News We Like to Share: How News Sharing on Social Networks Influences Voting Outcomes*

Kirill Pogorelskiy
University of Warwick†

Matthew Shum
Caltech‡

January 17, 2018

Abstract

We study the relationship between news sharing on social media and information aggregation by voting. Our context-neutral laboratory experimental treatments mimic the features of social networks in the presence of media bias to address concerns that voters obtaining their political news via social media may become more polarized in their voting behavior. Our results suggest substantial polarization: subjects selectively share news that is favorable to their party and do not account for biased news signals in their voting decisions. Overall, subjects behave as if news sharing and voting is expressive of their induced partisanship even though by design, their preferences have a common value component. Given these patterns of individual behavior, social networks amplify the underlying quality of the shared news: with unbiased media, they raise collective decision making efficiency, but efficiency deteriorates markedly in the presence of media bias, as news signals become less reliable.

JEL codes: C72, C91, C92, D72, D83, D85

Keywords: social networks, news sharing, voting, filter bubbles, media bias, polarization, lab experiment

**First draft: May 22, 2017. Current version is available at ssrn.com/abstract=2972231.* We thank Marina Agranov, Simon Anderson, Matthew Chao, Dan Bernhardt, Khai X. Chiong, Alejandro Robinson, Ruben Enikolopov, Guillaume Frechette, Jens Grosser, Marlene Guraieb, Asen Ivanov, Dimitri Landa, Aniol Llorente-Saguer, Matt Mitchell, Pietro Ortoleva, Tom Palfrey, Wolfgang Pesendorfer, Maria Petrova, Charlie Plott, Anja Prummer, Ronny Razin, Pedro Robalo, Daniel Sgroi, Erik Snowberg, Andis Sofianos, Francesco Squintani, Leeat Yariv, and audiences at QMUL, WBS DR@W forum, Political Economy Webinar, Florida State, NYU, Oxford CESS (Nuffield), Aix Marseille, the 2016 North American ESA Meetings, the Zhejiang University IO Conference, the University of Konstanz 2017 Workshop "Political Economy: Theory meets Empirics", the 15th Annual Media Economics Workshop in Barcelona, and the 2017 Lancaster Game Theory Conference for useful comments and discussions. Chris Crabbe provided expert assistance in designing the experimental software, and Andis Sofianos and Mahnaz Nazneen helped run the sessions. We are grateful to Warwick Economics for financial support, and to Umar Taj and Gallup Pakistan for administering our questions in their survey. This paper was previously circulated as "News Sharing and Voting on Social Networks: An Experimental Study."

†Department of Economics, k.pogorelskiy@warwick.ac.uk

‡Division of the Humanities and Social Sciences, mshum@caltech.edu

1 Introduction

Social media like Twitter and Facebook have become a primary source of news and political discourse for an increasing share of voters.[1] This tendency can have dramatic policy consequences, as social media, faced with abundance of shared information, filter their displayed content so that users only see news and advertising related to what they themselves or their friends have responded to. As one's friends tend to be similar to oneself, social media users may only encounter viewpoints similar to their own, effectively staying inside their personal "filter bubble", enforcing a biased view. As one media observer colorfully puts it, "personalization filters serve up a kind of invisible autopropaganda, indoctrinating us with our own ideas" (Pariser, 2011).

Do such features of social media affect voting behavior? Do elections aggregate information efficiently in the presence of social media? In June 2016, in a historic "Brexit" referendum, the UK voted to leave the European Union, contrary to predictions of opinion polls, economic experts, trading markets, and even the Brexit supporters themselves. The very unexpectedness of the outcome suggested a filter bubble explanation: younger voters, who use social media relatively more, and who overwhelmingly supported Remain, might have become convinced that the actual majority supports Remain and therefore turned out in smaller numbers than older voters, who use social media less, and mostly voted for Leave. In November 2016, a largely shocking victory of Donald Trump in the U.S. Presidential elections followed a similar pattern, and led to discussion of a "hidden majority" (outside the filter bubble of the mainstream media outlets) who voted Trump into office.[2]

To date there has been little research formally analyzing and quantifying the impact of social media on election outcomes. In this paper, we seek to fill this gap. Our main contribution is to integrate news sharing via social networks with a voting framework in a controlled laboratory setting. This allows us to examine how the combination of the limited communication possibilities in social media along with media bias, may distort news content shared in social networks and, ultimately, influence electoral outcomes.

We construct an artificial, context-neutral, social media environment in the laboratory, where we have control over the most important and subtle factors of the process: voters' preferences, social network, and media bias. We focus on information aggregation by voting in a two-candidate election. There are two states of the world and two candidates, each of whom can be the best choice depending on the realized state. There are two groups of voters who each support a different candidate, and even though both groups want to elect the "correct" candidate in the "correct" state, the supporters of the elected candidate enjoy a much larger payoff than the supporters of the non-elected candidate, provided that the "correct" candidate is elected. Voters obtain private

[1] According to a June 2015 Pew Research Center report, 61% of Americans aged 18-33 get political news from Facebook in a given week, in contrast with 39% of Baby Boomers (see http://www.journalism.org/2015/06/01/millennials-political-news). See also a more recent September 2017 version focused on the structure of social media use at http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/.

[2] See, for example, http://nymag.com/selectall/2016/11/donald-trump-won-because-of-facebook.html and http://www.nytimes.com/2016/11/15/opinion/mark-zuckerberg-is-in-denial.html.

signals about the state ("news") and can share them with other voters. Social media in this context restrict the communication protocol. Specifically, voters can only share their news with people that they are "connected" to in the social network, but not with those who are outside, as in the news sharing mechanism on Facebook.

In the experiment, we study treatments varying in two inter-related dimensions: social networks, which determine the type of news sharing audience, and media bias, which determines the accuracy of individual news pieces (private signals). Since communication affects information available at the voting stage, it is important to distinguish between different models of communication at the preceding news sharing (signalling) stage. Broadly, we design the experiment to not only measure the effects of news sharing on voting outcomes, but also shed light on voters' motivation in sharing news. We investigate two network treatments (complete network, in which everyone is connected to everyone else, and polarized network, in which each candidate's supporters are put together in a separate network component) relative to the control of the empty network. We also vary the degree of media bias across treatments.

A key finding from our data is that, across social network and media bias treatments, subjects consistently share news that is favorable to their preferred candidate more often than the news unfavorable to their candidate. So we clearly reject a model of communication in which voters naïvely reveal all information; they are at least somewhat sophisticated. But how sophisticated? To determine this, we conducted sessions in which the probability of not receiving a signal was zero, preventing the subjects from hiding unfavorable signals under the disguise of an empty signal. We find, surprisingly, that subjects' signalling strategies, on average, remain the same; in particular, the frequency of hiding unfavorable signals is the same across these two treatments. This suggests that subjects' news sharing behavior is not as sophisticated as predicted by a Bayesian equilibrium, in which backward-induction reasoning about how their news will affect others' votes guides behavior. Rather, they behave as if news sharing and voting is expressive of their induced partisanship even though by design, their preferences have a common value component. Voters appear to obtain intrinsic utility from sharing favorable signals, and disutility from sharing unfavorable signals, with little, if any, regard for how the signals are subsequently acted upon. This "selective news sharing" behavior among our experimental subjects confirms anecdotal features of news sharing on real-world social networks.[3]

We also find that social networks amplify the underlying quality of the shared news: Without media bias, social networks raise collective decision making efficiency. This positive effect, while predicted by the theory, is partly mechanical, because without media bias, the news shared by voters is sufficiently accurate about the realized state of the world. In some of our media bias treatments, however, unfavorable signals are actually more informative than the favorable ones (they move the posterior belief about the state farther away from the prior of 1/2), but they are also unlikely to occur due to biased signal sampling. From this point of view, media bias may

[3] E.g., An et al. (2014) report evidence of partisan news sharing among a subsample of Facebook and Twitter users with self-reported ideology, and Garz et al. (2018) report greater Facebook user engagement with congenial rather than with uncongenial, news about loss of immunity by German politicians investigated for crime offences.

not necessarily lead to less efficient electoral outcomes (and may actually lead to more efficient outcomes!), if voters correctly account for the sources of the signals and treat favorable signals differently from the unfavorable ones in their posteriors. However, our results strongly refute this possibility: when news quality deteriorates, we find that moderate media bias significantly reduces efficiency relative to any unbiased treatment, even the ones in which (in the case of the empty network) voters only have access to their own private signal. This suggests that voters by and large take signals at face value and fail to discount signal sources. This dimension of the failure of Bayesian updating seems different from what previous studies found, and implies that in the presence of bad information (e.g., "fake news"), social networks can be detrimental to information aggregation.

Related literature. Our results relate to several issues raised in the existing literature. First, a well-known study by DellaVigna and Kaplan (2007) uncovered the so-called "Fox news" effect, whereby television viewers who were exposed to the right-leaning Fox news stations tended to vote more for Republican candidates.[4] Against such a backdrop, social media may dilute the Fox news effect since people who obtain their news via social media are less likely to encounter viewpoints contrary to their own – that is, politically left-leaning people are less likely to be friends with the right-leaning people who constitute the core audience of the Fox news.[5] Consistently with this reasoning, we find that polarized networks are less efficient (in a Bayesian sense) than complete ones, especially in the presence of moderate media bias.

Second, newsfeed filtering by social media may increase political polarization. Biased, inaccurate views may be more persistent; rumors and half-truths may spread more quickly, since people who believe and spread rumors may be less likely to receive counterbalancing information contradicting these views. For example, Bakshy et al. (2015) find that Facebook users who indicate their ideology in their profiles tend to maintain relationships with ideologically similar friends, and that content-sharing is well-aligned with ideology. On Facebook, information sharing is largely done by "liking"; this terminology may tend to accentuate the tendency of users to share only news that is congruent with their preferences, consistently with the finding of selective signal sharing in our experiments.

Third, media bias at the source can directly affect the news and information that people share on social networks. Media source reputation concerns may endogenously generate the demand for slant, whereby news outlets bias their reports towards their consumers' prior beliefs.[6] Social media may amplify the existing bias if a biased news piece becomes more prominent as it is shared by a larger number of people.[7] Our results lend support to this interpretation of social media, as

[4]See also Martin and Yurukoglu (2017) and Enikolopov et al. (2011).

[5]See, however, Gentzkow and Shapiro (2011) who find relatively little online segregation by ideology, and Boxell et al. (2017) who question the hypothesis that social media is a primary driver of increasing polarization in the U.S.

[6]E.g., Gentzkow and Shapiro (2006). See also Mullainathan and Shleifer (2005). Other relevant theoretical studies include Bernhardt et al. (2008), Duggan and Martinelli (2011), Anderson and McLaren (2012), and Piolatto and Schuett (2015), among others.

[7]Empirical studies face significant challenge in identifying and measuring the media bias; see, e.g., Chiang and Knight (2011), Barberá et al. (2015). See also Groeling (2013) for an overview. Cross-platform aggregation indices capturing media power in a meaningful way have only started to appear very recently, e.g., Prat (2015); Kennedy and Prat (2017).

subjects by and large fail to take into account the sources of their signals.

Finally, there is a number of theoretical and lab experimental studies on information aggregation by voting that share some similarities with some of our treatments but either did not consider pre-vote communication within subgroups or modelled it as cheap talk; neither approach fully captures pertinent features of news sharing on social networks. We point out similarities and differences when we describe each our treatment in detail below.

Outline. The remainder of the paper is organized as follows. Section 2 describes the model primitives, the mapping of the primitives into experimental treatments (Subsections 2.1–2.2), and the experimental design and procedures (Subsection 2.3). Section 3 derives benchmark equilibrium predictions for each treatment. Section 4 presents the aggregate results and their interpretation. Section 5 continues with individual-level results and their interpretation. We supplement our lab experimental results with survey evidence from Pakistan as a robustness check in Section 6. The survey results are broadly consistent with our experimental findings. Section 7 concludes. Appendix A contains proofs of the theoretical propositions. Appendix B presents additional details. Appendix C contains experimental instructions.

2 The Model and Experimental Design

A substantive contribution of this paper is to study a model that integrates news sharing via social networks with a voting framework in a controlled laboratory setting. We extend the standard model of information aggregation by committee voting (Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1997, 1998)) to allow for news sharing on a social network prior to voting. In this model, two candidates, C_a and C_b, compete in an election. There are n voters, split evenly into two partisan groups, supporting candidates C_a and C_b, respectively. There are two equally likely states of the world, a, and b, with $\theta \in \{a, b\}$ the realized state, which can be interpreted as the identity of the "correct" candidate to elect. The voting rule is simple majority, with ties broken randomly. In treatments other than the control, voters are connected to each other in a social network. Communication among voters, described below, is network-restricted according to an exogenous network configuration.[8]

To focus on information aggregation, we assume a common value voting environment, in which voters are *weak partisans*, who are biased toward a particular candidate, but still care about choosing the "correct" candidate for the state, that is, candidate C_j in state $j, j \in \{a, b\}$.

Preferences. Voters' utilities are identical within their respective groups and depend on the realized state of the world, θ, and the elected candidate, C. The utility function of each weak

[8] Thus, our setting features network-restricted *deliberation* among voters. Relevant theoretical work incorporating versions of deliberation in committees include Coughlan (2000), Austen-Smith and Feddersen (2005, 2006), Gerardi and Yariv (2007), Dickson et al. (2008), Meirowitz (2007), and Schulte (2010) among others. Battaglini (2017) extends a related Poisson framework to the case of public protests and finds that with precise enough signals, social media improve information aggregation.

C_j-partisan, $j \in \{a, b\}$, is $u_j(\theta, C)$, normalized as follows.

$$u_j(j, C_j) = 1.5, \qquad u_j(-j, C_{-j}) = 0.5, \qquad u_j(j, C_{-j}) = u_j(-j, C_j) = 0.15 \tag{1}$$

Here, $-j = \{a, b\} \setminus j$. By default, a C_j-partisan prefers to have C_j elected, but she will vote for C_{-j} if sufficiently sure that the realized state makes C_{-j} the correct candidate.

Information. Voter preferences over candidates are common knowledge. Each voter may receive a private conditionally independent signal s about the realized state of the world (a "news" item) from a media source, or receive no signal. We let $s \in \{s_a, s_b, s_\emptyset\}$ to cover both cases. When signals are informative, a signal s_a indicating that state a is more likely to have occurred, is favorable to C_a-partisans, since C_a is more likely to be the "correct" choice, translating into a higher expected payoff of C_a-partisans, but it is unfavorable to C_b-partisans, since C_b is less likely to be the "correct" choice, translating into a lower expected payoff, and vice versa for an s_b signal. There are two different media sources that distribute the signals, each group of supporters listens to a different source. We do not explicitly model media, they only serve as a device to deliver private signals. Furthermore, each media source may be "biased" in favor of one of the candidates in the sense of drawing signals that favor one candidate more often than the other in both states of the world. That is, a C_b-partisan may read newspapers that report the favorable signal s_b with a high probability in both states (or even regardless of the state), and vice versa for a C_a-partisan.

Timing and actions. We consider a game in which the news sharing stage is followed by the voting stage according to the following timeline:

1. State $\theta \in \{a, b\}$ is realized (voters do not observe it).

2. News signals are drawn from distributions that vary across the media bias treatments, as described below, and voters privately observe their signals.

3. Each voter who has received a non-empty signal (i.e. $s \neq s_\emptyset$) decides on whether or not to share the signal with her neighbors on the network. She cannot selectively choose who to share the signal with: either the signal is shared with everyone in her network or with no one. She also cannot lie about her signal if she decides to share it (no cheap talk). After everyone's signal sharing choice is carried out, voters observe signals shared with them, if any.

4. Voters choose between voting for C_a and voting for C_b.[9]

5. The candidate supported by the majority of votes is elected, and payoffs are realized. If there is a tie, the winner is selected by a random coin flip.

[9] Abstention is not allowed here to keep the design simple, but remains an important future extension.

2.1 Social network treatments

The existing literature on communication within social networks has emphasized how learning and information efficiency of social networks varies depending on their structure.[10] Since we focus on how news sharing via a network affects voting rather than the network itself, we study three network configurations in our experiments, which are "extreme" cases of networks:

SN0: Empty network (control). This is the default setting for the Condorcet Jury Theorem, without deliberation among jurors. In this treatment, there is no communication stage – after observing their private signal, all players go to the voting stage.[11]

SN1: Polarized network. The social network consists of two fully-connected components, where each one contains only supporters of one candidate. During the communication stage, voters can only share signals publicly with everyone in their party. SN1 captures the idea that in an extremely polarized society, political support is concentrated at the opposite ends of the spectrum, and communication is predominantly between likewise-minded people (which is one of the reasons extreme polarization is considered dangerous). At the same time, voters within each component, at least potentially, could be persuaded to vote differently from their default partisan choice, if they are sufficiently confident that the state favors the other candidate. Thus there is room for deliberation.[12]

SN2: Complete network. Every voter is connected to everyone else, so there is a channel for electorate-wide communication, and voters can share signals publicly with everyone, including supporters of the opposing candidate. This setting can be related to jury voting with deliberation (e.g., when communication between jurors is free-form) or polls (e.g., when communication is limited to vote intention revelation).[13]

The crucial feature in SN1 and SN2, dictated by our focus on news sharing, is that, unlike previous studies, shared signals are not cheap talk but "hard" evidence, i.e. they are always believed.[14]

[10] Homophily in social networks has received a lot of attention since McPherson et al. (2001). Some recent studies include, e.g., Golub and Jackson (2012), Baccara and Yariv (2013), Halberstam and Knight (2016).

[11] Already the first experiment of this kind – Guarnaschelli et al. (2000) – allowed for pre-play deliberation in the form of straw-polls. More recent studies, e.g., Elbittar et al. (2017), allowed voters to acquire unbiased signals at a cost; however, their signals could not be shared with others. See Martinelli and Palfrey (2017) for a recent survey of experimental results in collective decision games.

[12] Several papers also studied heterogeneous voters: e.g., Maug and Yilmaz (2002) had voters put into separate subgroups and Mengel and Rivas (2017) had different members receiving signals of different accuracies. However, in those papers voters were not allowed to communicate to each other. Le Quement and Yokeeswaran (2015) studied subgroup deliberation via cheap talk under unanimity, and found that theoretically, subgroup deliberation Pareto dominates no deliberation.

[13] This treatment also shares some features of other experimental papers studying communication in a voting setting, including Goeree and Yariv (2011), Le Quement and Marcin (2016), Buechel and Mechtenberg (2016). See also Bouton et al. (2017). Unlike those papers, we use verifiable signals rather than versions of cheap talk.

[14] While this property of signal sharing bears similarity to signals studied in the information disclosure literature (see, in particular, Jin et al. (2017)) the structure of incentives in disclosure games is very different from our setting – there are no common values, no voting stage, and (usually) two players, only one of whom can disclose a signal. Notwithstanding these differences, Jin et al. (2017) find that in a two-person sender-receiver game, senders disclose more favorable information more often than less favorable information, which may indicate robustness of the selective news sharing across contexts.

2.2 Media bias treatments

Given the popular and policy debates on the bias and reliability of news circulated on real-world social networks, we consider two different media bias treatments, which differ in the quality of the news signals that subjects in our experiments receive, relative to the no bias control:

MB0: No media bias. Signals are unbiased and informative. There is no news (empty signal) with probability r.

