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

Renee Bowen[†]
Danil Dmitriev[‡]
Simone Galperti[§]

February 10, 2021

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. Our results hold without media bias or fake news, so eliminating these is not sufficient to reduce polarization. When fake news is included, we show that it can lead to polarization but *only* through misperceived selective sharing. News aggregators can curb polarization caused by shared news.

JEL codes: D82, D83, D90

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

[*] We thank S. Nageeb Ali, Myles Ellis, Harry Pei, Jacopo Perego, Joel Sobel, and participants in conferences and seminars at UCSD, PhDEI, SIOE, PennState, NBER POL, UC Berkeley, UC Davis, NYU, and MEDS 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

Social divisions have been linked to economic and political issues such as inequality, political gridlock, poor legislation, weak property rights, low trust, investment, and growth.[1] Recent decades have witnessed rising polarization in politics, media, and public opinions—especially in the United States.[2] "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" (Alesina et al., 2020). Economists have thus been studying the causes of belief polarization. Some have suggested a connection with the use of the Internet as a source of information.[3] Others have blamed *misinformation*—that is, fake news, bots, and media bias—leading to discussions about regulating social media.[4] Even if successful, will such regulations curb polarization? Can polarization simply result from how people consume and share information through their social connections, even without misinformation? Will the abundance and extensive sharing of information brought by technology lead to more or less polarization?

We provide a theoretical framework to answer these questions and discuss implications for policies aimed at curbing polarization. To this end, we build a simple benchmark model that incorporates key empirical findings about how people share first-hand information via social connections and absorb the resulting second-hand information. In a nutshell, people often *share selectively*; as a result, some people consume *unbalanced diets* of second-hand information; moreover, they tend to misinterpret this information because they *misperceive* how others share it selectively.[5] We study the consequences for learning and belief polarization. We find that misperceptions and quality of first-hand information (as opposed to quantity) play critical and subtle roles. Our results do not require preexisting differences in people's worldviews nor misinformation. Yet, they imply that selective sharing is one (and in a sense the only) channel through which fake news can lead to polarization. We suggest mechanisms whereby changes in people's information ecosystem and social connections brought on by the Internet may contribute to polarization.

In our baseline model, agents learn about a binary state of the world, A or B, over time. In every period, each agent gets an objective signal about the state with probability γ and

[1] See Zak and Knack (2001); Keefer and Knack (2002); Bartels (2008); Bishop (2009); McCarty et al. (2009); Gilens (2012); Barber and McCarty (2015).

[2] See Pew Research Center (2014, 2020); Desmet and Wacziarg (2018) Bertrand and Kamenica (2018).

[3] See Periser (2011); Flaxman et al. (2016); Sunstein (2017); Azzimonti and Fernandes (2018); Tucker et al. (2019); Zhuravskaya et al. (2020).

[4] See, for example, "Should the Government Regulate Social Media?", Wall Street Journal (June 25, 2019) and "Facebook Throws More Money at Wiping Out Hate Speech and Bad Actors", Wall Street Journal (May 15, 2018).

[5] For evidence of selective sharing, see Shin and Thorson (2017); Weeks et al. (2017); Shin et al. (2018); Pogorelskiy and Shum (2019); Levy (2020); Zhuravskaya et al. (2020). Unbalanced news diets are a distinctive aspect of so-called echo chambers or media bubbles, which appear in a wealth of evidence (see Levy and Razin, 2019a; Zhuravskaya et al., 2020, for a review). Bertrand and Kamenica (2018) also stress the importance of media diet driving social differences. Pogorelskiy and Shum (2019) provide evidence of misperception about selective sharing.

no signal otherwise (first-hand information). Signals are i.i.d. across agents and periods. We refer to their informativeness as *quality*. In every period, each agent can remain silent or share her signal with her social connections—called *friends*. She cannot tamper with her signals, but can select which to share. We assume that some agents—called *normal*—share every signal; other agents—called *dogmatic*—selectively share only signals supporting one state. To fix ideas, some people (possibly a tiny minority) may hold a dogmatic view on whether to vaccinate children and share only articles in its favor; others simply share any article on the topic. We refer to an agent's sources of second-hand signals as her echo chamber. If a majority of her dogmatic friends supports one state, it creates an unbalanced news diet. We model misperception of selective sharing in a way that renders the agents partially unresponsive to it (as found in Pogorelskiy and Shum (2019)) and is inspired by the psychology literature.[6] Each agent interprets all signals correctly, but thinks that they arrive with probability $\hat{\gamma} \neq \gamma$. This is akin to assuming that friends read the newspaper less or more often than they actually do. Thus, our agents have a common misspecified model of selective sharing, based on which they update beliefs using Bayes' rule. We also consider other misperceptions and show that they have similar implications.

We find that people's understanding of the selectivity of shared news turns out to be crucial. Without any misperception, unbalanced selective sharing alone cannot lead to polarization (Remark 1). This is an important qualification of a common intuition about the effects of echo chambers. If a person took at face value what her friends say supporting only one view, her opinion could be swayed accordingly. In reality, however, only a few friends may share information selectively and often people also get first-hand information. Even in the absence of these mitigating factors, if a person fully understands how her friends select what to share, she will adjust for it and her beliefs will not be distorted. It is unrealistic that people are completely naive about selective sharing, but experimental evidence suggests they do not fully take it into account either.

We analyze learning both in the short run (after one round of signals) and in the long run (after infinitely many rounds). In the short run, an agent's *expected* posterior can differ from her prior, even when she has many normal friends. The intuition is that the silence of a dogmatic friend indicates bad news for the state he supports. Yet, if an agent thinks that this friend is informed *less* often than he is ($\hat{\gamma} < \gamma$), she reads too little into his silence, thus enabling him to distort her posterior *towards* the state he supports. By contrast, if an agent thinks that her friends are informed *more* often than they are ($\hat{\gamma} > \gamma$), she reads too much into their silence, which distorts her posterior *away* from the state supported by a dogmatic friend. Thus, if a majority of such friends favors, say, state A, her expected posterior is distorted towards (away from) A if $\hat{\gamma} < \gamma$ ($\hat{\gamma} > \gamma$)—at least when the signal quality is sufficiently low. Perhaps unexpectedly, we find that even balanced echo chambers can distort an agent's posterior.

For the long run, abundant information can boost the distorting power of dogmatic friends—instead of curbing it—thereby exaggerating incorrect learning. We identify a precise quality threshold below which the agent's long-run belief assigns probability one to a given

[6] See Cross (1977); Svenson (1981); Odean (1998); Zuckerman and Jost (2001).

state, *irrespective* of the truth. For higher quality, her belief converges to the truth despite the echo-chamber effect. Long-run incorrect learning requires unbalanced dogmatic friends: If their majority favors A, the agent's belief converges to A if $\hat{\gamma} < \gamma$ and to B if $\hat{\gamma} > \gamma$. Note that the imbalance can be arbitrarily small, yet offset many unfiltered signals.

It is easy to see how these forces can cause polarization. If the echo chambers of some agents are unbalanced towards different states *and* information quality is sufficiently low, their beliefs can move apart on average in the short run and almost surely in the long run. One of our main contributions is to highlight the role of information quality. Indeed, we find that for intermediate qualities polarization can occur even if all echo chambers are unbalanced towards the same state. One may also expect that raising information quality would undoubtedly help curb polarization. However, it can actually *increase* polarization under some simple conditions, which we identify. The role of information quality also implies that polarization in echo chambers and in beliefs need not go hand in hand.

We find that the expansion of social connections can be another driver of polarization. This is not obvious, as more connections may provide greater scope for echo chambers to distort beliefs but also bring more information. Fixing its quality, we obtain conditions on how the internal structure of echo chambers has to change for polarization to weaken. For instance, it is possible that society is not polarized when people have small echo chambers, but becomes polarized when they have similarly divided, but larger, echo chambers. This could happen if social media recommend new friends in ways that depend only marginally (or not at all) on how they share information.

Our analysis goes to the heart of why new communication technologies and formats enabled by the Internet can increase polarization. They speed up the arrival of information and possibly lower its quality—for instance, tweets and social-media posts tend to be short. Moreover, overwhelmed by the stream of information, people may spread their attention across more sources, thereby absorbing less content from each. This may effectively lower the quality of consumed information. All of this can lead to polarization even without deliberate misinformation.

Finally, we return to the motivating question of what policies may reduce the effects of the Internet and news sharing more generally on polarization. Some of the drivers we highlight may be off limits, as they result from legitimate personal rights: making friends and sharing information as one pleases. Possible solutions may emerge from our focus on information quality. An immediate one is that news outlets provide higher-quality information, but this may be hard to incentivize and decentralize. Another solution is to exploit news aggregators. Although their reasons to exist may be different, we show how aggregators can provide higher-quality information even when they *lose* some information by summarizing facts. We identify a minimal degree of aggregation that suffices to remove polarization. The catch is that reaching this degree may require institutional aggregators that take into account the externalities caused by selective sharing. These insights provide a rationale for authorities to commit to releasing information only rarely in "digested" batches.

