M}{2} \rfloor}{M}\right)^2\right),\end{aligned}$$

where the last inequality follows from Hoeffding's inequality (Hoeffding (1963)). Therefore, our desired condition holds if

$$2M\left(q - \frac{\lfloor \frac{M}{2} \rfloor}{M}\right)^2 > -\ln(1-\bar{q}_{LR}).$$

Recalling that M is an odd number by assumption (i.e., $M = 2m+1$ for $m \in \mathbb{N}$), we have that

$$2M\left(q - \frac{\lfloor \frac{M}{2} \rfloor}{M}\right)^2 > 2M\left(q - \frac{1}{2}\right)^2.$$

		Imbalance: Small	Imbalance: Large
Size	Small	$n = 4$ $d_m = 4$ $d_M = 6$	$n = 4$ $d_m = 1$ $d_M = 9$
	Large	$n = 16$ $d_m = 16$ $d_M = 24$	$n = 16$ $d_m = 4$ $d_M = 36$

Table 2: Simulation Echo Chamber Composition

Therefore, it suffices that

$$2M\left(q - \frac{1}{2}\right)^2 > -\ln(1 - \bar{q}_{LR}).$$

B Further Simulations on the Speed of Belief Divergence

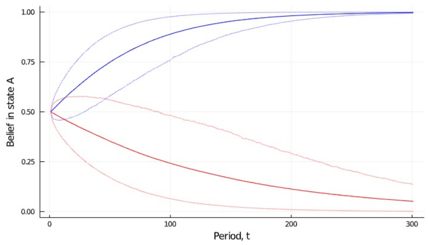
This appendix presents additional simulations to those in Section 6.1. We generated the distributions of belief trajectories using the same procedure and the following parameters:

- $\gamma = 0.95$;
- $\hat{\gamma} = 0.9\gamma$;
- $d_A^{\text{Alice}} = d_B^{\text{Bob}} = d_M$ and $d_B^{\text{Alice}} = d_A^{\text{Bob}} = d_m$, where
- Alice* and Bob* are again identical to Alice and Bob respectively, except that they use γ to update beliefs;
- $q_\ell = 0.7 \times 0.5 + 0.3 \times q_{LR}$ and $q_h = 0.3 \times 0.5 + 0.7 \times q_{LR}$.

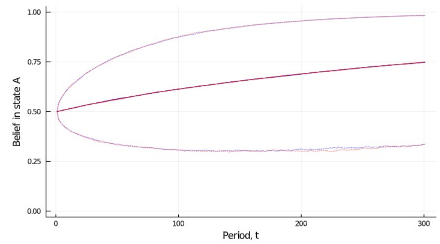
Figures 6 and 7 represent

- solid blue line: mean of the belief distribution of Alice and Alice*;

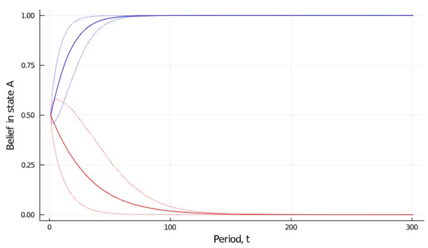
- light blue lines: 10% and 90% quantile of the belief distribution of Alice and Alice*;
- solid red line: mean of the belief distribution of Bob and Bob*;
- light red lines: 10% and 90% quantile of the belief distribution of Bob and Bob*.



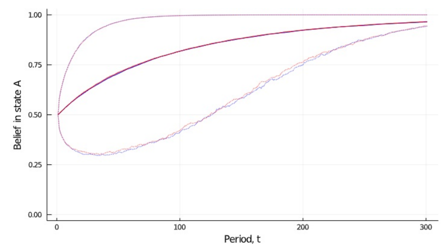
(a) Small echo chamber, large imbalance, misperception.



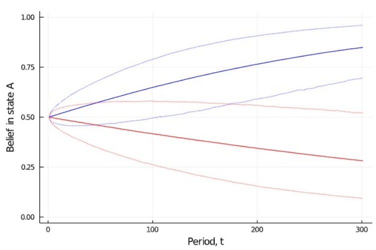
(b) Small echo chamber, large imbalance, *no* misperception.



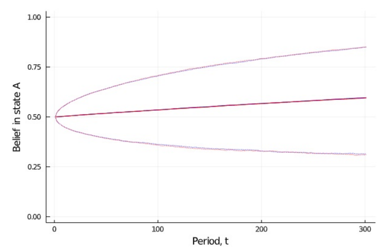
(c) Large echo chamber, large imbalance, misperception.



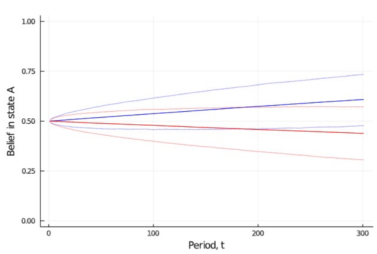
(d) Large echo chamber, large imbalance, *no* misperception.



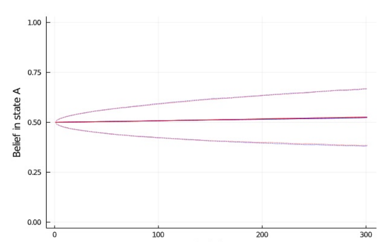
(e) Large echo chamber, small imbalance, misperception.



(f) Large echo chamber, small imbalance, *no* misperception.

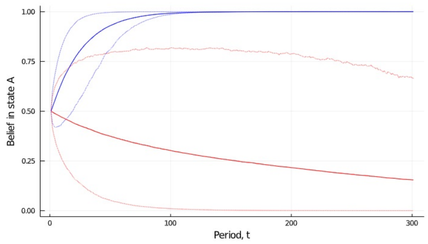


(g) Small echo chamber, small imbalance, misperception.

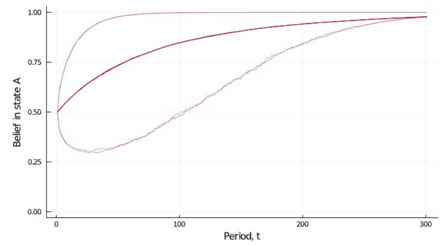


(h) Small echo chamber, small imbalance, *no* misperception.

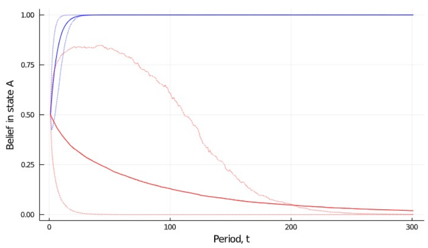
Figure 6: Belief paths for low quality of information (q_ℓ)



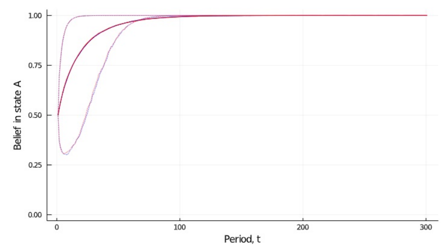
(a) Small echo chamber, large imbalance, misperception.



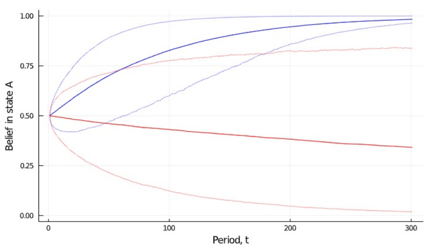
(b) Small echo chamber, large imbalance, *no* misperception.



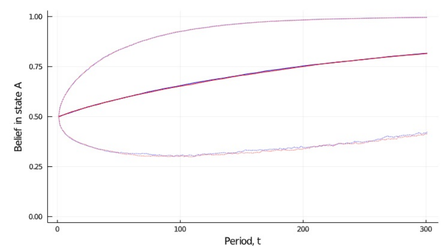
(c) Large echo chamber, large imbalance, misperception.



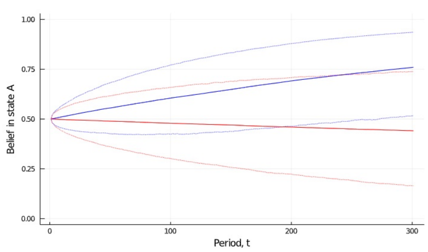
(d) Large echo chamber, large imbalance, *no* misperception.



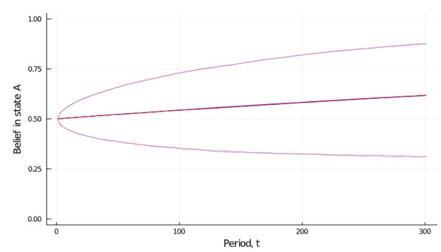
(e) Large echo chamber, small imbalance, misperception.



(f) Large echo chamber, small imbalance, *no* misperception.



(g) Small echo chamber, small imbalance, misperception.



(h) Small echo chamber, small imbalance, *no* misperception.

Figure 7: Belief paths for high quality of information (q_h)

References

Acemoglu, D., G. Como, F. Fagnani, and A. Ozdaglar (2013). Opinion fluctuations and disagreement in social networks. Mathematics of Operations Research 38(1), 1–27.

Acemoglu, D., A. Ozdaglar, and A. ParadehGheibi (2010). Spread of (Mis)information in Social Networks. Games and Economic Behavior 70, 194–227.

Alesina, A., A. Miano, and S. Stantcheva (2020). The Polarization of Reality. In AEA Papers and Proceedings, Volume 110, pp. 324–28.

Allcott, H., L. Braghieri, S. Eichmeyer, and M. Gentzkow (2020). The welfare effects of social media. American Economic Review 110(3), 629–76.

Allcott, H. and M. Gentzkow (2017). Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives 31(2), 211–36.

Andreoni, J. and T. Mylovanov (2012, February). Diverging Opinions. American Economic Journal: Microeconomics 4(1), 209–32.

Anesi, V. and T. Bowen (2021). Policy Experimentation in Committees: A Case against Veto Rights under Redistributive Constraints. American Economic Journal: Microeconomics.

Angelucci, C. and A. Prat (2021). Is journalistic truth dead? measuring how informed voters are about political news. Technical report, mimeo, Columbia University.

Athey, S., M. M. Mobius, and J. Pál (2017). The Impact of Aggregators on Internet News Consumption.

Austen-Smith, D. and J. S. Banks (1999). Positive Political Theory I: Collective Preference. University of Michigan Press.

Azzimonti, M. and M. Fernandes (2018, March). Social media networks, fake news, and polarization. Working Paper 24462, National Bureau of Economic Research.

Ba, C. and A. Gindin (2020). A Multi-Agent Model of Misspecified Learning with Overconfidence. Available at SSRN 3691728.

Baccara, M. and L. Yariv (2013). Homophily in Peer Groups. American Economic Journal: Microeconomics 5(3), 69–96.

Bakshy, E., S. Messing, and L. A. Adamic (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. Science 348, 1130–1132.

Barber, M. and N. McCarty (2015). Causes and Consequences of Polarization. Political Negotiation: A Handbook 37, 39–43.

Barberá, P. (2020). Social Media, Echo Chambers, and Political Polarization, Chapter 3, pp. 34–55. Cambridge University Press.

Bartels, L. M. (2008). Unequal Democracy: The Political Economy of the New Gilded Age. Princeton University Press.

Beinart, P. (2008). When politics no longer stops at the water's edge: Partisan polarization and foreign policy. Red and blue nation 2, 151–67.

Ben-Porath, E., E. Dekel, and B. Lipman (2018). Disclosure and Choice. Review of Economic Studies 85(3), 1471–1501.