MB1: Extreme partisan bias. Each of the two media sources is biased towards the candidate from a particular party so it is more likely to report the news item about that candidate than about his opponent. Each voter receives the signal from the media source biased towards her preferred candidate. Each voter may also receive no signal, with the same probability as before. Moreover, the probability of getting a "favorable" signal is the *same* in both states – so the signals are uninformative (not correlated with the realized state). This is why we call this case "extreme" bias. The interpretation is that biased media highlight good news for their candidate, and hush good news for the opponent, and they do so all the time.[15]

MB2: Moderate partisan bias. Same as MB1, but while "favorable" signals are more likely to be reported in both states, there is a difference in probabilities that depends on the realized state. Thus, signals are informative, and more so if they are "unfavorable". In fact one "unfavorable" signal carries more weight in the posterior about the realized state than several "favorable" signals.

In all MB treatments, the positive probability of not receiving any signal is crucial. If each voter receives a signal with probability one, then her decision not to reveal a signal will be "recovered" by other voters, and may be interpreted as an attempt to suppress information unfavorable to her candidate. In order to see if voters take these strategic considerations into account, we experimentally check this case below.

The signal accuracy details for all MB treatments are described in Table 1, computed using the actual experimental parameters.

2.3 Experimental procedures

Our experiment combines three social network treatments and three media bias treatments, as well as two additional treatments with no-signal probability $r = 0$ and a non-empty network. We ran 17 sessions at the Warwick Business School in June, July, and October 2016. Sessions lasted between 50 and 80 minutes. For each session, we recruited two to three ten-person groups (i.e. the electorate size $n = 10$), each group split into two equal-sized subgroups of C_a- and C_b-partisans. We kept the media bias fixed during a session, and varied the network, with the first 16 rounds of one network treatment, and the last 16 rounds of another one. We paid two random rounds from each network

[15] We do not consider a possibility of negative campaigning *in addition* to the good news bias, but it is a straightforward exercise to re-arrange the signals to get a model of negative campaigning *instead* of the positive one.

Table 1: Signal Structure

State	MB0 possible signals s_j			MB1 possible signals s_j			MB2 possible signals s_j		
	s_j	s_{-j}	$s_\emptyset$	s_j	s_{-j}	$s_\emptyset$	s_j	s_{-j}	$s_\emptyset$
j	$q(1-r)$	$(1-q)(1-r)$	r	$q(1-r)$	$(1-q)(1-r)$	r	$q_j^j(1-r)$	$(1-q_j^j)(1-r)$	r
$-j$	$(1-q)(1-r)$	$q(1-r)$	r	$q(1-r)$	$(1-q)(1-r)$	r	$(1-q_j^{-j})(1-r)$	$q_j^{-j}(1-r)$	r
j	0.56	0.24	0.2	0.56	0.24	0.2	0.72	0.08	0.2
$-j$	0.24	0.56	0.2	0.56	0.24	0.2	0.48	0.32	0.2

Notes: Conditional probability that a C_j-partisan, $j \in \{a, b\}$, receives each of the three possible private signals, s_j, s_{-j}, or $s_\emptyset$, conditional on each of the two possible state realizations, $j, -j$, for each MB treatment. Upper panel: formal expressions using our notation. Lower panel: actual values computed using experimental parameters $q = 0.7, r = 0.2, q_j^j = 0.9, q_j^{-j} = 0.4$.

treatment, with GBP payoffs for each player specified in Eq (1). Each session details and average payoffs are summarized in Table 11 in Appendix B. Table 1 lists signal accuracies in each media bias treatment. Benchmark equilibrium predictions for our parameters are described in Section 3 below.

Within a session, we kept a subject's "party" affiliation fixed for all rounds to avoid confusion and make subjects more attached to their party identity. Their member IDs and the actual composition of their fellow party members changed randomly every round (recall that we had two to three ten-person groups to reshuffle party members across groups every round). Subjects were made aware of this process, as explained in the experimental instructions available in Appendix C. In order to highlight biased signals in the media bias treatments, we used two-sector "roulette" wheels to deliver signals. The idea is illustrated by the initial interface screen in Figure 1. Wheel sectors have different colors, corresponding to s_a and s_b (blue and green in the actual interface). There are two possible wheel sector compositions that depend on which state has been selected, represented by the two wheels at the top, called Blue Wheel and Green Wheel. When ready to receive a signal, subjects are shown a covered wheel, which corresponds to the selected wheel in the no bias treatments, and to the wheel displayed directly below the selected wheel in the bias treatments (e.g., one of the two Greenish Wheels in Figure 1). To receive a signal, subjects spin the covered wheel, and their signal (if any) is the color of a randomly selected sector strip of the covered wheel. In case of an empty signal, they see a text saying "No signal".

At the end of the session we asked subjects to fill out a basic questionnaire. Table 12 in Appendix B lists the summary characteristics of their answers. In total, 450 subjects participated.

3 Equilibrium predictions with rational voters

In this section, we derive equilibrium predictions for each experimental treatment when voters are fully rational; the proofs and more detailed derivations for each case are presented in Appendix A. While the analysis techniques described below are quite general, we mostly focus on deriving

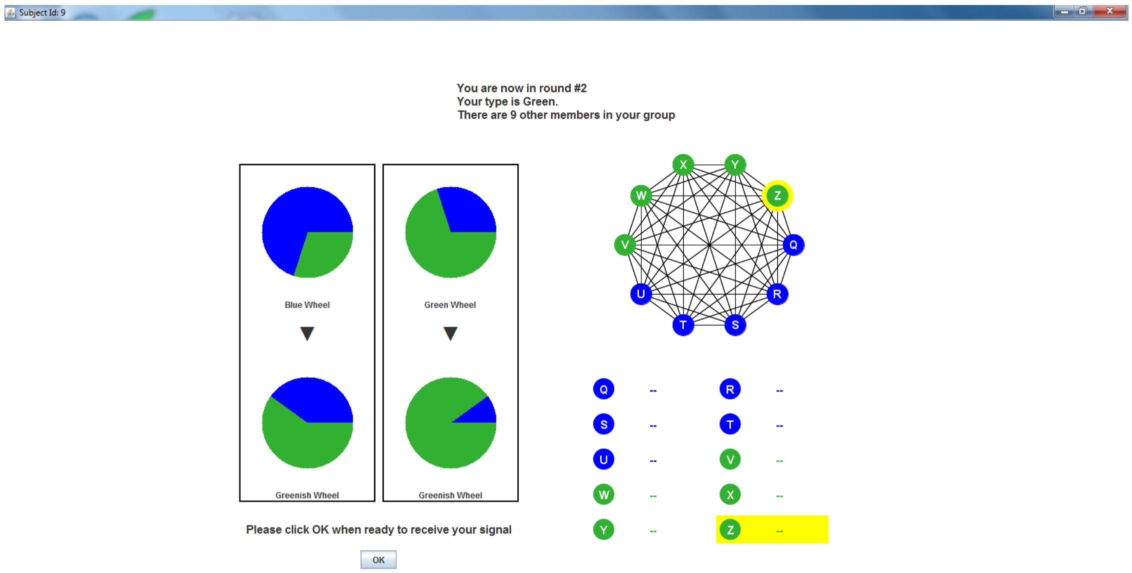


Figure 1: User Interface (MB2+SN2 Treatment, initial screen for a Green partisan.)

equilibrium predictions specific to the parameters of the experiment.[16]

We start by describing behavior at the voting stage, common to all treatments. Let p_i denote voter i's posterior belief that the realized state is a, conditional on all her available information, including her payoff type, her private signal, equilibrium messages received from others, and pivotality.[17] Then i prefers to have C_a elected if and only if her expected utility from this is higher than from electing C_b:

$$p_i u_i(a, C_a) + (1 - p_i)u_i(b, C_a) > (1 - p_i)u_i(b, C_b) + p_i u_i(a, C_b) \tag{2}$$

For the utility specification in (1), inequality (2) holds for C_a-partisans if and only if $p_i > t_a := \frac{7}{34} \approx 0.206$ and for C_b-partisans if and only if $p_i > t_b := 1 - \frac{7}{34} \approx 0.794$. Thresholds t_a and t_b indicate critical values of the posterior probability C_a-partisans and C_b-partisans should hold about $\theta = a$ so that their expected utility maximizing action becomes voting for C_a whenever their posterior exceeds the respective threshold. A *sincere Bayesian voting strategy*, in which players vote for the candidate who maximizes their expected utility can be described as follows. If i is a C_a-partisan, she votes for C_a whenever $p_i > t_a$, and if i is a C_b-partisan, she votes for C_a whenever $p_i > t_b$.

[16] The baseline case of SN0+MB0 has been studied in many papers on jury voting, starting from Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1998). One difference is the possibility of "no-signal" outcomes, which creates an endogenous group of "uninformed" voters. Without abstention allowed, these voters can play an important role in the efficiency of information aggregation. Austen-Smith and Feddersen (2005) consider this possibility for the case of three-member committees. See also Oliveros and Várdy (2015). A version of our SN2+MB0 treatment was analyzed theoretically in Schulte (2010) and Coughlan (2000). Both of those papers, as well as a more general treatment of communication games in Hagenbach et al. (2014) only focussed on full information revelation equilibria, which do not exist in our setting, so we provide a characterization of semi-pooling equilibria. SN1+MB0 and all MB2 treatments appear to be new in the literature.

[17] Voter i can be pivotal in two senses: at the voting stage, if her vote changes the outcome, and at the signalling stage, if her message moves the posterior belief of others enough to change their vote.

This voting strategy is always a best response at the voting stage. Moreover, it is a unique best response if players are pivotal with positive probability.[18]

With these preliminaries at hand, we now characterize Bayesian equilibria for each combination of media bias and social network treatments. To aid the reader, we have summarized the main empirical implications for each treatment combination in Table 2.

Table 2: Brief Summary of Equilibrium Predictions for each Treatment Combination

		Social network		
Media bias	**Prediction type**	Empty (SN0)	Polarized (SN1)	Complete (SN2)
Unbiased (MB0)	News sharing	n/a	All	Selective
	Info aggregation	No	Partial	Partial
		(Prop. 1)	(Prop. 4)	(Prop. 3)
Moderate (MB2)	News sharing	n/a	All	Selective
	Info aggregation	Partial	Partial	Partial
		(Prop. 2)	(Prop. 6)	(Prop. 5)
Extreme (MB1)	News sharing		Uninformative (anything goes)	
	Info aggregation		No (Subsection 3.4)	

Notes: All predictions are based on experimental parameters. 'Selective news sharing' means always sharing favorable signals and sharing unfavorable ones with a small enough probability specified in the respective propositions. 'Info aggregation' indicates if information aggregation happens in equilibrium. No information aggregation means that C_j-partisans vote for candidate $C_j, j \in \{a, b\}$ irrespective of their information. Partial information aggregation means that voters vote informatively under some combinations of news signals shared with them.

3.1 Empty Network, [SN0]

In the absence of social networks, no communication among voters is possible, and we are in the classical Condorcet Jury Theorem setting. Each player can only condition on her own signal and the equilibrium play to deduce the optimal voting strategy.

Unbiased media, [MB0]. Without media bias, signals are equally informative about either state. For each C_j-partisan, $j \in \{a, b\}$, conditional probabilities of private signals are specified in the respective cells of Table 1. Conditional on observing signals s_a and s_b, respectively, with our parameters and common prior of 1/2, player i's posterior is

$$p_i(\theta = a|s_a) = \frac{(1-r)q}{(1-r)q + (1-r)(1-q)} = \frac{1}{1 + \left(\frac{1-q}{q}\right)} = 0.7$$

$$p_i(\theta = a|s_b) = \frac{(1-r)(1-q)}{(1-r)(1-q) + (1-r)q} = \frac{1}{1 + \left(\frac{q}{1-q}\right)} = 0.3$$

If all players used sincere Bayesian voting strategies, then C_a-partisans would vote for C_a after either signal, as $p_i(\theta = a|\cdot) \geq 0.3 > t_a \approx 0.206$, and C_b-partisan would vote for C_b after either

[18] If players correctly update their beliefs about the state given all available signals but are not fully strategic, they should still be playing sincere Bayesian voting strategies.

signal, as $p_i(\theta = a|\cdot) \leq 0.7 < t_b \approx 0.794$. If everyone else is playing a sincere Bayesian voting strategy, since both parties are of the same size, a tie is expected, so every vote is pivotal. Hence in the case of no communication, sincere voting is also a Perfect Bayesian Nash equilibrium. However, in this equilibrium, voters ignore their private signals so there is no information aggregation.[19] These remarks yield the following proposition.

Proposition 1 ***[Unbiased media and empty social network]*** *With unbiased media and empty social network, voting is not informative. In the unique sincere voting equilibrium, all C_j-partisans vote for $C_j, j \in \{a, b\}$ regardless of their signal, and due to equal group sizes, the election results in tie.*

It should also be noted here that *there is no informative voting equilibrium*, i.e. there is no equilibrium in which everyone with a non-empty signal votes according to that signal (see Lemma 1 in Appendix A).

Moderate media bias, [MB2]. In the moderate bias case, the posterior beliefs depend not only on the number of revealed signals but also on their sources. For each C_j-partisan, $j \in \{a, b\}$, conditional probabilities of private signals are specified in the respective cells of Table 1. Note that probabilities are different for different partisans: conditional on getting a non-empty signal, $q_j^j > 1/2$ is the accuracy of C_j-partisan's "favorable" signal s_j conditional on her favorable state j, and $q_j^{-j} < 1/2$ is the accuracy of C_j-partisan's "unfavorable" signal s_{-j} conditional on her unfavorable state $-j$. A favorable signal is more likely than an unfavorable one in both states but there is correlation with the state. In other words, there is partisan bias but signals are weakly informative, allowing partisans to update from the prior of 1/2.

Suppose i is C_j-partisan. Conditional on observing signals s_j and s_{-j}, respectively, with our parameters, player i's posterior is

$$p_i(\theta = j|s_j) = \frac{(1-r)q_j^j}{(1-r)q_j^j + (1-r)(1-q_j^{-j})} = \frac{1}{1 + \left(\frac{1-q_j^{-j}}{q_j^j}\right)} = 0.6 \tag{3}$$

$$p_i(\theta = j|s_{-j}) = \frac{(1-r)(1-q_j^j)}{(1-r)(1-q_j^j) + (1-r)q_j^{-j}} = \frac{1}{1 + \left(\frac{q_j^{-j}}{1-q_j^j}\right)} = 0.2 \tag{4}$$

If players use sincere Bayesian voting strategies, then i would vote for C_j after signal s_j, but would vote for C_{-j} after signal s_{-j} since $p_i(\theta = j|s_{-j}) < t_j$. Importantly, the sources of the signals matter – favorable signals from each group should receive less weight in i's posterior than unfavorable signals from the same group. We have the following proposition.

[19] There are other equilibria, e.g., those in which everyone always votes for one of the candidates regardless of the state, and no single vote can change the outcome – but such equilibria are in weakly dominated strategies or they exhibit large asymmetry between voters of the same extended type. Following the literature, we do not consider them in the analysis.

Proposition 2 ***[Moderate media bias and empty network]*** *Given our experimental parameters, under MB2+SN0, voting is partially informative: In the sincere voting equilibrium, C_j-partisans, $j \in \{a, b\}$, vote according to their signal, and vote for C_j if they get no signal.*

According to Proposition 2, *moderate bias might be actually helpful in the case of no communication* – sincere voting becomes informative, unlike the no bias case. This arises because, under moderate bias, an unfavorable signal is unlikely to be observed, but when it *is* observed, it is a very strong signal to a voter that the candidate he doesn't favor is the correct choice (imagine Fox News endorsing a Democratic candidate). In this way, a moderate bias may play a similar beneficial role for information aggregation as correlation neglect (see, e.g., Levy and Razin (2015)).[20]

3.2 Complete Network, [SN2]

In the network treatments, there is a communication stage, during which players can exchange signals with others. Since both states are assumed equally likely and revealed signals are "hard" evidence, the posterior belief that the realized state is $\theta = a$ depends on the difference between the number of reported signals in favor of C_a, and, in case of moderate media bias, it also depends on the sources of the signals. The posterior also depends on the beliefs about non-reported signals: if some player i did not reveal a signal during the communication stage, other voters must consider whether i received an empty signal or, alternatively, might have received a non-empty signal and withheld it strategically.

The simplest signalling strategy is to reveal any signal. Fix player i and suppose everyone other than i is using a *fully revealing signalling strategy*, according to which they always share their non-empty signals during the communication stage, and use sincere Bayesian voting strategies at the voting stage. It turns out that in a complete network, there is no full information revelation equilibrium, i.e. i would sometimes prefer not to reveal her signal (see Lemmata 2 and 3 in Appendices A.1 and A.2). The reason is that in a complete network, if i reveals an unfavorable signal, it will also be shared with the supporters of the other candidate, and can be pivotal in pushing their posterior in the direction i does not like.

Since there is no fully revealing equilibrium, we consider a *semi-pooling equilibrium*, in which all agents fully reveal favorable signals and hide unfavorable signals by pooling with the uninformed. In this case, other players will have to form consistent beliefs about how likely each reported empty signal is meant to hide an unfavorable signal.

In the no bias case, we have the following proposition:

Proposition 3 ***[Unbiased media and complete network]*** *Given our experimental parameters, under MB0+SN2, there is no full information revelation equilibrium. However, there is a range of semi-pooling equilibria, in which all C_j-partisans, $j \in \{a, b\}$, with favorable signals s_j reveal them truthfully at the communication stage, and hide the unfavorable signals s_{-j}, with a commonly known*

[20] This might also resemble the classical result in Calvert (1985) for a single decision maker faced with the choice between a neutral and a biased expert. The mechanism there is different: Calvert assumes the bias structure in which extreme realizations of the state are always revealed.

equilibrium probability ν^*. At the voting stage, each player i has a potentially different posterior belief p_i.[21] *Given these posteriors, C_a-partisans vote for C_a as long as $p_i > t_a$, and otherwise vote for C_b. C_b-partisans vote for C_b as long as $p_i < t_b$, and otherwise vote for C_a. Each such equilibrium is characterized by fixing any $\nu^* \in (.922, 1]$. Information gets partially aggregated.*

In the moderate bias case, we obtain a similar result (see Proposition 5 in Appendix A.2). Comparing it with Proposition 3, we continue to observe a positive effect of the media bias in our setting: The range of beliefs that support semi-pooling equilibria widens from $(0.922, 1]$ to $(0.862, 1]$. An important difference is that proper belief updating in Proposition 5 requires taking into account the sources of the signals: that is, a signal favoring C_a is interpreted differently depending on whether it is reported by a C_a- or C_b-partisan – since revealed favorable signals receive less weight in the posterior than revealed unfavorable signals.

3.3 Polarized Network, [SN1]

In the polarized network, players can only reveal signals to members of their own group. In this case, they cannot directly affect beliefs in the other group – their messages are only pivotal for the posteriors hold in own group – so fully revealing unfavorable signals becomes incentive compatible.