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.[7] Our contribution to this literature is to highlight misperception of selective sharing as a driver of polarization. Another cause is heterogeneity in preferences.[8] Such heterogeneity would exacerbate the polarization we find. A third cause is biased or multidimensional information sources, where biases usually come from media competition over viewers.[9] In our analysis, the sources of first-hand information—which can be interpreted as media outlets—are not biased. Thus, removing all media biases may still not be enough to curb polarization.

This paper fits into the growing literature on model misspecification and social learning. The classic work of Berk (1966) provides a general model of individual misspecified learning in the long run. 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 how model misspecification impacts long-run learning in environments where agents learn from private signals and the actions of others. In particular, Bohren and Hauser (2018) study when agents with different, yet reasonable, models of the world have no limit beliefs (i.e., beliefs cycle) or different limit beliefs (disagreement). Although close in spirit, our disagreement results are driven by a fundamentally different mechanism, as all our agents have the same model of the world. We also emphasize the role of information quality and its implications for curbing polarization. Like Bohren and Hauser (2018), Mailath and Samuelson (2020) consider agents with different misspecified models of the world. Molavi et al. (2018) study long-run learning on social networks when non-Bayesian agents exhibit imperfect recall. They show that such agents may overweight evidence encountered early on, which can lead to mislearning. Unlike these authors, we assume that agents update beliefs via Bayes' rule.[10]

A key element of our model is the idea of an echo chamber as the group of friends from whom one receives information. This connects our work to a large literature that studies both Bayesian and non-Bayesian learning in networks.[11] One closely related paper is Levy and Razin (2019a), which shows that an updating heuristic called "Bayesian Peer Influence" can cause limit beliefs in networks to become polarized. However, their meaning of polarization is different from ours: There the entire society's *consensus* shifts towards a common extreme belief; here the agents' beliefs diverge to different extreme beliefs.

Recent empirical studies show that social media is an important source of news for people

[7] See, e.g., Levy and Razin (2019b); Hoffmann et al. (2019); Enke et al. (2019).

[8] See Dixit and Weibull (2007); Pogorelskiy and Shum (2019).

[9] 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 (2018).

[10] 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).

[11] 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 (2019).

and can lead beliefs and attitudes to diverge.[12] Other evidence by Boxell et al. (2018) suggests that the Internet does not drive polarization. Our model can predict in which environments we expect to see the Internet drive polarization. We, thus, contribute to this literature by providing a theoretical framework to better understand how social media can contribute to polarization and to guide future empirical investigation and policy discussion.

2 Model

We consider a stylized model of learning from information shared through social connections. Time t is discrete, where $t = 0, \dots, T$ and $T \leq \infty$. A state of the world $\omega \in \{A, B\}$ realizes at $t = 0$. For example, ω can represent whether preserving the environment requires higher national spending than the current level, 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:

$$\begin{aligned} \mathbb{P}(s_{it} = a | \omega = A) &= \mathbb{P}(s_{it} = b | \omega = B) = q, \\ \mathbb{P}(s_{it} = b | \omega = A) &= \mathbb{P}(s_{it} = a | \omega = B) = 1 - q, \end{aligned} \tag{1}$$

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.[13]

Selective Sharing. Agents share their first-hand information with other agents with whom they have a social connection. We call these social connections *friends.* We aim to capture key aspects of social information sharing suggested by experimental evidence (see Footnote 5). The first is selectivity. In our model, after receiving her own signal, an agent 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, which rules out fake news. Concretely, an agent can share a newspaper article, but cannot edit its content.

We introduce three types of agents characterized by their information-sharing behavior. An agent is *normal* if she shares any signal she receives, *A-dogmatic* if she shares only signals $s_{it} = a$, and *B-dogmatic* if she shares only signals $s_{it} = b$. One interpretation is that some agents dogmatically believe in their conviction that only one state is true and share only information that supports it. Formally, in each period the dogmatic types share their

[12] See, e.g., Allcott and Gentzkow (2017); Bursztyn et al. (2019); Mosquera et al. (2019); Levy (2020). See also Barberá (2020) and Zhuravskaya et al. (2020) for recent reviews of this literature.

[13] 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.

received signals as follows:

$$\sigma_A(s_{it}) = \begin{cases} \text{share} & \text{if } s_{it} = a \\ \text{stay silent} & \text{if } s_{it} = b, \end{cases} \qquad \sigma_B(s_{it}) \ = \begin{cases} \text{stay silent} & \text{if } s_{it} = a \\ \text{share} & \text{if } s_{it} = b. \end{cases}$$

Each agent's type is exogenous and known to her friends.[14]

A key aspect of social news sharing is that it contributes to creating heterogeneous information diets (Pew Research Center, 2014; Levy and Razin, 2019a).[15] Agent i's diet depends on the composition of friends she listens to, namely, the number $d_{Ai} \geq 0$ of A-dogmatic friends, $d_{Bi} \geq 0$ of B-dogmatic friends, and $n_i \geq 0$ 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. Finally, we refer to the majority (minority) of an agent's dogmatic friends as her *dogmatic majority (minority).*

Timing. Within each period $t \geq 1$ 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 share 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' 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 $\mu(\mathbf{s}_i^\infty) = \mathbf{plim}_{T\to\infty}\, \mu(\mathbf{s}^T)$. We will introduce a formal measure of belief polarization in Section 4. However, intuitively, polarization requires that beliefs move *systematically* apart between agents. It is well known that $\mu(\mathbf{s}_i^1)$ and $\mu(\mathbf{s}_j^1)$ for $i \neq j$ can differ in completely standard Bayesian models simply because agent i and j observe different signal realizations. Therefore, for the short run we adopt a more demanding condition for polarization that looks at differences between the expectations $\mathbb{E}\left[\mu(\mathbf{s}_i^1)\right]$ and $\mathbb{E}\left[\mu(\mathbf{s}_j^1)\right]$. Recall that both must equal the prior π in standard Bayesian models.[16]

At first glance, 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.

[14] In Section 6 we consider a more general model where dogmatic friends share their signals probabilistically and their type may not be perfectly known.

[15] People also have heterogeneous news diets because they choose to listen to different first-hand sources. We abstract from this aspect to focus on the effects of news sharing.

[16] As another interpretation, we can view each agent i and j 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 $\mathbb{E}\left[\mu(\mathbf{s}_i^1)\right]$ and $\mathbb{E}\left[\mu(\mathbf{s}_j^1)\right]$ approximate the empirical average belief of the respective group and may be used to assess intra-group polarization.

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

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

This is because if an agent fully 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, the agent gets some information every period, so her belief must satisfy standard properties of Bayesian updating.

Misperception. To break the impossibility implied by Remark 1, we again refer to the empirical evidence for guidance. Pogorelskiy and Shum (2019) suggest a third aspect specific to learning from shared news: Agents often misperceive the selectivity of second-hand information. When friends share their first-hand information, an agent simply has to absorb what she receives. But if they share nothing, she faces a more complex inference problem: Why did a friend remain silent? Did he get no signal? Did he suppress his signal? When does he do so? It is reasonable that her answer to any of these questions may be miscalibrated. In the baseline model, we consider the simplest form of miscalibration, which involves only one parameter: the arrival probability of signals. Formally, we let each agent think that the i.i.d. probability of getting a signal is $\hat{\gamma} \in (0,1]$. We will refer to $\hat{\gamma} < \gamma$ as under-estimating news arrival and to $\hat{\gamma} > \gamma$ as over-estimating news arrival. The agents continue to use Bayes' rule to calculate $\mu(\mathbf{s}^t)$, yet applied to this slightly misspecified model of the world. The rest of the model is unchanged. Note that we, as the external observer, will calculate $\mathbb{E}\left[\mu(\mathbf{s}^t)\right]$ and $\mu(\mathbf{s}^\infty)$ using the correct model of the world (i.e., γ not $\hat{\gamma}$).[17]

Discussion of the Model

We discuss the motivation for our modeling choices. The analysis beginning in Section 3 does not rely on anything mentioned here, so a reader may skip this section without confusion.