Berk, R. H. (1966, 02). Limiting Behavior of Posterior Distributions when the Model is Incorrect. Ann. Math. Statist. 37(1), 51–58.

Bertrand, M. and E. Kamenica (2018). Coming Apart? Cultural Distances in the United States over Time. NBER Working Papers 24771, National Bureau of Economic Research, Inc.

Bishop, B. (2009). The Big Sort: Why the Clustering of Like-Minded America is Tearing us Apart. Houghton Mifflin Harcourt.

Bohren, A. and D. Hauser (2018). Social Learning with Model Misspecification: A Framework and a Robustness Result. PIER Working Paper Archive 18-017, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Bohren, J. A. (2016). Informational Herding with Model Misspecification. Journal of Economic Theory 163(C), 222–247.

Boxell, L., M. Gentzkow, and J. Shapiro (2018). Greater Internet Use is Not Associated with Faster Growth of Political Polarization Among US Demographic Groups. Proceedings of the National Academy of Sciences 115(3).

Boykoff, M. T. (2008). Lost in translation? united states television news coverage of anthropogenic climate change, 1995–2004. Climatic Change 86(1), 1–11.

Brady, D. W., J. Ferejohn, and L. Harbridge (2008). Polarization and public policy: A general assessment. Red and blue nation 2, 185–216.

Bursztyn, L., G. Egorov, R. Enikolopov, and M. Petrova (2019). Social Media and Xenophobia: Evidence from Russia. NBER Working Paper No. 26567.

Cinelli, M., G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini (2021). The echo chamber effect on social media. Proceedings of the National Academy of Sciences 118(9).

Conroy-Krutz, J. and D. C. Moehler (2015). Moderation from Bias: A Field Experiment on Partisan Media in a New Democracy. The Journal of Politics 77(2), 575–587.

Cross, P. (1977). Not Can But Will College Teachers Be Improved? New Directions for Higher Education 17, 1–15.

Dasaratha, K. and K. He (2020). Network Structure and Naive Sequential Learning. Theoretical Economics 15(2), 415–444.

Del Vicario, M., A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi (2016). The Spreading of Misinformation Online. Proceedings of the National Academy of Sciences 113(3), 554–559.

DeMarzo, P., I. Kremer, and A. Skrzypacz (2019). Test Design and Minimum Standards. American Economic Review 109(6), 2173–2207.

DeMarzo, P., D. Vayanos, and J. Zweibel (2003). Persuasion Bias, Social Influence, and Unidimensional Opinions. Quarterly Journal of Economics 118(3), 909–968.

Desmet, K. and R. Wacziarg (2018). The Cultural Divide. CEPR Discussion Papers 12947, C.E.P.R. Discussion Papers.

Dixit, A. and J. Weibull (2007). Political Polarization. Proceedings of the National Academy of Sciences 104(18), 7351–7256.

Dye, R. (1985). Disclosure of Nonproprietary Information. Journal of Accounting Research 23(1), 123–145.

Edwards, W. (1968). Conservatism in Human Information Processing. Formal Representation of Human Judgment.

Egan, P. J. and M. Mullin (2017). Climate change: Us public opinion. Annual Review of Political Science 20, 209–227.

Enke, B. (2020). What You See Is All There Is. The Quarterly Journal of Economics 135(3), 1363–1398.

Enke, B., F. Schwerter, and F. Zimmermann (2020, January). Associative memory and belief formation. Working Paper 26664, National Bureau of Economic Research.

Enke, B. and F. Zimmermann (2017). Correlation Neglect in Belief Formation. The Review of Economic Studies 86(1), 313–332.

Epstein, D. and J. D. Graham (2007). Polarized politics and policy consequences. Rand Corporation.

Erev, I., T. S. Wallsten, and D. V. Budescu (1994). Simultaneous over-and underconfidence: The role of error in judgment processes. Psychological review 101(3), 519.

Esponda, I. and D. Pouzo (2016). Berk–Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models. Econometrica 84, 1093–1130.

Esponda, I., D. Pouzo, and Y. Yamamoto (2019). Asymptotic Behavior of Bayesian Learners with Misspecified Models. arXiv preprint arXiv:1904.08551.

Esponda, I. and E. Vespa (2018). Endogenous sample selection: A laboratory study. Quantitative Economics 9(1), 183–216.

Esponda, I., E. Vespa, and S. Yuksel (2021). Mental Models and Learning: The Case of Base-Rate Neglect. *Working paper*.

Esteban, J.-M. and D. Ray (1994). On the Measurement of Polarization. Econometrica 62(4), 819–851.

Evans, J. S. B. (1989). Bias in Human Reasoning: Causes and Consequences. Lawrence Erlbaum Associates, Inc.

Eyster, E. and M. Rabin (2010). Naïve Herding in Rich-Information Settings. American Economic Journal: Microeconomics 2, 221–243.

Eyster, E., M. Rabin, and G. Weizsäcker (2018). An Experiment On Social Mislearning. Rationality and Competition Discussion Paper Series 73, CRC TRR 190 Rationality and Competition.

Ferejohn, J., I. Katznelson, and D. Yashar (2020). Social Media and Democracy: The State of the Field, Prospects for Reform, Chapter *SSRC Anxieties of Democracy*, pp. ii. Cambridge University Press.

Fiedler, K. and P. Juslin (2006). Information Sampling and Adaptive Cognition. Cambridge: Cambridge University Press.

Fiorina, M. and S. Abrams (2008). Political polarization in the american public. Annual Review of Political Science 11, 56–588.

Flaxman, S., S. Goel, and J. M. Rao (2016, 03). Filter Bubbles, Echo Chambers, and Online News Consumption. Public Opinion Quarterly 80(S1), 298–320.

Frick, M., R. Iijima, and Y. Ishii (2020). Misinterpreting Others and the Fragility of Social Learning. Econometrica 88(6), 2281–2328.

Fudenberg, D. and G. Lanzani (2020). Which misperceptions persist? Available at SSRN.

Fudenberg, D., G. Lanzani, and P. Strack (2020). Limits Points of Endogenous Misspecified Learning. Available at SSRN.

Fudenberg, D., G. Romanyuk, and P. Strack (2017). Active Learning with a Misspecified Prior. Theoretical Economics 12(3), 1155–1189.

Funk, C. and A. Tyson (2021). 67% of Americans perceive a rise in extreme weather, but partisans differ over government efforts to address it. Pew Research Center.

Gabielkov, M., A. Ramachandran, A. Chaintreau, and A. Legout (2016). Social clicks: What and who gets read on twitter? In Proceedings of the 2016 ACM SIGMETRICS international conference on measurement and modeling of computer science, pp. 179–192.

Gagnon-Bartsch, T., M. Rabin, and J. Schwartzstein (2018). Channeled attention and stable errors. Harvard Business School Boston.

Galperti, S. (2019). Persuasion: The Art of Changing Worldviews. American Economic Review 109(3), 996–1031.

Gilens, M. (2012). Affluence and Influence: Economic Inequality and Political Power in America. Princeton University Press.

Golub, B. and M. O. Jackson (2010). Naïve Learning in Social Networks and the Wisdom of Crowds. American Economic Journal: Microeconomics 2(1), 112–49.

Golub, B. and M. O. Jackson (2012). How Homophily Affects the Speed of Learning and Best-Response Dynamics. The Quarterly Journal of Economics 127(3), 1287–1338.

Golub, B. and E. Sadler (2016). Learning in social networks. The Oxford Handbook of the Economics of Networks.

Halberstam, Y. and B. Knight (2016). Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence from Twitter. Journal of Public Economics 143, 73–88.

He, K. (2018). Mislearning from Censored Data: The Gambler's Fallacy in Optimal-Stopping Problems. arXiv preprint arXiv:1803.08170.

He, K. and J. Libgober (2020). Evolutionarily Stable (Mis) specifications: Theory and Applications. arXiv preprint arXiv:2012.15007.

Heidhues, P., B. Koszegi, and P. Strack (2018). Convergence in Misspecified Learning Models with Endogenous Actions. Available at SSRN 3312968.

Helmuth, B., T. Gouhier, S. Scyphers, and J. Mocarski (2016). Trust, tribalism and tweets: has political polarization made science a "wedge issue"? Annual Review of Political Science 3, 1–14.

Hoeffding, W. (1963). Probability Inequalities for Sums of Bounded Random Variables. Journal of the American Statistical Association 58(301), 13–30.

Hoffmann, F., K. Khalmetski, and M. Le Quement (2019). Disliking to Disagree. *Working paper.*

Hu, L., A. Li, and I. Segal (2019). The Politics of News Personalization. arXiv preprint arXiv:1910.11405.

Iyengar, S. and M. Krupenkin (2018). The Strengthening of Partisan Affect. Advances in Political Psychology 39, 201–218.

Iyengar, S., Y. Lelkes, M. Levendusky, N. Malhotra, and S. J. Westwood (2019). The Origins and Consequences of Affective Polarization in the United States. Annual Review of Political Science 22(1), 129–146.

Jehiel, P. (2018). Investment Strategy and Selection Bias: An Equilibrium Perspective on Overoptimism. American Economic Review 108(6), 1582–97.

Jin, G. Z., M. Luca, and D. Martin (2021). Is No News (Perceived As) Bad News? An Experimental Investigation of Information Disclosure. American Economic Journal: Microeconomics 13(2), 141–73.

Juslin, P., A. Winman, and P. Hansson (2007, 08). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. Psychological Review 114, 678–703.

Kahneman, D. (2011). Thinking, fast and slow. Macmillan.

Keefer, P. and S. Knack (2002). Polarization, Politics and Property Rights: Links between Inequality and Growth. Public choice 111(1-2), 127–154.

Kruger, J. (1999). Lake wobegon be gone! the" below-average effect" and the egocentric nature of comparative ability judgments. Journal of personality and social psychology 77(2), 221.

Layman, G. C., T. M. Carsey, and J. M. Horowitz (2006). Party polarization in american politics: Characteristics, causes, and consequences. Annu. Rev. Polit. Sci. 9, 83–110.

Levendusky, M. S. (2013). Why Do Partisan Media Polarize Viewers? American Journal of Political Science 57(3), 611–623.

Levy, G. and R. Razin (2018). Information diffusion in networks with the bayesian peer influence heuristic. Games and Economic Behavior 109, 262–270.

Levy, G. and R. Razin (2019). Echo chambers and their effects on economic and political outcomes. Annual Review of Economics 11(1), 303–328.

Levy, R. (2021, March). Social media, news consumption, and polarization: Evidence from a field experiment. American Economic Review 111(3), 831–70.

Li, Y. and H. Pei (2020). Misspecified Beliefs about Time Lags. *Working paper.*

Mailath, G. and L. Samuelson (2020). Learning under Diverse World Views: Model-Based Inference. American Economic Review 110(5), 1464–1501.

Mann, M. E. (2021). The New Climate War: the fight to take back our planet. Hachette UK.

Mason, L. (2015). "I Disrespectfully Agree": The Differential Effects of Partisan Sorting on Social and Issue Polarization. American Journal of Political Science 59(1), 128–145.

Masson-Delmotte, V., P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. e. Zhou (2021a). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, 2021, Cambridge University Press. In Press.

Masson-Delmotte, V., P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. e. Zhou (2021b). IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Technical report, Cambridge University Press. In Press.

Mayer, F. W. (2012). Stories of climate change: Competing narratives, the media, and us public opinion 2001–2010. Joan Shorenstein Center on the Press, Politics and Public Policy, Discussion Paper Series, D-72. http://shorensteincenter.org/wpcontent/uploads/2012/03/d72_mayer. pdf 21, 2015.

McCarty, N. (2019). Polarization: What everyone needs to know. Oxford University Press.