In the no bias case, we have the following proposition:

Proposition 4 ***[Unbiased media and polarized network]*** *Given our experimental parameters, under MB0+SN1, there is a full information revelation equilibrium, in which all voters with non-empty signals reveal them truthfully at the communication stage and believe with probability one that non-revealing agents are uninformed. All voters in a partisan group have identical party-specific posterior beliefs about the realized state. At the voting stage, each player votes for their favorite candidate as long as the set of signals they observe within their network is sufficiently strong for their candidate and otherwise vote their signal.*[22] *Information gets partially aggregated.*

In the moderate bias case, there is a similar result for full information revelation equilibrium (see Proposition 6 in Appendix A.2). Comparing the posterior beliefs from Proposition 6 with the posterior beliefs from the no bias analog, Proposition 4, we observe that the same difference in the number of revealed s_a and s_b signals produces a posterior that is closer to 1 or 0, i.e. sincere voting is more informative in the presence of media bias.[23]

Taken together, Propositions 3 and 4 (and Propositions 5 and 6, for the moderate bias case) highlight an important tradeoff: on the one hand, SN2, a complete network, enables exposure to a larger number of informative signals than SN1, simply due to its larger size; this can improve everyone's welfare. On the other hand, full information revelation fails under SN2 because partisans have incentives to hide unfavorable signals, which is not the case under SN1. Thus without media bias, *having all voters grouped in a polarized network by party affiliation can be welfare-improving*

[21]See Eq (8) in Appendix A.1 in which the difference in the number of signals s_a and s_b revealed by others, is supplemented by i's private signal.

[22]The complete details are given in the proof in Appendix A.1.

[23]This follows from comparing expressions (31)–(32) in Appendix A.2 with Eq (16) in Appendix A.1.

relative to a complete network. With moderate media bias, this could also be the case, but the positive effect of having more informative unfavorable signals is offset by their small probability of occurrence.

3.4 Extreme media bias, [MB1]

Finally, we consider the case of extreme media bias. These treatments feature uninformative signals: conditional on getting any signal, C_a-partisans are more likely to receive signals in favor of C_a, and C_b-partisans in favor of C_b, *independently of the state*. The posterior belief about the realized state conditional on any combination of signals coincides with the prior and equals 1/2. Hence any signalling strategy is rationalizable. While any voting strategy is rationalizable (everyone is always uninformed), the unique sincere Bayesian equilibrium requires every voter to vote for their preferred candidate as in Proposition 1.

4 Aggregate Treatment Effects

This section provides a first look at the main results and their interpretation, beginning by documenting the experimental treatment effects. The analysis then continues at the individual level in Section 5.

To present the treatment effects in the simplest way, we organize the discussion around pairwise comparisons with a single ten-person group decision as the unit of observation, and estimate mean differences using the standard t-test. Every round we have two to three simultaneous ten-person groups, and we treat their decisions as independent across groups and rounds in those comparisons. All indicated effects survive in a linear regression analysis of efficiency measures (group success rates or Bayesian efficiency scores) on treatment dummies and their interactions, with clustering at group-round, subject, or session level, which also serves as a way to mitigate the multiple-hypothesis testing concerns.

Group success rates. First, we look at the effects of social networks and media bias on group success rates. The group decision success is coded as 1 if the elected candidate corresponds to the true state, and 0 otherwise, and the group success rate is computed by averaging decision success across all group decisions in a given treatment.

Summary statistics are reported in Table 3. We report averages and standard errors based on all data (columns labeled "All"), as well as those based on restricted samples. In particular, since ties are resolved randomly and cannot be controlled by the group, it makes sense to exclude the tied cases (depending on one's prior, one can view ties as either favorable or unfavorable group outcomes). These restricted results are in columns labeled "No Ties". Finally, in columns labeled "No Ties, Last 12" we also exclude the first four rounds from consideration, since presumably there could be some initial learning/adjustment going on. Success rates are quite robust to these restrictions, and effects even get stronger. For that reason, in the remainder of the paper we will use all data (more conservative) in discussing the results.

Table 3: Group Success Rates By Treatment

Network	Pr.(No Signal)	No media bias								
		N	[**All**]		N	[**No Ties**]		N	[**No Ties, Last 12**]	
Empty	20%	176	.739	(.033)	142	.803	(.034)	111	.820	(.037)
Polarized	–	192	.818	(.028)	164	.866	(.027)	120	.867	(.031)
Complete	–	176	.881	(.025)	171	.883	(.025)	128	.914	(.025)
Polarized	0%	96	.833	(.038)	80	.888	(.036)	60	.867	(.044)
Complete	–	96	.844	(.037)	94	.851	(.037)	72	.819	.(046)
		Extreme Media Bias								
		N	[**All**]		N	[**No Ties**]		N	[**No Ties, Last 12**]	
Empty	20%	112	.420	(.047)	75	.373	(.056)	60	.400	(.064)
Polarized	–	144	.493	(.042)	102	.490	(.050)	76	.474	(.058)
Complete	–	256	.504	(.031)	167	.497	(.039)	171	.474	(.038)
		Moderate Media Bias								
		N	[**All**]		N	[**No Ties**]		N	[**No Ties, Last 12**]	
Polarized	20%	96	.729	(.046)	83	.771	(.046)	63	.825	(.048)
Complete	–	96	.573	(.051)	83	.566	(.055)	60	.600	(.064)

Notes: N is the number of group decision observations. Standard errors are in parentheses.

We start by considering the effects of social networks on outcomes, without media bias. For convenience, these results are summarized in Table 4, in the top row. There, we see that, relative to the empty-network benchmark, the success rates in polarized networks are higher by 7.9% ($p = 0.069$), and even more so, by 14.2%, in complete networks ($p = 0.001$). Moreover, the success rate with complete networks is 6.3% higher ($p = 0.091$) than in a polarized network. Thus, without media bias, our results suggest that social networks improve efficiency, and the finding that success rates are highest with complete networks indicates that this increase arises from the communication amongst voters, which is maximized in the complete network treatment.

However, we see that once media bias enters the picture, these benefits from social networks dissipate. Indeed, from the second row of Table 4, we see that the introduction of moderate media bias *lowers* the success rate by 15.6% points – in complete vs. polarized networks, lowering the success rate from 73% down to 57%. Extreme media bias lowers success rates across all network configurations, as should be expected, since signals are uninformative. These effects of media bias on success rates are summarized in Table 5. This suggests that social networks, per se, do not appear to harmfully impact voting behavior; rather the culprit is media bias. Communication per se appears to improve efficiency; however, communication of bad quality information (as in the media bias treatments) worsens efficiency.

Bayesian decision efficiency. Group success rates offer only an incomplete picture of efficiency, as they do not consider the possibility that due to sampling, the realized signals may not reflect the true state sufficiently to allow voters to reach the correct decision. For instance, if the true state was b but voters' signals, pooled together, indicated that state a was more likely and so the group elected candidate C_a, then the group success criterion would classify this as a mistake, even though in this case voters would not have had the requisite information to make the correct decision given their signals.

Table 4: Effects of Network on Group Success Rate

Bias	N	Polarized minus Empty		N	Complete minus Empty		N	Complete minus Polarized	
		Δ	p-val		Δ	p-val		Δ	p-val
No	368	.079*	(0.069)	352	.142***	(0.001)	368	.063*	(0.091)
Moderate							192	−.156**	(0.023)

Notes: Δ indicates the difference in average success rates between respective network configurations. Two-sided p-values in parentheses. Significance: *** < 0.01, ** < 0.05, * < 0.1

Table 5: Effects of Bias on Group Success Rate

Network	N	Extreme minus No		N	Moderate minus No		N	Moderate minus Extreme	
		Δ	p-val		Δ	p-val		Δ	p-val
Empty	288	−.319***	(0.000)						
Polarized	336	−.325***	(0.000)	288	−.089*	(0.100)	240	.236***	(0.000)
Complete	432	−.377***	(0.000)	272	−.308***	(0.000)	352	.069	(0.249)

Notes: Δ indicates the difference in average success rates between respective media biases. Two-sided p-values in parentheses. Significance: *** < 0.01, ** < 0.05, * < 0.1

To overcome these drawbacks, we use an alternative measure of “Bayesian efficiency”, in which we compare the actual group decisions with the choice a benevolent Bayesian social planner would have made, had she observed all voters’ private signals. One direct implication of this definition is that if signals are uninformative, as in our extreme bias treatment, any group choice is 100% efficient.

Table 6 illustrates the performance of this alternative efficiency measure. As before, we report averages and standard errors based on all data (columns labeled “All”), as well as those that exclude the first four rounds (“Last 12”). Excluding tied vote outcomes is not relevant here, since Bayesian efficiency does not depend on how the ties are resolved (if the Bayesian posterior equals 0.5, a tie is counted as a correct group decision, otherwise it is a mistake). Excluding the first four rounds does not change the overall picture.

Compared to group success rates, Bayesian efficiency in all unbiased treatments is higher. An important difference from results in Table 3 is that under moderate bias, the polarized network is now less efficient than complete. This indicates that the opposite result we obtained using group success rates, is largely due to sampling of the signals. Indeed, with moderate bias, in complete networks, the more informative unfavorable signals represent only 25.13% of the non-empty signals, and in polarized networks, unfavorable signals represent only 20.28% of the non-empty signals. At the same time, we continue to obtain a significant the drop in efficiency of network treatments, compared to the no bias, empty network case. The analogs of Tables 4 and 5 for Bayesian efficiency are, respectively, Tables 13 and 14 in Appendix B. Importantly, the moderate bias continues to exhibit a negative effect on efficiency in both polarized and complete networks.[24]

[24] Since extreme bias treatments are now 100% efficient, the differences from extreme bias in Table 14 have their signs reversed, as should be expected.

Table 6: Bayesian Efficiency By Treatment

Network	Pr.(No Signal)	No media bias					
		N	[**All**]		N	[**Last 12**]	
Empty	20%	176	.886	(.024)	132	.909	(.025)
Polarized	–	192	.948	(.016)	144	.951	(.018)
Complete	–	176	.983	(.010)	132	.985	(.011)
Polarized	0%	96	.990	(.010)	72	.986	(.014)
Complete	–	96	.990	(.010)	72	.986	(.014)
		Moderate Media Bias					
		N	[**All**]		N	[**Last 12**]	
Polarized	20%	96	.667	(.048)	72	.638	(.057)
Complete	–	96	.792	(.042)	72	.806	(.047)

Notes: N is the number of group decision observations. Standard errors are in parentheses. Under extreme media bias, any group decision is 100% bayesian efficient so those treatments all have an efficiency score of 1.00 and zero variance, and are not listed in the table.

We summarise both bias and network effects as follows.

Result 1 *A complete network improves efficiency (as measured by either success rate or Bayesian efficiency score), by 10–14% relative to the control if there is no media bias. However, with moderate media bias there are substantial efficiency losses: both complete and polarized networks are less efficient than the unbiased empty network, even though they can potentially reveal a larger number of informative signals. Furthermore, as Bayesian efficiency indicates, the losses are largely due to subjects' imperfect updating.*

News-sharing patterns. Next, we look at the news sharing patterns. Recall that in most treatments, 20% of the time a subject gets no signal in which case it cannot be shared, and in two treatments, subjects always receive informative private signals.

We look at three different subgroups of signals: First, we look at the average signal sending rate for all signals, clustered by subject (reported in Table 7 under "All signals sent"). Second, we look at average signal sending rates for signals favorable to the subject's party ("Fav. signals sent"), i.e. the ones that match the voters' preferred party color. Third, we look at average signal sending rates for signals unfavorable to the subject's party ("Unfav. signals sent").

In the polarized network, all subjects in the same subnetwork have aligned interests, so in theory, there should be full signal revelation. Table 7 shows clearly that most of the time, subjects do share their signals – roughly 83–90% of all signals (conditional on getting a non-empty signal). However, in all treatments, subjects share significantly less than 100% of the signals that they receive, so full signal revelation is not supported in our data.

Distinguishing the signals shared by their type, and starting with the case of no media bias, we see from the first two rows of Table 7 that *subjects are selective in the signals they share*: in a polarized network, favorable signals are shared 93.6% of the time, whereas unfavorable signals are shared 80.3% of the time. This difference in significant (in all discussed cases except the last two rows, $p \leq 0.002$). This pattern is even more evident in complete networks. This finding that

Table 7: Signal Sharing Patterns by Treatment

Network	Pr.(No Signal)	Bias	N	All signals sent		N	Fav. signals sent		N	Unfav. signals sent	
Polarized	20%	No	120	.869	(.019)	120	.936	(.017)	120	.803	(.029)
Complete	–	–	110	.825	(.021)	110	.906	(.021)	110	.738	(.032)
Polarized	0%	–	60	.861	(.026)	60	.936	(.023)	60	.782	(.042)
Complete	–	–	60	.854	(.025)	60	.938	(.026)	60	.774	(.042)
Polarized	20%	Extreme	90	.827	(.021)	90	.891	(.019)	89	.678	(.039)
Complete	–	–	120	.825	(.016)	120	.888	(.018)	119	.673	(.034)
Polarized	–	Moderate	60	.899	(.022)	60	.902	(.027)	57	.886	(.031)
Complete	–	–	60	.868	(.024)	60	.870	(.030)	59	.864	(.032)

Notes: N is the number of subjects, standard errors in parenthesis are clustered by subject. All rates are conditional on getting a non-empty signal.

subjects selectively share news favoring their party may mimic what happens in a real-world social media context, where the couching of information sharing in terms of "liking" (as on Facebook) may encourage users to only share information that supports their entities they favor.

Rows 3–4 of Table 7 present the treatments in which the probability of not obtaining a signal is zero; in these cases, subjects who don't report a signal are not able to hide behind a facade of being uninformed. Interestingly, we find that the signal sharing is practically the same in these treatments as in the results discussed above, where the probability of no signal was set to 20%. This also supports our "liking" interpretation of the signal sharing result; that there is something inherently distasteful about sharing an unfavorable signal which trumps the possibility that their non-report will be "unraveled".

Once we introduce media bias, the overall frequency with which subjects share signals does not change, hovering around 85–90% (as shown in the bottom four rows of Table 7). However, we see that extreme and moderate media bias have divergent effects on the frequency of sending unfavorable signals: under extreme media bias, the selective signal sending result remains, but under moderate media bias, the frequency of reporting favorable vs. unfavorable signals becomes indistinguishable, for both polarized and complete networks. This is largely due to a much smaller number of unfavorable signal realizations under the moderate bias, as explained earlier.

Finally, we also split the rates by round to see dynamics of signal sharing across time for each treatment; these are illustrated in Figures 6–7 in Appendix B. Except for the first few initial rounds (expiring around round 4), the average sharing rates do not appear to change much over time.

These findings on signal sharing are summarized as follows.

Result 2 *While subjects send out the majority of signals to those connected to them in a network in all treatments, full signal revelation is rejected. There is selective signal sharing: unfavorable signals are communicated less often than favorable ones. Adding media bias does not change the overall frequency of sharing favorable signals, however unfavorable signals are shared less (extreme bias) and more (moderate bias) relative to the control.*

Do voters follow their signals? Next, we look at how often voters actually follow their signals, conditional on getting a non-empty one. The statistics are reported in Table 8. Of course, the numbers there do not allow us to distinguish between the effects of own signal and group-revealed signals whenever the network is non-empty.[25]

Table 8: Voting Your Signal Patterns by Treatment

Network	**Bias**	N	**Any signal**		N	**Fav. signal**		N	**Unfav. signal**	
Empty	No	110	.780	(.020)	110	.887	(.021)	109	.676	(.028)
Polarized	–	180	.726	(.012)	180	.849	(.015)	180	.601	(.019)
Complete	–	170	.716	(.010)	170	.777	(.014)	170	.653	(.016)
Empty	Extreme	70	.698	(.017)	70	.850	(.021)	69	.370	(.032)
Polarized	–	90	.727	(.016)	90	.838	(.018)	89	.470	(.032)
Complete	–	120	.713	(.012)	120	.776	(.015)	119	.560	(.025)
Polarized	Moderate	60	.620	(.022)	60	.659	(.027)	57	.468	(.040)
Complete	–	60	.643	(.012)	60	.691	(.025)	59	.503	(.042)

Notes: N is the number of subjects, standard errors in parenthesis are clustered by subject. All rates are conditional on getting any signal. No bias, non-empty network treatments with $r = 0$ and $r = 0.2$ are pooled together.

Again, starting from the case of no media bias, we see at the top panel of Table 8, that subjects are more likely to vote according to their signal when it favors their party, and this holds across all types of networks. For instance, under no network, when their signal is favorable, subjects vote for their party 89% of the time, but they vote for the other party only 67.6% of the time when their signal favors the other party.

These results continue to obtain even after introducing media bias. Under both extreme and moderate media bias, subjects tend to vote according to their signal more, on average, when the signal favors their party. There are some noticeable differences across media bias treatments. Under extreme media bias, our finding that subjects vote their signal when it favors their party (but not otherwise) is accentuated. Under moderate media bias, however, the overall probability of voting congruently with their signal falls, relative to the other media bias treatments.

Overall, then, subjects tend to vote for their preferred party, and may ignore information if it is not favorable to their party; this suggests that subjects are behaving as if voting was expressive, and does not reflect the common value decision environment which we designed.

These findings are summarised as follows.

Result 3 *i) In all our treatments: Subjects vote according to their signal significantly more often when the signal is favorable to their party than when it is not. ii) Without bias, either network reduces voting in line with any signal, and polarized network reduces voting in line with an unfavorable signal, relative to no network. iii) Media bias (either moderate or extreme) significantly lowers voting in line with an unfavorable signal.*

[25] For the uninformed voters, however, we can estimate the effects of group-revealed signals.

5 Individual subject behavior

In this section, we continue the analysis and interpretation of the experimental results. The equilibrium predictions in Subsection 3 provide a rational benchmark. In Subsection 5.1, we go over individual-level experimental results and test those predictions.

5.1 Testing the rational benchmark

The theoretic predictions from Section 3 have clear implications for individual voting and signalling behavior of fully rational voters. Of course, real subjects are likely to exhibit some deviations from those predictions. The analysis that follows should be construed as highlighting typical deviations from the benchmark rather than a narrow test of the theory.

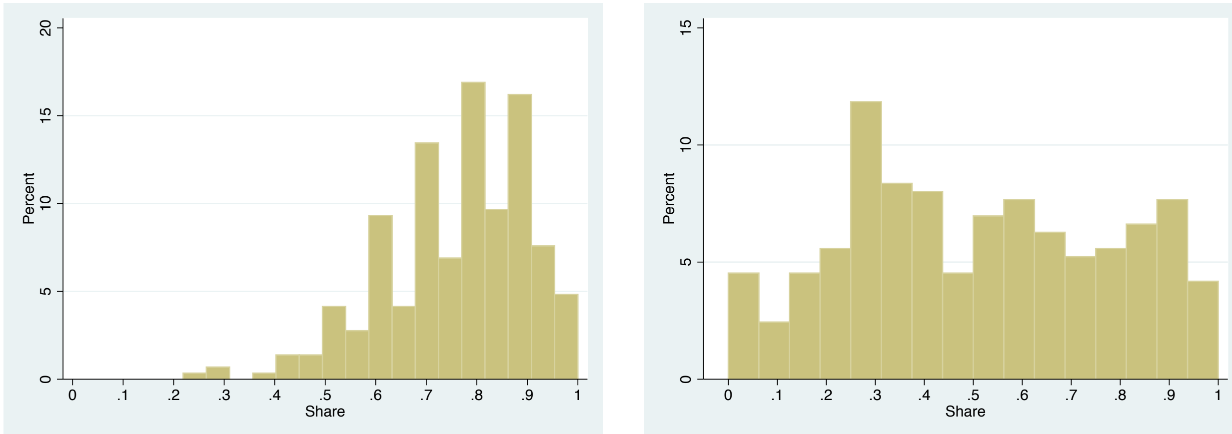
Looking at group success rates first, it is clear from Table 3 that already in the control case of no communication and no bias, aggregate efficiency is much higher than the theory benchmark of uninformative voting with 50% success rate. Relative to this benchmark, theory predicts positive effects of adding moderate media bias. Indeed, the success rates with the moderate bias case exceed 50%, especially for the polarized network. This result continues to hold when we switch to Bayesian efficiency in Table 6. However, it also illustrates that subjects are not perfectly updating their beliefs, and the overall performance, while still better than 50%, is significantly worse than in the no bias, empty network case.