We can interpret misperceptions about news arrival as follows. An agent may under- or over-estimate the probability that her friends receive signals. If $\hat{\gamma} < \gamma$, this may be a manifestation of the so-called "illusory superiority" or "better-than-average" heuristic,[18] which can lead an agent to think that others are *less* informed than she is even though everyone is equally informed. People often have unjustifiably favorable views of themselves relative to the population average or even in person-to-person comparisons on various characteristics, which may include how well informed they are or how good they are at getting and understanding information. By contrast, some agent may be insecure and think that her friends are *more* informed than she is, even though everyone is equally informed (i.e., $\hat{\gamma} > \gamma$). The case of $\hat{\gamma} < \gamma$ seems more consistent with introspection and the psychology literature, but we also analyze the case of $\hat{\gamma} > \gamma$ for completeness. An agent may also misperceive the

[17] In Section 6 we will consider misperceptions about the friends' types, their news-sharing behavior, or the information quality and show that they all lead agents to misinterpret silence in ways similar to $\hat{\gamma} \neq \gamma$.

[18] See, e.g., Cross (1977); Svenson (1981); Odean (1998); Zuckerman and Jost (2001).

probability of receiving her own signals. However, this turns out to have no effect because she cannot selectively share signals with herself.

The key feature of misperception is that the agent's view of the world rules out the true γ from the set of possibilities. This is a defining feature of models with misspecification (like those listed in the related literature). Thus, even if we allowed the agent to learn about the probability of signal arrivals, she would not converge to the truth.

For simplicity, we assumed that each agent cannot choose which friends to share her signal with: She either shares it with all friends or none. This is similar to posting a newspaper article on one's social-media page where all friends can see it. Also, our model is consistent with the possibility that an agent may knowingly receive a friend's shared signal through another friend. This is similar to knowing the origin of a re-tweet on Twitter.

Sharing a signal takes the form of verifiable information in our model. By ruling out fake news, we highlight the role of selective sharing in a baseline model to which these other aspects can be added. 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 negative news.[19]

We take the types of news-sharing behavior as given because they approximate the findings in the empirical literature (see, e.g., Pogorelskiy and Shum (2019)). Moreover, our focus is not understanding *why* people tend to share to a greater extent news that supports their convictions, but understanding its *consequences* for social learning. Future research may endogenize news-sharing behavior in settings similar to ours. With regard to how we model dogmatic agents, we can view such agents as having extreme beliefs that are very hard to change—perhaps because they are stubborn, narrow minded, or blindly follow and promote some ideas. Thus, we can model them as having degenerate prior beliefs in A or B, which do not change with new information. One can also interpret our model as situations where dogmatic agents can change their views, yet much more slowly than non-dogmatic agents.[20] As a result, how they selectively share information is very persistent. Note that for our results to hold it is enough to have a few dogmatic agents.

Finally, a brief comment is in order on the heterogeneity between agents that we allow. We assume that the prior π, the true and misperceived probability of receiving signals (γ and $\hat{\gamma}$), and the signal distribution (1) are the same for all normal agents. Only the composition of echo chambers can differ between them. Starting from a setting where they are all ex-ante identical and have the same model of the world helps to highlight 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. One can easily infer the consequences of such additional differences from our results in the next section.

[19] For example, see Ben-Porath et al. (2018) and DeMarzo et al. (2019).

[20] 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).

3 Single-Agent Learning

Before examining belief polarization among agents, we study how each individually updates her belief under the effects of selective sharing and misperceptions. Since we focus on a generic normal agent, we drop all i subscripts in this section.

3.1 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.

Our first result shows that, in the presence of misperception, selective news sharing can distort learning even if it does *not* give rise to unbalanced news diets. Specifically, if the agent under-estimates news arrival ($\hat{\gamma} < \gamma$), her expected posterior is distorted towards the state she deems more likely ex ante. This is reminiscent of updating distortions usually called confirmatory bias (Rabin (1998)). Conversely, if the agent over-estimates news arrival ($\hat{\gamma} > \gamma$), her expected posterior is distorted towards the state she deems less likely ex ante.

Proposition 1. *Fix any agent with a balanced echo chamber.*

1. *If $\hat{\gamma} < \gamma$, then* $\left(\mathbb{E}[\mu(\mathbf{s}^1)] - \pi\right)\left(\pi - \frac{1}{2}\right) > 0.$
2. *If $\hat{\gamma} > \gamma$, 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 explicitly write the agent's posterior after one period. Denote by a_A the number of a-signals her A-dogmatic friends received and by b_B the number of b-signals her B-dogmatic friends received. From her perspective, a_A is distributed as a Binomial random variable with probability $\hat{\gamma}(1-q)$ and sample size d_A, whereas b_B is distributed as a Binomial random variable with probability $\hat{\gamma}q$ and sample size d_B. The agent also receives $n+1$ independent private signals: n from her normal friends plus her own signal. Among these signals, let a_N and b_N denote the number of a-signals and b-signals, which are multinomial random variables with probabilities $\hat{\gamma}(1-q)$ and $\hat{\gamma}q$ and sample size $n+1$. Note that (a_A, b_B, a_N, b_N) summarizes the agent's information $\mathbf{s}^1$. By Bayes's rule her posterior belief is[21]

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

where

$$\begin{aligned} Q &\equiv \frac{1-q}{q}, \\ M &\equiv a_A + a_N - (b_B + b_N), \\ \hat{\Gamma} &\equiv \frac{\hat{\gamma}(1-q) + (1-\hat{\gamma})}{\hat{\gamma}q + (1-\hat{\gamma})}, \end{aligned}$$

[21] This representation is derived in the proof of Proposition 2.

$$S \equiv (d_B - b_B) - (d_A - a_A).$$

We can understand this expression as follows. The term Q^M captures the agent's interpretation of the received signals, which is always correct: By verifiability of information, the act of sharing a signal leaves no uncertainty regarding whether the signal was actually received—hence, $\hat{\gamma}$ is irrelevant. The term $\hat{\Gamma}^S$ captures how the agent incorrectly interprets the silence of her dogmatic friends. She observes silence from $d_B - b_B$ B-dogmatic friends and from $d_A - a_A$ A-dogmatic friends. She attributes each instance of silence to an unfavorable signal for the friend with probability $\hat{\gamma}$ or to no signal with probability $1 - \hat{\gamma}$. Note that $\hat{\Gamma}$ is a decreasing function of $\hat{\gamma}$. Thus, a higher $\hat{\gamma}$ increases $\hat{\Gamma}^S$ if $S < 0$ and decreases $\hat{\Gamma}^S$ if $S > 0$, thereby distorting the posterior downward or upward depending on S. It is therefore not immediate that the *average* distortion goes in any specific direction. For instance, the agent's misperception could inflate or deflate updating, but have no effect on average.

The agent's prior resolves this ambiguity. To see why, suppose she deems state A as very unlikely ex ante (small π). Consider $\hat{\gamma} < \gamma$. Silence of A-dogmatic friends induces her to update the probability that $\omega = A$ *downward*, while silence of B-dogmatic friends induces her to update it *upward*. In both cases, the agent updates less than she should because she excessively attributes silence to lack of news. However, this under-reaction has asymmetric consequences for A- and B-dogmatic friends. When π is small, B-dogmatic friends are relatively less likely to receive an unfavorable signal and thus remain silent than A-dogmatic friends are. Put differently, $\pi < \frac{1}{2}$ magnifies the under-reaction to the silence of B-dogmatic friends relative to A-dogmatic friends, which distorts updating downward. Figure 1a illustrates Proposition 1.