McCarty, N., K. T. Poole, and H. Rosenthal (2009). Does Gerrymandering Cause Polarization? American Journal of Political Science 53(3), 666–680.

Molavi, P., A. Tahbaz-Salehi, and A. Jadbabaie (2018). A Theory of Non-Bayesian Social Learning. Econometrica 86(2), 445–490.

Moser, S. C. (2014). Communicating adaptation to climate change: the art and science of public engagement when climate change comes home. Wiley Interdisciplinary Reviews: Climate Change 5(3), 337–358.

Mosquera, R., M. Odunowo, T. McNamara, X. Guo, and R. Petrie (2019). The Economic Effects of Facebook. Available at SSRN: https://ssrn.com/abstract=3312462 or http://dx.doi.org/10.2139/ssrn.3312462.

Mueller-Frank, M. (2013). A general framework for rational learning in social networks. Theoretical Economics 8(1), 1–40.

Mueller-Frank, M. and C. Neri (2021). A general analysis of boundedly rational learning in social networks. Theoretical Economics 16(1), 317–357.

Mullainathan, S. and A. Shleifer (2005). The Market for News. American Economic Review 95(4), 1031–1053.

Mullinix, K. (2016, 06). Partisanship and Preference Formation: Competing Motivations, Elite Polarization, and Issue Importance. Political Behavior 38, 383–411.

Nickerson, R. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. Review of General Psychology 2(2), 175–220.

Nisbett, R. E. and L. Ross (1980). Human Inference: Strategies and Shortcomings of Social Judgment.

Nivola, P. S. and D. W. Brady (2007). Red and Blue Nation?: Characteristics and Causes of America's Polarized Politics, Volume 1. Brookings Institution Press.

Nivola, P. S. and D. W. Brady (2008). Red and Blue Nation?: Consequences and Correction of America's Polarized Politics, Volume 2. Brookings Institution Press.

Nyarko, Y. (1991). Learning in Mis-Specified Models and the Possibility of Cycles. Journal of Economic Theory 55(2), 416–427.

Odean, T. (1998). Volume, Volatility, Price and Profit When All traders Are Above Average. Journal of Finance 53(6), 1887–1934.

Ookalkar, R., K. V. Reddy, and E. Gilbert (2019). Pop: Bursting News Filter Bubbles on Twitter Through Diverse Exposure. In Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing, CSCW '19, New York, NY, USA, pp. 18–22. Association for Computing Machinery.

Ortiz-Ospina, E. (2019). The rise of social media. Published on "Our World in Data" (accessed March 2, 2022).

Perego, J. and S. Yuksel (2016). Searching for Information and the Diffusion of Knowledge. Unpublished manuscript, New York University.

Perego, J. and S. Yuksel (2022). Media competition and social disagreement. Econometrica 90(1), 223–265.

Periser, E. (2011). The Filter Bubble: What the Internet is Hiding from You. Penguin, London.

Pew Research Center (2014, January). Political Polarization and Media Habits.

Pew Research Center (2020a, February). As Economic Concerns Recede, Environmental Protection Rises on the Public's Policy Agenda.

Pew Research Center (2020b, October). U.S. Media Polarization and the 2020 Election: A Nation Divided.

Pogorelskiy, K. and M. Shum (2019). News We Like to Share: How News Sharing on Social Networks Influences Voting Outcomes. Available at SSRN: https://ssrn.com/abstract=2972231 or http://dx.doi.org/10.2139/ssrn.2972231.

Quattrociocchi, W., A. Scala, and C. R. Sunstein (2016). Echo Chambers on Facebook. Available at SSRN: https://ssrn.com/abstract=2795110 or http://dx.doi.org/10.2139/ssrn.2795110.

Rabin, M. (1998). Psychology and economics. Journal of economic literature 36(1), 11–46.

Reeves, A., M. McKee, and D. Stuckler (2016). 'It's The Sun Wot Won It': Evidence of Media Influence on Political Attitudes and Voting from a UK Quasi-Natural Experiment. Social science research 56, 44–57.

Rogowski, J. C. and J. L. Sutherland (2016). How Ideology Fuels Affective Polarization. Political Behavior 38, 485–508.

Saad, L. (2021). Global Warming Attitudes Frozen Since 2016. Technical report, Gallup.

Samur, A. (2018). The History of Social Media: 29+ Moments. *Hootsuite* (accessed March 2, 2022).

Schmidt, A. L., F. Zollo, A. Scala, C. Betsch, and W. Quattrociocchi (2018). Polarization of the Vaccination Debate on Facebook. Vaccine 36(25), 3606–3612.

Schwartzstein, J. (2014). Selective attention and learning. Journal of the European Economic Association 12(6), 1423–1452.

Shin, J., L. Jian, K. Driscoll, and F. Bar (2018). The Diffusion of Misinformation on Social Media: Temporal Pattern, Message, and Source. Computers in Human Behavior 83, 278–287.

Shin, J. and K. Thorson (2017). Partisan Selective Sharing: The Biased Diffusion of Fact-Checking Messages on Social Media. Journal of Communication 67(2), 233–255.

Spiegler, R. (2020). Behavioral implications of causal misperceptions. Annual Review of Economics 12(1), 81–106.

Sunstein, C. (2017). Divided Democracy in the Age of Social Media. Princeton University Press.

Svenson, O. (1981). Are We all Less Risky and More Skillful Than Our Fellow Drivers? Acta Psychologica 47(2), 143–148.

Tucker, J., A. Guess, P. Barbera, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, and B. Nyhan (2019). Social Media, Political Polarization, and Political Disinformation: A Review of Scientific Literature. Available at SSRN: https://ssrn.com/abstract=3144139 or http://dx.doi.org/10.2139/ssrn.3144139.

Weeks, B. E., D. S. Lane, D. H. Kim, S. S. Lee, and N. Kwak (2017). Incidental Exposure, Selective Exposure, and Political Information Sharing: Integrating Online Exposure Patterns and Expression on Social Media. J. Computer-Mediated Communication 22, 363–379.

Zak, P. J. and S. Knack (2001). Trust and Growth. The economic journal 111(470), 295–321.

Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political Effects of the Internet and Social Media. Annual Review of Economics 12, 415–438.

Zuckerman, E. and J. Jost (2001). What Makes you Think you are so Popular? Social Psychology Quarterly 64(3), 207–223.

Online Appendix: Supplementary Material

(For Online Publication Only)

C Other Misspecifications

In this appendix, Section C.1 states and proves Propositions 1, 2, and 3 for the case $\hat{\gamma} > \gamma$, i.e., when agents overestimate selective sharing done by their friends. Sections C.2, C.3, and C.4 state and prove the formal results about long-run learning under each of misspecifications considered in Section 7. Together, Sections C.2, C.3, and C.4 imply Proposition 10.

C.1 Propositions 1, 2, and 3 for $\hat{\gamma} > \gamma$

Proposition (Proposition 1)**.** *Fix any agent with a balanced echo chamber and suppose $\hat{\gamma} > \gamma$. Then,*

$$\left(\mathbb{E}[\mu(\mathbf{s}^1)] - \pi\right)\left(\pi - \frac{1}{2}\right) < 0.$$

Proof sketch. The proof of Proposition 1 can be adopted here for the most part. The difference is towards the end, where $\hat{\gamma} > \gamma$ implies $\Gamma > \hat{\Gamma}$, which in turn is equivalent to $\Gamma^{y-x} > \hat{\Gamma}^{y-x}$. Hence, dividing the key inequality that determines the sign of Δ_{xy} by $\Gamma^x\hat{\Gamma}^y - \Gamma^y\hat{\Gamma}^x$ makes the subsequent inequality signs in the proof flip. As a result, we get that the derivative is positive for $\rho > 1$ (rather than negative). Intuitively, this implies that moving the prior from 50-50 towards a state will make the unconditional expected posterior of that state *lower* than the prior. ■

Proposition (Proposition 2)**.** *Fix any agent with an unbalanced echo chamber $e = (d_A, d_B, n)$ and suppose $\hat{\gamma} > \gamma$. Then there exists $q_{SR}(e, \gamma, \hat{\gamma}) > \frac{1}{2}$ such that, if $q < q_{SR}(e, \gamma, \hat{\gamma})$, then*

$$\left(\mathbb{E}[\mu(\mathbf{s}^1)] - \pi\right)(d_A - d_B) < 0.$$

Proof sketch. The proof of Proposition 2 can be adopted here in its entirety, since it does not use $\hat{\gamma} < \gamma$ until the last few lines. Recall the derivative of $\mathbb{E}[\mu|\omega]$ with respect to q at $q = \frac{1}{2}$ from equation (8):

$$\frac{\partial}{\partial q}\mathbb{E}[\mu|\omega]\bigg|_{q=\frac{1}{2}} = \frac{4\pi(1-\pi)}{2-\hat{\gamma}}(d_A - d_B)(\gamma - \hat{\gamma}).$$

For $d_A > d_B$ and $\hat{\gamma} > \gamma$, the derivative is negative at $q = \frac{1}{2}$, which means that $\mathbb{E}[\mu|\omega]$ is distorted towards B for low q. ■

Proposition (Proposition 3)**.** *Fix any agent with an unbalanced echo chamber $e = (d_A, d_B, n)$, and suppose $\hat{\gamma} > \gamma$. There exists $q_{LR}(e, \gamma, \hat{\gamma}) \in \left(\frac{1}{2}, 1\right)$ such that the following holds:*

1. *If $q < q_{LR}(e, \gamma, \hat{\gamma})$, then the agent's belief converges with probability 1 to δ_B if $d_A > d_B$, and to δ_A if $d_B > d_A$ (i.e., $\mu(\mathbf{s}^\infty) = \mathbb{1}_{d_B > d_A}$).*

2. *If $q > q_{LR}(e, \gamma, \hat{\gamma})$, then the agent's belief converges with probability 1 to δ_ω, where ω is the true state (i.e., $\mu(\mathbf{s}^\infty) = \mathbb{1}_{\omega=A}$.*

Proof sketch. The proof of Proposition 3 can be adopted here for the most part, up until arriving at the definition of the function $\tau(q)$:

$$\tau(q) = \frac{1}{2} + \frac{((2-\gamma)z(q,\hat{\gamma}) - \gamma)(d_A - d_B)}{2\gamma(2(1+n) + (d_A + d_B)(1 + z(q,\hat{\gamma})))}.$$

Recall that the agent's belief converges to δ_A if and only if $H(q;\omega) > \tau(q)$, where $H(q;\omega)$ equals q if $\omega = A$, and equals $1-q$ if $\omega = B$. Also recall that $\tau(q)$ is decreasing and concave for $q \in (\frac{1}{2}, 1)$, and $\tau'(\frac{1}{2}) = 0$.

