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

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

September 30, 2020

Abstract

We study social learning via news sharing. Each period agents receive the same quantity and quality of first-hand information and have the opportunity to share it with friends. Some agents (possibly a few) share information selectively. Selective sharing generates heterogeneous news diets across agents, who, however, are aware of it and update beliefs via Bayes' rule. We show that, contrary to standard learning results, agents' beliefs can diverge in this environment. This occurs if and only if agents hold misperceptions (even minor) about friends' access to first-hand information and if its *quality* is low. We show that abundant information can exacerbate belief polarization. That is, when the *quantity* of first-hand information grows indefinitely agents can hold opposite degenerate beliefs. Intuitively, polarization worsens with misperception and imbalance of news diets. Polarization can also worsen when information quality rises or when the agents' social networks expand, despite providing them with more information. Information aggregation can mitigate, and even eliminate, polarization.

*We thank our discussant, Myles Ellis, Joel Sobel, Nageeb Ali and participants in the UCSD Theory and Behavioral, PhDEI, and PennTheon workshops for helpful comments and suggestions

[†]UC San Diego and NBER, trbowen@ucsd.edu

[‡]UC San Diego, ddmitrie@ucsd.edu

[§]UC San Diego, sgalperti@ucsd.edu

1 Introduction

Social divisions have been linked to negative economic and political outcomes, including 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, particularly in the United States.[2] As Alesina et al. (2020) notes, "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." Economists have, thus, been investigating the determinants of belief polarization and we contribute to this line of research.

Several authors have made a connection between rising polarization of the American public and growing use of the Internet as a source of information (Periser, 2011; Sunstein, 2017; Azzimonti and Fernandes, 2018; Tucker et al., 2019). Some have highlighted the effects of systematic *misinformation*—fake news, bots, and bad actors—leading to discussions about regulating social media to minimize these elements.[3] But, even if successful, will such regulation solve the problem of polarization? Can polarization simply result from how people consume and share information in social networks, even without misinformation? Does the information abundance brought by such networks lead to more or less polarization? This paper provides a theoretical framework to answer these questions.

Building on recent evidence on how people use information on social networks, we analyze how they individually learn from first-hand and shared news and whether this can cause their beliefs to become polarized. We highlight that polarization depends on people's information diet, which is driven by the composition of friends in their network. But this is not all: The quality of information and misperceptions about how it is shared also play a key role. Our analysis suggests mechanisms whereby changes in people's information ecosystem brought by the Internet and the expansion of social networks may contribute to polarization. In addition, *low* quality of external information is crucial in generating belief polarization. Despite selective sharing relying on the same quality, it is capable to overcome correct learning only when people misperceive it and that information quality falls below a certain threshold. We also discuss implications for policies aimed at curbing polarization.

We base our theory on evidence highlighting specific ways in which information flows on social networks. First, people tend to share information *selectively*.[4] For example, they share only information that favors their preferred political candidate or views on the importance of vaccinating children. Second, although most people have friends who share different kinds of information, the resulting information diet is likely to feature some *imbalance*. This is a distinctive aspect of so-called echo chambers or media bubbles, which appear in a wealth of

[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).

[3] 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.

[4] See Shin and Thorson (2017); Weeks et al. (2017); Shin et al. (2018); Pogorelskiy and Shum (2019).

evidence (see Levy and Razin, 2019a, for a review).

It seems intuitive that unbalanced selective sharing of information can lead to polarization: If a person takes at face value what her friends say and they support one view, her opinion can be swayed accordingly. On brief reflection, however, it is clear this reasoning has some flaws. First, people often get first-hand information in addition to information received from friends, so this dampens any effect of selective sharing. Second, only a few of their friends may share information selectively, so exposure to selective sharing may be limited. Finally, even in the absence of these two mitigating factors, if a person fully understands how her friends select what to share, she will adjust for it (or ignore the received information altogether) and her beliefs will not be distorted.

What we assume about people's understanding of the selectivity of shared news turns out to be crucial. It is unrealistic to expect people to be completely naive about selective sharing, but experimental evidence suggests they do not fully take it into account either (Pogorelskiy and Shum, 2019). Consistent with this evidence, our theory allows for some *misperception* of selective sharing.

We incorporate selective sharing, echo chambers, and misperception into a simple model of learning from shared information. A binary state of the world, A or B, realizes in the first period. In every subsequent period, each agent directly observes an unbiased signal about the state with some probability γ and no signal with the remaining probability. The signals have the same independent distribution across agents and periods. We refer to their informativeness as the information quality. After observing her signal, each agent can share it with her friends (i.e., neighbors in the network) or remain silent. This rules out fake news in the form of fabricated signals. Selective sharing arises as follows: Some agents—called *normal*—share every signal; other agents—called *dogmatic*—share only signals supporting a specific state. To fix ideas, some people (possibly a tiny minority) may hold a dogmatic view on whether to vaccinate children and share only information in its favor; others simply share any information. We refer to the composition of normal and dogmatic friends of an agent as her echo chamber. We model the agents' misperception of selective sharing in a way that renders them partially unresponsive to it (as found in Pogorelskiy and Shum (2019)) and is inspired by the psychology literature.[5] In a nutshell, each agent correctly interprets all received signals, but assigns a mis-calibrated probability $\hat{\gamma} \neq \gamma$ to the arrival of signals. This is akin to assuming that her friends read the newspaper less or more often then they actually do. Thus, our agents have a common misspecified model of selective sharing, based on which they update beliefs according to Bayes' rule.

We analyze learning and belief polarization by studying how beliefs responds to one round of signals (short-run learning) as well as infinite sequences of signals (long-run learning). Consider any agent and suppose that, besides possibly many normal friends, she has more dogmatic friends who favor state A than B. For the short run, we find that her *expected* posterior belief differs from her prior—at least when the signal quality is sufficiently low. For the long run, we identify a precise information quality threshold below which almost surely

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

the agent's asymptotic belief assigns probability one always to the same state, *irrespective* of the truth. For higher quality, beliefs converge to the truth despite the echo-chamber effect. Both results depart from standard Bayesian learning. The distortion in the agent's posterior is always driven by the majority of her friends favoring A. However, perhaps counterintuitively, by *over*-estimating how often her friends are uninformed ($\hat{\gamma} < \gamma$), the agent reads too much into their signals, which distort her posterior towards A. By contrast, by *under*-estimating how often they are uninformed ($\hat{\gamma} > \gamma$), the agent reads too much into their silence, which distorts her posterior *away from* A. Moreover, instead of curbing this distorting power of dogmatic friends, abundant information can boost it and exaggerate incorrect learning. Note that the echo-chamber imbalance can be arbitrarily small, yet offset many unbiased signals. This happens when information has low quality, because, starting from this level, the misperception of shared signals and silence is more consequential. Naturally, belief distortion is possible for higher quality signals when echo-chamber imbalance increases or misperception increases.

It is easy to see how these forces can cause beliefs to polarize. If some agents have echo chambers 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 this role of information quality, in conjunction with echo chambers. It is obvious that if the signals perfectly revealed the state, it would be impossible for echo chambers to distort beliefs. But in reality information is noisy. By quantifying the minimal quality that ensures correct learning, we open the door to feasible interventions aimed at curbing polarization. Our analysis goes to the heart of why new technologies and formats of communication enabled by the Internet can increase polarization. They raised the frequency of information arrival and possibly lowered its quality as a result—for instance, tweets and social-media posts tend to be short. Moreover, overwhelmed by the huge abundance of information, people may spread their limited attention across more sources, thereby absorbing less accurate information from each of them. All of this can be a driver of polarization, even without deliberate misinformation. Given this, one may expect that raising information quality would undoubtedly help curb polarization in a network. However, we find that raising the information quality may *increase* polarization. We identify conditions on the network that are necessary and sufficient for this effect.

We find that the expansion of connections as networks grow can also be a driver of polarization. Larger networks provide better access to information but also greater scope for echo chambers to distort beliefs. We provide sharp conditions on the growth rates of the different types of friends under which network expansion curbs or exacerbates polarization. In short, curbing polarization requires that normal friends grow sufficiently *faster* than dogmatic friends. Thus, fixing information quality, 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.

Equipped with these results, we return to the motivating question of what policies may reduce the effects of the Internet and social media on polarization. We highlighted several drivers of polarization: selective news sharing in echo chambers, misperceptions, and low

information quality. The first seems hard to regulate without infringing personal freedom. The second may be addressed by improving digital literacy so that people can better take into account selective sharing. The last allows for perhaps easier and less intrusive interventions aimed at providing higher-quality information. Getting independent news sources to do so seems an immediate option, but it may be hard to incentivize and implement in a decentralized way. Given this, we show that an alternative route is to create institutional intermediaries that aggregate news for people, namely, that provide information in larger "digested" batches rather than many "raw" bits. Even if these less frequent news reports *lose* some of the information in the summarized facts, they can mitigate polarization by reducing the scope for echo chambers to cause people's beliefs to go awry. This also provides a rationale for authorities to commit to releasing information only rarely but of high quality.

Related Literature

The economics literature discusses at least three possible causes for belief polarization. One strand of papers that is most closely related (e.g., Levy and Razin, 2019b; Hoffmann et al., 2019; Enke et al., 2019) has studied polarization arising from behavioral biases. Our main contribution to this literature is to highlight the importance of the network structure coupled with selective sharing and misperception in generating polarization. Another strand studies polarization arising as a result of heterogeneity in preferences (see Dixit and Weibull (2007) and Pogorelskiy and Shum (2019)). Such heterogeneity of preferences would exacerbate the polarization we find, but is not required for generating our results about polarization. Finally, another reason for polarization that has been studied is biased or multidimensional information sources (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), where bias usually comes from media competition over viewers. In our model, the external information sources are assumed to be unbiased when reporting information; biases occur in how people share that information with each other. Thus, removing all media biases may still not be enough to curb polarization. Andreoni and Mylovanov (2012) show that one-dimensional opinions can diverge with two-dimensional information. We provide conditions under which one-dimensional opinions diverge with one-dimensional information, based on selective sharing and misperception.

This project also fits into the growing literature on the effects of model misspecification on social learning. Bohren (2016), Bohren and Hauser (2018), and Frick et al. (2019) analyze how model misspecification impacts long-term learning in environments where agents learn from private signals and the actions of other agents. In particular, Bohren and Hauser (2018) study when agents with different, yet reasonable, models have no limit beliefs (i.e. beliefs cycle) or have different limit beliefs (disagreement). The result about disagreement is close in spirit to what we find, but the driving mechanism is fundamentally different: In our model, all agents have the same misspecified model of the world and polarization results from exposure to selectively shared information through the social network. We also emphasize the important role played by the quality of information, which may suggest a simple way to

alleviate polarization by aggregating signals. Molavi et al. (2018) study long-run learning on social networks when non-Bayesian agents exhibit imperfect recall. They show that such agents may overweigh evidence encountered early on relative to later information, which can lead to mislearning in the limit. Again, while this conclusion is related to ours, unlike these authors, we assume that agents update beliefs via Bayes' rule, albeit with a misspecified model of the world.