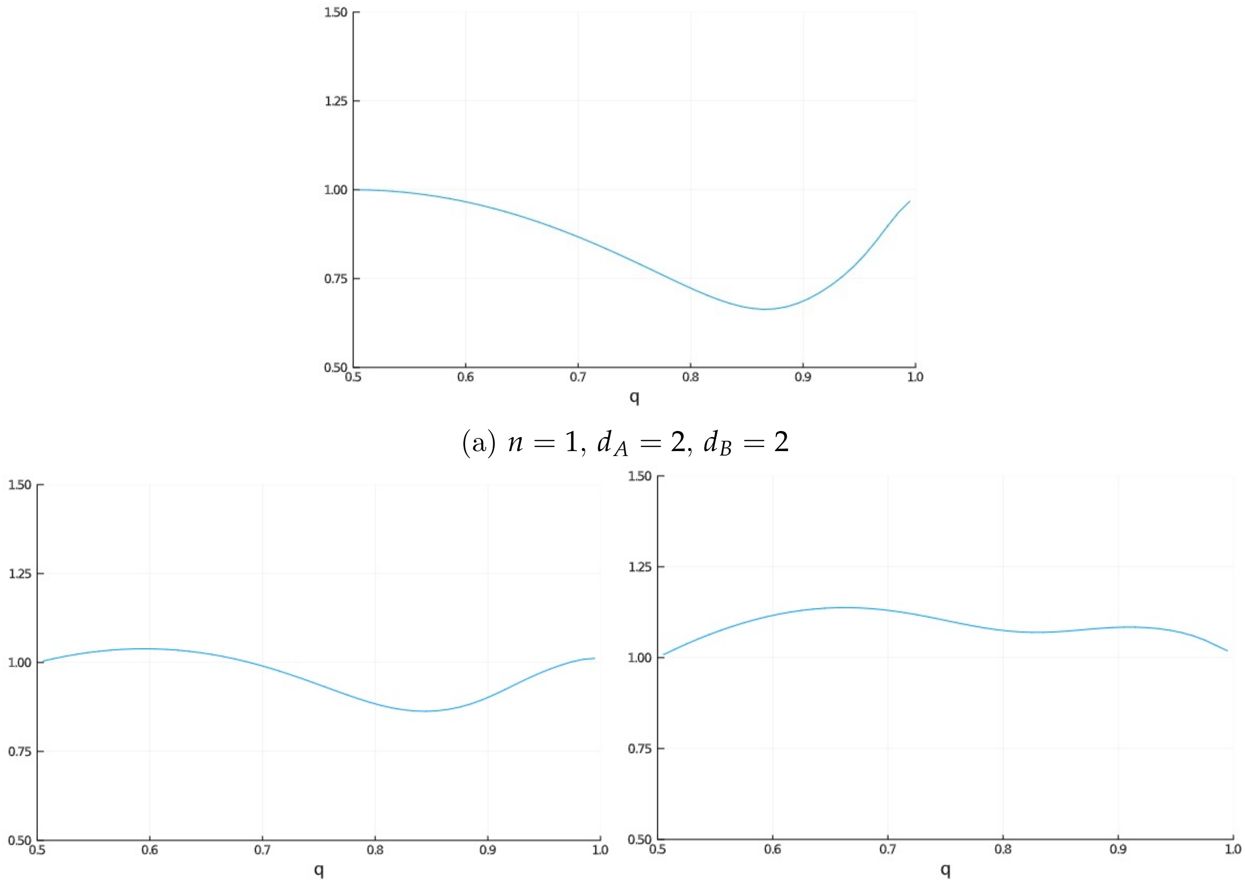
Our second result focuses on the effects of an imbalance in the agent's echo chamber. It states that if the agent under-estimates news arrival ($\hat{\gamma} < \gamma$), her expected posterior is distorted *towards* the conviction of her dogmatic majority. By contrast, if the agent over-estimates news arrival ($\hat{\gamma} > \gamma$), her expected posterior is distorted *against* her dogmatic majority. However, for the dogmatic majority to have such effects the information quality has to be sufficiently low.

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

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

Our proof actually shows that the expected posterior is distorted as stated also *conditional* on any true state of the world. Figures 1b and 1c illustrate Proposition 2.

Consider again how the agent updates (see (2)). If she has more A- than B-dogmatic friends, she will tend to receive more signals supporting state A than B. Yet, this does not imply that her posterior will be distorted towards A. To see why, it helps to consider extreme misperceptions. Suppose she severely under-estimates news arrival: $\gamma > \hat{\gamma} \approx 0$. She then interprets silence as almost certainly no news, rather than bad news for her dogmatic friends. Thus, she essentially ignores silence and updates based only on the shared signals, which tend to favor A. By contrast, suppose she severely over-estimates news arrival: $\gamma < \hat{\gamma} \approx 1$.



(a) $n = 1$, $d_A = 2$, $d_B = 2$

(b) $n = 1$, $d_A = 3$, $d_B = 2$

(c) $n = 1$, $d_A = 4$, $d_B = 2$

Figure 1: Graphs of the ratio between the expected posterior and the prior as a function of q, for different echo chambers (determined by the values of n, d_A and d_B). Other parameters are as follows: $\pi = 0.001$, $\gamma = 0.8$ and $\hat{\gamma} = 0.5$.

She then interprets silence as almost certainly bad news for her dogmatic friend, rather than no news. Thus, she reads too much into the silence of her dogmatic majority and incorrectly updates her belief *away* from their preferred state. Put differently, her dogmatic majority always drives her belief through selective sharing, but this can backfire and push her to believe that the state is B. The case of $\hat{\gamma} < \gamma$ seems more consistent with the common understanding of the effects of echo chambers. It is, however, interesting that these effects do not disappear when $\hat{\gamma} > \gamma$, but rather change direction.

Clearly, for such distortions to arise the information quality cannot be perfect (i.e., $q = 1$). Proposition 2 shows that for sufficiently low quality the imbalance between dogmatic friends—however small—always prevails over the information coming from normal friends and own signals. In some cases, it prevails for all $q \in \left(\frac{1}{2}, 1\right)$ (for an example, see Figure 1c).

Corollary 1. Fix any agent with an unbalanced echo chamber that satisfies $d_A > d_B$.

1. If $\hat{\gamma} < \gamma$ and $\pi > \frac{1}{2}$, then $\mathbb{E}[\mu(\mathbf{s}^1)] > \pi$ for all $q \in \left(\frac{1}{2}, 1\right)$.

2. If $\hat{\gamma} > \gamma$ and $\pi < \frac{1}{2}$, then $\mathbb{E}[\mu(\mathbf{s}^1)] < \pi$ for all $q \in \left(\frac{1}{2}, 1\right)$.

However, it is not true that an agent's echo chamber always distorts her learning towards her dogmatic majority, even if her under-reaction to silence favors that majority (for an example, see Figure 1b).

3.2 Long Run - Abundant Information

We showed that with one round of information echo chambers can systematically distort beliefs. One may expect these distortions to vanish when information becomes abundant (i.e., in the long run after many signals). In fact, the opposite can occur: Abundant information can exacerbate the effect of misperceived selective sharing and cause beliefs to be almost certainly incorrect, but only if information quality is sufficiently low. In this case, with probability 1 and irrespective of the true state, the agent's posterior converges to a degenerate belief on one state (denoted by δ_ω). This is the state favored by her dogmatic majority if $\hat{\gamma} < \gamma$ and by her dogmatic minority if $\hat{\gamma} > \gamma$.

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

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

2. *If $q < q_{LR}(e, \gamma, \hat{\gamma})$ and $\hat{\gamma} > \gamma$, then the agent's belief converges to δ_B with probability 1 (i.e., $\mu(\mathbf{s}^\infty) = 0$).*

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

It follows from the proof that for balanced echo chambers the agent's posterior always converges to δ_ω with probability 1, where ω is the true state. Thus, the distortion in Proposition 1 does not survive in the long run.

We can intuitively understand this result as the outcome of a non-trivial race between two kinds of information over time. The agent's first-hand information provides an increasingly accurate estimate of the state, which would result in perfect learning in a standard setting. The second-hand signals from her friends also provide more information, but are selected in ways she does not correctly take into account. It turns out that with low quality signals the distortion in each step of updating unveiled in Proposition 2 accumulates over time leading the posterior astray. Thus, only high-quality information eventually removes both the intrinsic distortions caused by selective sharing (Proposition 1 and Corollary 1) and the distortions caused by echo-chamber imbalance (Proposition 2 and Corollary 1). Another way to see the role of q is to reconsider the correct updating term Q^M and incorrect updating term $\hat{\Gamma}^S$ in formula (2). As q increases from $\frac{1}{2}$ (low informativeness) to 1 (high informativeness), Q^M falls from 1 to zero, thereby curtailing the misperception effect through $\hat{\Gamma}^S$. While in the short run this curtailment may be complete only at $q = 1$, in the long run it is always complete for a range of $q < 1$.

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

Proposition 4. *The threshold $q_{LR}(e, \gamma, \hat{\gamma})$ is strictly increasing in $|d_A - d_B|$ and $|\gamma - \hat{\gamma}|$ and decreasing in n.*

The threshold increases with the degree of echo-chamber imbalance and of misperception, as both strengthen the forces leading posteriors astray. The threshold decreases with the number of normal friends, as they provide more unfiltered information.

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

3.3 Making and Losing Friends

Advances in technology—such as the development 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 a normal agent makes or loses friends of any type, which changes the composition of her echo chamber. How does this affect the range of information qualities that result in incorrect learning?[22]

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, start from the first term in parentheses, which measures the net growth rate of the echo-chamber imbalance. If this is positive, then (3) requires normal friends to grow sufficiently *faster* in order to decrease q_{LR}. If instead more dogmatic connections reduce the echo-chamber imbalance, the number of normal friends can even *fall*—but not too much—without increasing q_{LR}. The second line of (3) takes into account what happens individually to the group of A- and B-dogmatic friends and hence to the flow of selected

[22] Appendix E provides a more general result that also covers the case of $\lambda_A d_A < \lambda_B d_B$.

signals the agent receives from each group. If the A-group grows more, then (3) requires an even larger growth of normal friends to decrease q_{LR}. If the B-group grows more, this partially compensates the change in the imbalance and hence requires a smaller growth of normal friends. One can show that q_{LR} always increases if scaling is proportional ($\lambda_A = \lambda_B = \lambda_N$). This could happen on social media, for instance, if how they suggest new connections is independent of what news people share. In short, an increase in the number of friends involves a trade-off between access to information and the scope for echo-chambers to distort beliefs.