Consider the case $\omega = B$, and hence $H(q;\omega) = 1-q$. Given $\hat{\gamma} > \gamma$, $H(q;\omega) < \tau(q)$ holds at $q = \frac{1}{2}$ and hence at all $q > \frac{1}{2}$ by the properties of $\tau(q)$. If $\omega = A$, on the other hand, we will have $H(q;\omega) < \tau(q)$ holding at $q = \frac{1}{2}$ and $H(q;\omega) > \tau(q)$ at $q = 1$. By the properties of $\tau(q)$, there exists a unique $q_{LR} \in (\frac{1}{2}, 1)$ such that $\operatorname{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 0$ if $q < q_{LR}$ and $\operatorname{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 1$ if $q > q_{LR}$. ■

C.2 Misspecification (I): Random Selective Sharing

Proposition 12. *Fix any agent with echo chamber $e = (d_A, d_B, n)$, true probabilities of selective sharing g and f, and perceived probabilities of selective sharing $\hat{g}$ and $\hat{f}$.*

- *If $d_A > d_B$ and $g - f > \hat{g} - \hat{f}$, there exists sufficiently small $q > \frac{1}{2}$ such that the agent's belief converges to δ_A with probability 1 (i.e., $\mu(\mathbf{s}^\infty) = 1$).*
- *If $d_A > d_B$ and $g - f < \hat{g} - \hat{f}$, there exists sufficiently small $q > \frac{1}{2}$ such that the agent's belief converges to δ_B with probability 1 (i.e., $\mu(\mathbf{s}^\infty) = 0$).*
- *In either case, there exists sufficiently large $q < 1$ such that the agent's belief converges to δ_ω with probability 1, where ω is the true state (i.e., $\mu(\mathbf{s}^\infty) = I_{\{\omega=A\}}$).*
- *If $d_A = d_B$, the agent's belief converges to δ_ω with probability 1, where ω is the true state (i.e., $\mu(\mathbf{s}^\infty) = I_{\{\omega=A\}}$).*

Proof. Adapt the terminology of Proposition 3's proof as follows. Let a_j^k be the number of signals $s = a$ that have been shared by agent i's friend j of type $k \in \{A, B, N\}$. Define b_j^k similarly for $s = b$.

Then, agent i's posterior that $\omega = A$ is

$$\mu(\mathbf{s}^T) = \frac{\pi}{\pi + (1-\pi) \cdot Q^M \cdot \left(\frac{(1-\gamma)+\gamma(q(1-\hat{g})+(1-q)(1-\hat{f}))}{(1-\gamma)+\gamma((1-q)(1-\hat{g})+q(1-\hat{f}))} \right)^S},$$

where

$$M = (a_i - b_i) + \sum_{j=1}^{n} (a_j^N - b_j^N) + \sum_{j=1}^{d_A} (a_j^A - b_j^A) - \sum_{j=1}^{d_B} (b_j^B - a_j^B),$$
$$S = \sum_{j=1}^{d_B} (T - a_j^B - b_j^B) - \sum_{j=1}^{d_A} (T - a_j^A - b_j^A).$$

For $\hat{\mathbf{p}} = (\hat{g}, \hat{f})$, define the function

$$z(q, \gamma, \hat{\mathbf{p}}) = \ln \left(\frac{(1-\gamma) + \gamma(q(1-\hat{g}) + (1-q)(1-\hat{f}))}{(1-\gamma) + \gamma((1-q)(1-\hat{g}) + q(1-\hat{f}))} \right) [\ln Q]^{-1}.$$

Similarly to Proposition 3, we have $\text{plim}_{T\to\infty}\,\mu(\mathbf{s}^T) = 1$ (resp. $\text{plim}_{T\to\infty}\,\mu(\mathbf{s}^T) = 0$) if and only if $\text{plim}_{T\to\infty}\frac{K(\mathbf{x},T;q,\gamma,\mathbf{p},\hat{\mathbf{p}})}{T} > 0$ (resp. < 0), where

$$\begin{aligned}
\underset{T\to\infty}{\text{plim}}\,\frac{K(\mathbf{x},T;q,\gamma,\mathbf{p},\hat{\mathbf{p}})}{T} &= \gamma(1+\nu n)(2H(q;\omega)-1) \\
&\quad + \gamma d_A(gH(q;\omega) - f(1-H(q;\omega))) - \gamma d_B(g(1-H(q;\omega)) - fH(q;\omega)) \\
&\quad + d_B(1-\gamma fH(q;\omega) - \gamma g(1-H(q;\omega))z(q,\gamma,\hat{\mathbf{p}}) \\
&\quad - d_A(1-\gamma gH(q;\omega) - \gamma f(1-H(q;\omega))z(q,\gamma,\hat{\mathbf{p}}) \\
&= -\gamma(1+\nu n) - \gamma f d_A - \gamma g d_B + ((1-\gamma g)d_B - (1-\gamma f)d_A)\, z(q,\gamma,\hat{\mathbf{p}}) \\
&\quad + \gamma\Big(2(1+\nu n) + g(d_A+d_B)(1+z(q,\gamma,\hat{\mathbf{p}})) \\
&\quad + f(d_A+d_B)(1-z(q,\gamma,\hat{\mathbf{p}}))\Big)H(q;\omega).
\end{aligned}$$

The required inequality is then

$$H(q;\omega) > (\text{resp.} <) \frac{\gamma(1+\nu n) + \gamma f\,(1-z(q,\gamma,\hat{\mathbf{p}}))\,d_A + \gamma g\,(1+z(q,\gamma,\hat{\mathbf{p}}))\,d_B + (d_A-d_B)z(q,\gamma,\hat{\mathbf{p}})}{\gamma\,(2(1+\nu n) + g(d_A+d_B)(1+z(q,\gamma,\hat{\mathbf{p}})) + f(d_A+d_B)(1-z(q,\gamma,\hat{\mathbf{p}})))},$$

which is equivalent to

$$H(q;\omega) > (\text{resp.} <) \frac{1}{2} + \frac{\left(\left(1-\frac{\gamma}{2}(f+g)\right)z(q,\gamma,\hat{\mathbf{p}}) - \frac{\gamma}{2}(g-f)\right)(d_A-d_B)}{\gamma\,(2(1+\nu n) + g(d_A+d_B)(1+z(q,\gamma,\hat{\mathbf{p}})) + f(d_A+d_B)(1-z(q,\gamma,\hat{\mathbf{p}})))}.$$

Fix state $\omega = B$ so that $H(q;B) = 1-q$. The inequality above takes form

$$q < (\text{resp.} >) \frac{1}{2} + \frac{\left(\frac{\gamma}{2}(g-f) - \left(1-\frac{\gamma}{2}(f+g)\right)z(q,\gamma,\hat{\mathbf{p}})\right)(d_A-d_B)}{\gamma\,(2(1+\nu n) + g(d_A+d_B)(1+z(q,\gamma,\hat{\mathbf{p}})) + f(d_A+d_B)(1-z(q,\gamma,\hat{\mathbf{p}})))}. \quad (12)$$

This implies that if $d_A = d_B$, then $\text{plim}_{T\to\infty}\,\mu(\mathbf{s}^T) = 0$.

It can be shown that

$$\lim_{q\to 1} z(q,\gamma,\hat{\mathbf{p}}) = 0 \qquad \text{and} \qquad \lim_{q\to\frac{1}{2}} z(q,\gamma,\hat{\mathbf{p}}) = \frac{\frac{\gamma}{2}(\hat{g}-\hat{f})}{(1-\gamma)+\frac{\gamma}{2}(2-\hat{g}-\hat{f})},$$

which increases in $\hat{g}$ and decreases in $\hat{f}$. Using this limit, condition (12) at $q = \frac{1}{2}$ becomes

$$\frac{1}{2} < (\text{resp.} >) \frac{1}{2} + \frac{\frac{1}{2}\Big((g-f) - (\hat{g} - \hat{f})\Big)(d_A - d_B)}{2(1+\nu n)(1 - \frac{\gamma}{2}(\hat{g} + \hat{f})) + g(d_A + d_B)(1 - \gamma(\hat{g} + \hat{f}))},$$

and at $q = 1$ it becomes

$$1 < (\text{resp.} >) \frac{(g-f)(d_A - d_B)}{(2(1+\nu n) + (g+f)(d_A + d_B))}.$$

Given $d_A > d_B$, the first condition holds with "<" whenever $(g - f) > (\hat{g} - \hat{f})$; the second holds with ">". By continuity, there exists q' and q'' that satisfy $\frac{1}{2} < q' \leq q'' < 1$, $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 1$ if $q < q'$, and $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 0$ if $q > q''$.

Now suppose $\omega = A$ so that $H(q; A) = q$. The key inequality takes form

$$q > (\text{resp.} <) \frac{1}{2} + \frac{\left(\left(1 - \frac{\gamma}{2}(f+g)\right) z(q,\gamma,\hat{\mathbf{p}}) - \frac{\gamma}{2}(g-f)\right)(d_A - d_B)}{\gamma\left(2(1+\nu n) + g(d_A + d_B)(1 + z(q,\gamma,\hat{\mathbf{p}})) + f(d_A + d_B)(1 - z(q,\gamma,\hat{\mathbf{p}}))\right)}. \quad (13)$$

At $q = \frac{1}{2}$, it takes form

$$\frac{1}{2} > (\text{resp.} <) \frac{1}{2} - \frac{\frac{1}{2}\left((g-f) - (\hat{g} - \hat{f})\right)(d_A - d_B)}{2(1+\nu n)(1 - \frac{\gamma}{2}(\hat{g} + \hat{f})) + g(d_A + d_B)(1 - \gamma(\hat{g} + \hat{f}))}$$

and at $q = 1$, it takes the form

$$1 > (\text{resp.} <) - \frac{(g-f)(d_A - d_B)}{2(1+\nu n) + (f+g)(d_A + d_B)}.$$

Given $(g - f) < (\hat{g} - \hat{f})$, the first condition holds with "<"; the second condition holds with ">". By continuity, there exists q' and q'' that satisfy $\frac{1}{2} < q' \leq q'' < 1$, $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 0$ if $q < q'$, and $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 1$ if $q > q''$.

■

C.3 Misspecification (II): Friends' Types

Proposition 13. *Fix any agent with echo chamber $e = (d_A, d_B, n)$ and misspecified number of dogmatic friends $\hat{d}_A \leq d_A$ and $\hat{d}_B \leq d_B$.*

- *If $d_A - d_B > \hat{d}_A - \hat{d}_B$, there exists sufficiently small $q > \frac{1}{2}$ such that the agent's belief converges to δ_A with probability 1 (i.e., $\mu(\mathbf{s}^\infty) = 1$).*
- *If $d_A - d_B < \hat{d}_A - \hat{d}_B$, there exists sufficiently small $q > \frac{1}{2}$ such that the agent's belief converges to δ_B with probability 1 (i.e., $\mu(\mathbf{s}^\infty) = 0$).*
- *In either case, there exists sufficiently large $q < 1$ such that the agent's belief converges to δ_ω with probability 1, where ω is the true state (i.e., $\mu(\mathbf{s}^\infty) = I_{\{\omega=A\}}$).*
- *If $d_A - d_B = \hat{d}_A - \hat{d}_B = 0$, the agent's belief converges to δ_ω with probability 1, where ω is the true state (i.e., $\mu(\mathbf{s}^\infty) = I_{\{\omega=A\}}$).*

Proof. Let the misspecified number of A-dogmatic and B-dogmatic friends be $\hat{d}_A = d_A - \hat{n}_A$ and $\hat{d}_B = d_B - \hat{n}_B$. Then agent's i posterior belief is

$$\mu(\mathbf{s}^T) = \frac{\pi}{\pi + (1-\pi) \cdot Q^M \cdot \left(\frac{(1-\gamma)+\gamma(1-q)}{(1-\gamma)+\gamma q}\right)^S},$$

where

$$M = (a_i - b_i) + \sum_{j=1}^{n} (a_j^N - b_j^N) + \sum_{j=1}^{d_A} a_j^A - \sum_{j=1}^{d_B} b_j^B,$$
$$S = \sum_{j=1}^{\hat{d}_B} (T - b_j^B) - \sum_{j=1}^{\hat{d}_A} (T - a_j^A).$$

Define the function

$$z(q, \gamma) = \ln\left(\frac{(1-\gamma)+\gamma(1-q)}{(1-\gamma)+\gamma q}\right) [\ln Q]^{-1}.$$