A key element of our model is the idea of an echo chamber as a network of friends. There is a large literature on both Bayesian (Acemoglu et al., 2010; Pogorelskiy and Shum, 2019; Spiegler, 2019) and non-Bayesian learning in networks (Golub and Jackson, 2010; DeMarzo et al., 2003; Azzimonti and Fernandes, 2018; Eyster and Rabin, 2010). One closely related paper in this literature is Levy and Razin (2019a), which considers the effect of what they call a "Bayesian Peer Influence" updating heuristic on the limit beliefs in the network. One of the main results shows how beliefs in society can become polarized as a result of agents using this updating rule. However, the meaning of polarization in that paper is different from ours: Instead of groups in society becoming polarized, it is the entire society's consensus that shifts towards extreme beliefs, whereas we provide conditions under which beliefs of agents in society diverge upon arrival of new information.

Recent empirical studies show that social media is an important source of news for people and can lead to divergence of beliefs and attitudes (e.g. Allcott and Gentzkow, 2017; Bursztyn et al., 2019; Mosquera et al., 2019; Levy, 2020). Still other empirical 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 investigations of this phenomenon.

The remainder of the paper is organized as follows. Section 2 presents our stylized model of information sharing in a network. Section 3 presents the main results on learning in the short and long run. Section 4 considers what happens when the network expands. Section 5 considers the polarization in the entire network. Section 6 demonstrates how aggregating signals can mitigate polarization. We conclude with a discussion of the results in Section 7.

2 Model

We consider a stylized model of learning from information shared through social connections. Time is discrete and denoted by $t = 0, \ldots, T$, where $T \leq \infty$. A state of the world ω realizes at $t = 0$ and can take value A or B. For example, ω can represent whether the preservation of the environment requires higher national spending than the current level, or whether vaccinations can be harmful for children. There is a fixed group of agents who seek to learn ω, as it matters for their decisions.

Each agent receives both first-hand information from original sources and second-hand information shared by other agents. For each $t \geq 1$, agent i's first-hand information is a private signal $s_{it} \in \{a, b\}$, which she receives with probability $\gamma \in (0, 1]$. With probability

$1-\gamma$ the agent receives no signal. Signals are partially informative:

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

where $\frac{1}{2}<q<1$. We refer to q as the information *quality*. The events of whether agent i receives a signal and its realization are i.i.d. across agents and time.[6]

Agents share their first-hand information with their *friends*, who are other agents with whom they have a social connection. We aim to capture key aspects of social information sharing suggested by experimental evidence (for example, Pogorelskiy and Shum, 2019). The first is selectivity. After receiving her own signal, an agent can share it with her friends or stay silent. If no information is received, then the agent stays silent. Thus, she can selectively suppress information, but cannot fabricate information, which rules out fake news. For example, an agent can share a newspaper article, but cannot edit its content. For simplicity, we also assume that she cannot choose with which friends to share her signal: She either shares it with all friends or none. By ruling out fake news and targeted sharing, we highlight the role of selective sharing in a baseline model to which these other aspects can be added. The verifiable nature of information sharing and the possibility of being uninformed 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.[7]

We introduce three types of agents who are distinguished by their information-sharing behavior. An agent is *normal* if she shares any signal she receives. By contrast, an agent is *A-dogmatic* if she shares only signals $s_{it}=a$ and *B-dogmatic* if she only shares signals $s_{it}=b$. One interpretation is that such agents share only information that supports their conviction, in which they dogmatically believe.[8] Formally, each period, conditional on receiving a signal, the dogmatic types share it as follows:

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

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

Each agent's type is exogenous and known to her friends. We take these types of news-sharing behavior as given because they approximate what the empirical evidence finds (Pogorelskiy

[6] In reality, people receive pieces of news that are correlated. However, strong evidence suggests that people 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 without qualitatively changing the results.

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

[8] We can think of dogmatic 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. For the sake of modeling, we can capture such agents as having degenerate prior beliefs in A or B, which do not change with new information (unless conclusive, i.e., $q=1$).

and Shum (2019)). Moreover, our focus is not understanding *why* people share only news that support their convictions, but understanding its *consequences* for social learning.[9]

Another key empirical aspect of social news sharing is that it contributes to creating some heterogeneity in information diets (Pew Research Center, 2014; Levy and Razin, 2019a).[10] This depends on the composition of friends. Suppose agent i has $d_{Ai} \geq 0$ A-dogmatic friends, $d_{Bi} \geq 0$ B-dogmatic friends, and $n_i \geq 0$ normal friends. We will refer to (d_{Ai}, d_{Bi}, n_i) as i's *echo chamber*, because this composition determines what information she listens to. An echo chamber—and hence an information diet—exhibits an imbalance if $d_{Ai} \neq d_{Bi}$.

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 studying the beliefs of normal agents. They all start with a common prior belief $\pi \in (0,1)$ that $\omega = A$. Given a sequence of received signals $\mathbf{s}^t$ up to t, let $\mu_i(\mathbf{s}^t)$ be agent i's Bayesian posterior that $\omega = A$. To examine learning in the short run, we will consider $\mu_i(\mathbf{s}^1)$; to examine learning in the long run and so the effects of abundant information, we will consider the (probability) limit of $\mu_i(\mathbf{s}^T)$ as $T \to \infty$. Intuitively, polarization occurs when beliefs move systematically apart between agents—a formal measure of group polarization appears in Section 5. It is well known that in the short run $\mu_i(\mathbf{s}^t)$ and $\mu_j(\mathbf{s}^t)$ for $i \neq j$ can move apart in completely standard Bayesian models because of different realizations of the signals of agent i and j. Such differences would not be considered as polarization. Therefore, for the short run we adopt a more demanding criterion to define polarization, namely, a difference between $\mathbb{E}\left[\mu_i(\mathbf{s}^1)\right]$ and $\mathbb{E}\left[\mu_j(\mathbf{s}^1)\right]$. Such a difference can never occur in standard Bayesian models, where $\mathbb{E}\left[\mu_i(\mathbf{s}^1)\right] = \pi = \mathbb{E}\left[\mu_j(\mathbf{s}^1)\right]$.

At first glance, one might think that selective sharing and unbalanced echo chambers should suffice to give rise to polarization. However, this is not the case.

Remark 1. Fix any agent i and echo chamber (d_{Ai}, d_{Bi}, n_i). For every $\gamma \in (0,1]$, we have

$$\mathbb{E}\left[\mu_i(\mathbf{s}^1)\right] = \pi \qquad \text{and} \qquad \operatorname*{plim}_{T\to\infty} \mu_i(\mathbf{s}^T) = I_{\{\omega=A\}},$$

where $I_{\{\omega=A\}}$ is the indicator function that equals 1 if $\omega = A$ and 0 otherwise.

The intuition is simple. If agent i 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, agent i gets some information every period. Her belief must then satisfy standard properties of Bayesian updating.

[9] Future research may endogenize news-sharing behavior in settings similar to ours.

[10] Of course, people also have heterogeneous news diets because they choose to listen to different first-hand sources. We abstract from this aspect to focus on effects of news sharing.

Misperception. Given Remark 1, in order for polarization to be possible, we have to modify the model laid out so far. We again refer to the empirical evidence for guidance. Pogorelskiy and Shum (2019) suggest a third aspect of learning from shared news: Agents often misperceive the selectivity of second-hand information. A minimal modification of our model that gives rise to such misperceptions is to let the agents assign an incorrect probability to the arrival of signals. This leads to possibly attributing silence to a lack of news incorrectly. Formally, we assume that each agent thinks that the i.i.d. probability of getting a signal is $\hat{\gamma} \in (0, 1]$, where $\hat{\gamma} \neq \gamma$. The rest of the model is unchanged. The agents continue to use Bayes' rule, yet applied to this slightly misspecified model of the world that replaces the objective γ with the subjective $\hat{\gamma}$.

We can interpret these misperceptions as follows. Selective sharing involves replacing some received signals with silence. Unless agents know precisely the probability with which signals arrive, there will be misperception. This misperception can take the form of over- or under-estimating the probability their friends are actually uninformed. This may happen because agents are mis-calibrated about the arrival probability of signals, including their own. Another possibility is that agents are mis-calibrated only about the probability that their friends are informed (i.e., they apply γ for their own signal arrival). If $\gamma > \hat{\gamma}$, this may be a manifestation of the so-called "illusory superiority" or "better-than-average" heuristic which leads an agent to think that others may be *less* informed than she is, even though everybody is equally informed (see, for example, Cross, 1977; Svenson, 1981; Odean, 1998; Zuckerman and Jost, 2001). 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 have weak self-esteem and think that her friends are *more* informed than she is, even though everybody is equally informed (i.e., $\gamma < \hat{\gamma}$). It turns out that whether agents apply $\hat{\gamma}$ or γ to their own signals does not affect our results.

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 are the same across all agents. Only the composition of echo chambers can differ between agents. This is because we want to start from a setting where all agents are ex-ante identical and have the same model of the world. The only possible driver of belief divergence is the difference in information diets due to echo chambers. It is intuitive that allowing other parameters to differ between agents can introduce other drivers of polarization. We will discuss some below.

3 Learning in the short and long run

To examine the effects of selective sharing and misperceptions in the short run, our first result looks at the expected posterior belief of agent i that results from one round of updating (i.e.,

$T = 1$).[11] Since the focus of this section is normal agent i, we drop all i subscripts. Recall that $\mu(\mathbf{s}^1)$ is the Bayesian posterior probability that a normal agent assigns to state A given all the information she obtains after one period, which we denote by $\mathbf{s}^1$ (i.e., her signal, her friends' signals, and their silence). We show that whenever there is an imbalance in i's echo chamber, i's expected posterior conditional on *any* state will be distorted when the signal quality q is sufficiently low. Besides obviously depending on the imbalance between A- and B-dogmatic friends, the direction of the distortion depends on whether the agent under- or over-estimates the probability that information is received.

Proposition 1. *Fix an agent such that $d_A > d_B$.*

1. *If $\hat{\gamma} < \gamma$, there exists $\bar{q}_1(d_A, d_B, n) > \frac{1}{2}$ such that if $q < \bar{q}_1(d_A, d_B, n)$, then for any $\omega \in \{A, B\}$, we have $\mathbb{E}[\mu(\mathbf{s}^1|\omega)] > \pi$.*
2. *If $\hat{\gamma} > \gamma$, there exists $\bar{q}_1(d_A, d_B, n) > \frac{1}{2}$ such that if $q < \bar{q}_1(d_A, d_B, n)$, then for any $\omega \in \{A, B\}$, we have $\mathbb{E}[\mu(\mathbf{s}^1|\omega)] < \pi$.*

An immediate implication of this result is that a defining property of Bayesian updating no longer holds in our setting. Since the expected posterior is distorted conditional on any true state of the world, the *unconditional* expected posterior will also be distorted in the same direction. This is in contrast to Bayesian updating with a correctly specified model of the world, where the expected posterior would be *equal* to the prior as noted in Remark 1.