We now turn to the analysis of individual voting strategies. Using the expressions for equilibrium posteriors from Propositions 1–6, we computed sincere Bayesian posteriors for each subject and each treatment (except for extreme bias, since the posterior coincides with the prior there), as a function of revealed and private signals in each group-round.[26]

To estimate how often subjects played sincere Bayesian strategies, we used two approaches. In the first approach, we coded for each subject their voting decision as "correct" if they always followed the sincere Bayesian strategy, i.e. a C_a-partisan with posterior $p_i \geq t_a$ (resp. $p_i < t_a$) voted for C_a (resp. C_b), and a C_b-partisan with posterior $p_i > t_b$ (resp. $p_i \leq t_b$) voted for C_a (resp. C_b), obtaining a binary indicator of sincere Bayesian voting. Then we averaged this indicator for each subject over all her decisions, to obtain a subject-level measure of the fraction of her voting decisions that are fully consistent with a sincere Bayesian strategy. This is illustrated in Figure 2(a). In the second approach, illustrated in Figure 2(b), we only looked at those cases in which the Bayesian posterior counseled a player to vote against their favorite candidate, and computed how often they voted accordingly. This is a more stringent measure of sincere Bayesian voting because to vote against own partisan bias, partisans must be sufficiently convinced that the state favors their less preferred choice.

From Figure 2(a), we see that a substantial proportion of subjects use voting strategies that look consistent with sincere Bayesian voting 60% of the time or more. However, a comparison of panels (a) and (b) in Figure 2 reveals that this seemingly high consistency is partly due to sincere Bayesian

[26] For the case of complete networks, in which the posterior also depends on the equilibrium beliefs, we used the lower bound on ν^*– the probability of not reporting a nonfavorable signal– wherever applicable.



(a) $N = 290$ subjects

(b) $N = 287$ subjects

Figure 2: Are subjects consistent with sincere Bayesian voting? Panel (a): using all votes; Panel (b): using only votes against own favorite candidate.
y-axis: percent of subjects; x-axis: share of subject's votes (out of 32 total vote decisions) consistent with sincere Bayesian voting strategies.

strategy prescribing to vote for one's favorite candidate. If we look exclusively at those cases in which the sincere Bayesian strategy prescribes voting against one's favorite candidate (in panel (b)), the consistency rate is markedly lower, and more than half the subjects fail to vote against their favored candidate when doing so is prescribed by the Bayesian strategy. Nevertheless, there is a non-negligible proportion of subjects using sincere Bayesian strategies. Moreover, it is precisely those subjects who score highly in panel (b) who also score highly in panel (a), as illustrated in Figure 8 in Appendix B, which depicts consistency with sincere Bayesian strategies only for those subjects who score more than 60% correct in Figure 2(b).

All in all, a significant proportion of subjects behave consistently with sincere Bayesian voting strategies. However there is a marked difference in consistency depending on whether the Bayesian voting strategy prescribes voting for or against a subject's favorite candidate.[27]

To obtain a more complete picture of individual behavior both at the voting and at the messaging stage, we look at subject's signal sharing patterns conditional on their voting types in Table 9.

Across the board, there is a gap in sharing probabilities between favorable and unfavorable signals, which depends very little on the network structure. In fact, individuals voting more consistently with Bayesian strategies engage in selective sharing relatively more than the less-consistently voting ones. While sharing favorable signals all the time, and sharing unfavorable signals only sometimes might, in principle, be in line with semi-pooling equilibria in complete networks, the probability of hiding unfavorable signals (from the last two columns, this varies from roughly 16–25%) is much lower than even the lowest bound on ν^* in these equilibria (cf. Propositions 3 and

[27]An alternative approach to this exercise would be to use subjects' actual rather than theoretical posteriors, estimated from the data. The difficulty is that with the difference in the number of revealed signals by other voters ranging between -9 and 9 in the worst case scenario, we do not have enough observations for posterior estimation.

Table 9: Signal Sharing Patterns by Voting Types

Voting type	**Network**	N	**Fav. signals sent**		N	**Unfav. signals sent**	
All	All	290	.914	(.012)	290	.787	(.019)
> 60% Bayesian	All	117	.941	(.017)	117	.754	(.030)
–	Polarized	112	.939	(.018)	112	.755	(.031)
–	Complete	113	.944	(.017)	113	.754	(.031)
≤ 60% Bayesian	All	173	.891	(.017)	173	.830	(.019)
–	Polarized	128	.906	(.017)	125	.839	(.024)
–	Complete	117	.876	(.020)	116	.820	(.025)

Notes: N is the number of subjects, standard errors in parenthesis are clustered by subject. All rates are conditional on getting a non-empty signal and exclude MB1 treatment. 60% Bayesian refers to those subjects who scored more than 60% correct in Figure 2(b).

5).

We summarize individual-level findings on vote strategies as follows.

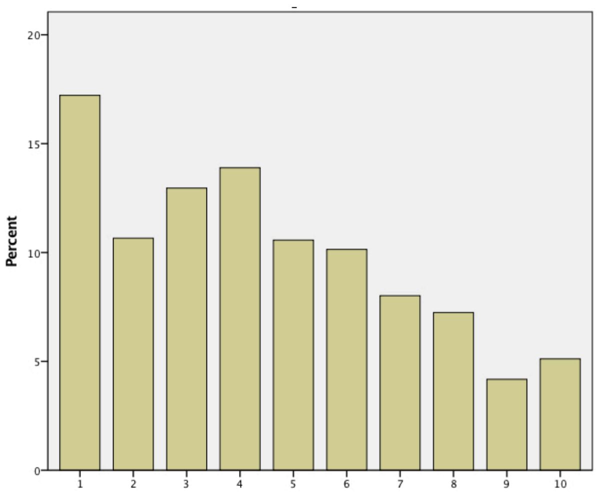
Result 4 *Subject voting behavior differs from theoretical predictions. We reject fully revealing and semi-pooling equilibrium predictions. While there is a significant proportion of subjects who actually use sincere Bayesian voting strategies, the majority of subjects use voting strategies biased towards their favorite candidate. The selective signal sharing behavior identified in Section 4, in which favorable signals are shared more often than unfavorable, holds across all individuals.*

6 Corroborating survey evidence from Pakistan

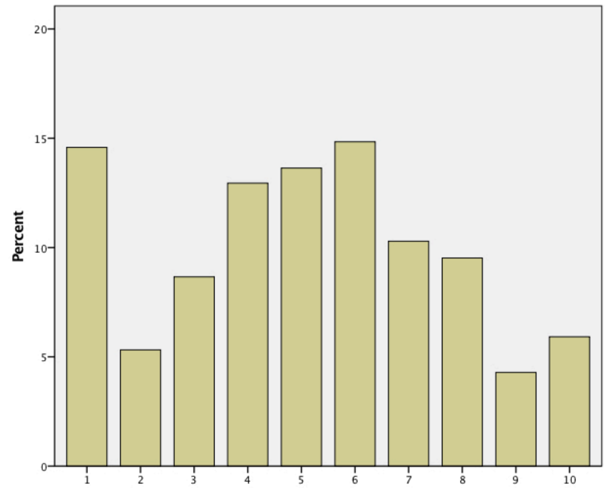
As a robustness check, to see how our results – obtained in a laboratory setting – translate to the real world, we present here survey evidence, collected for us by Gallup Pakistan. The goal was to check whether, like in our main findings, 1) subjects selectively share information that is favorable to their party more often than the unfavorable information; and 2) subjects tend to ignore information unfavorable to their party.

We introduced five questions on social media into a standard questionnaire administered by Gallup interviewers to a panel of Pakistani population, representative with respect to the province level and urban/rural split, during January 16th–20th, 2017. In the survey questions we asked how likely a respondent was to share a "favorable news" (that the favorite political candidate was a major force behind building a new hospital), to share an "unfavorable news" (that the favorite political candidate was accused of corruption), and to revise their opinion after an "unfavorable news" shared by a Facebook friend. We also asked how often respondents received news about politics and government from social media, and how trustworthy they thought the information from social media was.[28]

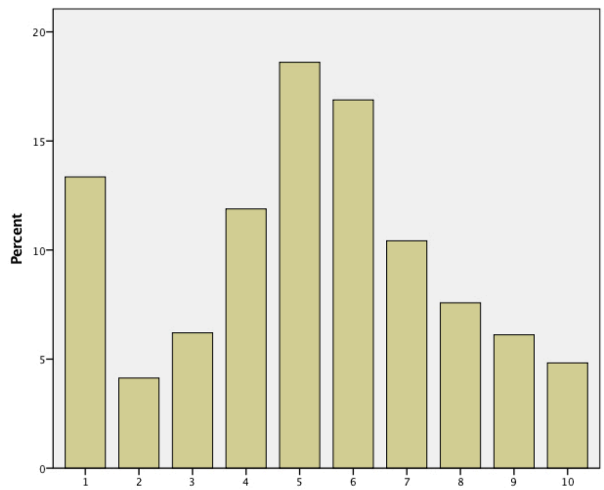
[28]The exact statements of the questions are available in Table 16 in Appendix B. The questionnaire was administered in an Urdu translation.



(a) How likely to share "bad" news (corruption); $N = 1,173$.



(b) How likely to share "good" news (hospital); $N = 1,166$.



(c) How likely to revise opinion if Facebook friend shared "bad" news; $N = 1,161$.

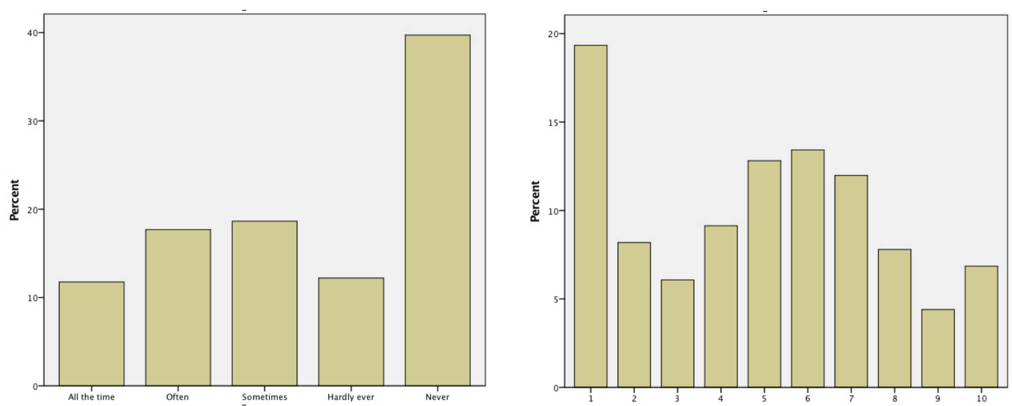
Figure 3: Distribution of survey responses. All figures exclude those who said they did not use social media. Scale: 1=Not at all likely, 10=Extremely likely.

Figures 3–4 present the survey responses graphically.[29]

Comparison of Figures 3(a) and 3(b) shows that amongst the social media users, favorable news is shared more often than unfavorable. Collapsing categories 1 to 5 into "not likely", and 6 to 10 into "likely", we see that "bad" news about the favorite candidate is "likely" to be shared by about 34.7% of social media users. In contrast, "good" news about the favorite candidate is "likely" to be shared by about 44.9% of social media users. Furthermore, less than half – only 45.8% – of social media users are "likely" to revise their opinion after a Facebook friend shared a "bad" news article. Both of these findings support the experimental results described earlier.

From Figure 4(a), we see that about 48.1% of respondents regularly get the news about politics

[29]The actual response frequencies for each category are in Table 16, and the main demographic characteristics of the survey respondents are in Table 15 in Appendix B.



(a) Get news about politics & gov't from social media (b) How trustworthy is information from social media

Figure 4: Distribution of survey responses about social media as a source of news (left panel) and degree of trust (right panel); $N = 1,803$. Scale: 1=not at all trustworthy, 10=extremely trustworthy.

and government from social media. In the full sample (including those who do not use social media), according to Figure 4(b), about 44.5% view the information from social media as "trustworthy", and the average media trustworthiness on the 10-point scale is 4.89 (in the experiment, average media trust is very similar – 4.80). The distribution of trustworthiness is different from the lab subjects, however (see Figure 5 for lab subject responses), and its shape is not driven by those who say they don't use social media, although such subjects contribute a lot to the peak at category 1 in Figure 4(b). These descriptive results are broadly in line with our experimental findings.

7 Concluding remarks

The proliferation of news and information targeting via social media has raised concerns that a voting populace obtaining a growing share of its information about competing electoral candidates from social networks may become more polarized, as targeting makes it less likely for voters to hear points of views contrary to their preferred positions. Complementing other recent studies that document and measure the extent of polarization in social networks, in this paper we use laboratory experiments to explore the effects of social networks and media bias on voter behavior.

Overall, our results provide some support to the idea that by filtering out unfavorable content social media may lead to polarization in voting behavior. Our results suggest substantial effects of polarization at the expense of efficient information aggregation by voting: voters publicly send out signals favorable to their party more often than signals unfavorable to their party, and in all treatments, they vote according to their private signal more often if the signal is favorable. Both signal sharing and voting behavior differ from the fully rational equilibrium predictions. Media bias lowers efficiency, and its negative effects are amplified in networks due to voters taking signals at face value, not accounting for signal sources. This suggests that improving the quality of information shared on social networks is of first order importance, compared to overcoming the effects of filter bubbles.

Our results that voters selectively share signals favorable to their party are different from findings in the "confirmation bias" and "information avoidance" literatures[30] – subjects in our experiment cannot choose which kind of news to receive, but rather can only decide which kind of news to relay to others. In real-world social networks this could be due to social preferences (e.g., I don't tell you bad news to keep you happy), but could also reflect far-sighted individual preferences (e.g., if all I do is share bad news for you, you won't want to keep me as a Facebook friend). In ongoing work, we are exploring how to enrich our experimental setting to allow for these types of effects.

The dimension of news sharing studied in this paper may prove especially important in light of recent Facebook announcements regarding the change in the newsfeed algorithm that would prioritize posts and news shared by users over that by advertisers and publishers.[31]

References

An, J., Quercia, D., and Crowcroft, J. (2014). Partisan Sharing: Facebook Evidence and Societal Consequences. In *Proceedings of the Second ACM Conference on Online Social Networks*, COSN '14, pages 13–24, New York, NY, USA. ACM.

Anderson, S. P. and McLaren, J. (2012). Media mergers and media bias with rational consumers. *Journal of European Economic Association*, 10(4):831–859.

Austen-Smith, D. and Banks, J. S. (1996). Information Aggregation, Rationality, and the Condorcet Jury Theorem. *American Political Science Review*, 90(1):34–45.

Austen-Smith, D. and Feddersen, T. J. (2005). Deliberation and Voting Rules. In *Social Choice and Strategic Decisions*, pages 269–316. Springer.

Austen-Smith, D. and Feddersen, T. J. (2006). Deliberation, Preference Uncertainty, and Voting Rules. *American Political Science Review*, 100(2):209–217.

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

Bakshy, E., Messing, S., and Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R. (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science*, 26(10):1531–1542.

Battaglini, M. (2017). Public Protests and Policy Making. *Quarterly Journal of Economics*, 132(1):485–549.

Bernhardt, D. M., Krasa, S., and Polborn, M. (2008). Political polarization and the electoral effects of media bias. *Journal of Public Economics*, 92(5):1092–1104.

Bouton, L., Llorente-Saguer, A., and Malherbe, F. (2017). Unanimous rules in the laboratory. *Games and Economic Behavior*, 102:179–198.

Boxell, L., Gentzkow, M., and Shapiro, J. M. (2017). Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences*, 114(40):10612–10617.

Buechel, B. and Mechtenberg, L. (2016). The Swing Voter's Curse in Social Networks. Working paper.

Calvert, R. L. (1985). The Value of Biased Information: A Rational Choice Model of Political Advice. *Journal of Politics*, 47(2):530–555.

[30] See Golman et al. (2017) for a detailed overview. Esponda and Vespa (2014) document the difficulty subjects face in extracting information from hypothetical events in a voting environment. While this bias can be present in our setup, it cannot explain the selective news sharing observed in the polarized network treatments, where there is a fully revealing equilibrium.

[31] See http://www.latimes.com/business/la-fi-tn-facebook-shares-20180112-story.html.

Chiang, C.-F. and Knight, B. (2011). Media Bias and Influence: Evidence from Newspaper Endorsements. *Review of Economic Studies*, 78(3):795–820.

Coughlan, P. J. (2000). In Defense of Unanimous Jury Verdicts: Mistrials,Communication, and Strategic Voting. *American Political Science Review*, 94(2):375–393.

DellaVigna, S. and Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting. *Quarterly Journal of Economics*, 122(3):1187–1234.

Dickson, E. S., Hafer, C., and Landa, D. (2008). Cognition and Strategy: A Deliberation Experiment. *Journal of Politics*, 70(4):974–989.

Duggan, J. and Martinelli, C. (2011). A Spatial Theory of Media Slant and Voter Choice. *Review of Economic Studies*, 78:640–666.

Elbittar, A., Gomberg, A., Martinelli, C., and Palfrey, T. R. (2017). Ignorance and bias in collective decisions. *Forthcoming, Journal of Economic Behavior & Organization*.

Enikolopov, R., Petrova, M., and Zhuravskaya, E. (2011). Media and political persuasion: Evidence from Russia. *American Economic Review*, 101(7):3253–3285.

Esponda, I. and Vespa, E. (2014). Hypothetical Thinking and Information Extraction in the Laboratory. *American Economic Journal: Microeconomics*, 6(4):180–202.

Feddersen, T. J. and Pesendorfer, W. (1997). Voting Behavior and Information Aggregation in Elections with Private Information. *Econometrica*, 65(5):1029–1058.

Feddersen, T. J. and Pesendorfer, W. (1998). Convicting the Innoncent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting. *American Political Science Review*, 92(1):23–35.

Garz, M., Sörensen, J., and Stone, D. F. (2018). Congeniality and News Demand: Evidence from Facebook. Working paper.

Gentzkow, M. and Shapiro, J. M. (2006). Media bias and reputation. *Journal of Political Economy*, 114(2):280–316.

Gentzkow, M. and Shapiro, J. M. (2011). Ideological segregation online and offline. *Quarterly Journal of Economics*, 126(4):1799–1839.

Gerardi, D. and Yariv, L. (2007). Deliberative Voting. *Journal of Economic Theory*, 134(1):317–338.

Goeree, J. K. and Yariv, L. (2011). An Experimental Study of Collective Deliberation. *Econometrica*, 79(3):893–921.

Golman, R., Hagmann, D., and Loewenstein, G. (2017). Information Avoidance. *Journal of Economic Literature*, 55(1):96–135.

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

Groeling, T. (2013). Media Bias by the Numbers: Challenges and Opportunities in the Empirical Study of Partisan News. *Annual Review of Political Science*, 16(1):129–51.

Guarnaschelli, S., McKelvey, R. D., and Palfrey, T. R. (2000). An Experimental Study of Jury Decision Rules. *American Political Science Review*, 94(2):407–423.