We now ask a different question: Given the existing information quality, what changes in an agent's friends are sufficient to overcome her echo-chamber's power to distort beliefs and re-establish correct learning? Specifically, given any $\hat{q} < q_{LR}$ 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 the following holds:*

1. *for $\hat{\gamma} < \gamma$,*

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

2. *for $\hat{\gamma} > \gamma$,*

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

Propositions 5 and 6 may have several practical implications. For instance, the growth 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., before q_{LR} falls below $\hat{q}$). Alternatively, algorithms designed by social-media platforms often control how people form new connections. Knowing the effects of echo chambers' composition on people's learning can inform the design of such algorithms so as to limit the distortions of selective news sharing.

4 Belief Polarization in Society

We build on the previous results to examine belief polarization in society. We will treat the set of normal agents as our society of interest, which we denote by $\mathcal{N}$. We exclude dogmatic agents based on the interpretation that they have degenerate beliefs, which therefore do not respond to new information.

We begin by defining a measure of belief polarization. Polarization does not simply mean heterogeneous beliefs but rather the existence of groups with sharply different beliefs

(Esteban and Ray (1994)), which usually emerge over time. For this reason, we start by examining 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, every $\mathbf{e}$ gives rise to a distribution of long-run beliefs *across the agents* in $\mathcal{N}$, which is characterized by the agents who converge to having a degenerate belief on state $\omega \in \{A, B\}$: Let

$$\mathcal{N}_\omega(\mathbf{e}) = \{i \in \mathcal{N} : \mu(\mathbf{s}_i^\infty) = \delta_\omega\}.$$

We then define *long-run polarization* as[23]

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

Note that $\Pi(\mathbf{e})$ takes values in $[0, 1]$ and attains its maximum when $|\mathcal{N}_A(\mathbf{e})| = |\mathcal{N}_B(\mathbf{e})|$. Given the true state ω, we will call $\mathcal{N}_\omega(\mathbf{e})$ the set of "eventually correct" agents and $\mathcal{N}_{-\omega}(\mathbf{e})$ the set of "eventually incorrect" agents.

Our previous results imply that selective information sharing can cause beliefs to polarize in the long run if and only if it is combined with misperceptions. Indeed, $\Pi(\mathbf{e}) = 0$ if there are no misperceptions (Remark 1), or if all echo chambers are balanced. Otherwise, Proposition 3 implies the following.

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 $\Pi(\mathbf{e}) > 0$.

This formalizes the common narrative that, if some agents in society have echo chambers that skew their news diets in opposite directions—where the imbalances can be small—then their beliefs can polarize.[24] However, our results qualify this narrative: Belief polarization requires sufficiently low quality of information and some misperception of the effects of echo chambers, but does not require fake news nor that people look at the world in fundamentally incompatible ways. Our results also show that oppositely unbalanced echo chambers are not necessary for polarization to arise. By Propositions 3 and 4, if we take any society $\mathcal{N}$ such that $d_{Ai} \geq d_{Bi}$ 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 this society, if the information quality is extremely low or high, long-run polarization is zero because either everybody is eventually incorrect or everybody is eventually correct. But for intermediate information

[23] Note that by standard continuity arguments

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

[24] Rich evidence shows that 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)).

quality some agents will be eventually correct despite their unbalanced 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, this need not be true. The following result provides a necessary and sufficient condition for polarization to be non-monotonic in q.[25] To this end, let $\mathcal{D}_\omega$ be the set of agents who will be eventually incorrect if their long-run belief agrees with their ω-dogmatic friends, but the true state is *not* ω:

$$\mathcal{D}_A = \{i \in \mathcal{N} : (d_{Ai} - d_{Bi})(\gamma - \hat{\gamma}) > 0\}$$

and $\mathcal{D}_B$ is defined similarly by swapping d_{Ai} and d_{Bi} (Proposition 3).

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, $\Pi(\mathbf{e})$ is decreasing in q over $\left(\frac{1}{2}, 1\right)$ if and only if $|\mathcal{D}_{-\omega}| \leq \frac{1}{2}\left(|\mathcal{N}| + 1\right)$. Otherwise, $\Pi(\mathbf{e})$ is single peaked.*

To see the intuition, it helps to consider how $\mathcal{N}_B(\mathbf{e})$ and $\mathcal{N}_A(\mathbf{e})$ change as q increases. Assume the true state is B and $\hat{\gamma} < \gamma$. Fix some $q \in \left(\frac{1}{2}, 1\right)$. Then, $\mathcal{N}_B(\mathbf{e})$ contains the agents for whom (1) $q > q_{LR}$ and so learn correctly, *or* (2) $d_B > d_A$ and $q < q_{LR}$ and so have $\mu(\mathbf{s}^\infty) = 0$ irrespective of the true state; $\mathcal{N}_A(\mathbf{e})$ contains all agents for whom $d_A > d_B$ and $q < q_{LR}$. As q increases, all agents in $\mathcal{N}_B(\mathbf{e})$ will remain there: For agents in group (1) nothing changes; agents in group (2) may stay there or pass to group (1). By contrast, agents in $\mathcal{N}_A(\mathbf{e})$ will switch to $\mathcal{N}_B(\mathbf{e})$ one by one:[26] When q becomes larger than q_{LR} for an agent in $\mathcal{N}_A(\mathbf{e})$, her echo chamber no longer distorts her long-run belief, which now converges to δ_B. Thus, if the eventually incorrect agents outnumber the eventually correct agents initially (i.e., for $q \approx \frac{1}{2}$), then as q increases it will cause a gradual migration into the set of eventually correct agents and polarization will initially *increase* and then decrease towards zero.

To recap, our results suggest that increasing the quality of first-hand information for the agents can be a way to counteract the power of echo chambers to polarize beliefs. However, such quality increases may need to be significant to actually curb polarization.

Shifting perspective, one may wonder how changes in the distribution of echo chambers in society affects the distribution of long-run beliefs and hence polarization. Propositions 5 and 6 show that the answer is not obvious—even in our stylized model—as both changes in the echo-chamber imbalances and changes in the number of normal friends matter. Fix any q and $\mathbf{e}$ such that $\Pi(\mathbf{e}) > 0$. Suppose the agents make and lose friends, which results in $\mathbf{e}'$. By Proposition 6, if for each agent her normal friends grow sufficiently more than her echo-chamber imbalance, then $\Pi(\mathbf{e}') = 0$. Thus, technology advances that expand the agents'

[25] The same result holds for other, 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 $\alpha^* \approx 1.6$. One can show that this form is single peaked in ν, which is key for the non-monotonicity in q.

[26] 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.

echo chambers can curb belief polarization. But the opposite can also happen: An agent can learn correctly in a small echo chamber, but incorrectly after her echo chamber expands, which can cause her belief to polarize from others. Our results highlight that what matters is the composition of echo chambers, not their absolute size. These observations may offer a new perspective on the evidence showing that polarization seems to be more pronounced for demographic groups that are least likley to use the Internet and social media (Zhuravskaya et al., 2020). Their echo chambers may be smaller, but also more unbalanced.

Our analysis also suggests that polarization in echo chambers need not lead to polarization in beliefs. Imagine two societies characterized by $\mathbf{e}$ and $\mathbf{e}'$, where each distribution is evenly divided in terms of echo-chamber imbalances (i.e., $|\mathcal{D}_A| = |\mathcal{D}_B| = \frac{1}{2}|\mathcal{N}|$). The only difference is that, for all agents, $\mathbf{e}$ involves small imbalances and $\mathbf{e}'$ large imbalances. In terms of echo chambers, we may view $\mathbf{e}$ as *less* polarized than $\mathbf{e}'$. Yet, we can have $\Pi(\mathbf{e}) > \Pi(\mathbf{e}')$. This may be counterintuitive, but becomes clear once we take into account the role of information quality that we highlight, which may be high in the society with $\mathbf{e}'$ and low in the society with $\mathbf{e}$. Importantly, our results provide tools to handle this complexity and predict what happens based on the observable characteristics of a society summarized by $\mathbf{e}$. Such predictions can also guide policy interventions.

Finally, our theory also offers insights about 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. Removing redundancies, assume $e_i \neq e_j$ if $i \neq j$. We can summarize the beliefs within each group with their empirical average and then use these summary statistics to quantify intra-group polarization in society. Importantly, if group i is large enough, its empirical average belief in the short run 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

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

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.[27]

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 belief polarization in the short run. In contrast to the long run, where this requires low information quality, for the short run polarization can arise even for high information quality (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.