Note that the direction of the bias depends critically on whether agents under- or over-estimate the probability with which friends receive signals. Specifically, Part 1 of Proposition 1 states that if $\hat{\gamma} < \gamma$ (agents *under-estimate* friends' information), then the agent's expected belief is biased towards the conviction of the *majority* of her dogmatic friends. Conversely, Part 2 states that if $\hat{\gamma} > \gamma$ (agents *over-estimate* friends' information), then the agent's expected belief is biased towards the conviction of the *minority* of her dogmatic friends. In both cases, agents beliefs are systematically biased because of misperception.

To give some intuition for the result, we begin by describing agent i's posterior after one period. Let her echo chamber be (d_A, d_B, n). Denote by a_A the number of a-signals that her A-dogmatic friends received and by b_B the number of b-signals that her B-dogmatic friends received. From the agent's 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. Agent i also receives a private signal and n independent signals from her normal friends. This is equivalent to receiving $n+1$ independent private signals. Let a_N and b_N denote the number of a-signals and b-signals that are contained within them, respectively. The quantities a_N and b_N are multinomial random variables with probabilities $\hat{\gamma}(1-q)$ and $\hat{\gamma}q$ respectively, and sample size $N = n+1$. Note that agent i's information $\mathbf{s}^1$ is summarized by (a_A, b_B, a_N, b_N). Given this, by Bayes's rule

[11] Note that in what follows we, as the external observer, will calculate $\mathbb{E}\left[\mu_i(\mathbf{s}^1)\right]$ and $\operatorname{plim}_{T\to\infty} \mu_i(\mathbf{s}^T)$ using the correct model of the world (i.e., γ not $\hat{\gamma}$).

agent i's posterior belief is[12]

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

where

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

We can understand the parts of this expression as follows. First, Q^M captures the agent's interpretation of the received signals, which is always correct. Agent i receives $a_A + a_N$ a-signals from A-dogmatic and normal friends. She receives $b_B + b_N$ b-signals from B-dogmatic and normal friends, which counteract the a-signals (and hence enter negatively in the exponent M). Note that this part does not depend on $\hat{\gamma}$. This is because, by the assumption of verifiable information, the act of sharing a signal leaves no uncertainty on whether the signal was actually received—hence, $\hat{\gamma}$ is irrelevant.

To form her belief, the agent also has to interpret the silence of her dogmatic friends. This is captured by Γ^S. Agent i observes silence from $d_B - b_B$ B-dogmatic friends and from $d_A - a_A$ A-dogmatic friends, each of which she (incorrectly) attributes to an unfavorable signal for the friend with probability $\hat{\gamma}$ or to no signal with probability $1 - \hat{\gamma}$. Note that Γ is a decreasing function of $\hat{\gamma}$. Thus, a higher $\hat{\gamma}$ raises Γ^S if $S < 0$ and lowers Γ^S if $S > 0$, thereby distorting the posterior downward or upward depending on the realized 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 still be correct on average.

If the agent has more A- than B-dogmatic friends ($d_A > d_B$), 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 is useful to consider extreme misperceptions. Suppose $\gamma > \hat{\gamma} \approx 0$, namely, the agent severely *under-estimates* the probability her friends are informed. In this case, she interprets silence as almost certainly no news, rather than negative news for her dogmatic friend's preferred state. Thus, she essentially ignores silence and updates based only on the shared signals, which tend to be more in favor of state A. By contrast, suppose $\gamma < \hat{\gamma} \approx 1$, namely, the agent severely *over-estimates* the probability her friends are informed. In this case, she interprets silence as almost certainly negative news for her dogmatic friend, rather than no news. Thus, she reads too much into the silence of the majority of her dogmatic friends and incorrectly updates her belief *away* from their preferred state A. Put differently, the A-majority of the agent's dogmatic friends always drives her belief through selective sharing, but this can backfire and systematically push her to believe that the state is B.

[12]This representation is derived in the proof of Proposition 1.

Either way, the agent's belief is systematically biased because of misperception. Clearly, the quality of the signals cannot be perfect (i.e., $q = 1$) for i's expected posterior to deviate from her prior. Yet, the result shows that the imbalance between dogmatic friends—however small—can always overcome the unbiased information coming from normal friends and i's own signal when the agents get signals of sufficiently low quality. In particular, the misperceived selective sharing dominates even though dogmatic friends' signals become almost uninformative.

Proposition 1 has immediate consequences for belief polarization in societies. If i has more A-dogmatic than B-dogmatic friends and i' has more B-dogmatic than A-dogmatic friends—where again the imbalances can be small—then in expectation their posteriors will move apart, at least if the quality of signals is small. Thus, our result suggests that the composition of one's echo chamber can be a driver of belief polarization in spite of all agents having the same access to first-hand information and the same model of the world. This is consistent with the narrative that social media can give rise to "echo chambers" where people are exposed to an unbalanced diet of opinions and as a result develop polarized beliefs. However, our result qualifies this narrative in two ways: First, it does not require the presence of fake news; second, it stresses that echo chambers also require misperception in order to distort beliefs from the truth.

Our first result showed that one round of information arrival is enough to give rise to polarization, at least in expectation. However, it leaves open the possibility that when information becomes abundant (i.e., in the long run after many signals) polarization disappears. In fact, the opposite can occur. As long as information quality is sufficiently low, abundant information can exacerbate the effect of misperceived selective sharing and cause i's beliefs to be almost certainly incorrect. To show this, we consider the limit in probability of the posterior belief $\mu(\mathbf{s}^T)$ as $T \to \infty$, which we denote by $\operatorname{plim}_{T\to\infty} \mu(\mathbf{s}^T)$. If q is sufficiently small, i's posterior belief is pushed towards the degenerate belief on the state that is favored by the majority or minority of her dogmatic friends, with probability 1 and irrespective of the true state. Whether long-run beliefs are distorted towards the minority or the majority of dogmatic friends once again depends on if $\hat{\gamma}$ is greater or less than γ. Denote, $I_{\{\omega=A\}}$ as the indicator function that equals 1 if $\omega = A$ and 0 otherwise.

Proposition 2. *Fix any $d_A > d_B$, and any $\gamma \in (0,1)$. Irrespective of what the true state is, we have:*

1. *If $\hat{\gamma} < \gamma$, there exists $\bar{q}_2(d_A, d_B, n) > \frac{1}{2}$ such that if $q < \bar{q}_2(d_A, d_B, n)$, then the agent's asymptotic belief converges to state $\omega = A$ with probability 1.*

2. *If $\hat{\gamma} > \gamma$, there exists $\bar{q}_2(d_A, d_B, n) > \frac{1}{2}$ such that if $q < \bar{q}_2(d_A, d_B, n)$, then the agent's asymptotic belief converges to state $\omega = B$ with probability 1.*

Consider the first part of Proposition 2. Intuitively, with many rounds of information arrival agent i's own signals provide a more accurate estimate of the state, which would result in perfect learning in a standard setting. However, i combines her signals with the signals from her friends, which provide more information, but this information is biased in

ways she does not correctly take into account. Once again, the outcome of this race between the two kinds of information is a priori not clear. Yet, with low quality signals, the distortion introduced in each step of updating unveiled in Proposition 1 accumulates over time leading the posterior astray with certainty.

To see this more precisely, consider updating at each period t. Dogmatic friends tend to provide more $s = a$ than $s = b$ signals. Normal friends provide unbiased signals. Who wins the race as $T \to \infty$? Recall the correct updating term Q^M and incorrect updating term Γ^S. When q is close to $1/2$ (low informativeness), the correct updating term Q^M is close to 1 and thus the misperception boost (or curtailment) to informativeness Γ^S dominates. By contrast, when q is close to 1 (maximal informativeness), the correct updating term Q^M is close to zero and hence the misperception boost or curtailment to informativeness Γ^S is diminished.

Proposition 3. *Fix any $\gamma \in (0,1)$. Then, the bound $\bar{q}_2(d_A, d_B, n)$ is increasing in $|d_A - d_B|$ and $|\gamma - \hat{\gamma}|$, and decreasing in n.*

Proposition 3 provides intuitive, yet important, comparative statics on the lower bound of informativeness that avoids incorrect learning and hence polarization. This lower bound increases with the magnitude of the echo-chamber imbalance (i.e., the absolute difference between the number of A-dogmatic and B-dogmatic friends) and with the degree of misperception. It decreases with number of normal friends, as they provide the agents with more non-selected information. Therefore, given q, it is possible that if an agent belongs to a social group with a large but moderately unbalanced number of dogmatic friends, she learns correctly as information becomes abundant. By contrast, if the agent belongs to a social group with a small but severely unbalanced number of dogmatic friends, their effect may prevail for the same q, causing her belief to polarize relative to other agents.

With regard to polarization, Propositions 1 and 2 can shed further light on the role of echo chambers, contributing to and qualifying the ongoing debate about them. In sum, the results highlight the importance of the composition of an echo chamber, rather than its absolute size, and of misperceptions about how information is selectively shared in it. To the extent that interacting online promotes the formation of more unbalanced social groups and misperceptions about selective sharing, it may contribute to belief polarization even when the quality of information is unchanged.

Propositions 1 and 2 show that imbalance in selective sharing is sufficient for polarization to arise. But is it necessary? The answer is yes. Intuitively, if i has an equal number of A-dogmatic and B-dogmatic friends, they offset each other when sharing information selectively, even in the presence of misperception. Note that even a minor imbalance with misperceptions will make polarization possible.

4 Network expansion

We wish to consider network expansion. For the purpose of this section we consider the case of $\hat{\gamma} < \gamma$, which implies that an agent's belief is distorted towards the majority of

dogmatic friends in her echo chamber. Consider agent i once again and suppose that due to advances in technology (e.g. development of social media), the agent is now connected with more agents of all types, while still receiving the same quantity of information on her own. Specifically, assume that the number of normal, A-dogmatic and B-dogmatic friends expands according to proportions $(\lambda_A, \lambda_B, \lambda_N)$, respectively. Proposition 4 characterizes the impact on the informativeness required to avoid polarization.

Proposition 4. *Fix an agent such that $d_A > d_B$, then $\bar{q}_2(\lambda_A d_A, \lambda_B d_B, \lambda_N n) \leq \bar{q}_2(d_A, d_B, n)$ if and only if*

$$(\lambda_N - 1) \geq \left(\frac{\lambda_A d_A - \lambda_B d_B}{d_A - d_B} - 1\right) \cdot \left(1 + \frac{1}{n}\right)$$

Proposition 4 states that the range of information quality q leading to incorrect learning shrinks if and only if normal friends grow sufficiently fast. Intuitively, there is a trade-off between access to information and the scope for echo-chambers to bias beliefs. Note that if scaling is proportional (i.e. $\lambda_A = \lambda_B = \lambda_N$) then $1 < 1 + \frac{1}{n}$ and hence the range of information quality leading to incorrect beliefs *expands*. Thus a proportional scaling of the network leads to more opportunities for polarization.

This result suggests that when there are changes in the network environment that increase connectivity of people (keeping proportions of different types of friends the same), but do not increase the amount of information people get on their own, it can contribute to belief polarization. For a fixed information quality q, it is possible that a person whose beliefs are not biased in a smaller network becomes polarized when the network expands.