Hagenbach, J., Koessler, F., and Perez-Richet, E. (2014). Certifiable Pre-play Communication: Full Disclosure. *Econometrica*, 82(3):1093–1131.

Halberstam, Y. and Knight, B. (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, 143(3):73–88.

Jin, G. Z., Luca, M., and Martin, D. (2017). Is no News (Perceived as) Bad News? An Experimental Investigation of Information Disclosure. Working paper.

Kennedy, P. and Prat, A. (2017). Where Do People Get Their News? Working paper.

Le Quement, M. T. and Marcin, I. (2016). Communication and voting in heterogeneous committees: An experimental study. Working paper.

Le Quement, M. T. and Yokeeswaran, V. (2015). Subgroup deliberation and voting. *Social Choice and Welfare*, pages 155–186.

Levy, G. and Razin, R. (2015). Correlation Neglect, Voting Behavior, and Information Aggregation. *American Economic Review*, 105(4):1634–1645.

Martin, G. J. and Yurukoglu, A. (2017). Bias in Cable News: Persuasion and Polarization. *American Economic Review*, 107(9):2565–2599.

Martinelli, C. and Palfrey, T. R. (2017). Communication and Information in Games of Collective Decision: A Survey of Experimental Results. Working paper.

Maug, E. and Yilmaz, B. (2002). Two-Class Voting: A Mechanism for Conflict Resolution. *American Economic Review*, 92(5):1448–1471.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27:415–444.

Meirowitz, A. (2007). In defense of exclusionary deliberation: Communication and voting with private beliefs and values. *Journal of Theoretical Politics*, 19(3):301–327.

Mengel, F. and Rivas, J. (2017). Common value elections with private information and informative priors: Theory and experiments. *Games and Economic Behavior*, 104:190–221.

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

Oliveros, S. and Várdy, F. (2015). Demand for slant: how abstention shapes voters' choice of news media. *Economic Journal*, 125(587):1327–1368.

Pariser, E. (2011). *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. The Penguin Press.

Piolatto, A. and Schuett, F. (2015). Media competition and electoral politics. *Journal of Public Economics*, 130:80–93.

Prat, A. (2015). Media power. *Journal of Political Economy, forthcoming*.

Schulte, E. (2010). Information aggregation and preference heterogeneity in committees. *Theory and Decision*, 69:97–118.

Appendix A Proofs and additional theoretical results

An (extended) player type specifies party preference (C_a or C_b) as well as observed private signal realization ($s_a, s_b, s_\emptyset$). A (pure) message strategy μ (applicable in treatments other than SN0) is a mapping from the set of types into the message space. Given our restrictions on communication protocols, the messages allowed are either truthful signal revelation or silence: $\mu : \{C_a, C_b\} \times \{s_a, s_b, s_\emptyset\} \to \{s_a, s_b, s_\emptyset\}$ and $\mu(\cdot, s_j) \neq s_{-j}$ for $j \in \{a, b\}$. That is, signals are "hard" evidence: those with empty signals cannot pretend they got a non-empty one, and those with non-empty signals cannot pretend they got a signal different from the one they have. We will also consider mixed message strategies of a special kind, called *semi-pooling*, in which players always reveal signals that match their party preference, but if they receive a signal that does not match their preference, then with some probability they hide it by reporting an empty signal, as if they are uninformed. A (mixed) voting strategy σ is a mapping from the set of types into the unit interval, representing the probability of voting for candidate C_a.

A.1 The case of no media bias

Lemma 1 *Under no communication and no media bias, there is no informative equilibrium.*

Proof. Since there is no abstention and each voter is independently uninformed with probability $r = 0.2$, a fully informative equilibrium is not always possible − the uninformed also have to vote. The "best" informative equilibrium one could hope for is the one in which all the informed voters vote their signals and the uninformed voters either mix or vote their bias. To keep things simple, let us assume first that everyone is always informed. Suppose everyone but $i \in C_a$ votes their signal and $s_i = s_b$. In a Bayesian Nash equilibrium, players should take into account that their vote only matters when they are pivotal (and form correct beliefs about how others vote). With our parameters and under assumed voting strategy of others, i is pivotal in two cases: (i) when five other voters voted for C_a and four for C_b, or (ii) four other voters voted for C_a and five for C_b. Together with i's signal, this means that in (i), there are 5 signals s_a and 5 signals s_b, and in (ii), there are 4 signals s_a and 6 signals s_b. Conditioning on pivotality (i.e. either of the two cases happening)

$$
\begin{aligned}
p_i(\theta = a|s_b, \text{piv}) &= \frac{\Pr(\theta = a, 5s_a, 5s_b) + \Pr(\theta = a, 4s_a, 6s_b)}{\Pr(\theta = a, 5s_a, 5s_b) + \Pr(\theta = a, 4s_a, 6s_b) + \Pr(\theta = b, 5s_a, 5s_b) + \Pr(\theta = b, 4s_a, 6s_b)} \\
&= \frac{\frac{1}{2}(1-q)\frac{9!}{5!4!}q^5(1-q)^4 + \frac{1}{2}(1-q)\frac{9!}{4!5!}q^4(1-q)^5}{\frac{1}{2}(1-q)\frac{9!}{5!4!}q^5(1-q)^4 + \frac{1}{2}(1-q)\frac{9!}{4!5!}q^4(1-q)^5 + \frac{1}{2}q\frac{9!}{4!5!}q^4(1-q)^5 + \frac{1}{2}q\frac{9!}{5!4!}q^5(1-q)^4} \\
&= \frac{q^5(1-q)^5 + q^4(1-q)^6}{q^5(1-q)^5 + q^4(1-q)^6 + q^5(1-q)^5 + q^6(1-q)^4} \quad // \text{ divide by } q^4(1-q)^4 \neq 0 \\
&= \frac{q(1-q) + (1-q)^2}{q(1-q) + (1-q)^2 + q(1-q) + q^2} = \frac{1-q}{1} = 0.3 > t_a
\end{aligned}
$$

Hence even though state b is more likely, i still prefers to vote for C_a, i.e. not to vote informatively. Now, if some voters may be uninformed, the analysis becomes more complicated − the posterior calculations require taking the expectation over each possible combination of uninformed and informed votes. However, since the empty signals are iid across voters, not correlated with the state, and there is no communication that could reveal the partisanship of the uninformed, the overall result does not change. ■

Lemma 2 *Under complete network and no media bias, there is no full information revelation equilibrium.*

Proof. Let $k := \#s_a - \#s_b$, $k \in \{-n, \ldots, n\}$ be the realization of the difference in the number of revealed s_a and s_b signals, and suppose k_{-i} is the reported difference in the number of s_a and s_b signals by players other than i. For each player i, $k \equiv k_{-i} + \mathbb{1}_{s_a} - \mathbb{1}_{s_b}$ is the difference in the number of s_a and s_b signals with i's signal included, which i will use to form her posterior belief that $\theta = a$. Conditional on k_{-i} this posterior is

$$p_i(\theta = a|k_{-i}) = \frac{\overbrace{((1-r)q)^{\#_{-i}(s_a)}((1-r)(1-q))^{\#_{-i}(s_b)}r^{n-1-\#_{-i}(s_a)-\#_{-i}(s_b)}}^{\text{denote as } X}}{X + ((1-r)(1-q))^{\#_{-i}(s_a)}((1-r)q)^{\#_{-i}(s_b)}r^{n-1-\#_{-i}(s_a)-\#_{-i}(s_b)}}$$
$$= \frac{1}{1+q^{-k_{-i}}(1-q)^{k_{-i}}} = \frac{1}{1+\left(\frac{1-q}{q}\right)^{k_{-i}}} \tag{5}$$

where on the first line, $\#_{-i}(s_j)$ is the total number of signals $s_j, j \in \{a,b\}$ revealed by players other than i. Suppose i got an s_a signal. Assuming that others always reveal their signals and believe that non-revelation means no signal with probability 1, and use sincere Bayesian voting strategies at the voting stage, i's message is pivotal with our parameters when $k_{-i} = 1$: If i withholds her signal, other voters would believe that i got no signal with probability 1 and have posterior $t_a < p_{-i}(\theta = a|1) = 0.7 < t_b$. So all C_a-partisans would favor voting for C_a, and all C_b-partisan would favor voting for C_b, creating a tie. But if i reveals s_a, $p_{-i}(\theta = a|2) \approx .845 > t_b > t_a$ so everyone would favor voting for C_a. In this case, both partisans prefer to reveal s_a. i's message is also pivotal when $k_{-i} = -2$, because in this case, if i withholds her signal, other voters would believe that i got no signal with probability 1 and favor voting for C_b, since $p_{-i}(\theta = a| -2) \approx .156 < t_a < t_b$, whereas if i reveals s_a, $t_a < p_{-i}(\theta = a| -1) = .3 < t_b$ and so all C_a-partisans, would favor voting for C_a, and all C_b-partisan would favor voting for C_b, creating a tie. Thus if i is C_a-partisan, she prefers to reveal signal s_a. However, if i is C_b-partisan, she prefers to keep silent because her own posterior is below t_b so she'd rather have the majority voting for C_b. Similarly, if i got an s_b signal, and assuming that others always reveal their signals and believe that non-revelation means no signal with probability 1, i's message would be pivotal for $k_{-i} = -1$, pushing the posterior below t_a and breaking the tie in favor of C_b. In this case, both partisans prefer to reveal s_b. i's message is also pivotal when $k_{-i} = 2$: if i withholds her signal, the majority would vote for C_a, whereas if i reveals s_b, $t_a < p_{-i}(\theta = a| + 1) = .7 < t_b$ and so all C_a-partisans would vote for C_a and C_b-partisans for C_b, creating a tie. If i is C_b-partisan, she prefers to reveal signal s_b. However, if i is C_a-partisan, she prefers to keep silent because her own posterior remains above t_a so she'd rather have the majority voting for C_a. Therefore, a fully revealing message strategy is not incentive compatible under complete network.[32] ■

Proof of Proposition 3. Suppose players use a (possibly mixed) semi-pooling message strategy, according to which they always reveal favorable signals and hide unfavorable signals with some state-independent probability $0 < \nu \leq 1$. Belief consistency requires that upon receiving an empty signal, denoted $\tilde{s}^j_\emptyset$, from a C_j-partisan, all other players believe that this signal is actually an unfavorable signal to C_j (rather than a true empty signal $s_\emptyset$) with probability

$$\mu_{-j}(\tilde{s}^j_\emptyset, \nu) \equiv \Pr(s = s_{-j}|\tilde{s}^j_\emptyset) = \frac{\frac{1}{2}(1-r)(1-q)\nu + \frac{1}{2}(1-r)q\nu}{\frac{1}{2}(1-r)(1-q)\nu + \frac{1}{2}(1-r)q\nu + r} = \frac{\frac{1}{2}(1-r)\nu}{\frac{1}{2}(1-r)\nu + r} \tag{6}$$

Of course, if they receive a non-empty signal, they believe it, since signals are hard evidence. Note also that $1 - \mu_{-j}(\tilde{s}^j_\emptyset, \nu) \equiv \Pr(s = s_\emptyset|\tilde{s}^j_\emptyset)$.[33] Due to symmetry, $\mu_a(\tilde{s}^b_\emptyset, \cdot) = \mu_b(\tilde{s}^a_\emptyset, \cdot)$, thus we can omit the subscript and simply write $\mu = \mu(\nu)$, where μ is an increasing function of ν.

Fix player i and consider a mixed semi-pooling message strategy, described above. Let $\#_{-i}(\tilde{s}^j_\emptyset)$ be the number of empty signals reported by C_j-partisans other than i, and $\#_{-i}(s_j)$ the total number of signals s_j revealed by players other than i, $j \in \{a,b\}$. The expected number of signals s_{-j} hidden by C_j-partisans, being the expectation of a binomial random variable, is

$$h_{-j}(\#_{-i}(\tilde{s}^j_\emptyset)) = \sum_{\ell=0}^{\#_{-i}(\tilde{s}^j_\emptyset)} \ell \binom{\#_{-i}(\tilde{s}^j_\emptyset)}{\ell} (\mu)^\ell (1-\mu)^{\#_{-i}(\tilde{s}^j_\emptyset)-\ell} \equiv \mu \cdot \#_{-i}(\tilde{s}^j_\emptyset) \tag{7}$$

[32] Schulte (2010, Proposition 2) gives a necessary and sufficient condition on individual preference heterogeneity for a full revelation equilibrium to exist in this setting.

[33] Since favorable signals are fully revealed, in equilibrium, players put probability zero on the event that an empty signal from a C_j-partisan is a hidden favorable signal s_j.

Let $k_{-i} = \#_{-i}(s_a) - \#_{-i}(s_b)$, $k_{-i}(\tilde{s}_\emptyset) = \#_{-i}(\tilde{s}^b_\emptyset) - \#_{-i}(\tilde{s}^a_\emptyset)$, and $\pi_{-i}(\tilde{s}_\emptyset) := h_a(\#_{-i}(\tilde{s}^b_\emptyset)) - h_b(\#_{-i}(\tilde{s}^a_\emptyset)) \equiv \mu \cdot k_{-i}(\tilde{s}_\emptyset)$. i's posterior that $\theta = a$ conditional on signals revealed (and non-revealed) by others is

$$
\begin{aligned}
&p_i(\theta = a | k_{-i}, \pi_{-i}(\tilde{s}_\emptyset)) = \\
&\quad = \frac{\overbrace{((1-r)q)^{\#_{-i}(s_a)+h_a(\#_{-i}(\tilde{s}^b_\emptyset))}((1-r)(1-q))^{\#_{-i}(s_b)+h_b(\#_{-i}(\tilde{s}^a_\emptyset))} r^{n-1-h_b(\#_{-i}(\tilde{s}^a_\emptyset))-h_a(\#_{-i}(\tilde{s}^b_\emptyset))}}^{\text{denote as } X}}{X + ((1-r)(1-q))^{\#_{-i}(s_a)+h_a(\#_{-i}(\tilde{s}^b_\emptyset))}((1-r)q)^{\#_{-i}(s_b)+h_b(\#_{-i}(\tilde{s}^a_\emptyset))} r^{n-1-h_b(\#_{-i}(\tilde{s}^a_\emptyset))-h_a(\#_{-i}(\tilde{s}^b_\emptyset))}} \\
&\quad = \frac{1}{1 + \left(\frac{1-q}{q}\right)^{\#_{-i}(s_a)+h_a(\#_{-i}(\tilde{s}^b_\emptyset))-\#_{-i}(s_b)-h_b(\#_{-i}(\tilde{s}^a_\emptyset))}} = \frac{1}{1 + \left(\frac{1-q}{q}\right)^{k_{-i}+\pi_{-i}(\tilde{s}_\emptyset)}} \\
&\quad = \frac{1}{1 + \left(\frac{1-q}{q}\right)^{k_{-i}+\mu k_{-i}(\tilde{s}_\emptyset)}} \qquad (8)
\end{aligned}
$$

Note that since $q = 0.7 > 0.5$, p_i is increasing in μ for $k_{-i}(\tilde{s}_\emptyset) > 0$ and decreasing in μ for $k_{-i}(\tilde{s}_\emptyset) < 0$. This implies the same dynamics for p_i as a function of equilibrium probability ν, since μ is increasing in ν, as follows from (6). i's decision whether or not to reveal her signal is going to affect the posterior held by others, $p_{-i}(\theta = a|k', \pi'(\tilde{s}_\emptyset))$, through a change in one of the numbers that they observe and condition upon: k' or $\pi'(\tilde{s}_\emptyset)$. Namely, if i reveals her signal, k' will be updated; if i hides her signal, $\pi'(\tilde{s}_\emptyset)$ will be updated.

Suppose i is a C_j-partisan. if i reveals, then instead of $k' = k_{-i}$, others will observe $k'' := k_{-i} + \mathbb{1}_{\{s_i=s_a\}} - \mathbb{1}_{\{s_i=s_b\}}$. C_j-partisans always reveal a favorable signal s_j under our semi-pooling strategy. If i receives an unfavorable signal s_{-j} and hides it, this will affect $\pi'(\tilde{s}_\emptyset)$ in the posterior of others: instead of $\pi'(\tilde{s}_\emptyset) = \pi_{-i}(\tilde{s}_\emptyset)$, others will observe $\pi''(\tilde{s}_\emptyset) = x_j(\tilde{s}_\emptyset)$, where $x_b(\tilde{s}_\emptyset) := \pi_{-i}(\tilde{s}_\emptyset) + \mu\mathbb{1}_{\{s_i=s_a\}}$, $x_a(\tilde{s}_\emptyset) := \pi_{-i}(\tilde{s}_\emptyset) - \mu\mathbb{1}_{\{s_i=s_b\}}$. Thus the effect of hiding an unfavorable signal on the others' posterior depends on μ. Exact posterior changes only matter around the two critical thresholds, t_a and t_b. Whatever i does with an unfavorable signal, either k' or $\pi'(\tilde{s}_\emptyset)$ will be updated and observed by others; and revealing an unfavorable signal has a larger effect (positive for s_a, negative for s_b) on the others' posterior than hiding it:

$$p_{-i}(\theta = a|k_{-i} - \mathbb{1}_{\{s_i=s_b\}}, \pi'(\tilde{s}_\emptyset)) < p_{-i}(\theta = a|k_{-i}, x_a(\tilde{s}_\emptyset)) \qquad (9)$$

and

$$p_{-i}(\theta = a|k_{-i} + \mathbb{1}_{\{s_i=s_a\}}, \pi'(\tilde{s}_\emptyset)) > p_{-i}(\theta = a|k_{-i}, x_b(\tilde{s}_\emptyset)) \qquad (10)$$

(with weak inequalities for a pure semi-pooling strategy).

In equilibrium, it must be incentive compatible for i to use the semi-pooling strategy ν, if she believes that the others also use it at the messaging stage and use Bayesian sincere strategies at the voting stage. Since revealing favorable signals is incentive compatible for any ν, the actual restrictions on equilibrium ν come from comparing the effect of hiding vs. revealing an unfavorable signal when i is pivotal. Due to (9), for $i \in C_a$ the respective pivotality condition is i) $p_{-i}(\theta = a|k'', \pi'(\tilde{s}_\emptyset)) < t_j < p_{-i}(\theta = a|k', \pi''(\tilde{s}_\emptyset))$, and due to (10), for $i \in C_b$ it is ii) $p_{-i}(\theta = a|k'', \pi'(\tilde{s}_\emptyset)) > t_j > p_{-i}(\theta = a|k', \pi''(\tilde{s}_\emptyset))$. As long as $\nu < 1$, there may be one weak inequality in both cases. If $t_j = t_a$, then in case i), $i \in C_a$ wants to reveal the unfavorable signal s_b, but in case ii), $i \in C_b$ wants to hide the unfavorable signal s_a. For any belief ν it is possible to affect the vote by revealing the unfavorable signal, since signals are hard evidence, so case i) does not restrict ν. However, for given $k', k'(\tilde{s}_\emptyset)$, in case ii) there is a range of ν for which hiding the signal will not work: The other players believe that an empty signal means "unfavorable" signal with too high a probability, thereby "undoing" the hiding. If $t_j = t_b$, the situation is reversed: in case i) $i \in C_a$ prefers to to hide her signal, whereas in case ii), $i \in C_b$ prefers to reveal her signal. To ensure incentive compatibility, it is sufficient to consider these conditions i)–ii) only at the critical values of μ at which the others' posterior, computed using an appropriately modified Eq (8), equals threshold t_j.