[27] The same result holds for the general measures of polarization axiomatized by Esteban and Ray (1994).

5 Mitigating Polarization

Ferejohn et al. (2020) note that challenges to shaping the character of democratic institutions include "managing the development of media and information technologies to ensure they enhance, rather than degrade, robust pluralism and civil political engagement." We take a step in that direction in this section.

How could a social planner address polarization generated by shared news? Selective sharing and misperceptions seem hard to influence, as they belong to each individual's private life and personal freedom. It may instead be easier to influence people's echo chambers and, in particular, the quality of their first-hand information. With regard to echo chambers, Section 3.3 described how influencing the rate at which people connect with friends on social-media platforms may help avoid incorrect learning and hence belief polarization.

Acting on the quality of information seems the least intrusive intervention. One obvious way is to *directly* increase q at the source. This may be difficult, however, due to technological or economic reasons. For instance, it may involve incentivizing or forcing newspapers to spend more on reporters, data gathering, and fact checking. Other ways may still exist to increase the quality of information that people ultimately receive without changing q of the primitive signals s_{it}.

The last decades have witnessed the expansion of so-called *news aggregators*, namely, online platforms that summarize the news for their users. Examples include *The Drudge Report*, *Apple News*, or *Yahoo! News*. This may have several explanations: Aggregators may help people handle the overload of daily news given time or attention constraints, or may help pool news from different sources into one convenient access point. By filtering and summarizing news, aggregators throw away some information relative to the totality of the aggregated signals. Nonetheless, the resulting output can have higher information quality than each aggregated signal *individually*, which is the key observation for our purposes. Through the lens of our theory news aggregators can serve another function, which is to curb polarization by undermining the distortions of selective news sharing.[28]

There are many ways to aggregate signals. To make our point we consider the following simple form, which has the advantage that one can easily provide a sufficient degree of aggregation to ensure no polarization. Divide time into blocks of M periods, where M is an odd number. For every agent i and $t = 1, 2, \ldots$, define $\hat{s}^i_{Mt}$ as a new signal that is released to i at the end of each time block 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,

[28] 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.

$\hat{s}^i_{Mt}$ has higher quality than each s_{it}. To see this, suppose $M = 3$ and $\omega = A$. We have

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

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.

The remaining question is how much aggregation is enough to curb polarization. The next proposition gives an answer in terms of a sufficient finite number M of aggregated periods.[29] Let $\hat{\Pi}$ be the limit polarization when signals s_{it} are replaced with $\hat{s}^i_{Mt}$.

Proposition 8. *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, $\hat{\Pi}(\mathbf{e})$ equals zero if*

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

A few remarks are in order. Note that $\hat{s}^i_{Mt}$ essentially reports whether the sample average of a signals is above $\frac{1}{2}$ or not. By the Law of Large Numbers, as $M \to \infty$ that average is $q > \frac{1}{2}$ if $\omega = A$ and $1 - q < \frac{1}{2}$ if $\omega = B$ with probability 1. In other words, with infinite aggregation, $\hat{s}^i_{Mt}$ can learn the state and then report it to the agents. Clearly, this implies correct learning, but is not how news aggregators work in reality. We can still conclude that partial news aggregation can help curb polarization, because we showed that undoing the effects of misperceived selective sharing in the long run does not require perfect information quality.

Another observation is that our aggregators summarize the primitive signals for each individual, but have a common degree of aggregation M. Since incorrect learning is caused by news sharing, how much we aggregate agent i's signals has to take into account the information quality required for her friends to learn correctly. This is why our finite threshold for M is in terms of the quality $\bar{q}_{LR}$ of the agents for whom the effects of misperceived selective sharing are the hardest to overcome. This is where Proposition 4 can guide how to adjust M as echo chambers change. Also, if we knew that a subgroup of agents shares signals only among themselves—essentially forming a sub-society $\mathcal{N}' \subset \mathcal{N}$—it would be possible to pick a lower M tailored to this group. These points suggest that if agents choose degrees of news aggregation for themselves based on their individual reasons, they may not internalize the effects of the news they then share and create too little aggregation from society's viewpoint. This may call for institutional intermediaries or platforms that aggregate news taking into account these externalities.

[29] The threshold in Proposition 8 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.

6 Extensions: Other Misperceptions

We now consider other ways in which agents may misperceive information. This clarifies the main mechanism through which selective news sharing can lead to polarization: the combination of unbalanced echo chambers and incorrect interpretation of friends' silence.

Throughout this section, the true properties of first-hand information as well as timing remain as in the baseline model. We now assume that all agents correctly assign probability γ to the arrival of first-hand information in each period (i.e., $\hat{\gamma} = \gamma$) in order to isolate the effects of other misperceptions. We consider three:

(I) *Agents misperceive the probabilities with which friends shares signals.* To model this, we allow for probabilistic selective sharing. Normal agents share any first-hand signal s_{it} with probability $\nu \in (0, 1]$ and stay silent with probability $1 - \nu$. An 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. B-dogmatic agents are like A-dogmatic agents, except for swapping probabilities of sharing a and b signals. Note that our baseline model corresponds to $\nu = g = 1$ and $f = 0$. We continue to assume that all agents of a specific type are the same. Misperception (I) means that each agent knows all her friends' types, but 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) *Agents misclassify some of their friends' types.* With three types, there are in principle many possible misclassifications. For conciseness, we consider the case where dogmatic friends are misclassified as normal. The sharing behavior is deterministic as in the baseline model ($\nu = g = 1$ and $f = 0$). Let $\hat{n}_A$ be the number of A-dogmatic friends that an agent misclassifies as normal; define $\hat{n}_B$ similarly. That is, the agent treats these friends as always sharing any signal they receive, while in reality they share only signals favorable to one state.

(III) *Agents misperceive the quality of first-hand information.* Each agent thinks that the probability with which a signal matches the state is $\hat{q} \in \left(\frac{1}{2}, 1\right)$ instead of the true q given in equation (1). Note that this misperception differs conceptually from all other misperceptions considered in this paper, which are about how friends share news.

Proposition 9. *Each of misperceptions (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 misperception (see Online Appendix A for a detailed analysis). 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 the differences between these misperceptions, they all cause incorrect learning and polarization through the same fundamental mechanism as in the baseline model. That is, the agents interpret silence incorrectly by misunderstanding how much of it depends on lack rather than suppression of information. This is also the only mechanism through which misperceptions of information quality cause polarization: If $\gamma = 1$ and selective sharing unravels, the agents always learn correctly in the long run despite $\hat{q} \neq q$. For this misperception to cause polarization the agents must over-estimate the information quality, that is, $\hat{q} > q$.

One may interpret the case of $\hat{q} > q$ as related to the idea of "fake news:" Such news are false or very uninformative (low q), yet people mistakenly take them as reliable and informative (high $\hat{q}$). Our results then suggest that this form of 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 may have become especially powerful in the age of social media. This may provide a rationale for fact-checking as a way to realign $\hat{q}$ with q.

Finally, another takeaway in common with the baseline model is the role of low and high quality in enabling and preventing polarization, respectively. This further supports our insights about the ability to mitigate polarization by aggregating news.

7 Concluding Remarks

We studied if and when learning from shared news can lead to belief polarization. Our positive answer is consistent with some common narratives about news sharing, yet highlights several qualifications. Selective sharing alone does not lead to polarization, even if it gives rise to unbalanced news diets. It has to be combined with some misperception that causes people to misinterpret when others do *not* share information. Moreover, this key mechanism leads to polarization if (and only if) the quality of first-hand information is sufficiently low. Our insights about the importance of information quality (in contrast to quantity) and echo-chamber imbalances shed light on how policies that aim to improve news quality or diversify people's diet of shared news can curb or inflate polarization. We hope this advances our understanding of some of the mechanisms behind polarization in modern societies.

Our analysis goes to the heart of why new communication technologies and formats enabled by the Internet may contribute to 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 and imprecise, and overwhelmed by the information abundance, people may spread their limited attention across more sources and hence absorb less content from each. Third, the Internet has offered bad actors a megaphone to spread fake news, and we found that it is the selective sharing of fake news—not fake news per se—that can distort beliefs. However, our results suggest that even in the absence of fabricated news, specific aspects of how people share and process information online would still cause polarization. To the extent that its causes are legitimate behaviors within people's rights to free speech and self-determination, our analysis offers a new perspective on whether the Internet—and in particular social media—may be held accountable for polarization and what regulations may address it.