The previous proposition deals with how large λ_N needs to be in order to decrease an individual's $\bar{q}_2$. This reduces the range of q in which the individual gets biased, but may still not be enough if the true q is far lower. The next proposition explores how large λ_N needs to be in order to drop $\bar{q}_2$ below any given value of q, so as to un-bias a given agent. Put another way: Given some $\hat{q}$, what λ_N suffices to ensure correct learning?

Proposition 5. *Fix $d_A > d_B$ and $\hat{q} \in \left(\frac{1}{2}, \bar{q}_2(d_A, d_B, n, \gamma)\right)$. Then $\bar{q}_2(\lambda d_A, \lambda d_B, \lambda_N n, \gamma) < \hat{q}$ if*

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

Given echo chamber (d_A, d_B, n) and growth rates of friends (which may be determined by data or algorithms), this result predicts if and when $\bar{q}_2$ crosses $\hat{q}$. This points to how link-formation algorithms may possibly be designed to minimize polarization. i.e. the result suggests the growth rate of normal friends required. Note that the right side of the inequality is increasing in d_A and λ, decreasing in d_B and $\hat{q}$. Moreover, the right side of the inequality is decreasing in n when it is positive (when $\hat{q} < \bar{q}_2$).

5 Network polarization

In this section, we move from the analysis of individual agents to the analysis of the network as a whole. Our focus is the set of normal agents in the network (i.e., excluding dogmatic ones). Interpreting dogmatic agents as those whose beliefs are degenerate implies that only the beliefs of normal agents respond to arriving information. Let the set of all normal agents be denoted by $\mathcal{N}$.

Before proceeding, we define a measure of polarization for our network. Without loss of generality, we will define it for beliefs in state A. Define polarization in period t as the average sum of differences of beliefs across all agents in $\mathcal{N}$:

$$\Pi^t = \frac{2}{|\mathcal{N}|^2} \sum_{i,j \in \mathcal{N}} \left| \mu_i(\mathbf{s}_i^t) - \mu_j(\mathbf{s}_j^t) \right|.$$

This expression captures the extent to which disagreement is present in the network. The scaling factor of 2 is introduced in order to ensure $\Pi^t \in [0,1]$.

We focus on asymptotic polarization of beliefs (as $T \to \infty$), when all normal agents' beliefs have already converged. Let $\mathcal{N}_B$ and $\mathcal{N}_A$ denote the sets of normal agents whose asymptotic beliefs put probability 1 on states B and A, respectively. We will also refer to these sets as "eventually incorrect" and "eventually correct" populations, depending on which state is true.[13] Define the limit polarization Π as

$$\Pi \equiv \operatorname*{plim}_{t \to \infty} \Pi^t = \frac{2}{|\mathcal{N}|^2} \cdot 2|\mathcal{N}_B||\mathcal{N}_A| \cdot |1-0| = \frac{4|\mathcal{N}_B||\mathcal{N}_A|}{|\mathcal{N}|^2}$$

The value of Π varies from 0 to 1, attaining its maximum when $\mathcal{N}_B$ and $\mathcal{N}_A$ have the same cardinality.

The first result in this section concerns the dynamics of $\mathcal{N}_B$ and $\mathcal{N}_A$ as q increases from $1/2$ to 1. Assume the true state is B and fix some $q \in (1/2, 1)$. Then, $\mathcal{N}_B$ consists of two sets of agents: (1) all agents for whom $\bar{q}_{2i}$ is below q (these agents learn correctly), *and* (2) all agents for whom B-dogmatic friends dominate A-dogmatic friends and $\bar{q}_{2i}$ is above q (these agents place probability 0 on $\omega = A$ in the long run irrespective of true state). Meanwhile, $\mathcal{N}_A$ consists of all agents for whom A-dogmatic friends dominate B-dogmatic friends and whose $\bar{q}_{2i}$ is above q.

As q increases, all agents that are already in $\mathcal{N}_B$ will remain there. Each agent for whom $\bar{q}_{2i}$ is below q will continue to be below q, and other agents for whom B-dogmatic friends dominate A-dogmatic friends may remain in that category or pass into the first category and still remain in $\mathcal{N}_B$. However, agents in $\mathcal{N}_A$ may switch to the other set. When q becomes larger than $\bar{q}_{2i}$ for an agent in $\mathcal{N}_A$, he or she is no longer biased by A-dogmatic friends, and thus will place probability 0 on $\omega = A$ in the limit instead of 1, passing into $\mathcal{N}_B$. This argument is formalized in the following lemma.

[13] If B is the true state, then $\mathcal{N}_B$ is the "eventually correct" population, whereas $\mathcal{N}_A$ is the "eventually incorrect" population—vice versa if the true state is A.

Lemma 1. *As q increases, the set of "eventually correct" agents is weakly expanding and the set of "eventually incorrect" agents is weakly contracting.*

What implications does this have for network polarization? A common intuition would suggest that as people receive better information, disagreements should decline. Lemma 1 places some limits on this intuition: if for low q the set of "eventually incorrect" agents is sufficiently large, polarization in society will be low. However, as q increases, it will cause a gradual shift into the set of "eventually correct" agents, and polarization in society will actually *increase* initially. Once the set of "eventually correct" agents outnumbers the set of "eventually incorrect" ones, polarization will start to decline as q increases. This implies that polarization in a social network may temporarily increase as quality of information is slowly improving.

The following proposition summarizes the intuition above and provides a necessary and sufficient condition for such non-monotonicity of network polarization to take place. Let $\mathcal{N}_B(q)$ and $\mathcal{N}_A(q)$ denote the sets of agents whose beliefs converge to state B and A, respectively, for a given value of q. Additionally, let $\mathcal{D}_B$ and $\mathcal{D}_A$ denote sets of agents who have a B- and A-slanted echo chambers, respectively. The remaining agents do not have any imbalance in their echo chamber.

Proposition 6. *Fix network with $\bar{q}_{2i} \neq \bar{q}_{2j}$ for all i, j and ω. Π is decreasing in q over $\left(\frac{1}{2}, 1\right)$ iff $|\mathcal{D}_{-\omega}| \leq \frac{1}{2}(|\mathcal{N}| + 1)$. Otherwise, Π is single peaked.*

6 Mitigating polarization

We have shown that selective sharing and misperceptions are enough to generate polarization of beliefs in a network of agents. How could a social planner combat this polarization, if he or she could affect only the information the agents receive (taking selective sharing and misperception as given)? That is, we want to see whether it is possible to eliminate polarization by changing only the external information structure of the agents. One obvious way to do this is to *directly* increase the quality of information q at the source. However, this may not be feasible due to technological or economic reasons. For example, this may involve forcing newspapers to spend more on reporters, data gathering, and fact checking. However, it is possible to increase the quality of information that the agents receive without changing the quality of the primitive signals s_{it}. This involves *signal aggregation*.

Signal aggregation consists in summarizing a set of signals into a single message. This summary can, of course, be done in many ways, emphasizing some aspects of the original signals and omitting others. It is important to note, however, that signal aggregation involves some loss of information relative to the totality of the aggregated signals. Nonetheless, the resulting message can have higher quality than each aggregated signal individually. It is this distinction that renders signal aggregation useful for our goal of reducing polarization, despite the loss of information. As such, signal aggregation involves a trade-off between

slowing short-run learning and debiasing long-run learning, which is a novel aspect of our learning environment.

Of course, there are many ways to aggregate signals. We consider the following simple form to make our point. Let M be an odd number and divide time into blocks of M periods. Define $\hat{s}^i_{Mt}$ as the aggregated signal that is released to agent i at the end of each time block and reports whether more $s = a$ or more $s = b$ signals realized over the previous M periods:

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

A natural question is why a social planner would prefer to coarsen each agent's information structure in such a way. Clearly, $\hat{s}^i_{Mt}$ conveys less information than do the aggregated M signals together. However, on the one hand, if all agents observe only $\hat{s}^i_{Mt}$ every other M periods, they have *fewer* opportunities to selectively share information with their friends. But this is irrelevant in the long run. On the other hand—and more importantly—$\hat{s}^i_{Mt}$ has higher quality than each individual s_{it}. As a result, by substituting s_{it} information structure with $\hat{s}^i_{Mt}$ for each agent, we are worsening information quality for each agent individually, but we are also reducing the influence of selective sharing on agents' beliefs. Hence, this might restore the standard result of convergence to the truth in asymptotic learning. To see this, suppose $M = 3$, consider the first block of three periods, and assume $\omega = A$. We have

$$\begin{aligned} \mathbb{P}(\hat{s}^i_3 = 0|\omega = A) &= \mathbb{P}\left(\sum_{t=1}^{3} s_{it} \leq 1 \middle| \omega = A\right) \\ &= q^3 + 3q^2(1-q) \\ &= q \cdot \Big(q^2 + 2q(1-q) + q(1-q)\Big) \\ &> q \cdot \Big(q^2 + 2q(1-q) + (1-q)^2\Big) \\ &= q = \mathbb{P}(s_{it} = a|\omega = A). \end{aligned}$$

Thus, $\hat{s}^i_3$ is more informative than each s_{it} of quality q. In contrast to standard models where the quality of information does not matter in the long run, we saw that in our model a sufficiently high quality of information can allow all agents to learn correctly, thereby removing polarization. The remaining question is then whether aggregating signals according to $\hat{s}^i_{Mt}$ can achieve that quality level. The next proposition shows that this is always possible. Denote $\hat{\Pi}$ as the resulting limit polarization when signal s_{it} is replaced with aggregate signal $\hat{s}^i_{Mt}$ every M periods.

Proposition 7. *Fix any network and q such that $\Pi > 0$. There exists M such that, if each agent i observes signals $\hat{s}^i_{Mt}$, then the resulting limit polarization $\hat{\Pi}$ equals zero.*

7 Discussion

We explored conditions under which learning from shared news may lead to belief polarization. We develop a model reflecting two important facts about news sharing in networks: unbalanced selective transmission of information and misperception. We include misperception to account for the evidence that people are neither completely naive nor fully aware of selective sharing. We found that with these features polarization occurs if (and only if) information quality is sufficiently low. Thus, our results emphasize the importance of distinguishing between quality and quantity of information, a distinction that is irrelevant in standard models.

With this understanding, we also investigated ways to mitigate polarization. On the one hand, our analysis of the effects of network expansion suggests ways to influence network dynamics so as to limit the scope for polarization to occur. On the other hand, we show that aggregating the signals that agents receive may mitigate polarization. A benevolent authority may use this insight to stop the effects of misperceived selective sharing and debias beliefs. In practice, one way to achieve this is to create institutional intermediaries or platforms that aggregate news, namely, that provide people with information in larger digested batches rather than many raw bits. Even if these less frequent news reports *lose* some of the information in the summarized news, they can mitigate polarization by reducing the scope for echo chambers to cause people's beliefs to go awry.

Our analysis goes to the heart of why new technologies and formats of communication enabled by the Internet may increase polarization. Tweets and social-media posts tend to be short and hence of low quality. Moreover, overwhelmed by the abundance of information, people may spread their limited attention across more sources, thereby absorbing less accurate information from each of them.