There are two critical values for each threshold: Either 1) $p_{-i}(\theta = a|k'', \pi'(\tilde{s}_\emptyset)) = t_j$ or 2) $p_{-i}(\theta = a|k', \pi''(\tilde{s}_\emptyset)) = t_j$. For $t_j = t_a$, $i \in C_b$, and case ii), if $k_{-i}(\tilde{s}_\emptyset) > 0$, it is condition 2) that defines the relevant critical value of μ, and if $k_{-i}(\tilde{s}_\emptyset) < 0$, it is condition 1) that defines the critical value of μ. For $t_j = t_b$, $i \in C_a$, and case i), if $k_{-i}(\tilde{s}_\emptyset) > 0$, it is condition 1) that defines the relevant critical value of μ, and if $k_{-i}(\tilde{s}_\emptyset) < 0$, it is condition 2) that defines the relevant critical value of μ.

So for fixed k_{-i}, $k_{-i}(\tilde{s}_\emptyset)$, there are four possibilities, and the corresponding critical values can be expressed via the following equations:

$$\mu^*_{i)1)}(s_b, C_a) = \frac{\ln\left(\frac{1}{t_b} - 1\right) - (k_{-i} - \mathbb{1}_{\{s_i = s_b\}}) \ln \frac{1-q}{q}}{k_{-i}(\tilde{s}_\emptyset) \ln \frac{1-q}{q}} \tag{11}$$

$$\mu^*_{i)2)}(s_b, C_a) = \frac{\ln\left(\frac{1}{t_b} - 1\right) - k_{-i} \ln \frac{1-q}{q}}{(k_{-i}(\tilde{s}_\emptyset) - \mathbb{1}_{\{s_i = s_b\}}) \ln \frac{1-q}{q}} \tag{12}$$

$$\mu^*_{ii)1)}(s_a, C_b) = \frac{\ln\left(\frac{1}{t_a} - 1\right) - (k_{-i} + \mathbb{1}_{\{s_i = s_a\}}) \ln \frac{1-q}{q}}{k_{-i}(\tilde{s}_\emptyset) \ln \frac{1-q}{q}} \tag{13}$$

$$\mu^*_{ii)2)}(s_a, C_b) = \frac{\ln\left(\frac{1}{t_a} - 1\right) - k_{-i} \ln \frac{1-q}{q}}{(k_{-i}(\tilde{s}_\emptyset) + \mathbb{1}_{\{s_i = s_a\}}) \ln \frac{1-q}{q}} \tag{14}$$

The critical values of ν, denoted ν^* are obtained by reversing (6):

$$\nu = \frac{2\mu r}{(1-\mu)(1-r)}. \tag{15}$$

It is straightforward to show that any $\nu \geq \nu^*$ is also incentive compatible. Thus we obtain a series of critical values ν^* that depend on i's partisanship, her signal, and different combinations of k_{-i}, and $k_{-i}(\tilde{s}_\emptyset)$, which define a consistency range for ν. We directly compute the consistency range for each case. A semi-pooling equilibrium probability ν must be in the intersection of these consistency ranges across all cases; direct computation yields that this range of ν is $(.922, 1]$.

At the voting stage, signals revealed and non-revealed become common knowledge, but individuals may have different posteriors, since some may have hidden their private signals and others got no signals. Players believe that each empty signal reported by a C_j-partisan is an unfavorable one with probability $\mu(\nu^*)$, given in (6). Since ν^* is incentive-compatible for all possible communication outcomes, sincere Bayesian voting remains a best response even conditional on vote pivotality. ■

Proof of Proposition 4. Fix player i, who is a C_j-partisan, and assume that all C_{-j}-partisans use fully revealing strategies. i's posterior about the state should be conditional on $k_{-i}(C_j)$, the reported difference in the number of s_a and s_b signals by C_j-partisans other than i (which i can observe), and it takes the following form:

$$\begin{aligned} p_i(\theta = a | k_{-i}(C_j)) &= \frac{\overbrace{((1-r)q)^{\#_{-i}(s_a(C_j))}((1-r)(1-q))^{\#_{-i}(s_b(C_j))} r^{\frac{n}{2}-1-\#_{-i}(s_a(C_j))-\#_{-i}(s_b(C_j))}}^{\text{denote as } X}}{X + ((1-r)(1-q))^{\#_{-i}(s_a(C_j))}((1-r)q)^{\#_{-i}(s_b(C_j))} r^{\frac{n}{2}-1-\#_{-i}(s_a(C_j))-\#_{-i}(s_b(C_j))}} \\ &= \frac{1}{1 + q^{-k_{-i}(C_j)}(1-q)^{k_{-i}(C_j)}} = \frac{1}{1 + \left(\frac{1-q}{q}\right)^{k_{-i}(C_j)}} \end{aligned} \tag{16}$$

where on the first line, $\#_{-i}(s_j(C_j))$ is the total number of signals $s_j, j \in \{a, b\}$ revealed by players other than i in group C_j. A full equilibrium description also requires players to form beliefs about the signals revealed in the other group conditional on their private signal as well as the signals revealed by others in their group, i.e., on $k(C_j) \equiv k_{-i}(C_j) + \mathbb{1}_{s_i = s_a} - \mathbb{1}_{s_i = s_b}$, to be used at the voting stage. Let

$$\mu(k_{-i}(C_{-j}) | \theta = a) = \sum_{\alpha=0}^{\frac{n}{2}} \sum_{\beta=0}^{\frac{n}{2}-\alpha} \frac{(\frac{n}{2})!(\alpha - \beta)}{\alpha!\beta!(\frac{n}{2} - \alpha - \beta)!} ((1-r)q)^\alpha ((1-r)(1-q))^\beta r^{\frac{n}{2}-\alpha-\beta}$$

be the expected difference in the number of revealed signals s_a and s_b in group C_{-j}, denoted $k_{-i}(C_{-j})$, conditional on state $\theta = a$, assuming C_{-j}-partisans are using fully revealing strategies, and let

$$\mu(k_{-i}(C_{-j})|\theta = b) = \sum_{\alpha=0}^{\frac{n}{2}} \sum_{\beta=0}^{\frac{n}{2}-\alpha} \frac{(\frac{n}{2})!(\alpha - \beta)}{\alpha!\beta!(\frac{n}{2} - \alpha - \beta)!}((1-r)(1-q))^{\alpha}((1-r)q)^{\beta} r^{\frac{n}{2}-\alpha-\beta}$$

be the same quantity conditional on state $\theta = b$. Since signals have the same accuracy in both states and both groups have the same size $n/2$, $\mu(k_{-i}(C_{-j})|\theta = a) = -\mu(k_{-i}(C_{-j})|\theta = b)$. Given beliefs μ and posterior $p_i(\theta = a|k(C_j))$, player i expects group C_{-j} to have a common posterior $p_{-j}(\theta = a|k(C_j))$:

$$\begin{aligned}
p_{-j}(\theta = a|k(C_j)) &= \frac{p_i(\theta = a|k(C_j))}{1 + \left(\frac{1-q}{q}\right)^{\mu(k_{-i}(C_{-j})|\theta=a)}} + \frac{1 - p_i(\theta = a|k(C_j))}{1 + \left(\frac{1-q}{q}\right)^{\mu(k_{-i}(C_{-j})|\theta=b)}} \\
&= \frac{p_i(\theta = a|k(C_j))}{1 + \left(\frac{1-q}{q}\right)^{\mu(k_{-i}(C_{-j})|\theta=a)}} + \frac{1 - p_i(\theta = a|k(C_j))}{1 + \left(\frac{1-q}{q}\right)^{-\mu(k_{-i}(C_{-j})|\theta=a)}} \\
&= p_i(\theta = a|k(C_j)) \left(\frac{1}{1 + \left(\frac{1-q}{q}\right)^{\mu(k_{-i}(C_{-j})|\theta=a)}} - \frac{1}{1 + \left(\frac{1-q}{q}\right)^{-\mu(k_{-i}(C_{-j})|\theta=a)}} \right) \\
&\quad + \frac{1}{1 + \left(\frac{1-q}{q}\right)^{-\mu(k_{-i}(C_{-j})|\theta=a)}} \\
&\approx 0.5901 \cdot p_i(\theta = a|k(C_j)) + 0.2049
\end{aligned} \tag{17}$$

where the last line is obtained using our parameters ($n = 10, q = 0.7, r = 0.2$, which imply $\mu(k_{-i}(C_{-j})|\theta = a) = 1.6$). So if i is C_a-partisan, assuming all others are using fully revealing message strategies and sincere Bayesian voting strategies, i's signal of s_a is pivotal when $k_{-i}(C_a) = -2$, implying $k(C_a) = -1$. In this case, as well as for $k(C_a) \in \{0, \ldots, 5\}$, i's vote for C_a is pivotal and i expects a tie. i's signal of s_b is pivotal when $k_{-i}(C_a) = -1$, however then $k(C_a) = -2$ and i's vote is not pivotal because the majority is expected to vote for C_b. In this case, i can vote either way, in particular, vote for C_b, as prescribed by a sincere informative voting strategy. Similarly, if i is C_b-partisan, assuming all others are using fully revealing message strategies and sincere Bayesian voting strategies, i's signal of s_b is pivotal when $k_{-i}(C_b) = 2$, implying $k(C_b) = 1$. In this case, as well as for $k(C_b) \in \{-5, \ldots, 0\}$, i's vote for C_b is pivotal and i expects a tie. i's signal of s_a is pivotal when $k_{-i}(C_b) = 1$, however then $k(C_b) = 2$ and i's vote is not pivotal because the majority is expected to vote for C_a. In this case, i can vote either way, in particular, vote for C_a, as prescribed by a sincere informative voting strategy. Hence there is a fully revealing equilibrium in which all voters vote sincerely and, sometimes, informatively – grouping voters by their preference biases enables information aggregation within groups. Notice also that all C_j-partisans share a common posterior

$$p_j(\theta = a|k(C_j)) = \frac{1}{1 + \left(\frac{1-q}{q}\right)^{k(C_j)}} \tag{18}$$

■

A.2 The case of moderate media bias

Proof of Proposition 2. Suppose i is C_j-partisan. If players use sincere Bayesian voting strategies, then i would vote for C_j after signal s_j, but would vote for C_{-j} after signal s_{-j} since $p_i(\theta = j|s_{-j}) < t_j$ (as follows from (3)–(4)). After an empty signal, i just votes her bias. If everyone else is voting sincerely and informatively (with the exception of the uninformed), i's vote is pivotal when the difference between the number of s_a and s_b signals amongst others, denoted k_{-i}, is in $\{-1, 0, 1\}$. This leaves the following possibilities to consider: i) extra signal s_j in k_{-i} is reported by a C_j-partisan; ii) extra signal s_j in k_{-i} is reported by a C_{-j}-partisan; iii) extra signal s_{-j} in k_{-i} is reported by a C_j-partisan; iv) extra signal

s_{-j} in k_{-i} is reported by a C_{-j}-partisan; or v) there is no extra signal. if i got s_j signal (recall that i is C_j-partisan) she should vote her signal in all cases i)–v). If i got s_{-j} signal, in cases iii) – v) she should vote C_{-j}. However, in cases i) and ii), she should vote C_j. Since i does not observe which of i)–v) takes place due to no communication, and conditional on observing s_{-j} and pivotality, events iii)–iv) are more likely than events i)–ii), i's voting informatively remains incentive compatible. ■

Lemma 3 *Under complete network and moderate media bias, there is no full information revelation equilibrium.*

Proof. Let $\#s_j(C_j), \#s_{-j}(C_j)$ be the total number of s_j and s_{-j} signals reported by C_j-partisans, $j \in \{a, b\}$. Fix player i, and let $k_{-i}(\text{f}) := \#_{-i}(s_a(C_a)) - \#_{-i}(s_b(C_b))$ be the difference in the number of reported favorable signals s_a by C_a-partisans other than i and reported favorable signals s_b by C_b-partisans other than i. Let $k_{-i}(\text{uf}) := \#_{-i}(s_a(C_b)) - \#_{-i}(s_b(C_a))$ be the difference in the number of reported unfavorable signals s_a by C_b-partisans other than i and reported unfavorable signals s_b by C_a-partisans other than i. Consider a fully revealing message strategy, according to which everyone is always sharing their non-empty signals during the communication stage. i's posterior that $\theta = a$ conditional on signals revealed by others is

$$\begin{aligned}
p_i(\theta = a|k_{-i}(\text{f}), k_{-i}(\text{uf})) &= \frac{\overbrace{(q_a^a)^{\#_{-i}(s_a(C_a))}(q_b^a)^{\#_{-i}(s_a(C_b))}(1-q_a^a)^{\#_{-i}(s_b(C_a))}(1-q_b^a)^{\#_{-i}(s_b(C_b))}}^{\text{denote as } Z}}{Z + (1-q_a^b)^{\#_{-i}(s_a(C_a))}(1-q_b^b)^{\#_{-i}(s_a(C_b))}(q_a^b)^{\#_{-i}(s_b(C_a))}(q_b^b)^{\#_{-i}(s_b(C_b))}} \\
&= \frac{1}{1 + \left(\frac{1-q_a^b}{q_a^a}\right)^{\#_{-i}(s_a(C_a))}\left(\frac{1-q_b^b}{q_b^a}\right)^{\#_{-i}(s_a(C_b))}\left(\frac{q_a^b}{1-q_a^a}\right)^{\#_{-i}(s_b(C_a))}\left(\frac{q_b^b}{1-q_b^a}\right)^{\#_{-i}(s_b(C_b))}} \\
&= \frac{1}{1 + \left(\frac{2}{3}\right)^{\#_{-i}(s_a(C_a))}\left(\frac{1}{4}\right)^{\#_{-i}(s_a(C_b))}\left(\frac{1}{4}\right)^{-\#_{-i}(s_b(C_a))}\left(\frac{2}{3}\right)^{-\#_{-i}(s_b(C_b))}} \\
&= \frac{1}{1 + \left(\frac{2}{3}\right)^{k_{-i}(\text{f})}\left(\frac{1}{4}\right)^{k_{-i}(\text{uf})}} \qquad (19)
\end{aligned}$$

The transition from line 2 to line 3 follows by substituting $q_a^a = q_b^b = 0.9$ and $q_b^a = q_a^b = 0.4$ from our parameters. Denote $k_a(s_a) := k_{-i}(\text{f}) + \mathbb{1}_{s_a}$, $k_a(s_b) := k_{-i}(\text{uf}) - \mathbb{1}_{s_b}$ the reported difference in the number of s_a and s_b signals with i's signal included, if i is C_a-partisan; and $k_b(s_a) := k_{-i}(\text{uf}) + \mathbb{1}_{s_a}$, $k_b(s_b) := k_{-i}(\text{f}) - \mathbb{1}_{s_b}$ the reported difference in the number of s_a and s_b signals with i's signal included, if i is C_b-partisan. Denote $\boldsymbol{k} := (k_a(s_a), k_a(s_b), k_b(s_a), k_b(s_b))$ the vector of decision-relevant revealed signal differences. i's message is pivotal for different values of $k_{-i}(\text{f}), k_{-i}(\text{uf})$, as described in Table 10 (by revealing her signal, i pushes the others' posterior over one of the two critical thresholds, t_a and t_b).

Similarly to the no bias case, Table 10 shows that both partisan types sometimes have incentives not to reveal an unfavorable signal, so there is no full information revelation equilibrium. ■

Proposition 5 ***[Moderate media bias and complete network]*** *Given our experimental parameters, under MB2+SN2, there is no full information revelation equilibrium. However, there is a range of semi-pooling equilibria, in which all C_j-partisans, $j \in \{a, b\}$, with favorable signals s_j reveal them truthfully at the communication stage, and hide the unfavorable signals s_{-j} with a commonly known equilibrium probability ν^*. At the voting stage, each player i has a potentially different posterior p_i, which depends not only on the number and type of signals but also their* sources *– whether the signal comes from a voter who favors their candidate or the opposing candidate.*[34] *Given these posteriors, C_a-partisans vote for C_a as long as $p_i > t_a$, and otherwise vote C_b. C_b-partisans vote for C_b as long as $p_i < t_b$, and otherwise vote C_a. Each such equilibrium is characterized by fixing any $\nu^* \in (0.862, 1]$. Information gets partially aggregated.*

Proof of Proposition 5. Suppose players use a (possibly mixed) semi-pooling message strategy, in which they always reveal favorable signals and hide unfavorable signals with some state-independent

[34] See Eq (22), in which either $k_{-i}(\text{f})$, the difference in the number of revealed favorable signals s_a and s_b, or $k_{-i}(\text{uf})$, the difference in the number of revealed unfavorable signals s_a and s_b, is supplemented by i's private signal.