While we focused on mitigating polarization, our theory also sheds light on how malevolent actors can leverage news sharing among people and misperceptions to exacerbate polarization. Obvious ways include using fake news to directly lower information quality or expanding echo chambers' imbalances. A more subtle way is to release bits of true but low-quality news with high frequency (like Tweets) so as to leverage the power of misperceived selective sharing. This may also serve to draw attention away from high-quality information sources. A better understanding of what malevolent actors may try to do could offer guidelines for preemptive countermeasures.

Several directions remain for future research. We briefly mention two. We took selective news-sharing behavior as given and fixed, modeling its key aspects found in the empirical evidence. In reality, people choose what to share strategically—for example, to persuade friends to take an action. As long as it involves suppressing specific information, our insights about its consequences should be valid and may in turn be a steppingstone to understanding what drives selective sharing in the first place.

Finally, the role of unbalanced echo chambers in our analysis begs the question of what happens if we allow social links to form endogenously. In this process, people may follow their demand for information or other socio-economic forces (identity, class, race, ideology, work career, etc.).[30] On the one hand, they may tend to link with like-minded friends, which may create a vicious cycle where belief polarization and echo chambers' imbalances reinforce each other. On the other hand, they may be more likely to link with reliable sources of objective information, which would have opposite implications. Which tendency prevails is ultimately an empirical question. We hope our framework can guide further theoretical and empirical investigations of the relation between evolving social relations and polarization.

Renee Bowen, UC San Diego and NBER
Danil Dmitriev, UC San Diego
Simone Galperti, UC San Diego

[30] Recent studies on homophily in social networks include, for example, Golub and Jackson (2012), Baccara and Yariv (2013), and Halberstam and Knight (2016).

Appendix

A Proof of Proposition 1

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

Without loss of generality, we can ignore the normal friends and assume that $n = 0$. 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 the agent is not misspecified with respect to them, 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

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

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

$$\begin{aligned}
\mathbb{E}[\mu-\mu^*] \;\; &= \pi \textstyle\sum_{a_A,b_B} \left(\tfrac{d!}{a_A!(d-a_A)!}\cdot\tfrac{d!}{b_B!(d-b_B)!}\gamma^{a_A+b_B}q^{a_A+b_B}(\gamma(1-q)+(1-\gamma))^{2d-a_A-b_B}\right) \\
&\qquad \times \left(\tfrac{1-q}{q}\right)^{b_B}\left(\tfrac{\gamma(1-q)+(1-\gamma)}{\gamma q+(1-\gamma)}\right)^{b_B-d}\left(\tfrac{1+\rho Q^{a_A-b_B}\Gamma^{a_A-b_B}}{1+\rho Q^{a_A-b_B}\hat{\Gamma}^{a_A-b_B}}-1\right) \\
&= \pi \textstyle\sum_{a_A,b_B} \left(\tfrac{d!}{a_A!(d-a_A)!}\cdot\tfrac{d!}{b_B!(d-b_B)!}\gamma^{a_A+b_B}q^{a_A+b_B}(\gamma(1-q)+(1-\gamma))^{2d-a_A-b_B}\right) \\
&\qquad \times\Gamma^{-d}\left(\tfrac{Q^{b_B}\Gamma^{b_B}+\rho Q^{a_A}\Gamma^{a_A}}{Q^{b_B}\hat{\Gamma}^{b_B}+\rho Q^{a_A}\hat{\Gamma}^{a_A}}Q^{b_B}\hat{\Gamma}^{b_B}-Q^{b_B}\Gamma^{b_B}\right) \\
&= \pi\Gamma^{-d}\textstyle\sum_{0\le x\le y\le d}\left(\tfrac{d!d!}{x!(d-x)!y!(d-y)!}\gamma^{x+y}q^{x+y}(\gamma(1-q)+(1-\gamma))^{2d-x-y}\right) \qquad (6)\\
&\qquad \times\left(\tfrac{Q^y\Gamma^y+\rho Q^x\Gamma^x}{Q^y\hat{\Gamma}^y+\rho Q^x\hat{\Gamma}^x}Q^y\hat{\Gamma}^y-Q^y\Gamma^y+\tfrac{Q^x\Gamma^x+\rho Q^y\Gamma^y}{Q^x\hat{\Gamma}^x+\rho Q^y\hat{\Gamma}^y}Q^x\hat{\Gamma}^x-Q^x\Gamma^x\right).
\end{aligned}$$

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 $y > x$.

Suppose $\hat{\gamma} < \gamma$. 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.

If $\hat{\gamma} > \gamma$ holds, than the argument above applies in a symmetric way with all inequalities flipping after dividing by $\Gamma^x\hat{\Gamma}^y-\Gamma^y\hat{\Gamma}^x$, which is negative. It will follow that moving the prior from 50-50 towards a state will make the unconditional expected posterior of that state *lower* than the prior.

B Proof of Proposition 2

We will prove that there exists $q_{SR} > \frac{1}{2}$ such that, if $q \in \left(\frac{1}{2}, q_{SR}\right)$, then $\mathbb{E}\left[\mu|\omega\right] > \pi$ or $\mathbb{E}\left[\mu|\omega\right] < \pi$ for any ω depending on the signs of $d_A - d_B$ and $\gamma - \hat{\gamma}$. To this end, we will first find the derivative of $\mathbb{E}\left[\mu|\omega\right]$ with respect to q at $q = \frac{1}{2}$ and then show how its sign depends on $d_A - d_B$ and $\gamma - \hat{\gamma}$. Using continuity of $\mathbb{E}\left[\mu|\omega\right]$ in q and the fact that $\mathbb{E}\left[\mu|\omega\right] = \pi$ 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 $\mathbb{E}\left[\mu|\omega, x_A, x_B, x_N\right]$ 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

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

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

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

We now find $\frac{\partial}{\partial q}\mathbb{E}\left[\mu|\omega, x_A, x_B, x_N\right]$ 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.[31]

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

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

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

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. ■

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* $q = \frac{1}{2}$,

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

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

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

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

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

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

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

■

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

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

Thus, if $d_A > d_B$ and $\gamma > \hat{\gamma}$, then the derivative is positive, which means that $\mathbb{E}[\mu|\omega]$ is distorted towards A for low q. If instead $d_A > d_B$ and $\hat{\gamma} > \gamma$, then the derivative is negative, which means that $\mathbb{E}[\mu|\omega]$ is distorted towards B for low q.

C Proof of Proposition 3

Recall that agent i has d_A A-dogmatic, d_B B-dogmatic, and n normal friends and that $d_A > d_B$.[32]

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

- 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

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

[32] 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.

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

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

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

$$H(q;\omega) > \text{(resp. } <\text{)}\;\; \tau(q) = \frac{1}{2} + \frac{((2-\gamma)z(q,\hat{\gamma}) - \gamma)\,(d_A - d_B)}{2\gamma\,(2(1+n) + (d_A + d_B)(1 + z(q,\hat{\gamma})))}. \qquad (8)$$

Note that

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

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

In Online Appendix B, 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. Suppose $\omega=B$ and hence $H(q;B)=1-q$. If $\hat{\gamma}<\gamma$, condition (8) 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 $q>q_{LR}$.[33] If $\hat{\gamma}>\gamma$, condition (8) 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 $\text{plim}_{T\to\infty}\mu(\mathbf{s}^T)=0$ for all $q\in\left(\frac{1}{2},1\right)$.

Now suppose $\omega=A$ and hence $H(q;A)=q$. If $\hat{\gamma}<\gamma$, condition (8) 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 $q\in\left(\frac{1}{2},1\right)$. If $\hat{\gamma}>\gamma$, condition (8) holds with "$<$" at $q=\frac{1}{2}$ and with "$>$" at $q=1$. By the properties of $\tau(q)$, there exists a unique $q_{LR}\in\left(\frac{1}{2},1\right)$ such that $\text{plim}_{T\to\infty}\mu(\mathbf{s}^T)=0$ if $q<q_{LR}$ and $\text{plim}_{T\to\infty}\mu(\mathbf{s}^T)=1$ if $q>q_{LR}$.

D 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 $\omega=B$ and $\hat{\gamma}<\gamma$. 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

$$q-\frac{1}{2}=\frac{\left(\gamma-(2-\gamma)z(q,\hat{\gamma})\right)(d_A-d_B)}{2\gamma\left(2(1+n)+(d_A+d_B)(1+z(q,\hat{\gamma}))\right)} \tag{9}$$

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}. 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.

Now consider the case of $\omega=A$ and $\hat{\gamma}>\gamma$. The value of q_{LR} is determined by the equation

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

[33] 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.

We can rewrite it as

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

By similar arguments, the right-hand side of the former condition is increasing in d_A; the right-hand side of the latter condition is strictly decreasing in n, d_B, and γ and increasing in $\hat{\gamma}$. The properties of q_{LR} follow similarly.

E Complete 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 10. *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 is (9) or (10) depending on which state results in incorrect learning.