Similar to Proposition 3, we have $\text{plim}_{T\to\infty} \mu(s^T) = 1$ (resp. $\text{plim}_{T\to\infty} \mu(s^T) = 0$) if and only if $\text{plim}_{T\to\infty} \frac{K(\mathbf{x},T;q,\gamma)}{T} > 0$ (resp. < 0), where

$$\begin{aligned}
\text{plim}_{T\to\infty} \frac{K(\mathbf{x},T;q,\gamma)}{T} &= \gamma(1+n)(2H(q,\omega)-1) + \gamma d_A H(q,\omega) - \gamma d_B(1-H(q,\omega)) + \\
&\quad + \left(\hat{d}_B(1-\gamma(1-H(q,\omega))) - \hat{d}_A(1-\gamma H(q,\omega))\right) z(q,\gamma) \\
&= -\gamma(1+n) - \gamma d_B + \hat{d}_B(1-\gamma)z(q,\gamma) - \hat{d}_A z(q,\gamma) + \\
&\quad + \gamma\left(2(1+n) + (d_A+d_B) + (\hat{d}_A+\hat{d}_B)z(q,\gamma)\right) H(q,\omega).
\end{aligned}$$

The required inequality is then

$$H(q;\omega) > (\text{resp. } <) \frac{\gamma(1+n) + \gamma d_B + \gamma \hat{d}_B z(q,\gamma) + (\hat{d}_A - \hat{d}_B)z(q,\gamma)}{\gamma\left(2(1+n) + (d_A+d_B) + (\hat{d}_A+\hat{d}_B)z(q,\gamma)\right)},$$

which is equivalent to

$$H(q;\omega) > (\text{resp. } <) \frac{1}{2} + \frac{-\frac{\gamma}{2}(d_A-d_B) + \left(1-\frac{\gamma}{2}\right)(\hat{d}_A - \hat{d}_B)z(q,\gamma)}{\gamma\left(2(1+n) + (d_A+d_B) + (\hat{d}_A+\hat{d}_B)z(q,\gamma)\right)}.$$

Note that if $\hat{d}_A - \hat{d}_B = d_A - d_B = 0$, this inequality holds with "$>$" when $\omega = A$ and "$<$" when $\omega = B$, implying $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = I_{\{\omega=A\}}$.

Fix state $\omega = B$ so that $H(q;B) = 1-q$. Then the inequality above takes form

$$q < (\text{resp. } >) \frac{1}{2} + \frac{\gamma(d_A-d_B) - (2-\gamma)(\hat{d}_A-\hat{d}_B)z(q,\gamma)}{2\gamma(2(1+n) + (d_A+d_B) + (\hat{d}_A+\hat{d}_B)z(q,\gamma))}. \tag{14}$$

It can be shown that

$$\lim_{q\to 1} z(q,\gamma) = 0 \qquad \text{and} \qquad \lim_{q\to\frac{1}{2}} z(q,\gamma) = \frac{\gamma}{2-\gamma}.$$

Using this limit, condition (14) at $q = \frac{1}{2}$ becomes

$$\frac{1}{2} < (\text{resp. } >)\frac{1}{2} + \frac{\gamma\Big((d_A - d_B) - (\hat{d}_A - \hat{d}_B)\Big)}{2\gamma\left(2(1+n) + (d_A + d_B) + (\hat{d}_A + \hat{d}_B)\frac{\gamma}{2-\gamma}\right)},$$

and at $q = 1$ it becomes

$$1 < (\text{resp. } >)\frac{1}{2} + \frac{\gamma(d_A - d_B)}{2\gamma(2(1+n) + (d_A + d_B))}.$$

The first inequality holds with "<" if and only if $\hat{d}_A - \hat{d}_B < d_A - d_B$; the second holds with ">". By continuity, there exists q' and q'' that satisfy $\frac{1}{2} < q' \le q'' < 1$, $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 1$ if $q < q'$ and $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 0$ if $q > q''$.

Now suppose $\omega = A$ so that $H(q; A) = q$. The key inequality takes form

$$q > (\text{resp. } <)\frac{1}{2} + \frac{-\frac{\gamma}{2}(d_A - d_B) + \left(1 - \frac{\gamma}{2}\right)(\hat{d}_A - \hat{d}_B)z(q,\gamma)}{\gamma\left(2(1+n) + (d_A + d_B) + (\hat{d}_A + \hat{d}_B)z(q,\gamma)\right)}. \tag{15}$$

At $q = \frac{1}{2}$, it takes form

$$\frac{1}{2} > (\text{resp. } <)\frac{1}{2} - \frac{\gamma\Big((d_A - d_B) - (\hat{d}_A - \hat{d}_B)\Big)}{2\gamma\left(2(1+n) + (d_A + d_B) + (\hat{d}_A + \hat{d}_B)\frac{\gamma}{2-\gamma}\right)},$$

and at $q = 1$, it takes form

$$1 > (\text{resp. } <)\frac{1}{2} - \frac{\gamma(d_A - d_B)}{2\gamma(2(1+n) + (d_A + d_B))}.$$

The first inequality holds with "<" whenever $\hat{d}_A - \hat{d}_B > d_A - d_B$; the second holds with ">". By continuity, there exists q' and q'' that satisfy $\frac{1}{2} < q' \le q'' < 1$, $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 0$ if $q < q'$ and $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 1$ if $q > q''$.

■

C.4 Misspecification (III): Information Quality

Proposition 14. *Fix any agent with echo chamber* $e = (d_A, d_B, n)$ *and any misspecified information quality* $\hat{q} > \frac{1}{2}$.

- *If* $d_A > d_B$ *and* $\gamma < 1$, *there exists sufficiently small* $q \in \left(\frac{1}{2}, \hat{q}\right)$ *such that the agent's belief converges to* δ_A *with probability 1 (i.e.,* $\mu(\mathbf{s}^\infty) = 1$*) and sufficiently large* $q < 1$ *such that the agent's belief converges to* δ_ω *with probability 1, where* ω *is the true state (i.e.,* $\mu(\mathbf{s}^\infty) = I_{\{\omega=A\}}$*).*
- *If either* $d_A = d_B$ *or* $\gamma = 1$, *the agent's belief converges to* δ_ω *with probability 1, where* ω *is the true state (i.e.,* $\mu(\mathbf{s}^\infty) = I_{\{\omega=A\}}$*).*

Proof. Fix echo chamber $e = (d_A, d_B, n)$. Keep notations the same as in the proof of Proposition 3. After T periods, the agent's posterior in state A is

$$\mu(\mathbf{s}^T) = \frac{\pi}{\pi + (1-\pi) \cdot \left(\frac{1-\hat{q}}{\hat{q}}\right)^M \cdot \left(\frac{\gamma(1-\hat{q})+(1-\gamma)}{\gamma\hat{q}+(1-\gamma)}\right)^S},$$

where

$$M = (a_i - b_i) + \sum_{j=1}^{n}(a_j^N - b_j^N) + \sum_{j=1}^{d_A} a_j^A - \sum_{j=1}^{d_B} b_j^B,$$
$$S = \sum_{j=1}^{d_B}(T - b_j^B) - \sum_{j=1}^{d_A}(T - a_j^A).$$

Define the function

$$z(\hat{q}, \gamma) = \ln\left(\frac{(1-\gamma)+\gamma(1-\hat{q})}{(1-\gamma)+\gamma\hat{q}}\right)\left[\ln\left(\frac{1-\hat{q}}{\hat{q}}\right)\right]^{-1}.$$

Similar to Proposition 3, we have $\mathbf{plim}_{T\to\infty}\,\mu(\mathbf{s}^T) = 1$ (resp. $\mathbf{plim}_{T\to\infty}\,\mu(\mathbf{s}^T) = 0$) if and only if

$\operatorname{plim}_{T\to\infty} \frac{K(\mathbf{x},T;q,\hat{q},\gamma)}{T} > 0$ (resp. < 0), where

$$\operatorname*{plim}_{T\to\infty} \frac{K(\mathbf{x},T;q,\hat{q},\gamma)}{T} = -\gamma(1+n+(1+z(\hat{q},\gamma))d_B) - (d_A - d_B)z(\hat{q},\gamma) + \\ + \gamma(2(1+n)+(d_A+d_B)(1+z(\hat{q},\gamma)))H(q;\omega).$$

The required inequality is then

$$H(q;\omega) > (\text{resp. } <) \frac{\gamma(1+n)+\gamma(1+z(\hat{q},\gamma))d_B + (d_A - d - B)z(\hat{q},\gamma)}{\gamma\,(2(1+n)+(d_A+d_B)(1+z(\hat{q},\gamma)))},$$

which is equivalent to

$$H(q;\omega) > (\text{resp. } <) \frac{1}{2} + \frac{((2-\gamma)z(\hat{q},\gamma)-\gamma)(d_A-d_B)}{\gamma\,(2(1+n)+(d_A+d_B)(1+z(\hat{q},\gamma)))}.$$

Note that if $d_A = d_B$ or $\gamma = 1$ (which implies $z(\hat{q},\gamma) = 1$), then the inequality holds with "$>$" when $\omega = A$ and "$<$" when $\omega = B$, implying $\operatorname{plim}_{T\to\infty} \mu(\mathbf{s}^T) = I_{\{\omega=A\}}$.

Fix state $\omega = B$ so that $H(q;B) = 1-q$. Then the inequality above takes form

$$q < (\text{resp. } >) \frac{1}{2} + \frac{((2-\gamma)z(\hat{q},\gamma)-\gamma)(d_A-d_B)}{\gamma\,(2(1+n)+(d_A+d_B)(1+z(\hat{q},\gamma)))}.$$

At $q = \frac{1}{2}$, it takes form

$$\frac{1}{2} < (\text{resp. } >) \frac{1}{2} + \frac{(\gamma-(2-\gamma)z(\hat{q},\gamma))(d_A-d_B)}{\gamma\,(2(1+n)+(d_A+d_B)(1+z(\hat{q},\gamma)))},$$

and at $q = 1$, it takes form

$$1 < (\text{resp. } >) \frac{1}{2} + \frac{(\gamma-(2-\gamma)z(\hat{q},\gamma))(d_A-d_B)}{\gamma\,(2(1+n)+(d_A+d_B)(1+z(\hat{q},\gamma)))}.$$

As shown in the Online Appendix E, $z(\hat{q},\gamma)$ is a (weakly) decreasing function that achieves maximum at $\hat{q} = \frac{1}{2}$, with value of $\frac{\gamma}{2-\gamma}$. Thus, $\gamma - (2-\gamma)z(\hat{q},\gamma) > 0$ for any $\hat{q} > \frac{1}{2}$. Given this and $d_A > d_B$, the first inequality above holds with "$<$"; the second holds with "$>$". By continuity, there exist q' and q'' that satisfy $\frac{1}{2} < q' \leq q'' < 1$, $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 1$ if $q < q'$ and $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 0$ if $q > q''$.

Now suppose $\omega = A$ so that $H(q; A) = q$. The key inequality then is

$$q > \text{(resp. <)} \frac{1}{2} + \frac{((2-\gamma)z(\hat{q},\gamma) - \gamma)(d_A - d_B)}{\gamma\,(2(1+n) + (d_A + d_B)(1 + z(\hat{q},\gamma)))}.$$

At $q = \frac{1}{2}$, this inequality takes form

$$\frac{1}{2} > \text{(resp. <)} \frac{1}{2} + \frac{((2-\gamma)z(\hat{q},\gamma) - \gamma)(d_A - d_B)}{\gamma\,(2(1+n) + (d_A + d_B)(1 + z(\hat{q},\gamma)))},$$

and at $q = 1$, it takes form

$$1 > \text{(resp. <)} \frac{1}{2} + \frac{((2-\gamma)z(\hat{q},\gamma) - \gamma)(d_A - d_B)}{\gamma\,(2(1+n) + (d_A + d_B)(1 + z(\hat{q},\gamma)))}.$$

Given $(2-\gamma)z(\hat{q},\gamma) - \gamma < 0$ and $d_A > d_B$, both inequalities hold with "$>$". Therefore, for any $q > \frac{1}{2}$, $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 1$.