Table 10: Individual i's Message Pivotality

IC	Partisan type	Signal $s_i \neq s_\emptyset$	Others' fav. diff. $k_{-i}(\text{f})$	Others' unfav. diff. $k_{-i}(\text{uf})$	Others' posterior if i hides s_i $p_{-i}(\theta = a \mid k_{-i}(\text{f}), k_{-i}(\text{uf}))$	Others' posterior if i reveals s_i $p_{-i}(\theta = a \mid \boldsymbol{k})$
$*$	C_a	s_a	-4	0	.165	.229
$*$	–	–	-1	1	.727	.800
$*$	–	–	0	-1	.200	.273
$*$	–	–	3	0	.771	.835
$*$	C_a	s_b	-4	1	.441	.164
$*$	–	–	-3	0	.229	.069
	–	–	-3	2	.825	.542
$*$	–	–	-2	0	.308	.100
	–	–	-2	2	.877	.640
$*$	–	–	-1	0	.400	.143
	–	–	-1	2	.914	.727
$*$	–	–	0	0	.500	.200
	–	–	0	1	.800	.500
$*$	–	–	1	-1	.273	.086
	–	–	1	1	.857	.600
$*$	–	–	2	-1	.360	.123
	–	–	2	1	.900	.692
$*$	–	–	3	-1	.458	.174
	–	–	3	1	.931	.771
	–	–	4	0	.835	.558
	C_b	s_a	-4	0	.164	.441
	–	–	-3	-1	.069	.229
$*$	–	–	-3	1	.542	.826
	–	–	-2	-1	.100	.308
$*$	–	–	-2	1	.640	.877
	–	–	-1	-1	.143	.400
$*$	–	–	-1	1	.727	.914
	–	–	0	-1	.200	.500
$*$	–	–	0	0	.500	.800
	–	–	1	-2	.086	.272
$*$	–	–	1	0	.600	.857
	–	–	2	-2	.123	.360
$*$	–	–	2	0	.692	.900
	–	–	3	-2	.174	.458
$*$	–	–	3	0	.771	.931
$*$	–	–	4	-1	.558	.835
$*$	C_b	s_b	-3	0	.228	.165
$*$	–	–	0	1	.800	.727
$*$	–	–	1	-1	.273	.200
$*$	–	–	4	0	.835	.771

Notes: Critical thresholds on others' posterior are $t_a = 7/34$ for C_a members, and $t_b = 27/34$ for C_b members. "$*$" in column IC denotes incentive-compatible revelation assuming everyone else is fully revealing and voting sincerely.

probability $0 < \nu \leq 1$. Belief consistency requires that upon observing an empty signal, $\tilde{s}^j_\emptyset$, reported by a C_j-partisan, all other players believe that this signal is actually an unfavorable signal to C_j with probability

$$\mu_{-j} \equiv \Pr(s = s_{-j}|\tilde{s}^j_\emptyset) = \frac{\frac{1}{2}(1-r)(1-q^j_j)\nu + \frac{1}{2}(1-r)q^{-j}_j\nu}{\frac{1}{2}(1-r)(1-q^j_j)\nu + \frac{1}{2}(1-r)q^{-j}_j\nu + r}$$
$$= \frac{\frac{1}{2}(1-r)\nu(1-q^j_j+q^{-j}_j)}{\frac{1}{2}(1-r)\nu(1-q^j_j+q^{-j}_j)+r} \tag{20}$$

Due to symmetry, $\mu_a(\tilde{s}^b_\emptyset) = \mu_b(\tilde{s}^a_\emptyset)$, thus we can omit the subscript and simply write μ. Fix player i and denote $\#_{-i}(\tilde{s}^j_\emptyset)$ the number of empty signals reported by C_j-partisans other than i. The expected number of unfavorable signals s_{-j} hidden by C_j-partisans amongst $\#_{-i}(\tilde{s}^j_\emptyset)$ reported empty signals is $h_{-j}(\#_{-i}(\tilde{s}^j_\emptyset)) = \mu \cdot \#_{-i}(\tilde{s}^j_\emptyset)$ (using (7)). Let $k_{-i}(\tilde{s}_\emptyset) := \#_{-i}(\tilde{s}^b_\emptyset) - \#_{-i}(\tilde{s}^a_\emptyset)$ be the difference in the number of empty signals reported by C_b partisans and C_a partisans; with our parameters, $k_{-i}(\tilde{s}_\emptyset) \in \{-5\mathbb{1}_{i\in C_b}, -4, \ldots, 4, 5\mathbb{1}_{i\in C_a}\}$.[35] Let $\pi_{-i}(\tilde{s}_\emptyset) := h_a(\#_{-i}(\tilde{s}^b_\emptyset)) - h_b(\#_{-i}(\tilde{s}^a_\emptyset)) = \mu\left(\#_{-i}(\tilde{s}^b_\emptyset) - \#_{-i}(\tilde{s}^a_\emptyset)\right) \equiv \mu k_{-i}(\tilde{s}_\emptyset)$ be the difference in expected unfavorable signals. Let $k_{-i}(\text{f}) := \#_{-i}(s_a(C_a)) - \#_{-i}(s_b(C_b))$ be the difference in the number of reported favorable signals s_a by C_a-partisans other than i and reported favorable signals s_b by C_b-partisans other than i; with our parameters, $k_{-i}(\text{f}) \in \{-5\mathbb{1}_{i\in C_a}, -4, \ldots, 4, 5\mathbb{1}_{i\in C_b}\}$. Let $k_{-i}(\text{uf}) := \#_{-i}(s_a(C_b)) - \#_{-i}(s_b(C_a))$ be the difference in the number of reported unfavorable signals s_a by C_b-partisans other than i and reported unfavorable signals s_b by C_a-partisans other than i; with our parameters, $k_{-i}(\text{uf}) \in \{-5\mathbb{1}_{i\in C_b}, -4, \ldots, 4, 5\mathbb{1}_{i\in C_a}\}$. By definition,

$$k_{-i}(\text{f}) - k_{-i}(\text{uf}) = k_{-i}(\tilde{s}_\emptyset) + \mathbb{1}_{i\in C_b} - \mathbb{1}_{i\in C_a} \tag{21}$$

Thus while we'll keep using $k_{-i}(\tilde{s}_\emptyset)$ as a shorthand notation for the difference in the number of empty signals, it is not an independent quantity and can be obtained from the respective differences in the number of favorable and unfavorable signals. i's posterior that $\theta = a$ conditional on signals revealed (and non-revealed) by others becomes

$$p_i(\theta = a|k_{-i}(\text{f}), k_{-i}(\text{uf}), \pi_{-i}(\tilde{s}_\emptyset)) =$$
$$= \frac{\overbrace{(q^a_a)^{\#_{-i}(s_a(C_a))}(q^a_b)^{\#_{-i}(s_a(C_b))+h_a(\#_{-i}(\tilde{s}^b_\emptyset))}(1-q^a_a)^{\#_{-i}(s_b(C_a))+h_b(\#_{-i}(\tilde{s}^a_\emptyset))}(1-q^a_b)^{\#_{-i}(s_b(C_b))}}^{\text{denote as } Z}}{Z + (1-q^b_a)^{\#_{-i}(s_a(C_a))}(1-q^b_b)^{\#_{-i}(s_a(C_b))+h_a(\#_{-i}(\tilde{s}^b_\emptyset))}(q^b_a)^{\#_{-i}(s_b(C_a))+h_b(\#_{-i}(\tilde{s}^a_\emptyset))}(q^b_b)^{\#_{-i}(s_b(C_b))}}$$
$$= \frac{1}{1 + \left(\frac{1-q^b_a}{q^a_a}\right)^{\#_{-i}(s_a(C_a))}\left(\frac{1-q^b_b}{q^a_b}\right)^{\#_{-i}(s_a(C_b))+h_a(\#_{-i}(\tilde{s}^b_\emptyset))}\left(\frac{q^b_a}{1-q^a_a}\right)^{\#_{-i}(s_b(C_a))+h_b(\#_{-i}(\tilde{s}^a_\emptyset))}\left(\frac{q^b_b}{1-q^a_b}\right)^{\#_{-i}(s_b(C_b))}}$$
$$= \frac{1}{1 + \left(\frac{2}{3}\right)^{\#_{-i}(s_a(C_a))}\left(\frac{1}{4}\right)^{\#_{-i}(s_a(C_b))+h_a(\#_{-i}(\tilde{s}^b_\emptyset))}\left(\frac{1}{4}\right)^{-\#_{-i}(s_b(C_a))-h_b(\#_{-i}(\tilde{s}^a_\emptyset))}\left(\frac{2}{3}\right)^{-\#_{-i}(s_b(C_b))}}$$
$$= \frac{1}{1 + \left(\frac{2}{3}\right)^{k_{-i}(\text{f})}\left(\frac{1}{4}\right)^{k_{-i}(\text{uf})+\pi_{-i}(\tilde{s}_\emptyset)}} = \frac{1}{1 + \left(\frac{2}{3}\right)^{k_{-i}(\text{f})}\left(\frac{1}{4}\right)^{k_{-i}(\text{uf})+\mu k_{-i}(\tilde{s}_\emptyset)}} \tag{22}$$

The remaining equilibrium analysis is very similar to the case of no bias and complete network, with a few extra complications, since p_i now depends on $k_{-i}(\text{f})$ and $k_{-i}(\text{uf})$ separately – players have to distinguish between signal sources. We take off from the expression for player i's posterior, obtained in (22):

$$p_i(\theta = a|k_{-i}(\text{f}), k_{-i}(\text{uf}), \pi_{-i}(\tilde{s}_\emptyset)) = \frac{1}{1 + \left(\frac{2}{3}\right)^{k_{-i}(\text{f})}\left(\frac{1}{4}\right)^{k_{-i}(\text{uf})+\mu k_{-i}(\tilde{s}_\emptyset)}} \tag{23}$$

Note that p_i is increasing in μ for $k_{-i}(\tilde{s}_\emptyset) > 0$ and decreasing in μ for $k_{-i}(\tilde{s}_\emptyset) < 0$. This implies the same dynamics for p_i as a function of equilibrium probability ν, since μ is increasing in ν, as follows

[35] The shorthand notation $z\mathbb{1}_{i\in C_j}$ means that z should be only considered when i is a C_j-partisan, $j \in \{a, b\}$ to cover both possible cases.

from (20). i's decision whether or not to reveal her signal is going to affect the posterior held by others, $p_{-i}(\theta = a|k'(\text{f}), k'(\text{uf}), \pi'(\tilde{s}_\emptyset))$, through a change in one of the numbers that they observe and condition upon: $k'(\text{f})$, $k'(\text{uf})$, or $\pi'(\tilde{s}_\emptyset)$. Namely, if i reveals her signal, $k'(\text{f})$ or $k'(\text{uf})$ will be updated; if i hides her signal, $\pi'(\tilde{s}_\emptyset)$ will be updated.

Suppose i is a C_j-partisan. C_j-partisans always reveal a favorable signal s_j, which affects $k'(\text{f})$ in the posterior of others: if i reveals, then instead of $k'(\text{f}) = k_{-i}(\text{f})$, others will observe $k''(\text{f}) = k_j(s_j)$, where $k_a(s_a) := k_{-i}(\text{f}) + \mathbb{1}_{\{s_i=s_a\}}$, and $k_b(s_b) := k_{-i}(\text{f}) - \mathbb{1}_{\{s_i=s_b\}}$. If i receives an unfavorable signal s_{-j} and reveals it, this will affect $k'(\text{uf})$ in the posterior of others: if i reveals, then instead of $k'(\text{uf}) = k_{-i}(\text{uf})$, others will observe $k''(\text{uf}) = k_j(s_{-j})$, where $k_a(s_b) := k_{-i}(\text{uf}) - \mathbb{1}_{\{s_i=s_b\}}$ and $k_b(s_a) := k_{-i}(\text{uf}) + \mathbb{1}_{\{s_i=s_a\}}$. If i receives an unfavorable signal s_{-j} and hides it, this will affect $\pi'(\tilde{s}_\emptyset)$ in the posterior of others: if i hides, then instead of $\pi'(\tilde{s}_\emptyset) = \pi_{-i}(\tilde{s}_\emptyset)$, others will observe $\pi''(\tilde{s}_\emptyset) = x_j(\tilde{s}_\emptyset)$, where $x_b(\tilde{s}_\emptyset) := \pi_{-i}(\tilde{s}_\emptyset) + \mu\mathbb{1}_{\{s_i=s_a\}}$, $x_a(\tilde{s}_\emptyset) := \pi_{-i}(\tilde{s}_\emptyset) - \mu\mathbb{1}_{\{s_i=s_b\}}$. Thus the effect of hiding an unfavorable signal on the others' posterior depends on μ. Exact posterior changes only matter around the two critical thresholds, t_a and t_b. Whatever i does with an unfavorable signal, either $k''(\text{uf})$ or $\pi''(\tilde{s}_\emptyset)$ will be updated and observed by others; and revealing an unfavorable signal has a larger effect (positive for s_a, negative for s_b) on the others' posterior than hiding it:

$$p_{-i}(\theta = a|k'(\text{f}), k_a(s_b), \pi'(\tilde{s}_\emptyset)) < p_{-i}(\theta = a|k'(\text{f}), k'(\text{uf}), x_a(\tilde{s}_\emptyset)) \tag{24}$$

and

$$p_{-i}(\theta = a|k'(\text{f}), k_b(s_a), \pi'(\tilde{s}_\emptyset)) > p_{-i}(\theta = a|k'(\text{f}), k'(\text{uf}), x_b(\tilde{s}_\emptyset)) \tag{25}$$

(with weak inequalities for a pure semi-pooling strategy). In equilibrium, it must be incentive compatible for i to use the semi-pooling strategy ν, if she believes that the others also use it at the messaging stage and use Bayesian sincere strategies at the voting stage. Since revealing favorable signals is incentive compatible for any ν, the actual restrictions on equilibrium ν come from comparing the effect of hiding vs. revealing an unfavorable signal when i is pivotal. Due to (24), for $i \in C_a$ the respective pivotality condition is i) $p_{-i}(\theta = a|k'(\text{f}), k''(\text{uf}), \pi'(\tilde{s}_\emptyset)) < t_j < p_{-i}(\theta = a|k'(\text{f}), k'(\text{uf}), \pi''(\tilde{s}_\emptyset))$, and due to (25), for $i \in C_b$ it is ii) $p_{-i}(\theta = a|k'(\text{f}), k''(\text{uf}), \pi'(\tilde{s}_\emptyset)) > t_j > p_{-i}(\theta = a|k'(\text{f}), k'(\text{uf}), \pi''(\tilde{s}_\emptyset))$. As long as $\nu < 1$, there may be one weak inequality in both cases. If $t_j = t_a$, then in case i), $i \in C_a$ wants to reveal the unfavorable signal s_b, but in case ii), $i \in C_b$ wants to hide the unfavorable signal s_a. For any belief ν it is possible to affect the vote by revealing the unfavorable signal, since signals are verifiable, so case i) does not restrict ν. However, for given $k(\text{f}), k(\text{uf}), k(\tilde{s}_\emptyset)$, in case ii) there is a range of ν for which hiding the signal will not work: The other players believe that an empty signal means "unfavorable" signal with too high a probability, thereby "undoing" the hiding. If $t_j = t_b$, the situation is reversed: in case i) $i \in C_a$ prefers to to hide her signal, whereas in case ii), $i \in C_b$ prefers to reveal her signal. To ensure incentive compatibility, it is sufficient to consider these conditions i)–ii) only at the critical values of μ at which the others' posterior, computed using an appropriately modified Eq (22), equals threshold t_j.

There are two critical values for each threshold: Either 1) $p_{-i}(\theta = a|k'(\text{f}), k''(\text{uf}), \pi'(\tilde{s}_\emptyset)) = t_j$ or 2) $p_{-i}(\theta = a|k'(\text{f}), k'(\text{uf}), \pi''(\tilde{s}_\emptyset)) = t_j$. For $t_j = t_a$, $i \in C_b$, and case ii), if $k_{-i}(\tilde{s}_\emptyset) > 0$, it is condition 2) that defines the relevant critical value of μ, and if $k_{-i}(\tilde{s}_\emptyset) < 0$, it is condition 1) that defines the critical value of μ. For $t_j = t_b$, $i \in C_a$, and case i), if $k_{-i}(\tilde{s}_\emptyset) > 0$, it is condition 1) that defines the relevant critical value of μ, and if $k_{-i}(\tilde{s}_\emptyset) < 0$, it is condition 2) that defines the relevant critical value of μ. So for fixed $k_{-i}(\text{f})$, $k_{-i}(\text{uf})$, there are four possibilities, and the corresponding critical values can be expressed via the following equations:

$$\mu^*_{i)1)}(s_b, C_a) = \frac{\ln\left(\frac{1}{t_b} - 1\right) - k_{-i}(\mathrm{f}) \ln \frac{2}{3} - (k_{-i}(\mathrm{uf}) - \mathbb{1}_{\{s_i = s_b\}}) \ln \frac{1}{4}}{k_{-i}(\tilde{s}_\emptyset) \ln \frac{1}{4}} \tag{26}$$

$$\mu^*_{i)2)}(s_b, C_a) = \frac{\ln\left(\frac{1}{t_b} - 1\right) - k_{-i}(\mathrm{f}) \ln \frac{2}{3} - k_{-i}(\mathrm{uf}) \ln \frac{1}{4}}{(k_{-i}(\tilde{s}_\emptyset) - \mathbb{1}_{\{s_i = s_b\}}) \ln \frac{1}{4}} \tag{27}$$

$$\mu^*_{ii)1)}(s_a, C_b) = \frac{\ln\left(\frac{1}{t_a} - 1\right) - k_{-i}(\mathrm{f}) \ln \frac{2}{3} - (k_{-i}(\mathrm{uf}) + \mathbb{1}_{\{s_i = s_a\}}) \ln \frac{1}{4}}{k_{-i}(\tilde{s}_\emptyset) \ln \frac{1}{4}} \tag{28}$$

$$\mu^*_{ii)2)}(s_a, C_b) = \frac{\ln\left(\frac{1}{t_a} - 1\right) - k_{-i}(\mathrm{f}) \ln \frac{2}{3} - k_{-i}(\mathrm{uf}) \ln \frac{1}{4}}{(k_{-i}(\tilde{s}_\emptyset) + \mathbb{1}_{\{s_i = s_a\}}) \ln \frac{1}{4}} \tag{29}$$

The critical values of ν, denoted ν^* are obtained by reversing (20):

$$\nu = \frac{2\mu r}{(1-\mu)(1-r)(1-q_j^j + q_j^{-j})}, \tag{30}$$

It is straightforward to show that any $\nu \geq \nu^*$ is also incentive compatible. Thus we obtain a series of critical values ν^* that depend on i's partisanship, her signal, and different combinations of $k_{-i}(\mathrm{f})$, $k_{-i}(\mathrm{uf})$, and $k_{-i}(\tilde{s}_\emptyset)$, which define a consistency range for ν. We directly compute the consistency range for each case. A semi-pooling equilibrium probability ν must be in the intersection of these consistency ranges across all cases; direct computation yields that this range of ν is $(.862, 1]$. ■

Proposition 6 ***[Moderate media bias and polarized network]*** *Given our experimental parameters, under MB2+SN1, there is a full information revelation equilibrium, in which all voters with non-empty signals reveal them truthfully at the communication stage and believe with probability 1 that non-revealing agents are uninformed. At the voting stage, all C_j-partisans, $j \in \{a, b\}$, have identical posterior beliefs, in which unfavorable signals receive more weight relative to favorable signals.*[36] *C_a-partisans vote for C_a as long as $p_i > t_a$, and otherwise vote C_b. C_b-partisans vote for C_b as long as $p_i < t_b$, and otherwise vote C_a. Information gets partially aggregated.*

Proof of Proposition 6. The analysis is completely analogous to the case of polarized network and no bias, with some modifications regarding the expressions for the posteriors. Namely, (16) becomes

$$p_i(\theta = a | \#_{-i}(s_a(C_a)), \#_{-i}(s_b(C_a))) = \frac{1}{1 + \left(\frac{q_a^b}{1-q_a^a}\right)^{\#_{-i}(s_b(C_a))} \left(\frac{1-q_a^b}{q_a^a}\right)^{\#_{-i}(s_a(C_a))}} \tag{31}$$

$$p_i(\theta = a | \#_{-i}(s_a(C_b)), \#_{-i}(s_b(C_b))) = \frac{1}{1 + \left(\frac{q_b^b}{1-q_b^a}\right)^{\#_{-i}(s_b(C_b))} \left(\frac{1-q_b^b}{q_b^a}\right)^{\#_{-i}(s_a(C_b))}} \tag{32}$$

for $i \in C_a$ and $i \in C_b$, respectively. In the full revelation equilibrium, C_j-partisans have a common posterior with i's non-empty signal s_i added to $\#_{-i}(s_i(C_j))$ under the conditioning operator: Let $\#(s_a(C_j)) := \#_{-i}(s_a(C_j)) + \mathbb{1}_{\{s_i = s_a\}}$ and $\#(s_b(C_j)) := \#_{-i}(s_b(C_j)) - \mathbb{1}_{\{s_i = s_b\}}$. Players form beliefs about the expected number of revealed signals of each type in the other group.