Another class of policies may be to intervene in network dynamics. Indeed, in our results expanding networks can help curb polarization if they also shrink the imbalance in echo chambers. We provide tools to precisely quantify how much the imbalance has to shrink and how long it will take, given the evolution of networks. These conditions can be easily tested using data about one's composition of friends, their news-sharing habits, and their growth rates. To the extent that social-network platforms influence this evolution with their algorithms suggesting new friends, our results may guide how to design them so as to render echo chambers more balanced.

Finally, our theory sheds new light on how malevolent actors who thrive on social polarization can benefit from how news is shared in social networks. One obvious way is to use fake news to directly lower the information quality or render echo chambers even more unbalanced. A more subtle way is to release bits of news with high frequency so as to leverage the power of misperceived selective sharing and to draw attention away from high-quality information sources. These insights may be useful to guide policies aimed at dealing with malevolent actors. Even though removing fake news is important for addressing polarization, our analysis suggests that it is not sufficient.

Our analysis opens several directions for future research. So far, we have exogenously

assumed selective sharing by having dogmatic agents who do not update their beliefs and so do not change how they selectively transmit their signals. However, in reality people often choose what information to share so as to promote their views or beliefs, which are not set in stone. One can interpret the current model as describing situations where dogmatic agents can change their beliefs, yet they do so much more slowly than normal agents. As a result, how they selectively share information is very persistent, which should leave the insights of our results unchanged.

Our results about the beliefs of a single agent can be used to further examine the distribution of beliefs in the network as a whole. We conjecture that there is a relationship between how the composition of echo chambers is distributed in the network and how beliefs are distributed across agents. For instance, if half the population has a majority of dogmatic-right friends, half a majority of dogmatic-left friends, and information quality is sufficiently low, then the beliefs of each half should move apart as information arrives. This is consistent with evidence showing that people on the left and on the right of the political spectrum tend to have more like-minded friends than not, and with the view that this may be a cause of the ongoing polarization (e.g., see Pew Research Center (2014)). Note that, according to our analysis, such polarization does not require that people look at the world in fundamentally incompatible ways, but only that they have different news diets—which may be easier to address.

Finally, the role of the network structure in our analysis begs the question of what would change if we allowed for endogenous network formation. On the one hand, people may tend to form more links with like-minded friends, which may enlarge the imbalance in their echo chambers. This would reinforce and magnify the effects of selective sharing and misperceptions we highlight. On the other hand, people may be more likely to link with normal friends than dogmatic friends, which would have opposite implications. Either way, this is ultimately an empirical question: Once we measure the rates of link formation with different types of friends, we can use our model to predict their consequences. Nonetheless, understanding the incentives to form links so as to obtain information is also important to answer these questions and guide further empirical investigations of polarization.

References

Acemoglu, D., A. Ozdaglar, and A. ParadejGheibi (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.

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

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

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.

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.

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.

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*.

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.

Frick, M., R. Iijima, and Y. Ishii (2019). Misinterpreting Others and the Fragility of Social Learning. Cowles Foundation Discussion Papers 2160, Cowles Foundation for Research in Economics, Yale University.

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.

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

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*.

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.

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 (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.

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.

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

Appendix

A Proof of Proposition 1

We consider agent i who has d_A A-dogmatic friends, d_B B-dogmatic friends and n normal friends. We drop all i-subscripts since we are focusing on a single agent.

We will prove that there exists $\bar{q}_1 > \frac{1}{2}$ such that if $q \in \left(\frac{1}{2}, \bar{q}_1\right)$, then $\mathbb{E}\left[\mu_i|\omega\right] > \pi$ or $\mathbb{E}\left[\mu_i|\omega\right] < \pi$ for any ω, depending on the signs of $(d_A - d_B)$ and $(\gamma - \hat{\gamma})$. For that, we will first find the derivative of $\mathbb{E}\left[\mu_i|\omega\right]$ with respect to q at $q = \frac{1}{2}$, and then show how the sign of the derivative of $\mathbb{E}\left[\mu_i|\omega\right]$ w.r.t. q at $q = \frac{1}{2}$ depends on $(d_A - d_B)$ and $(\gamma - \hat{\gamma})$. Using continuity of the expected posterior in q and the fact that at $q = \frac{1}{2}$ we have $\mathbb{E}\left[\mu_i|\omega\right] = \pi$, we will arrive at the desired conclusion.

We begin by deriving a normal agent's posterior belief. Given a fixed realization $\mathbf{s} = (a_A, b_B, a_N, b_N)$, by Bayes's rule an agent's posterior belief in state A is given by:

$$\mu_i(\mathbf{s}) = \frac{\pi X_A}{\pi X_A + (1-\pi) X_B}, \tag{2}$$

where

$$\begin{aligned} X_A &= \hat{\gamma}^{a_A+b_B+a_N+b_N}(1-\hat{\gamma})^{N-a_N-b_N} \\ &\quad \times q^{a_A+a_N}(1-q)^{b_B+b_N}(\hat{\gamma}q + (1-\hat{\gamma}))^{d_B-b_B}(\hat{\gamma}(1-q) + (1-\hat{\gamma}))^{d_A-a_A}, \\ X_B &= \hat{\gamma}^{a_A+b_B+a_N+b_N}(1-\hat{\gamma})^{N-a_N-b_N} \\ &\quad \times (1-q)^{a_A+a_N} q^{b_B+b_N}(\hat{\gamma}(1-q) + (1-\hat{\gamma}))^{d_B-b_B}(\hat{\gamma}q + (1-\hat{\gamma}))^{d_A-a_A}. \end{aligned}$$

To understand each term consider X_A. It is the conditional probability of observing $\mathbf{s}$, given that the true state is A. Specifically:

- $q^{a_A+a_N}$ is the probability of getting $a_A + a_N$ signals $s = a$ from A-dogmatic and normal friends;
- $(1-q)^{b_B+b_N}$ is the probability of getting $b_B + b_N$ signals $s = b$ from B-dogmatic and normal friends;
- $(\hat{\gamma}q + (1-\hat{\gamma}))^{d_B-b_B}$ is the probability of observing $d_B - b_B$ B-dogmatic friends staying silent, as it is either a genuine silence (with prob. $1-\hat{\gamma}$) or a suppressed signal $s = a$ (with prob. $\hat{\gamma}q$);
- $(\hat{\gamma}(1-q) + (1-\hat{\gamma}))^{d_A-a_A}$ is the probability of observing $d_A - a_A$ A-dogmatic friends staying silent, as it is either a genuine silence (with prob. $1-\hat{\gamma}$) or a suppressed signal $s = b$ (with prob. $\hat{\gamma}(1-q)$).

For X_B, the probabilities q and $1-q$ are reversed because the true state is switched to B.

Rearranging Equation 2 gives:

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

To compute the state-conditional expected posterior after one round of updating, $\mathbb{E}[\mu_i|\omega]$, it is useful to use iterated expectation and condition on the set of friends that actually receive the

signal. Let $\mathbb{E}\left[\mu_i|\omega, x_A, x_B, x_N\right]$ be the expected posterior belief in state A conditional on the fact that the true state is ω, that x_A A-dogmatic friends received a signal (and others didn't), that x_B B-dogmatic friends received a signal, and that x_N normal friends received a signal. For simplicity, let x_N also include the agent's own signal. Then we can represent the state-conditional expected posterior as follows:

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

The derivative of $\mathbb{E}[\mu_i|\omega]$ w.r.t. q is given by

$$\frac{\partial}{\partial q}\mathbb{E}\left[\mu_i|\omega\right] = \sum_{x_A=0}^{d_A}\sum_{x_B=0}^{d_B}\sum_{x_N=0}^{N}\frac{d_A!d_B!N!}{x_A!(d_A-x_A)!x_B!(d_B-x_B)!x_N!(N-x_N)!}\cdot \\ \cdot\gamma^{x_A+x_B+x_N}(1-\gamma)^{d_A+d_B+N-x_A-x_B-x_N}\frac{\partial}{\partial q}\mathbb{E}\left[\mu_i|\omega, x_A, x_B, x_N\right]. \tag{3}$$

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

Lemma 2.

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

Proof of Lemma 2. Let $H(q;\omega)$ denote the probability of getting signal $s = a$ conditional on receiving a signal. That means that $H(q;A) = q$ and $H(q;B) = 1-q$. Additionally, let $a_N \leq x_N$ be the number of informed normal friends that get signal $s = a$. Then we can represent $\mathbb{E}\left[\mu_i|\omega, x_A, x_B, x_N\right]$ as follows:

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

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

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

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

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

Note that the sum

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

is symmetric around $a_N=\frac{x_N}{2}$: for each positive term, there is an identical term with a negative sign. Thus, it must be equal to 0. We can then write

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

■

It remains to evaluate $(\partial/\partial q)\mathbb{E}\left[\mu_i|a_N,x_A,x_B,x_N\right]\Big|_{q=\frac{1}{2}}$. The following result provides a first intermediate step.

Lemma 3.

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

where

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

Proof of Lemma 3. Let $a_D\leq x_A+x_B$ be the total number of $s=a$ signals that A- and B-dogmatic friends have received. Using $b_B=x_B-b_B$ and $b_N=x_N-a_N$, we can write the agent's posterior belief as

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

Note that the dogmatic friends which haven't received a signal (d_A-x_A A-dogmatic ones and d_B-x_B B-dogmatic ones) are still included in the posterior belief of the agent, as she doesn't know whether they didn't get a signal or they suppressed their signal.

The expected posterior belief conditional on a_N,x_A,x_B,x_N and ω can be written as

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

Using binomial symmetry, we can rewrite the sum as follows:

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

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

The next lemma provides a second intermediate step to evaluate $(\partial/\partial q)\mathbb{E}\left[\mu_i|a_N, \omega = B, x_A, x_B, x_N\right]$ at $q = \frac{1}{2}$.

Lemma 4. *At* $q = \frac{1}{2}$,

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

Proof of Lemma 4. For simplifying subsequent algebra, define a function $z(q, \hat{\gamma}) \equiv \ln(\Gamma)\left[\ln(Q)\right]^{-1}$. Taking the derivative of $f(k, q, a_N)$ w.r.t. q we have:

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

Setting $q = 1/2$ gives

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

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

$$\begin{aligned}\tfrac{\partial}{\partial q}\left(f(a_D, q, a_N) + f(x_A + x_B - a_D, q, a_N)\right) &= \left(\tfrac{1}{2}\right)^{x_A+x_B-1}\Big(0 + 2\pi(1-\pi)\Big((x_A - x_B + 2(2a_N - x_N)) + (x_A - x_B - (d_A - d_B))\tfrac{\hat{\gamma}}{2-\hat{\gamma}}\Big)\Big)\\ &= \left(\tfrac{1}{2}\right)^{x_A+x_B-1}2\pi(1-\pi)\Big(2(2a_N - x_N) + \tfrac{2}{2-\hat{\gamma}}(x_A - x_B) - \tfrac{2\hat{\gamma}}{2-\hat{\gamma}}(d_A - d_B)\Big).\end{aligned}$$

■

We now further simplify $\frac{\partial}{\partial q}\mathbb{E}\left[\mu_i|\omega, x_A, x_B, x_N\right]$ evaluated at $q = 1/2$:

Lemma 5.