■

D A Model of Endogenous Selective Sharing

We consider a simple model in which information is strategically shared to influence the outcome of a policy-making process. Agents are connected in a fixed network, and learning occurs with misperception, i.e. $\hat{\gamma} < \gamma$. We use the model to explore implications for polarization and gridlock.

Time t is discrete, where $t = 0, \ldots, \infty$. A payoff-relevant state of the world $\omega \in \{A, B\}$ realizes at $t = 0$. A group of normal agents $\mathcal{N}$ must make a public policy decision $x \in \{x_A, x_B\}$ in the long-run. By "long-run" we simply mean that the decision (and hence payoffs) need not be immediate and beliefs have sufficient time to converge before the decision is taken.[26] An example is whether

[26] A one period, single receiver version of the endogenous sharing model presented in this section is closely related to Jin et al. (2021).

to spend on climate mitigation or adaptation: An agent's preference for this policy may depend on whether or not she believes the effects of global warming have already begun (see Section 6).

Payoffs As in the baseline model, normal agent $i \in \mathcal{N}$ has a set of A-dogmatic friends, B-dogmatic friends, and normal friends. For notational convenience, we denote these as D_{Ai}, D_{Bi} and N_i respectively. Normal agents receive standard quadratic loss utility $u_i(x) = -(x - y(\omega))^2$ when policy x is implemented, where $x_A = 1$, $x_B = 0$, $y(A) = 1$, and $y(B) = 0$. Dogmatic agents have state-independent preferences; they receive payoff 0 if their preferred policy is implemented and -1 if their least preferred is implemented.[27] A status quo policy x_0 exists and all agents receive $u_i(x_0) = -3/4$ if the status quo prevails. With reference to our application, the status quo can be thought of as doing nothing about climate change.

Policymaking Policy x is decided in the long run by a decisive coalition of normal agents $\mathcal{C}$. We make standard assumptions on $\mathcal{C}$ (Austen-Smith and Banks, 1999): 1) The set of decisive coalitions is *proper,* i.e., every pair of decisive coalitions has a nonempty intersection—$C_1, C_2 \in \mathcal{C}$ implies $C_1 \cap C_2 \neq \emptyset$; 2) $\mathcal{C}$ is *monotonic*, i.e., any superset of a decisive coalition is itself decisive—$C_1 \in \mathcal{C}$ and $C_1 \subseteq C_2$ imply $C_2 \in \mathcal{C}$. These assumptions admit common decision rules such as simple majority, all super majority rules, unanimity, and decision rules with veto players.

Information As before, in each period $t \geq 1$ agent i receives first-hand information about ω in the form of a private signal $s_{it} \in \{a, b\}$ with probability $\gamma \in (0, 1]$. With probability $1 - \gamma$ she receives no signal, which we denote by $s_{it} = 0$. Signals are partially informative, with information quality q as in the baseline model. The events of receiving a signal and its realization are i.i.d. across agents and time.

Sharing The set of agents in i's *echo chamber* is $E_i = D_{Ai} \cup D_{Bi} \cup N_i$. After receiving signal s_{it}, agent i sends a costless message m_{jit} to all friends $j \in E_i$. Consequently, i also receives message

[27]Preferences for dogmatic agents can also be interpreted as state-dependent with degenerate beliefs.

m_{ijt} from all friends $j \in E_i$.[28] Denote the message space as $M = \{-1, 0, 1\}$. Message -1 means that an a-signal was shared, message 1 means that a b-signal was shared and message 0 means that no signal was shared. Message m_{ijt} is verifiable, so signals cannot be fabricated, only omitted. Message $m_{ijt} = 0$ can be received from friend j either because j received no signal, or j received a signal, but decided not to share. Agent i cannot distinguish between these two events. Let m_{it} be the entire profile of messages that agent i receives in period t, inclusive of her own signal. To focus on selective sharing of dogmatic agents, we assume that normal agents are not strategic in signal sharing and share all signals they receive. Dogmatic agents share strategically. As a result of sharing, all agents receive information about ω from second-hand information shared by other agents, in addition to their own signal received from original sources.

Learning As in our baseline model, each normal agent assigns prior probability π to the state of the world being A. The beliefs of dogmatic agents do not affect decision-making given that their preferences are state-independent so we focus on the beliefs of normal agents. Both dogmatic and normal agents have the misperception that signals arrive at rate $\hat{\gamma} < \gamma$ and use this in calculating their own posteriors and posteriors of other agents.

After one round of sharing, agent i has posterior belief μ_{i1} that depends on messages m_{i1} received in period 1. We drop i and t subscripts in what follows as we are considering a single normal agent after one round of sharing. From equation (2) in the main text, agent i's posterior belief is:

$$\mu(m) \quad = \quad \frac{\pi}{\pi + (1-\pi) Q^M \hat{\Gamma}^S},$$

where Q and $\hat{\Gamma}$ are as before. Adapting notation to strategic messages, we have

$$M \equiv -\sum_{j \in E \cup i} m_j, \quad \text{and} \quad S \equiv \left(d_B - \sum_{j \in D_B} |m_j| \right) - \left(d_A - \sum_{j \in D_A} |m_j| \right).$$

[28]As in the baseline model, there is no choice of who receives signals in an agent's network. This is similar to posts in Facebook or Twitter.

Timing Each period messages are received, sharing occurs and beliefs are updated. We assume normal agents vote once and at the end of a sufficiently long horizon of learning such that only their long-run beliefs matter for sharing strategies. Payoffs occur after voting by normal agents.

Strategies and Equilibrium A sharing (message) strategy for dogmatic agent j is $\sigma_j : \{a, 0, b\} \to \{-1, 0, 1\}$ and a profile of messages of all dogmatic agents in society is σ. For simplicity we focus on symmetric strategies such that all A dogmatic agents use the same sharing strategy. Similarly, all B dogmatic agents use the same strategies. These strategies condition on the set of echo chambers in the network $\mathbf{e}$. Strategies may also condition on beliefs of normal agents, however, we consider long-run outcomes for which limit beliefs are entirely determined by the structure of the echo chamber, quality of information and misperception. Assume that normal agent i votes sincerely in the sense that if $\mu_i(\mathbf{s}^\infty) = 0$, then agent i votes for x_B, and if $\mu_i(\mathbf{s}^\infty) = 1$, then agent i votes for x_A. We consider Symmetric Bayes' Perfect Equilibrium, such that dogmatic agents' sharing strategies each period are best responses, and beliefs obey Bayes' rule when possible, with respect to agents' misspecified model. We refer to this as an equilibrium.

Proposition 15 (Equilibrium)**.** *There is an equilibrium in weakly dominant strategies where A-dogmatic agents share if and only if an a-signal is received, and B-dogmatic agents share if and only if a b-signal is received.*

Proof. Consider an A-dogmatic agent—the argument is similar for B-dogmatic agents. Sincere voting by normal agents implies that an A-dogmatic agent wishes to convince as many normal agents as possible that the state is A. Recall that the A-dogmatic agent prefers the status quo to x_B so there is never an incentive to convince agents to vote for B. Given this, suppose this A-dogmatic agent receives a b-signal. In this case, it is weakly dominant for him to share message $m = 0$, that is, to stay silent. Given any sharing strategy of the other dogmatic agents, sharing $m = 0$ rather than $m = b$ either does not change the long-run beliefs of the receiving normal agents, or changes it from believing that the state is B to believing that the state is A, because $m = 0$ suppresses evidence in favor of $\omega = B$. This leads to weakly more agents voting for A, and weakly less voting for B. This may change nothing in the outcome or transition it from policy B to status quo, or from status quo to policy A, both of which the agent prefers. Suppose instead that the A-dogmatic agent receives an a-signal. In this case, it is weakly dominant to share message $m = a$ rather than $m = 0$. Given all others' sharing strategies, sharing $m = a$ (weakly) increases the chances that in the long run the receiving agents believe that the state is A rather than B. Similarly, this will lead to the outcome changes that the agent prefers.

■

D.1 Polarization and Gridlock

As in the main model, for each agent i there exists an information quality threshold $q_{LRi}(\mathbf{e}_i, \gamma, \hat{\gamma})$ such that agent i learns correctly in the long run. Denote $\Delta_i = \text{sign}(d_{Ai} - d_{Bi})q_{LRi}$. Note that by this definition, $\Delta_i \notin (-1/2, 1/2)$ for any i because $q > 1/2$. Then

$$i \in \begin{cases} \mathcal{N}_B & \text{if } \Delta_i < -q, \\ \mathcal{N}_\omega & \text{if } -q < \Delta_i < q, \\ \mathcal{N}_A & \text{if } \Delta_i > q, \end{cases}$$

where $\mathcal{N}_\omega$ is the set of normal agents whose long-run belief puts probability 1 on state ω.

Pivotal Agents Order normal agents such that $i < j$ if $\Delta_i \leq \Delta_j$. Following Anesi and Bowen (2021), we identify two key members of the society called the A- and B- pivots.

Definition 1 (Anesi and Bowen (2021))**.** Agent $i \in \mathcal{N}$ is a pivot for decisive coalition $\mathcal{C}$ if $\{j \in \mathcal{N} : \Delta_j < \Delta_i\} \notin \mathcal{C}$ and $\{j \in \mathcal{N} : \Delta_j > \Delta_i\} \notin \mathcal{C}$. The set of pivots for $\mathcal{C}$ is denoted $P(\mathcal{C})$, and we refer to $\mathbb{A} \equiv \min P(\mathcal{C})$ and $\mathbb{B} \equiv \max P(\mathcal{C})$ as the A- and B- pivots, respectively.

Observe that $\mathbb{A} \leq \mathbb{B}$, and coalitions $\{1, \dots, \mathbb{B}\}$ and $\{\mathbb{A}, \dots, n\}$ must be decisive. So $\mathbb{A}$ is pivotal for implementing policy x_A and $\mathbb{B}$ is pivotal for implementing policy x_B. This leads to the following immediate result that relates the quality of information q to the implemented policy.

Proposition 16 (Gridlock)**.** *Suppose the state is B, then*

$$x = \begin{cases} x_B & \text{if } \Delta_{\mathbb{B}} < q \\ x_A & \text{if } \Delta_{\mathbb{A}} > q \\ x_0 & \text{if } \Delta_{\mathbb{A}} < q < \Delta_{\mathbb{B}} \end{cases}$$

Suppose the state is A, then

$$x = \begin{cases} x_B & \text{if } \Delta_{\mathbb{B}} < -q \\ x_A & \text{if } \Delta_{\mathbb{A}} > -q \\ x_0 & \text{if } \Delta_{\mathbb{A}} < -q < \Delta_{\mathbb{B}} \end{cases}$$

Thus, gridlock (i.e., $x = x_0$) occurs if $\Delta_{\mathbb{A}} < q < \Delta_{\mathbb{B}}$ and $\omega = B$, or if $\Delta_{\mathbb{A}} < -q < \Delta_{\mathbb{B}}$ and $\omega = A$.