Case 1: $\hat{\gamma} < \gamma$. 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:[34]

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

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

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

[34] 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.

Now 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:

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

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

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

Case 2: $\hat{\gamma} > \gamma$. Suppose $\lambda_A d_A - \lambda_B d_B > 0$. Then incorrect learning can occur in state A under both the original and the new echo chamber. Then, $q_{LR}(e,\gamma,\hat{\gamma}) < q_{LR}(e',\gamma,\hat{\gamma})$ if the following holds:

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

For $\hat{\gamma} > \gamma$, the difference $(2-\gamma)z(q,\hat{\gamma})-\gamma$ is positive when q is close to $\frac{1}{2}$. It also must be positive to have $q_{LR} > \frac{1}{2}$. Thus, we can divide both sides by it and simplify in the same way as above, obtaining the sufficient condition

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

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

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

Following the same steps as before, we obtain the sufficient condition

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

■

F Proof of Proposition 6

Case 1: $\omega = B$ and $\hat{\gamma} < \gamma$. In this case, q_{LR} is defined by condition (9). 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:

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

Case 2: $\omega = A$ and $\hat{\gamma} > \gamma$. In this case, q_{LR} is defined by condition (10). Fixing again d_A, d_B, n, λ and $\hat{q}$, we need to find λ_N such that

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

Since the right-hand side is increasing in $z(q,\hat{\gamma})$, we obtain a sufficient condition by imposing the inequality for the highest value of $z(q,\hat{\gamma})$, which is $\frac{\hat{\gamma}}{2-\hat{\gamma}}$. Rearranging gives the following:

$$\lambda_N > \frac{\left(\frac{\hat{\gamma}}{\gamma} - 2\hat{q}\right)d_A - \left(\frac{\hat{\gamma}}{\gamma} - 2(1-\hat{q})\right)d_B}{(2\hat{q}-1)n(2-\hat{\gamma})}\lambda - \frac{1}{n}$$

G 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

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

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, $|\mathcal{D}_B| \leq \frac{1}{2}(|\mathcal{N}| + 1)$.

H Proof of Proposition 8

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

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

Therefore, it suffices that

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

Online Appendix: Additional Proofs

(For Online Publication Only)

A Other Misperceptions

In this appendix, we state and prove the formal results about long-run learning under each of misperceptions considered in Section 6. Together, they imply Proposition 9.

A.1 Misperception (I): Random Selective Sharing

Proposition 11. *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 $\operatorname{\mathbf{plim}}_{T\to\infty}\mu(\mathbf{s}^T)=1$ (resp. $\operatorname{\mathbf{plim}}_{T\to\infty}\mu(\mathbf{s}^T)=0$) if and only if $\operatorname{\mathbf{plim}}_{T\to\infty}\frac{K(\mathbf{x},T;q,\gamma,\mathbf{p},\hat{\mathbf{p}})}{T}>0$ (resp. <0), where

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

The required inequality is then

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

which is equivalent to

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

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

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

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

It can be shown that

$$\lim_{q\to1}z(q,\gamma,\hat{\mathbf{p}})=0\qquad\text{and}\qquad\lim_{q\to\frac{1}{2}}z(q,\gamma,\hat{\mathbf{p}})=\frac{\frac{\gamma}{2}(\hat{g}-\hat{f})}{(1-\gamma)+\frac{\gamma}{2}(2-\hat{g}-\hat{f})},$$

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

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

and at $q=1$ it becomes

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

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 $\frac{1}{2}<q'\le q''<1$, $\operatorname{\mathbf{plim}}_{T\to\infty}\mu(\mathbf{s}^T)=1$ if $q<q'$, and $\operatorname{\mathbf{plim}}_{T\to\infty}\mu(\mathbf{s}^T)=0$ 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 $\frac{1}{2} < q' \leq q'' < 1$, $\mathbf{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 0$ if $q < q'$, and $\mathbf{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 1$ if $q > q''$.

■

A.2 Misperception (II): Friends' Types

Proposition 12. *Fix any agent with echo chamber $e = (d_A, d_B, n)$ and misperceived 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 perceived 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 $\mathbf{plim}_{T\to\infty}\mu(s^T) = 1$ (resp. $\mathbf{plim}_{T\to\infty}\mu(s^T) = 0$) if and only if $\mathbf{plim}_{T\to\infty}\frac{K(\mathbf{x},T;q,\gamma)}{T} > 0$ (resp. < 0), where

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

The required inequality is then

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

which is equivalent to

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

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 $\mathbf{plim}_{T\to\infty}\mu(\mathbf{s}^T) = I_{\{\omega=A\}}$.

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

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

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

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

and at $q = 1$ it becomes

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

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 $\frac{1}{2} < q' \leq q'' < 1$, $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 1$ 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

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

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 $\frac{1}{2} < q' \leq q'' < 1$, $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 0$ if $q < q'$ and $\text{plim}_{T\to\infty} \mu(\mathbf{s}^T) = 1$ if $q > q''$.

■

A.3 Misperception (III): Information Quality

Proposition 13. *Fix any agent with echo chamber $e = (d_A, d_B, n)$ and any perceived 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 $\text{plim}_{T\to\infty}\,\mu(\mathbf{s}^T)=1$ (resp. $\text{plim}_{T\to\infty}\,\mu(\mathbf{s}^T)=0$) if and only if $\text{plim}_{T\to\infty}\frac{K(\mathbf{x},T;q,\hat{q},\gamma)}{T}>0$ (resp. <0), where

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

The required inequality is then

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

which is equivalent to

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

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 $\text{plim}_{T\to\infty}\,\mu(\mathbf{s}^T)=I_{\{\omega=A\}}$.

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

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

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

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

and at $q=1$, it takes form

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

As shown in the Online Appendix B, $z(\hat{q},\gamma)$ 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 $\frac{1}{2}<q'\le q''<1$, $\text{plim}_{T\to\infty}\,\mu(\mathbf{s}^T)=1$ 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 then is

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

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

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

and at $q = 1$, it takes form

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

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

■

B Properties of $\tau(q)$

We will prove that $\tau(q)$ in condition (8) 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:

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

Note that $\lim_{q \to \frac{1}{2}} z\left(q, \hat{\gamma}\right) = \frac{\hat{\gamma}}{2-\hat{\gamma}} > 0 = z(1, \hat{\gamma})$. 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 $\lim_{q \to \frac{1}{1}} g'\left(q\right) = 0$ because

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

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

Proof. Since

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

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

$$z_{qq}(q,\hat{\gamma}) = K(q)\Bigg[\left(-\tfrac{(1-\hat{\gamma})(1-2q)}{q^2(1-q)^2} + \left(\tfrac{1-\hat{\gamma}}{q(1-q)} + \hat{\gamma}^2\right) z_q(q,\hat{\gamma})\right) \ln\left(\tfrac{1-q}{q}\right) \left(\hat{\gamma}^2 q(1-q) + (1-\hat{\gamma})\right)$$

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

$$= K(q)\Bigg[\left(\tfrac{(1-\hat{\gamma})(2q-1)}{q^2(1-q)^2} + \left(\tfrac{1-\hat{\gamma}}{q(1-q)} + \hat{\gamma}^2\right) z_q(q,\hat{\gamma})\right) \ln\left(\tfrac{1-q}{q}\right) \left(\hat{\gamma}^2 q(1-q) + (1-\hat{\gamma})\right) +$$

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

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

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

we can write

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

Let

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

and

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

Then we have

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

Note that

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

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 $q \in \left(\frac{1}{2}, 1\right)$.

■

References

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. 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.

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

Azzimonti, M. and M. Fernandes (2018). Social Media Networks, Fake News, and Polarization. *Working paper.*

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.

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.

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).

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

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 Directiosn for Higher Education 17, 1–15.

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

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.

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

Enke, B., F. Zimmermann, and F. Schwerter (2019). Associative Memory and Belief Formation. *Working paper.*

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.

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.

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., 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.

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.

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.

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.

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

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

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

Levy, G. and R. Razin (2019a). Echo Chambers and Their Effects on Economic and Political Outcomes. Annual Review of Economics forthcoming.

Levy, G. and R. Razin (2019b). Information Diffusion in Networks with the Bayesian Peer Influence Heuristic. *Working paper.*

Levy, R. (2020). Social Media, News Consumption and Polarization: Evidence from a Field Experiment. *Working paper.*

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.

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. TahbazSalehi, and A. Jadbabaie (2018). A Theory of NonBayesian Social Learning. Econometrica 86(2), 445–490.

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.

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

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.

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.

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

Perego, J. and S. Yuksel (2018). Media Competition and Social Disagreement. *Working paper.*

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

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

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

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.

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.

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. (2019). Behavioral Implications of Causal Misperceptions. *Working paper.*

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.