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

Proof of Lemma 5. From Lemma 3 and 4 we have

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

The simplification above is allowed by the fact that the terms in the sum do not depend on a_D, and 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 simply a binomial expansion of $\left(\frac{1}{2}+\frac{1}{2}\right)^{x_A+x_B}=1$.

Returning to $\frac{\partial}{\partial q}\mathbb{E}\left[\mu_i|\omega,x_A,x_B,x_N\right]$:

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

Here the simplification is allowed by they symmetry of $\sum_{a_N=0}^{x_N}\frac{x_N!}{a_N!(x_N-a_N)!}(2a_N-x_N)$ around $a_N=\frac{x_N}{2}$. ■

Finally, we return to the derivative of the state-conditional expected posterior $\mathbb{E}[\mu_i|\omega]$. Using Lemma 5, Equation(3) simplifies to:

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

Rearranging gives

$$\begin{aligned}\frac{\partial}{\partial q}\mathbb{E}\left[\mu_i|\omega\right] &\overset{q=\frac{1}{2}}{=} = \frac{4\pi(1-\pi)}{2-\hat{\gamma}}\left[\sum_{x_A=0}^{d_A}\frac{d_A!}{x_A!(d_A-x_A)!}\gamma^{x_A}(1-\gamma)^{d_A-x_A}x_A-\sum_{x_B=0}^{d_B}\frac{d_B!}{x_B!(d_B-x_B)!}\gamma^{x_B}\right.\\ &\qquad\left.(1-\gamma)^{d_B-x_B}x_B-\hat{\gamma}(d_A-d_B)\right]\\ &= \frac{4\pi(1-\pi)}{2-\hat{\gamma}}\left[\mathbb{E}[x_A]-\mathbb{E}[x_B]-\hat{\gamma}(d_A-d_B)\right]\\ &= \frac{4\pi(1-\pi)}{2-\hat{\gamma}}(d_A-d_B)(\gamma-\hat{\gamma}).\end{aligned}$$

If either $d_A > d_B$ and $\gamma > \hat{\gamma}$ or $d_A < d_B$ and $\gamma < \hat{\gamma}$ hold, then the derivative is positive. That is, for any true state ω, $d_A > d_B$ and $\hat{\gamma} < \gamma$ together imply that the state-conditional expected posterior is biased towards state A (for low q). And if $d_A < d_B$ and $\hat{\gamma} > \gamma$ hold, then the beliefs are still biased towards A (against the majority of dogmatic friends).

Note that due to symmetry, similar conclusions hold for the posterior belief in state B. That is, if $d_B > d_A$ and $\hat{\gamma} < \gamma$, then the beliefs are biased towards state B (for low q). And if $d_B < d_A$ and $\hat{\gamma} > \gamma$, the beliefs are still biased towards B (against the majority of dogmatic friends).

This completes the proof.

B Proof of Proposition 2

Suppose $\hat{\gamma} < \gamma$. Recall that i has d_A A-dogmatic friends, d_B B-dogmatic friends and n normal friends.

Denote the number of signals $s = a$ received by:

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

Denote the number of signals $s = b$ received by the same agents symmetrically:

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

Then the number of empty signals $s = \emptyset$ received by the same agents is given by:

- agent i – it is $(T - a_i - b_i)$
- dogmatic-left friends of agent i – it is $(T - a_j^A - b_j^A)$, $j \in \{1, 2, \ldots, d_A\}$
- dogmatic-right friends of agent i – it is $(T - a_j^B - b_j^B)$, $j \in \{1, 2, \ldots, d_B\}$
- normal friends of agent i – it is $(T - a_j^N - b_j^N)$, $j \in \{1, 2, \ldots, n\}$

Additionally, i's A-dogmatic friend j stayed silent b_j^A times, whereas her B-dogmatic friend k stayed silent a_j^B times. Thus, i's posterior belief will be given by:

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

where

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

This belief converges to one with probability 1 as $T \to \infty$ if and only if

$$\left(\frac{1-q}{q}\right)^M \left(\frac{\hat{\gamma}(1-q) + (1-\hat{\gamma})}{\hat{\gamma}q + (1-\hat{\gamma})}\right)^S$$

converges to zero with probability 1 as $T \to \infty$. This is equivalent to requiring that the natural log of this expression converges to $-\infty$ with probability 1 as $T \to \infty$, where the natural log equals $\ln\left(\frac{1-q}{q}\right) K(\mathbf{x}, T; q, \hat{\gamma})$, where

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

after we define

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

and

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

Given $\ln\left(\frac{1-q}{q}\right) < 0$, we require that $K(\mathbf{x}, T; q, \hat{\gamma})$ converges to $+\infty$ with probability 1 as $T \to \infty$. Now note that

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

Let $Q(\omega)$ be the conditional probability of signal $s = a$ when the true state is ω. It equals q if $\omega = A$ and $(1-q)$ if $\omega = B$.

By the Law of Large Numbers, we have that

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

Given this, $\operatorname{plim}_{T\to\infty}K(\mathbf{x},T;q,\hat{\gamma})=+\infty$ if and only if

$$-(1+n+(1+z(q,\hat{\gamma}))d_B)-\frac{z(q,\hat{\gamma})}{\gamma}(d_A-d_B)+\Big(2(1+n)+(d_A+d_B)(1+z(q,\hat{\gamma}))\Big)Q(\omega)>0,$$

which is equivalent to

$$Q(\omega)>\frac{1+n+(1+z(q,\hat{\gamma}))d_B+\frac{z(q,\hat{\gamma})}{\gamma}(d_A-d_B)}{2(1+n)+(d_A+d_B)(1+z(q,\hat{\gamma}))}:=\bar{q}_2(q).$$

Note that

$$\bar{q}_2(q)=\frac{1}{2}+\frac{((2-\gamma)z(q,\hat{\gamma})-\gamma)\,(d_A-d_B)}{2\gamma\,(2(1+n)+(d_A+d_B)(1+z(q,\hat{\gamma})))}$$

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

$$\begin{aligned}
\lim_{q\to\frac{1}{2}}\bar{q}_2(q) &= \frac{1}{2}+\frac{\left((2-\gamma)\frac{\hat{\gamma}}{2-\hat{\gamma}}-\gamma\right)(d_A-d_B)}{2\gamma\left(2(1+n)+(d_A+d_B)\left(1+\frac{\hat{\gamma}}{2-\hat{\gamma}}\right)\right)}\\
&= \frac{1}{2}+\frac{(\hat{\gamma}-\gamma)\,(d_A-d_B)}{\gamma(2-\hat{\gamma})(2(2-\hat{\gamma})(1+n)+2(d_A+d_B))}.
\end{aligned}$$

Consider the situation where the true state is $\omega=B$. Then $Q(\omega)=1-q$, so the inequality between $Q(\omega)$ and $\bar{q}_2(q)$ becomes

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

At $q=\frac{1}{2}$, the inequality becomes

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

If $d_A > d_B$ and $\hat{\gamma} < \gamma$, then the inequality holds. This implies that there exists $\bar{q}_2$ such that if $q \in \left(\frac{1}{2}, \bar{q}_2\right)$, the agent's beliefs converge to state A when the true state is B.

We will show further in the appendix that $\bar{q}_2(q)$ is convex in q. This is what Lemmas 6 and 7 and the section around them deal with.

Given that $\bar{q}_2(q)$ is convex in q and that $\lim_{q \to 1} \bar{q}_2(q) < 1$, it follows that $\bar{q}_2$ is the unique upper bar for q for which the beliefs converge to state A.

On the other hand, if $d_A > d_B$ but $\hat{\gamma} > \gamma$, then the inequality doesn't hold. Together with convexity and $\lim_{q \to 1} \bar{q}_2(q) < 1$, this implies that the agent's beliefs converge to state B when the true state is B for any $q > \frac{1}{2}$.

Now consider the case where the true state is $\omega = A$. Then $Q(\omega) = q$, so the inequality between $Q(\omega)$ and $\bar{q}_2(q)$ becomes

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

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

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

If $d_A > d_B$ and $\hat{\gamma} < \gamma$, then this inequality holds. Together with convexity and $\lim_{q \to 1} \bar{q}_2(q) < 1$, this implies that the agent's beliefs converge to state A for any $q > \frac{1}{2}$.

On the other hand, if $d_A > d_B$ but $\hat{\gamma} > \gamma$, then the inequality doesn't hold. Together with convexity and $\lim_{q \to 1} \bar{q}_2(q) < 1$, this implies that there exists a unique upper bound $\bar{q}_2'$ such that if $q \in \left(\frac{1}{2}, \bar{q}_2'\right)$, the agent's beliefs converge to state B when the true state is A.

If $\gamma = \hat{\gamma}$, then $Q(\omega) = \bar{q}_2(q)$ holds at $q = \frac{1}{2}$. Together with convexity and $\lim_{q \to 1} \bar{q}_2(q) < 1$, this implies that the agent's beliefs converge to the true state regardless of which state is true.

Proof that $\bar{q}_2$ is unique.

We would like to prove that $\bar{q}_2(q)$ is convex in q for all $q > \frac{1}{2}$. This, along with $\bar{q}_2(\frac{1}{2}) > \frac{1}{2}$ and $\bar{q}_2(1) < 1$, will imply that $\bar{q}_2(q) = q$ has a unique solution. For the purposes of this argument, we will still assume $d_A > d_B$ and a corresponding functional form of $\bar{q}_2(q)$.

First, we need to prove the following lemma:

Lemma 6. *$z(q,\gamma)$ is weakly decreasing for any $q \in \left(\frac{1}{2}; 1\right)$.*

Proof of Lemma 6. Consider the derivative of $z(q,\hat{\gamma})$ w.r.t. q. After simplifying, we obtain:

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

Note that $\lim_{q\to\frac{1}{2}} z(q,\gamma) = \frac{\gamma}{2-\gamma} > 0 = z(1,\gamma)$. As $z(q,\gamma)$ is continuously differentiable for $q \in \left(\frac{1}{2},1\right)$, it is enough to prove that there are no local maxima on that interval in order to show that $\frac{\partial z}{\partial q} \leq 0$ holds on that interval. At an intermediate local maximum, $\frac{\partial z}{\partial q} = 0$ must hold. Consider:

$$\begin{aligned}
\frac{\partial z}{\partial q} = 0 \;\Rightarrow\; \left(\frac{1-\gamma}{q(1-q)}+\gamma^2\right)z(q,\gamma)-(2-\gamma)\gamma &= 0 \\
z(q,\gamma) &= \frac{\gamma(2-\gamma)}{\gamma^2+\frac{1-\gamma}{q(1-q)}} \\
&\leq \frac{\gamma(2-\gamma)}{\gamma^2+\frac{1-\gamma}{\frac{1}{4}}} \\
&= \frac{\gamma(2-\gamma)}{(2-\gamma)^2} \\
&= \frac{\gamma}{2-\gamma}
\end{aligned}$$

Hence, $z(q,\gamma) \leq \frac{\gamma}{2-\gamma}$ must hold at any intermediate local maximum in $\left(\frac{1}{2},1\right)$. This immediately rules our the possibility that $z(q,\gamma)$ is increasing at $q = \frac{1}{2}$, since otherwise it would need to achieve a local maximum with value above $\frac{\gamma}{2-\gamma}$.