In the result below, we consider the reasonable case of $\mathbb{A} < |\mathcal{N}|/2$ and $\mathbb{B} > |\mathcal{N}|/2$, i.e., more than half the population of normal agents is to the right of the A-pivot and more than half the population is to the left of the B-pivot. Note that this holds for most decision rules commonly used in policy making, such as super majority rules (including a filibuster rule), presidential veto, or veto rules such as what is used in the UN. The cases of simple-majority and dictatorship are ruled out, because these imply $\mathbb{A} = \mathbb{B}$. In these cases, gridlock never happens by Proposition 16.

Proposition 17 (Polarization and Gridlock)**.** *Fix a society of $\mathcal{N}$ normal agents and their echo chambers. Suppose $\mathbb{A} < |\mathcal{N}|/2$ and $\mathbb{B} > |\mathcal{N}|/2$. There exist a threshold $\overline{\Pi}$ such that below $\overline{\Pi}$ gridlock does not occur, but above $\overline{\Pi}$ gridlock occurs.*

Proof. Recall that $\Pi \in [0, 1]$ according to our definition, where $\Pi = 1$ if $|\mathcal{N}_A| = |\mathcal{N}_B|$ and $\Pi = 0$ if $|\mathcal{N}_A| \in \{0, |\mathcal{N}|\}$.

Fix any state ω and suppose that there is no polarization (i.e., $\Pi = 0$), then clearly gridlock does not occur. More formally, if $|\mathcal{N}_\omega| = |\mathcal{N}|$, it means that all normal agents correctly learn that the state is ω and implement the policy x_ω. This implies that $\Delta_{\mathbb{A}} > -q$ if $\omega = A$ and $\Delta_{\mathbb{B}} < q$ if

$\omega = B$. If $|\mathcal{N}_\omega| = 0$, it means that all normal agents incorrectly learn that the state is $-\omega$ and implement the policy $x_{-\omega}$. This implies that $\Delta_{\mathbb{A}} > q$ if $\omega = B$ and $\Delta_{\mathbb{B}} < -q$ if $\omega = A$. Either way, gridlock does not happen.

Fix any state ω and suppose that $\Pi = 1$, then gridlock does occur. Indeed, if $\omega = A$, we must have $|\mathcal{N}_B| = |\mathcal{N}|/2$ agents with $\Delta_i < -q$ and $|\mathcal{N}_A| = |\mathcal{N}|/2$ agents with $\Delta_i > -q$. Since we ordered the normal agents by their Δ_i and we assumed $\mathbb{A} < |\mathcal{N}|/2$ and $\mathbb{B} > |\mathcal{N}|/2$, we must have that $\Delta_{\mathbb{A}} < -q < \Delta_{\mathbb{B}}$, which implies gridlock. If instead $\omega = B$, we must have $|\mathcal{N}_B| = |\mathcal{N}|/2$ agents with $\Delta_i < q$ and $|\mathcal{N}_A| = |\mathcal{N}|/2$ agents with $\Delta_i > q$. For the same reasons as before, we must have that $\Delta_{\mathbb{A}} < q < \Delta_{\mathbb{B}}$, which implies gridlock.

Given this, we only need to show that (i) if gridlock occurs at polarization level $\Pi < 1$, it will also happen at any $\Pi' > \Pi$, and (ii) if gridlock does not occur at polarization level $\Pi > 0$, it will also not happen at any $\Pi' < \Pi$. We prove (i) first. Suppose $\omega = A$, polarization is $\Pi < 1$, and $\Delta_{\mathbb{A}} < -q < \Delta_{\mathbb{B}}$, which means that the A-pivot is in $\mathcal{N}_B$ and the B-pivot is in $\mathcal{N}_A$. Suppose $|\mathcal{N}_A| > |\mathcal{N}_B| > 0$. Since $\Pi' > \Pi$ requires $|\mathcal{N}_A| > |\mathcal{N}'_A| \geq |\mathcal{N}'_B| > |\mathcal{N}_B|$, we must have $\mathcal{N}_B \subset \mathcal{N}'_B$, so we still have that the A-pivot is in $\mathcal{N}'_B$. For the B-pivot to be in $\mathcal{N}'_B$, we would need $|\mathcal{N}'_B| > |\mathcal{N}_A|$, which would imply $\Pi' < \Pi$ in contrast to our assumption that $\Pi' > \Pi$. Therefore, we still have the B-pivot is in $\mathcal{N}'_A$, so we still have gridlock. The other cases follow similarly.

We now prove (ii). Suppose $\omega = A$, polarization is $\Pi > 0$, and either $\Delta_{\mathbb{A}} > -q$ or $\Delta_{\mathbb{B}} < -q$, which means that the A-pivot is in $\mathcal{N}_A$ or the B-pivot is in $\mathcal{N}_B$. If the A-pivot is in $\mathcal{N}_A$, it means that $|\mathcal{N}_A| > |\mathcal{N}|/2 > |\mathcal{N}_B|$. Then, for $\Pi' < \Pi$ we must have either $|\mathcal{N}'_A| > |\mathcal{N}_A|$ (in which case the A-pivot is in $\mathcal{N}'_A$) or $|\mathcal{N}'_B| > |\mathcal{N}_A|$ (in which case the B-pivot is in $\mathcal{N}'_B$). Either way, there is no gridlock. The case in which we start with the B-pivot is in $\mathcal{N}_B$ works similarly. The other cases follow similarly. ∎

This last result implies that if polarization is non-monotonic in information quality q (as we point out in the paper), then gridlock also depends on q in a non-monotonic way: Gridlock may not occur for low and high q, but it may occur for intermediate q.

E Properties of $\tau(q)$

We will prove that $\tau(q)$ in condition (9) is concave for $q \in \left(\frac{1}{2}, 1\right)$ and that $\tau'\left(\frac{1}{2}\right) = 0$. Recall that we assume $d_A > d_B$. We can write

$$\tau(q) = \frac{1 + n + (1 + z(q, \hat{\gamma}))d_B + \frac{z(q,\hat{\gamma})}{\gamma}(d_A - d_B)}{2(1+n) + (d_A + d_B)(1 + z(q, \hat{\gamma}))} = \frac{A + Bz(q, \hat{\gamma})}{C + Dz(q, \hat{\gamma})} = \frac{B}{D} + \frac{AD - BC}{D(C + Dz(q, \hat{\gamma}))},$$

where

$$A = 1 + n + d_B,\ B = d_B + \frac{d_A - d_B}{\hat{\gamma}},\ C = 2 + 2n + d_A + d_B,\ D = d_A + d_B.$$

Also, we have that

$$AD - BC = -\left(\frac{d_A + d_B}{\hat{\gamma}} + \frac{(2 - \hat{\gamma})(1 + n)}{\hat{\gamma}}\right)(d_A - d_B),$$

which is strictly negative. Therefore, $\tau(q)$ is concave if and only if $g(q)$ is convex, where

$$g(q) = \frac{1}{C + Dz(q, \hat{\gamma})}.$$

We will prove this in steps.

Lemma 6. *$z_q(q, \hat{\gamma}) \leq 0$ for $q \in \left(\frac{1}{2}, 1\right)$ and $\lim_{q \to \frac{1}{2}} z_q(q, \hat{\gamma}) = 0$.*

Proof. Consider the derivative of $z(q, \hat{\gamma})$ with respect to q:

$$\begin{aligned} \frac{\partial}{\partial q} \frac{\ln\left(\frac{\hat{\gamma}(1-q)+(1-\hat{\gamma})}{\hat{\gamma}q+(1-\hat{\gamma})}\right)}{\ln\left(\frac{1-q}{q}\right)} &= \frac{-\frac{\hat{\gamma}(2-\hat{\gamma})}{\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})} \cdot \ln\left(\frac{1-q}{q}\right) + \ln\left(\frac{\hat{\gamma}(1-q)+(1-\hat{\gamma})}{\hat{\gamma}q+(1-\hat{\gamma})}\right) \cdot \frac{1}{q(1-q)}}{\ln^2\left(\frac{1-q}{q}\right)} \\ &= \frac{-\frac{\hat{\gamma}(2-\hat{\gamma})}{\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})} \cdot \ln\left(\frac{1-q}{q}\right) + \ln\left(\frac{1-q}{q}\right) \cdot \frac{z(q,\hat{\gamma})}{q(1-q)}}{\ln^2\left(\frac{1-q}{q}\right)} \\ &= \frac{\left(\frac{1-\hat{\gamma}}{q(1-q)} + \hat{\gamma}^2\right) z(q, \hat{\gamma}) - (2 - \hat{\gamma})\hat{\gamma}}{\ln\left(\frac{1-q}{q}\right) \cdot (\hat{\gamma}^2 q(1-q) + (1 - \hat{\gamma}))}. \end{aligned} \tag{16}$$

Note that $\lim_{q\to\frac{1}{2}} z(q,\hat{\gamma}) = \frac{\hat{\gamma}}{2-\hat{\gamma}} > 0 = z(1,\hat{\gamma})$. This immediately implies that $\lim_{q\to\frac{1}{2}} z_q(q,\hat{\gamma}) = 0$.

As $z(q,\hat{\gamma})$ is continuously differentiable for $q \in \left(\frac{1}{2},1\right)$, it is enough to prove that there are no local maximum on $\left(\frac{1}{2},1\right)$ in order to show that $z_q(q,\hat{\gamma}) \leq 0$ holds on this interval. At an intermediate local maximum, $z_q(q,\hat{\gamma}) = 0$ must hold. This requires that

$$\left(\frac{1-\hat{\gamma}}{q(1-q)} + \hat{\gamma}^2\right) z(q,\hat{\gamma}) - (2-\hat{\gamma})\hat{\gamma} = 0$$

and hence

$$\begin{aligned} z(q,\hat{\gamma}) &= \frac{\hat{\gamma}(2-\hat{\gamma})}{\hat{\gamma}^2 + \frac{1-\hat{\gamma}}{q(1-q)}} \\ &\leq \frac{\hat{\gamma}(2-\hat{\gamma})}{\hat{\gamma}^2 + \frac{1-\hat{\gamma}}{\frac{1}{4}}} = \frac{\hat{\gamma}}{2-\hat{\gamma}}. \end{aligned} \tag{17}$$

This rules out that $z(q,\hat{\gamma})$ is increasing at $q = \frac{1}{2}$, since it would need to achieve a local maximum with value above $\frac{\hat{\gamma}}{2-\hat{\gamma}}$. Now note that the right-hand side of (17) is strictly decreasing in q over $\left(\frac{1}{2},1\right)$. If $z(q,\hat{\gamma})$ was to decrease at first (as q rises from $\frac{1}{2}$) and then increase before going down to 0, the value of $z(q,\hat{\gamma})$ at the corresponding local maximum would be necessarily above the right-hand side of (17), which is a contradiction. One final case is that $z(q,\hat{\gamma})$ is decreasing at first, passing through a local minimum, and then is increasing until $q = 1$. This would mean that the value at the local minimum is less than $z(1,\hat{\gamma})$, which is equal to 0. Since $z(q,\hat{\gamma}) > 0$ for $q \in (\frac{1}{2},1)$ and $\hat{\gamma} \in (0,1)$, this case is also impossible. We conclude that $z(q,\hat{\gamma})$ is weakly decreasing over $\left(\frac{1}{2},1\right)$. ■

This implies that $\lim_{q\to\frac{1}{1}} g'(q) = 0$ because

$$g'(q) = -\frac{Dz_q(q,\hat{\gamma})}{(C + Dz(q,\hat{\gamma}))^2}.$$

Lemma 7. *$g(q)$ is convex.*

Proof. Since

$$g''(q) = \frac{2D^2\left(z_q(q,\hat{\gamma})\right)^2 - D(C+Dz(q,\hat{\gamma}))z_{qq}(q,\hat{\gamma})}{(C+Dz(q,\hat{\gamma}))^3},$$

the result follows if we can prove that $z_{qq}(q,\hat{\gamma}) < 0$ for all $q \in \left(\frac{1}{2},1\right)$.