Consider again the implication of $\frac{\partial z}{\partial q} = 0$:

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

Note that the right-hand side is strictly decreasing in q for $q \in \left(\frac{1}{2};1\right)$. If $z(q,\gamma)$ was to decrease at first (as q rises from $\frac{1}{2}$) and then increase before going down to 0, the value of $z(q,\gamma)$ at the corresponding local maximum would be necessarily above the right-hand side. This is a contradiction, and thus $z(q,\gamma)$ cannot be strictly increasing for any $q \in \left(\frac{1}{2};1\right)$.

One final case to rule out is that $z(q,\gamma)$ is decreasing at first, passing through a local minimum, and then is increasing all the way until $q = 1$. That, however, would mean that the value at the local minimum is less than $z(1,\gamma)$, which is equal to 0. Since $z(q,\gamma) > 0$ for any $q \in (\frac{1}{2};1)$ and $\gamma \in (0;1)$, this case is also impossible.

These considerations prove that $z(q,\gamma)$ is weakly decreasing for any $q \in \left(\frac{1}{2};1\right)$. ■

Given Lemma 6, consider:

$$\begin{aligned}\bar{q}_2(q) &= \frac{1+n+(1+z(q,\hat{\gamma}))d_B + \frac{z(q,\hat{\gamma})}{\gamma}(d_A - d_B)}{2(1+n)+(d_A+d_B)(1+z(q,\hat{\gamma}))} \\ &= \frac{A+Bz(q,\hat{\gamma})}{C+Dz(q,\hat{\gamma})}\end{aligned}$$

where

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

Therefore, we can represent

$$\bar{q}_2(q) = \frac{B}{D} + \frac{A - \frac{BC}{D}}{C + D\cdot z(q,\gamma)} = \frac{B}{D} + \frac{AD-BC}{D(C+Dz(q,\hat{\gamma}))}.$$

Note that

$$\begin{aligned}AD - BC &= (1+n+d_B)(d_A+d_B) - \left(d_B + \frac{d_A-d_B}{\gamma}\right)(2+2n+d_A+d_B) \\ &= (1+n)(d_A-d_B) - \frac{d_A+d_B}{\gamma}(d_A-d_B) - \frac{2(1+n)}{\gamma}(d_A-d_B) \\ &= \left(1+n-\frac{d_A+d_B}{\gamma} - \frac{2(1+n)}{\gamma}\right)(d_A-d_B).\end{aligned}$$

This is strictly positive if $d_A < d_B$ and strictly negative if $d_A > d_B$.

$\bar{q}_2(q)$ is convex (concave) in q if and only if $\frac{AD-BC}{D(C+Dz(q,\hat{\gamma}))}$ is convex (concave) in q. As Lemma 7 shows below,

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

It then follows that $\bar{q}_2(q)$ is convex in q if and only if $AD - BC > 0$. Otherwise, $\bar{q}_2(q)$ is (weakly) concave in q. Additionally, note that $g'\left(\frac{1}{2}\right) = 0$ because of $z_q\left(\frac{1}{2},\hat{\gamma}\right) = 0$. Hence, regardless of whether $\bar{q}_2(q)$ is concave or convex in q for all $q \in \left(\frac{1}{2},1\right)$, it will follow that the upper bar $\bar{q}_2$ is unique.

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

Proof of Lemma 7. Note:

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

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

If we are able to prove that $z_{qq}(q,\hat{\gamma}) < 0$ for all $q \in \left(\frac{1}{2}, 1\right)$, then we will have proven that $g''(q) > 0$, implying that $\bar{q}_2(q)$ is convex for all $q \in \left(\frac{1}{2}, 1\right)$.

Recall from Lemma 6:

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

Let $K(q) = \frac{1}{\ln^2\left(\frac{1-q}{q}\right)\left(\hat{\gamma}^2 q(1-q)+(1-\hat{\gamma})\right)^2}$. Then:

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

Let

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

Then we can write

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

Consider $C_1(q)$:

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

Therefore,

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

Similarly, consider:

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

Therefore, we can write:

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

Let

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

and

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

Then we have

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

Note:

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

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, observe:

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

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

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

Therefore, we can conclude

$$D_1(q) \leq \frac{1-z(q,\hat{\gamma})}{2-z(q,\hat{\gamma})}E(q) < 0 \text{ for any } q \in \left(\frac{1}{2},1\right).$$

Returning to $D_2(q)$, note:

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

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 that it 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.

Combining $D_1(q) < 0$ and $D_2(q) < 0$ (for any $q \in \left(\frac{1}{2}, 1\right)$ and recalling

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

we can conclude that $z_{qq}(q, z(q,\hat{\gamma})) < 0$ for any $q \in \left(\frac{1}{2}, 1\right)$.[14]

Returning to the function $\bar{q}_2(q)$, recall that we showed it is convex in q if $z(q,\hat{\gamma})$ is concave. Given that $z_{qq}(q, z(q,\hat{\gamma})) < 0$, it follows that $\bar{q}_2(q)$ is convex.

■

Proof of Proposition 3

Here we will prove that $\bar{q}_2$ is increasing in d_A and γ and decreasing in d_B, n and $\hat{\gamma}$.

Consider the case where the true state is $\omega = B$ and $\hat{\gamma} < \gamma$, with $d_A > d_B$ (for any bias to take place). The value of $\bar{q}_2$ is determined by the equation

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

We can rewrite it as

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

The right-hand side is strictly decreasing in d_B, n and $\hat{\gamma}$ (while the left-hand side is unaffected), which implies that the fixed point of this equation, $\bar{q}_2$, is also decreasing in these variables. The right-hand sie is also increasing in γ, so $\bar{q}_2$ must be increasing in γ.

Finally, consider d_A. Return to the equation

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

Denote the right-hand side by h and take the derivative w.r.t. d_A:

$$\begin{aligned}
\frac{\partial h}{\partial d_A} &= \frac{\frac{z(q,\hat{\gamma})}{\gamma}(2+2n+(d_A+d_B)(1+z(q,\hat{\gamma}))) - \left(1+n+(1+z(q,\hat{\gamma}))d_B + \frac{z(q,\hat{\gamma})}{\gamma}(d_A - d_B)\right)(1+z(q,\hat{\gamma}))}{(2+2n+(d_A+d_B)(1+z(q,\hat{\gamma})))^2} \\
&= \frac{\frac{2z(q,\hat{\gamma})}{\gamma}d_B + (1+n)\left(\frac{z(q,\hat{\gamma})}{\gamma} - 1 - z(q,\hat{\gamma})\right) - (1+z(q,\hat{\gamma}))^2 d_B}{(2+2n+(d_A+d_B)(1+z(q,\hat{\gamma})))^2} \\
&= \frac{\left((2-\gamma)z(q,\hat{\gamma}) - \gamma(1+z^2(q,\hat{\gamma}))\right)d_B + (1+n)\left((1-\gamma)z(q,\hat{\gamma}) - \gamma\right)}{\gamma(2+2n+(d_A+d_B)(1+z(q,\hat{\gamma})))^2} < 0,
\end{aligned}$$

since $(2-\gamma)z(q,\hat{\gamma}) < \gamma < \gamma(1+z^2(q,\hat{\gamma}))$ and $(1-\gamma)z(q,\hat{\gamma}) < \gamma$. Given that h is decreasing in d_A, it follows that $1 - \bar{q}_2$ is decreasing in d_A, which in turn implies that $\bar{q}_2$ is increasing in d_A.

[14] This is due to the fact that $K(q) > 0$.

Now consider the case where the true state is $\omega = A$ and $\hat{\gamma} > \gamma$, with $d_A > d_B$ (for any bias to take place). The value of $\bar{q}_2$ is determined by the equation

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

We can rewrite it as

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

As previously, the right-hand side is strictly decreasing in n and d_B and increasing in d_A. It is now increasing in $\hat{\gamma}$ and decreasing in γ. The arguments to show these are exactly symmetrical to what was done in the case $\omega = B$ and $\hat{\gamma} < \gamma$.

C Proof of Proposition 4

To prove the result, we need to consider the implicit equation that defines $\bar{q}_2$. If the true state is A, it is:

$$q = \bar{q}_2(q) \;\Leftrightarrow\; q = \frac{1 + n + (1 + z(q, \hat{\gamma}))d_B + \frac{z(q,\hat{\gamma})}{\gamma}(d_A - d_B)}{2(1+n) + (d_A + d_B)(1 + z(q, \hat{\gamma}))} = \frac{1}{2} + \frac{((2-\gamma)z(q,\hat{\gamma})-\gamma)(d_A - d_B)}{2\gamma(2(1+n)+(d_A+d_B)(1+z(q,\hat{\gamma})))}$$

In state A, bias in beliefs occurs when $\hat{\gamma} > \gamma$ and $d_A > d_B$, or $\hat{\gamma} < \gamma$ and $d_A < d_B$. In either case, we have $((2-\gamma)z(q, \hat{\gamma}) - \gamma)(d_A - d_B) > 0$, which will be important in a moment.

Consider the right-hand side of the equation above when d_A, d_B and n increase by a factor of $\lambda > 1$:

$$\begin{aligned}
\frac{1}{2} + \frac{((2-\gamma)z(q,\hat{\gamma})-\gamma)(\lambda d_A - \lambda d_B)}{2\gamma(2(1+\lambda n)+(\lambda d_A+\lambda d_B)(1+z(q,\hat{\gamma})))} &> \frac{1}{2} + \frac{((2-\gamma)z(q,\hat{\gamma})-\gamma)(\lambda d_A - \lambda d_B)}{2\gamma(2(\lambda+\lambda n)+(\lambda d_A+\lambda d_B)(1+z(q,\hat{\gamma})))} \\
&= \frac{1}{2} + \frac{((2-\gamma)z(q,\hat{\gamma})-\gamma)(d_A - d_B)}{2\gamma(2(1+n)+(d_A+d_B)(1+z(q,\hat{\gamma})))} \\
&= \bar{q}_2(q).
\end{aligned}$$

The inequality above is true because $((2-\gamma)z(q, \hat{\gamma}) - \gamma)(d_A - d_B) > 0$.

Hence, the right-hand side of the equation $q = \bar{q}_2(q)$ increases when d_A, d_B and n all increase by a factor of $\lambda > 1$, which means that we need a higher q on the left-hand side to satisfy the equation. Therefore, if $\bar{q}_2(d_A, d_B, n, \gamma)$ solves $q = \bar{q}_2(q)$, then

$$\bar{q}_2(\lambda d_A, \lambda d_B, \lambda n) > \bar{q}_2(d_A, d_B, n).$$

A similar result can be shown in the case of state B, given that it is largely symmetric.