Using (16) and letting $K(q) = \frac{1}{\left[\ln\left(\frac{1-q}{q}\right)\right]^2(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma}))^2}$, we have

$$\begin{aligned}
z_{qq}(q,\hat{\gamma}) = K(q)&\Bigg[\left(-\tfrac{(1-\hat{\gamma})(1-2q)}{q^2(1-q)^2} + \left(\tfrac{1-\hat{\gamma}}{q(1-q)}+\hat{\gamma}^2\right)z_q(q,\hat{\gamma})\right)\ln\left(\tfrac{1-q}{q}\right)\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right)\\
&-\left(\left(\tfrac{1-\hat{\gamma}}{q(1-q)}+\hat{\gamma}^2\right)z(q,\hat{\gamma})-\hat{\gamma}(2-\hat{\gamma})\right)\left(\tfrac{-1}{q(1-q)}\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right)+\ln\left(\tfrac{1-q}{q}\right)\hat{\gamma}^2(1-2q)\right)\Bigg]\\
= K(q)&\Bigg[\left(\tfrac{(1-\hat{\gamma})(2q-1)}{q^2(1-q)^2} + \left(\tfrac{1-\hat{\gamma}}{q(1-q)}+\hat{\gamma}^2\right)z_q(q,\hat{\gamma})\right)\ln\left(\tfrac{1-q}{q}\right)\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right)+\\
&+\left(\left(\tfrac{1-\hat{\gamma}}{q(1-q)}+\hat{\gamma}^2\right)z(q,\hat{\gamma})-\hat{\gamma}(2-\hat{\gamma})\right)\left(\tfrac{1}{q(1-q)}\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right)+\ln\left(\tfrac{1-q}{q}\right)\hat{\gamma}^2(2q-1)\right)\Bigg].
\end{aligned}$$

Let

$$\begin{aligned}
C_1(q) &= \frac{(1-\hat{\gamma})(2q-1)}{q^2(1-q)^2} + \left(\frac{1-\hat{\gamma}}{q(1-q)}+\hat{\gamma}^2\right)z_q(q,\hat{\gamma}),\\
C_2(q) &= \left(\frac{1-\hat{\gamma}}{q(1-q)}+\hat{\gamma}^2\right)z(q,\hat{\gamma})-\hat{\gamma}(2-\hat{\gamma}),\\
C_3(q) &= \frac{1}{q(1-q)}\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right)+\ln\left(\frac{1-q}{q}\right)\hat{\gamma}^2(2q-1).
\end{aligned}$$

Then we can write

$$z_{qq}(q,\hat{\gamma}) = K(q)\left[C_1(q)\ln\left(\frac{1-q}{q}\right)\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right)+C_2(q)C_3(q)\right]$$

Using the expression of $z_q(q,\hat{\gamma})$, we can write $C_1(q)$ as

$$\begin{aligned}
C_1(q) &= \frac{(1-\hat{\gamma})(2q-1)}{q^2(1-q)^2} + \frac{\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})}{q(1-q)} \cdot \frac{\left(\frac{1-\hat{\gamma}}{q(1-q)} + \hat{\gamma}^2\right) z(q,\hat{\gamma}) - \hat{\gamma}(2-\hat{\gamma})}{\ln\left(\frac{1-q}{q}\right)\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right)} \\
&= \frac{(1-\hat{\gamma})(2q-1)\ln\left(\frac{q}{1-q}\right) + \hat{\gamma}(2-\hat{\gamma})q(1-q) - \left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right) z(q,\hat{\gamma})}{q^2(1-q)^2 \ln\left(\frac{q}{1-q}\right)}
\end{aligned}$$

and therefore

$$\begin{aligned}
C_1(q)\ln\left(\frac{1-q}{q}\right)\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right) &= -\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right) \cdot \\
&\cdot \frac{(1-\hat{\gamma})(2q-1)\ln\left(\frac{q}{1-q}\right) + \hat{\gamma}(2-\hat{\gamma})q(1-q) - \left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right) z(q,\hat{\gamma})}{q^2(1-q)^2}
\end{aligned}$$

Using

$$\begin{aligned}
C_2(q)C_3(q) &= \frac{\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right) z(q,\hat{\gamma}) \cdot \left[\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right) + \ln\left(\frac{1-q}{q}\right)\hat{\gamma}^2(2q-1)q(1-q)\right]}{q^2(1-q)^2} \\
&\quad -\hat{\gamma}(2-\hat{\gamma})\frac{\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right)q(1-q) + \ln\left(\frac{1-q}{q}\right)\hat{\gamma}^2(2q-1)q^2(1-q)^2}{q^2(1-q)^2},
\end{aligned}$$

we can write

$$\begin{aligned}
\frac{z_{qq}(q,\hat{\gamma})q^2(1-q)^2}{K(q)} &= \left(2\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right) + \ln\left(\frac{1-q}{q}\right)\hat{\gamma}^2(2q-1)(1-q)\right) \cdot \\
&\qquad \cdot \left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right) z(q,\hat{\gamma}) \\
&\quad + \left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right)\left[(1-\hat{\gamma})(2q-1)\ln\left(\frac{1-q}{q}\right) - 2\hat{\gamma}(2-\hat{\gamma})q(1-q)\right] \\
&\quad + \ln\left(\frac{q}{1-q}\right)\hat{\gamma}^3(2-\hat{\gamma})(2q-1)q^2(1-q)^2 \\
&= 2\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right)^2 z(q,\hat{\gamma}) + \ln\left(\frac{q}{1-q}\right)\hat{\gamma}^3(2-\hat{\gamma})(2q-1)q^2(1-q)^2 \\
&\quad - \left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right)\ln\left(\frac{q}{1-q}\right)(2q-1)\left[\hat{\gamma}^2(1-q)z(q,\hat{\gamma}) + (1-z(q,\hat{\gamma}))\right] \\
&\quad -2\left(z(q,\hat{\gamma})^2 q(1-q) + (1-z(q,\hat{\gamma}))\right) z(q,\hat{\gamma})(2-z(q,\hat{\gamma}))q(1-q).
\end{aligned}$$

Let

$$D_1(q) = 2\left(z(q,\hat{\gamma})^2 q(1-q) + (1-z(q,\hat{\gamma}))\right) z(q,\hat{\gamma})$$
$$-\ln\left(\tfrac{q}{1-q}\right)(2q-1)(1-z(q,\hat{\gamma})) - 2z(q,\hat{\gamma})(2-z(q,\hat{\gamma}))q(1-q)$$

and

$$D_2(q) = z(q,\hat{\gamma})^3(2-z(q,\hat{\gamma}))q^2(1-q)^2$$
$$-\left(z(q,\hat{\gamma})^2 q(1-q) + (1-z(q,\hat{\gamma}))\right) z(q,\hat{\gamma})^2(1-q)z(q,\hat{\gamma})$$

Then we have

$$\frac{z_{qq}(q,z(q,\hat{\gamma}))q^2(1-q)^2}{K(q)} = \left(z(q,\hat{\gamma})^2 q(1-q) + (1-z(q,\hat{\gamma}))\right) D_1(q) + \ln\left(\frac{q}{1-q}\right)(2q-1)D_2(q). \tag{18}$$

Note that

$$\begin{aligned} D_1(q) &\leq 2\left(z(q,\hat{\gamma})^2 q(1-q) + (1-z(q,\hat{\gamma}))\right) \tfrac{z(q,\hat{\gamma})}{2-z(q,\hat{\gamma})} \\ &\quad -\ln\left(\tfrac{q}{1-q}\right)(2q-1)(1-z(q,\hat{\gamma})) - 2z(q,\hat{\gamma})(2-z(q,\hat{\gamma}))q(1-q) \\ &= \tfrac{1}{2-z(q,\hat{\gamma})}\Big[2z(q,\hat{\gamma})^3 q(1-q) + 2z(q,\hat{\gamma})(1-z(q,\hat{\gamma})) \\ &\quad -\ln\left(\tfrac{q}{1-q}\right)(2q-1)(1-z(q,\hat{\gamma}))(2-z(q,\hat{\gamma})) - 2z(q,\hat{\gamma})(2-z(q,\hat{\gamma}))^2 q(1-q)\Big] \\ &= \frac{1-z(q,\hat{\gamma})}{2-z(q,\hat{\gamma})}E(q), \end{aligned}$$

where $E(q) = 2z(q,\hat{\gamma})(1-4q(1-q)) - \ln\left(\frac{q}{1-q}\right)(2q-1)(2-z(q,\hat{\gamma}))$. Differentiating this expression with respect to q, we get

$$\begin{aligned} E'(q) &= 2z(q,\hat{\gamma}) \cdot 4(2q-1) - \frac{1}{q(1-q)}(2q-1)(2-z(q,\hat{\gamma})) - 2\ln\left(\frac{q}{1-q}\right)(2-z(q,\hat{\gamma})) \\ &= (2q-1)\left(4z(q,\hat{\gamma}) - \frac{2-z(q,\hat{\gamma})}{q(1-q)}\right) - 2\ln\left(\frac{q}{1-q}\right)(2-z(q,\hat{\gamma})) \\ &< (2q-1)\left(4z(q,\hat{\gamma}) - 4(2-z(q,\hat{\gamma}))\right) - 2\ln\left(\frac{q}{1-q}\right)(2-z(q,\hat{\gamma})) < 0 \end{aligned}$$

for $q \in \left(\frac{1}{2}, 1\right)$. Therefore, $E(q) < E\left(\frac{1}{2}\right)$ for any $q \in \left(\frac{1}{2}, 1\right)$, where

$$E\left(\frac{1}{2}\right) = 2z(q,\hat{\gamma})\left(1 - 4 \cdot \frac{1}{4}\right) - \ln(1)\left(2 \cdot \frac{1}{2} - 1\right)(2 - z(q,\hat{\gamma})) = 0.$$

Therefore, we can conclude that $D_1(q) < 0$ for $q \in \left(\frac{1}{2}, 1\right)$.

Returning to $D_2(q)$, note that

$$\begin{aligned} D_2(q) &= z(q,\hat{\gamma})^2(1-q)\left[z(q,\hat{\gamma})(2-z(q,\hat{\gamma}))q^2(1-q) - \left(z(q,\hat{\gamma})^2 q(1-q) + (1-z(q,\hat{\gamma}))\right)z(q,\hat{\gamma})\right] \\ &< z(q,\hat{\gamma})^2(1-q)\left[z(q,\hat{\gamma})(2-z(q,\hat{\gamma}))q(1-q) - \left(z(q,\hat{\gamma})^2 q(1-q) + (1-z(q,\hat{\gamma}))\right)z(q,\hat{\gamma})\right] \end{aligned}$$

The expression in the brackets is the negative of the numerator in $z_q(q, z(q,\hat{\gamma}))$. Given that $z_q(q, z(q,\hat{\gamma}))$ is negative and its expression includes $\ln\left(\frac{1-q}{q}\right)$, it follows that the numerator has to be positive. This implies that the expression above is negative, and therefore, $D_2(q)$ must be negative as well.

Using $D_1(q) < 0$ and $D_2(q) < 0$ for $q \in \left(\frac{1}{2}, 1\right)$ and (18), we can conclude that $z_{qq}(q, z(q,\hat{\gamma})) < 0$ for $q \in \left(\frac{1}{2}, 1\right)$.

■