Now consider arbitrary $\lambda_A > 1$, $\lambda_B > 1$ and $\lambda_N > 1$ and look at the expression for $\bar{q}_2$. Once again, assume true state is A:

$$\bar{q}_2(\lambda_A d_A, \lambda_B d_B, \lambda_N n) = \frac{1}{2} + \frac{((2-\gamma)z(q, \hat{\gamma}) - \gamma)\,(\lambda_A d_A - \lambda_B d_B)}{2\gamma\,(2(1+\lambda_N n) + (\lambda_A d_A + \lambda_B d_B)(1 + z(q, \hat{\gamma})))}$$

The inequality $\bar{q}_2(d_A, d_B, n) \leq \bar{q}_2(\lambda_A d_A, \lambda_B d_B, \lambda_N n)$ will hold if and only if the following holds:

$$\begin{aligned}
\frac{((2-\gamma)z(q,\hat{\gamma})-\gamma)\,(\lambda_A d_A - \lambda_B d_B)}{2\gamma\,(2(1+\lambda_N n)+(\lambda_A d_A+\lambda_B d_B)(1+z(q,\hat{\gamma})))} &\geq \frac{((2-\gamma)z(q,\hat{\gamma})-\gamma)\,(d_A-d_B)}{2\gamma\,(2(1+n)+(d_A+d_B)(1+z(q,\hat{\gamma})))} \\
\frac{((2-\gamma)z(q,\hat{\gamma})-\gamma)\,(\lambda_A d_A - \lambda_B d_B)}{2(1+\lambda_N n)+(\lambda_A d_A+\lambda_B d_B)(1+z(q,\hat{\gamma}))} &\geq \frac{((2-\gamma)z(q,\hat{\gamma})-\gamma)\,(d_A-d_B)}{2(1+n)+(d_A+d_B)(1+z(q,\hat{\gamma}))}
\end{aligned}$$

First, consider the case where $\hat{\gamma} > \gamma$ (with $d_A > d_B$). Then $(2-\gamma)z(q,\hat{\gamma})-\gamma > 0$ in the neighborhood of $q = \frac{1}{2}$, and we can divide by it.

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

Now consider the case where $\hat{\gamma} < \gamma$ (with $d_B > d_A$). Then $(2-\gamma)z(q,\hat{\gamma})-\gamma < 0$, so the inequality sign changes when we divide by the expression. Following similar steps, we will get:

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

If the true state is B, the same proof delivers a symmetric result. We omit it here.

D Proof of Proposition 5

Assume that the true state is A and we have $\hat{\gamma} < \gamma$ and $d_B > d_A$. Recall that $q = \bar{q}_2(d_A, d_B, n)$ is defined by

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

Fix d_A, d_B, n, λ and $\hat{q}$. We need to find λ_N such that

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

Note that the right-hand side is decreasing in $z(q,\hat{\gamma})$. So a sufficient condition would be to impose the inequality for the lowest value of $z(q,\hat{\gamma})$, which is 0. This gives us the following inequality:

$$\begin{aligned}
\hat{q} &\geq \frac{1}{2} + \frac{\frac{1}{2}\lambda(d_B - d_A)}{2 + 2\lambda_N n + \lambda(d_A + d_B)} \\
2\hat{q} - 1 &\geq \frac{\lambda(d_B - d_A)}{2 + 2\lambda_N n + \lambda(d_A + d_B)} \\
(2\hat{q}-1)(2 + 2\lambda_N n + \lambda(d_A + d_B)) &\geq \lambda(d_B - d_A) \\
(2\hat{q}-1)(2 + \lambda(d_A + d_B)) + 2(2\hat{q}-1)\lambda_N n &\geq \lambda(d_B - d_A) \\
\lambda_N &\geq \frac{2(1-\hat{q})\lambda d_B - 2(2\hat{q}-1) - 2\hat{q}\lambda d_A}{2(2\hat{q}-1)n} \\
\lambda_N &\geq \frac{(1-\hat{q})\lambda d_B - \hat{q}\lambda d_A - (2\hat{q}-1)}{(2\hat{q}-1)n} \\
\lambda_N &\geq \frac{(1-\hat{q})d_B - \hat{q}d_A}{(2\hat{q}-1)n}\lambda - \frac{1}{n} \\
\lambda_N &\geq \frac{d_B - \hat{q}(d_A + d_B)}{(2\hat{q}-1)n}\lambda - \frac{1}{n}.
\end{aligned}$$

Now consider the case where true state is B and we have $\hat{\gamma} < \gamma$ and $d_A > d_B$. Fix d_A, d_B, n, λ and $\hat{q}$. We need to find λ_N such that

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

Following the same steps as above (just switching d_A and d_B), we will get the following sufficient condition:

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

E Proof of Proposition 6

First, we will quickly prove Lemma 1.

Proof of Lemma 1. Consider agent i who is "eventually correct" at a given $q = \hat{q}$. There are two possibilities: either $\bar{q}_{2i} > \hat{q}$ or $\bar{q}_{2i} < \hat{q}$ (we omit the knife-edge case). If $\bar{q}_{2i} > \hat{q}$, then the agent must have dogmatic imbalance towards the correct state. Increasing q beyond $\bar{q}_{2i}$ will lead to the agent learning correctly and having beliefs converge to the same state as before. Hence, the agent will not leave the set of "eventually correct" agents as q increases. If $\bar{q}_{2i} < \hat{q}$, then the agent is already learning correctly, and increasing q further will not alter her asymptotic beliefs. Thus, she will remain "eventually correct" as q goes up. All of this implies that the set of "eventually correct" agents is not contracting as q increases.

Now consider agent j who is "eventually incorrect" at a given $q = \hat{q}$. This can only occur if the agent is learning incorrectly, meaning she has a dogmatic imbalance towards the wrong state that is sufficiently large, i.e. $\bar{q}_{2i} > \hat{q}$. Increasing q beyond $\bar{q}_{2i}$ will make the agent begin to learn correctly, which means she will leave the "eventually incorrect" set of agents and will now be in the "eventually correct" set. This finishes the proof of the lemma.

■

Fix some $q = \hat{q}$ and considers sets $\mathcal{N}_A(q)$ and $\mathcal{N}_B(q)$. For definiteness, let $\omega = A$ be the true state. Let $\bar{q}_{min}(\hat{q}) = \min_{i \in \mathcal{N}} \{\bar{q}_{2i} \mid \bar{q}_{2i} > \hat{q}\}$ be the lowest $\bar{q}_{2i}$ among agents who are "eventually incorrect" at $q = \hat{q}$. As q increases and reaches $\bar{q}_{min}(\hat{q})$, that agent will flip from being "eventually incorrect" to being "eventually correct". Since $\omega = A$ is the true state, this implies

$$|\mathcal{N}_A(\bar{q}_{min}(\hat{q}))| = |\mathcal{N}_A(\hat{q})| + 1 \quad \text{and} \quad |\mathcal{N}_B(\bar{q}_{min}(\hat{q}))| = |\mathcal{N}_B(\hat{q})| - 1.$$

Consider the network polarization $\Pi(q)$ at $q = \hat{q}$ and $q = \bar{q}_{min}(\hat{q})$:

$$\Pi(\hat{q}) = \frac{4}{|\mathcal{N}|} \cdot |\mathcal{N}_A(\hat{q})||\mathcal{N}_B(\hat{q})|$$
$$\Pi(\bar{q}_{min}(\hat{q})) = \frac{4}{|\mathcal{N}|} \cdot (|\mathcal{N}_A(\hat{q})| + 1)\,(\mathcal{N}_B(\hat{q}) - 1)$$

Note that $\Pi(\hat{q}) \geq \Pi(\bar{q}_{min}(\hat{q}))$ if and only if

$$|\mathcal{N}_A(\hat{q})||\mathcal{N}_B(\hat{q})| \geq (|\mathcal{N}_A(\hat{q})| + 1)\,(\mathcal{N}_B(\hat{q}) - 1)\,,$$

which is equivalent to

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

Hence, the network polarization weakly decreases with q if and only if the set of "eventually incorrect" agents initially (at $q = \hat{q}$) is smaller than the set of "eventually correct" agents plus one. Since $\mathcal{N}_B(q)$ is weakly contracting in q, a necessary and sufficient condition for $\Pi(q)$ to be always weakly decreasing in q is that $|\mathcal{N}_B(\frac{1}{2})| = |\mathcal{D}_R|$ is weakly smaller than $|\mathcal{N}| - |\mathcal{D}_B| + 1$, which is equivalent to $|\mathcal{D}_B| \leq \frac{1}{2}(|\mathcal{N}| + 1)$.

Therefore, the network polarization $\Pi(q)$ is weakly decreasing in q over $q \in \left(\frac{1}{2}, 1\right)$ if and only if $|\mathcal{D}_{-\omega}| \leq \frac{1}{2}(|\mathcal{N}| + 1)$.

F Proof of Proposition 7

The proof of this proposition is short, and requires us to prove that by aggregating enough signals, we can push quality of information above any given threshold. In other words, if $\hat{s}^i_{Mt}$ denotes an M-aggregated signal, we need to prove

$$\lim_{M \to \infty} \mathbb{P}(\hat{s}^i_{Mt} = a | \omega = A) = \lim_{M \to \infty} \mathbb{P}(\hat{s}^i_{Mt} = b | \omega = B) = 1.$$

Suppose that we aggregate M (odd number) signals together, and offer agents the following information structure:

$$\hat{s}^i_{Mt} = \begin{cases} 0, \text{ if } \sum_{t=1}^{M} s_{it} < \frac{M}{2} \\ 1, \text{ if } \sum_{t=1}^{M} s_{it} > \frac{M}{2} \end{cases}$$

For simplicity, assume that the true state is $\omega = A$, and consider the quality of signal $\hat{s}^i_{Mt}$:

$$\begin{aligned}
\mathbb{P}\left(\hat{s}^i_{Mt} = a \middle| \omega = L\right) &= \mathbb{P}\left(\sum_{t=1}^{M} s_{it} < \frac{M}{2} \middle| \omega = L\right) \\
&= \mathbb{P}\left(\frac{1}{M}\sum_{t=1}^{M} s_{it} < \frac{1}{2} \middle| \omega = L\right) \\
&= 1 - \mathbb{P}\left(\frac{1}{M}\sum_{t=1}^{M} s_{it} \geq \frac{1}{2} \middle| \omega = L\right) \\
&= 1 - \mathbb{P}\left(\frac{1}{M}\sum_{t=1}^{M} s_{it} - (1-q) \geq \frac{1}{2} - (1-q) \middle| \omega = L\right) \\
&= 1 - \mathbb{P}\left(a_M - (1-q) \geq \frac{1}{2} - (1-q) \middle| \omega = L\right) \\
&\overset{\text{as } M\to\infty}{\to} 1 - 0 = 1,
\end{aligned}$$

where the last line is due to the weak Law of Large Numbers on a Binomial random variable a_M with parameters M and $(1-q)$. Hence, the limit of signal's quality as $M \to \infty$ is equal to 1, meaning that there must exist a finite number of signals such that the quality surpasses the bound $\bar{q}_1$. For such aggregated signal, information quality is too high for Proposition 2 to apply, which implies that there is no divergence of beliefs in the network.