Beliefs of C_a-partisans about the expected number of signals s_a (first line) and s_b (second line) revealed

[36] The posteriors are given by Eq (31) – (32), supplemented by i's signal.

in group C_b in state $\theta = a$:

$$\mu_{C_a}(\#(s_a(C_b))|\theta = a) = \sum_{\alpha=0}^{\frac{n}{2}} \sum_{\beta=0}^{\frac{n}{2}-\alpha} \frac{\alpha(\frac{n}{2})!}{\alpha!\beta!(\frac{n}{2}-\alpha-\beta)!}((1-r)q_b^a)^\alpha((1-r)(1-q_b^a))^\beta r^{\frac{n}{2}-\alpha-\beta} \tag{33}$$

$$\mu_{C_a}(\#(s_b(C_b))|\theta = a) = \sum_{\alpha=0}^{\frac{n}{2}} \sum_{\beta=0}^{\frac{n}{2}-\alpha} \frac{\beta(\frac{n}{2})!}{\alpha!\beta!(\frac{n}{2}-\alpha-\beta)!}((1-r)q_b^a)^\alpha((1-r)(1-q_b^a))^\beta r^{\frac{n}{2}-\alpha-\beta} \tag{34}$$

and in state $\theta = b$:

$$\mu_{C_a}(\#(s_a(C_b))|\theta = b) = \sum_{\alpha=0}^{\frac{n}{2}} \sum_{\beta=0}^{\frac{n}{2}-\alpha} \frac{\alpha(\frac{n}{2})!}{\alpha!\beta!(\frac{n}{2}-\alpha-\beta)!}((1-r)(1-q_b^b))^\alpha((1-r)q_b^b)^\beta r^{\frac{n}{2}-\alpha-\beta} \tag{35}$$

$$\mu_{C_a}(\#(s_b(C_b))|\theta = b) = \sum_{\alpha=0}^{\frac{n}{2}} \sum_{\beta=0}^{\frac{n}{2}-\alpha} \frac{\beta(\frac{n}{2})!}{\alpha!\beta!(\frac{n}{2}-\alpha-\beta)!}((1-r)(1-q_b^b))^\alpha((1-r)q_b^b)^\beta r^{\frac{n}{2}-\alpha-\beta} \tag{36}$$

Similarly, beliefs of C_b-partisans about the expected number of signals s_a (first line) and s_b (second line) revealed in group C_a in state $\theta = a$:

$$\mu_{C_b}(\#(s_a(C_a))|\theta = a) = \sum_{\alpha=0}^{\frac{n}{2}} \sum_{\beta=0}^{\frac{n}{2}-\alpha} \frac{\alpha(\frac{n}{2})!}{\alpha!\beta!(\frac{n}{2}-\alpha-\beta)!}((1-r)q_a^a)^\alpha((1-r)(1-q_a^a))^\beta r^{\frac{n}{2}-\alpha-\beta} \tag{37}$$

$$\mu_{C_b}(\#(s_b(C_a))|\theta = a) = \sum_{\alpha=0}^{\frac{n}{2}} \sum_{\beta=0}^{\frac{n}{2}-\alpha} \frac{\beta(\frac{n}{2})!}{\alpha!\beta!(\frac{n}{2}-\alpha-\beta)!}((1-r)q_a^a)^\alpha((1-r)(1-q_a^a))^\beta r^{\frac{n}{2}-\alpha-\beta} \tag{38}$$

and in state $\theta = b$:

$$\mu_{C_b}(\#(s_a(C_a))|\theta = b) = \sum_{\alpha=0}^{\frac{n}{2}} \sum_{\beta=0}^{\frac{n}{2}-\alpha} \frac{\alpha(\frac{n}{2})!}{\alpha!\beta!(\frac{n}{2}-\alpha-\beta)!}((1-r)(1-q_a^b))^\alpha((1-r)q_a^b)^\beta r^{\frac{n}{2}-\alpha-\beta} \tag{39}$$

$$\mu_{C_b}(\#(s_b(C_a))|\theta = b) = \sum_{\alpha=0}^{\frac{n}{2}} \sum_{\beta=0}^{\frac{n}{2}-\alpha} \frac{\beta(\frac{n}{2})!}{\alpha!\beta!(\frac{n}{2}-\alpha-\beta)!}((1-r)(1-q_a^b))^\alpha((1-r)q_a^b)^\beta r^{\frac{n}{2}-\alpha-\beta} \tag{40}$$

Due to group symmetry with respect to size, $n/2$, and signal accuracy in favorite and unfavorite states,

$$\begin{aligned} \mu_{C_a}(\#(s_b(C_b))|\theta = b) &= \mu_{C_b}(\#(s_a(C_a))|\theta = a) \qquad & \mu_{C_a}(\#(s_a(C_b))|\theta = b) &= \mu_{C_b}(\#(s_b(C_a))|\theta = a) \\ \mu_{C_a}(\#(s_a(C_b))|\theta = a) &= \mu_{C_b}(\#(s_b(C_a))|\theta = b) \qquad & \mu_{C_a}(\#(s_b(C_b))|\theta = a) &= \mu_{C_b}(\#(s_a(C_a))|\theta = b) \end{aligned}$$

Let $\#(s_a(C_j)) := \#_{-i}(s_a(C_j)) + \mathbb{1}_{\{s_i = s_a\}}$ and $\#(s_b(C_j)) := \#_{-i}(s_b(C_j)) - \mathbb{1}_{\{s_i = s_b\}}$. Given beliefs about signal distributions in the other party, μ_{C_j}, and the posterior $p_i(\theta = a|\#(s_a(C_j)), \#(s_b(C_j)))$, player $i \in C_j$ expects group C_{-j} to have a common posterior $p_{-j}^{C_j}(\theta = a)$. Namely, for $i \in C_a$,

$$\begin{aligned} p_b^{C_a}(\theta = a) = &\frac{p_i(\theta = a|\#(s_a(C_a)), \#(s_b(C_a)))}{1 + \left(\frac{q_b^b}{1-q_b^a}\right)^{\mu_{C_a}(\#(s_b(C_b))|\theta=a)} \left(\frac{1-q_b^b}{q_b^a}\right)^{\mu_{C_a}(\#(s_a(C_b))|\theta=a)}} \\ &+ \frac{1 - p_i(\theta = a|\#(s_a(C_a)), \#(s_b(C_a)))}{1 + \left(\frac{q_b^b}{1-q_b^a}\right)^{\mu_{C_a}(\#(s_b(C_b))|\theta=b)} \left(\frac{1-q_b^b}{q_b^a}\right)^{\mu_{C_a}(\#(s_a(C_b))|\theta=b)}} \\ &\approx 0.488 \cdot p_i(\theta = a|\#(s_a(C_a)), \#(s_b(C_a))) + 0.288 \end{aligned} \tag{41}$$

and for $i \in C_b$,

$$
\begin{aligned}
p_a^{C_b}(\theta = a) = & \frac{p_i(\theta = a|\#(s_a(C_a)), \#(s_b(C_a)))}{1 + \left(\frac{q_a^b}{1-q_a^a}\right)^{\mu_{C_b}(\#(s_b(C_a))|\theta=a)} \left(\frac{1-q_a^b}{q_a^a}\right)^{\mu_{C_b}(\#(s_a(C_a))|\theta=a)}} \\
& + \frac{1 - p_i(\theta = a|\#(s_a(C_a)), \#(s_b(C_a)))}{1 + \left(\frac{q_a^b}{1-q_a^a}\right)^{\mu_{C_b}(\#(s_b(C_a))|\theta=b)} \left(\frac{1-q_a^b}{q_a^a}\right)^{\mu_{C_b}(\#(s_a(C_a))|\theta=b)}} \\
& \approx 0.488 \cdot p_i(\theta = a|\#(s_a(C_a)), \#(s_b(C_a))) + 0.224
\end{aligned} \tag{42}
$$

where the last lines are obtained using our parameters ($n = 10, q_j^j = 0.9, q_j^{-j} = 0.4, r = 0.2$). Players then vote taking into account pivotality. ■

Appendix B Additional details

Table 11: Session summary

Session	Date/Time	Bias	Pr.(No Signal)	Network		# Subjects	Avg. payoff, £
				First 16 rounds	Last 16 rounds		
1	6/17/16, 11:00	Extreme	20%	Complete	Complete	20	11.00
2	6/17/16, 13:30	No	–	Complete	None	20	14.13
3	6/17/16, 15:30	No	–	None	Complete	30	14.42
4	6/17/16, 17:30	Extreme	–	Complete	Complete	20	11.93
5	6/24/16, 09:00	No	–	Polarized	None	30	13.17
6	6/24/16, 11:00	Extreme	–	Polarized	Complete	20	14.00
7	6/24/16, 13:30	No	–	Polarized	Complete	30	12.75
8	6/24/16, 15:30	Extreme	–	None	Polarized	20	12.13
9	6/24/16, 17:30	Extreme	–	Polarized	None	20	12.13
10	7/1/16, 09:00	No	–	None	Polarized	30	12.32
11	7/1/16, 11:00	Extreme	–	Complete	Polarized	30	11.88
12	7/1/16, 13:30	No	–	Complete	Polarized	30	13.58
13	7/1/16, 15:30	Extreme	–	None	Complete	30	11.50
14	10/18/16, 09:00	No	0%	Complete	Polarized	30	13.58
15	10/18/16, 11:00	Moderate	20%	Polarized	Complete	30	11.08
16	10/18/16, 13:30	Moderate	–	Complete	Polarized	30	11.50
17	10/18/16, 15:30	No	0%	Polarized	Complete	30	14.00

Notes: "None", "Complete", and "Polarized" refer to the network treatment for each half of each session. In each round we had two to three independent ten-person groups of subjects. "Extreme" refers to the media bias treatment with uninformative signals. "Moderate" refers to the media bias treatment with probabilities of receiving a favorable signal (conditional on getting any signal) being 0.9 in the favorable state and 0.6 in the unfavorable state. In sessions, denoted 1 and 4 we could not have varied the network treatment due to minor software issues with the polarized network display we discovered during tests shortly before the sessions were about to start, so to ensure an uninterrupted process we decided to keep the complete network treatment throughout those sessions.

Table 12: Summary of lab participant characteristics

Trait	N	Mean	Std. Dev.
Age, years	449	21.91	4.49
Male, %	450	39.56	48.95
English native, %	449	52.11	50.46
Previous experiments, #	449	6.74	9.12
Deception experiments, #	436	1.70	2.66
More quantitative, %	449	68.37	46.55
Media trust	449	4.80	1.80

Notes: We had 450 subjects total. One subject accidentally quit the software before finishing the post-treatment questionnaire, so for most characteristics $N = 449$. "Deception experiments" is the average number of experiments involving deception amongst those who participated in at least one such experiment. "More quantitative" refers to the question whether a subject considers themselves as a more or less quantitative person. "Media trust" is about how trustworthy they think the information from social media is, on a scale from 1 ("Not at all trustworthy") to 10 ("Completely trustworthy").

Table 13: Effects of Network on Efficiency Score

Bias	N	Polarized minus Empty		N	Complete minus Empty		N	Complete minus Polarized	
		Δ	p-val		Δ	p-val		Δ	p-val
No	368	.062**	(0.034)	352	.097***	(0.000)	368	.035*	(0.064)
Moderate							192	.125*	(0.052)

Notes: Δ indicates the difference in average bayesian efficiency scores between respective network configurations. Two-sided p-values in parentheses. Significance: *** < 0.01, ** < 0.05, * < 0.1

Table 14: Effects of Bias on Efficiency Score

Network	N	Extreme minus No		N	Moderate minus No		N	Moderate minus Extreme	
		Δ	p-val		Δ	p-val		Δ	p-val
Empty	288	.114***	(0.000)						
Polarized	336	.052***	(0.001)	288	$-$.281***	(0.000)	240	$-$.333***	(0.000)
Complete	432	.017*	(0.083)	272	$-$.191***	(0.000)	352	$-$.208***	(0.000)

Notes: Δ indicates the difference in average bayesian efficiency scores between respective media biases (note that efficiency is 1.00 under extreme bias). Two-sided p-values in parentheses. Significance: *** < 0.01, ** < 0.05, * < 0.1

Table 15: Summary characteristics of the Gallup Pakistan survey respondents

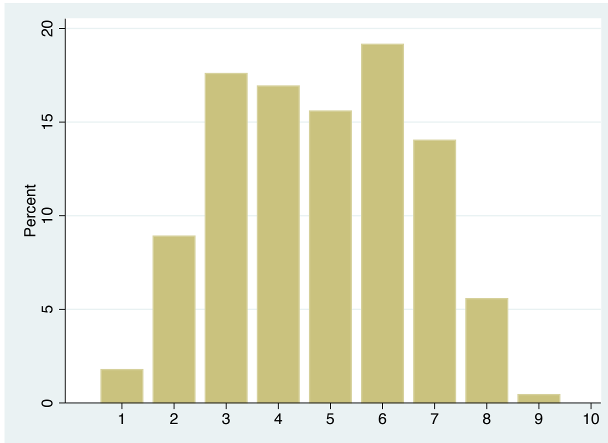
Variable	Label	Values	Percent
ur	Location type		
		Rural	69.99
		Urban	30.01
$d1$	Gender		
		Male	50.14
		Female	49.86
$d2\_1$	Age (years)		
		< 29	34.00
		$30 - 50$	53.13
		$51 - 65$	12.76
		> 65	0.11
$d3$	Education		
		Illiterate	11.37
		Literate but no formal	9.76
		Up to Primary	14.37
		Middle school	16.53
		Matric	19.69
		Intermediate	10.65
		Graduate	13.98
		Postgraduate	2.66
		Professional/Doctor	0.33
		No response	0.67
$d4\_1$	Monthly household income		
		$\leq 7,000$ Rs.	18.19
		$7,001 - 10,000$ Rs.	14.48
		$10,001 - 15,000$ Rs.	24.35
		$15,001 - 30,000$ Rs.	32.50
		$\geq 30,001$ Rs.	10.48
$d5$	Native tongue		
		Urdu	20.58
		Punjabi	52.19
		Sindhi	9.10
		Pashto	11.54
		Balochi	1.05
		Saraekee	3.44
		Others + No response	2.11
$d6$	Religion		
		Muslim	97.28
		Christian	1.72
		Hindu	0.33
		No response	0.67
zPR	Province		
		Punjab	61.12
		Sindh	25.07
		KPK	6.49
		Balochistan	7.32

Notes: Survey administered by Gallup Pakistan between 1/16/2017 and 1/20/2017 to a panel of size $N = 1,803$.

Table 16: Responses to questions about social media

Question	Options	Percent
How often do you get the news about politics and government from social media (e.g., Facebook, Twitter, and the like)?	All the time	11.76
	Often	17.69
	Sometimes	18.64
	Hardly ever	12.20
	Never	39.71
Consider a hypothetical situation in which you have come across a news article that says that your favorite political candidate has been accused of corruption. How likely are you to share this article on social media like Facebook and Twitter?	1=not at all likely	11.20
	2	6.93
	3	8.43
	4	9.04
	5	6.88
	6	6.60
	7	5.21
	8	4.71
	9	2.72
	10=extremely likely	3.33
	Don't use social media	34.94
Consider a hypothetical situation in which you have come across a news article that says that your favorite political candidate has been a major force behind building a new hospital. How likely are you to share this article on social media like Facebook and Twitter?	1=not at all likely	9.43
	2	3.44
	3	5.60
	4	8.37
	5	8.82
	6	9.60
	7	6.66
	8	6.16
	9	2.77
	10=extremely likely	3.83
	Don't use social media	35.33
Consider a hypothetical situation in which your Facebook friend has shared a news article that says that your favorite political candidate has been accused of corruption. How likely are you to revise your opinion about the candidate?	1=not at all likely	8.60
	2	2.66
	3	3.99
	4	7.65
	5	11.98
	6	10.87
	7	6.71
	8	4.88
	9	3.94
	10=extremely likely	3.11
	Don't use social media	35.61
In your opinion, how trustworthy is the information that you get from social media?	1=not at all	19.25
	2	8.15
	3	6.05
	4	9.10
	5	12.76
	6	13.37
	7	11.92
	8	7.76
	9	4.38
	10=extremely trustworthy	6.82
	No response	0.44

Notes: Survey administered by Gallup Pakistan between 1/16/2017 and 1/20/2017 to a panel of size $N = 1,803$.



(a) How trustworthy is information from social media

Figure 5: Lab subject responses ($N = 449$). Scale: 1=not at all trustworthy, 10=extremely trustworthy.

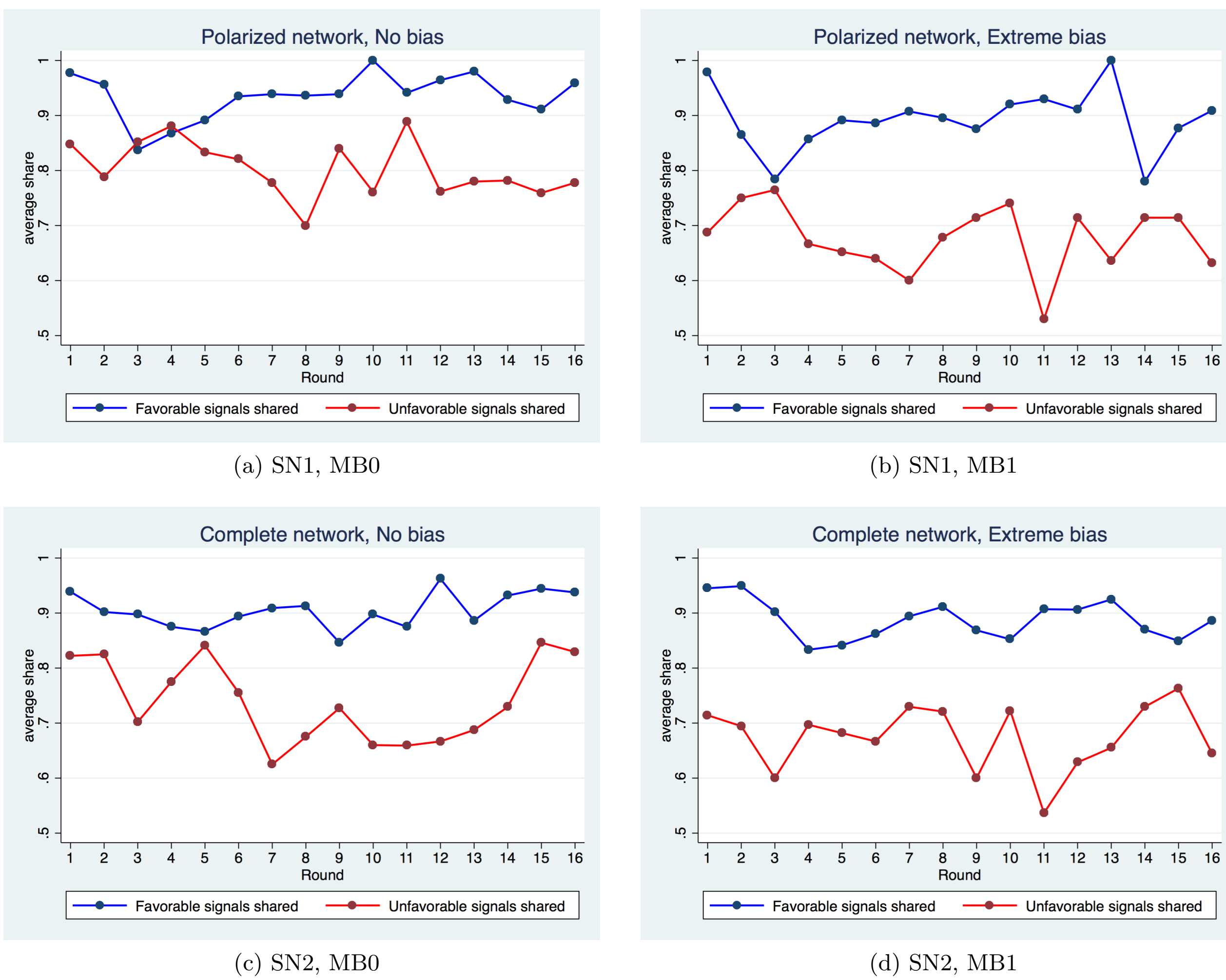
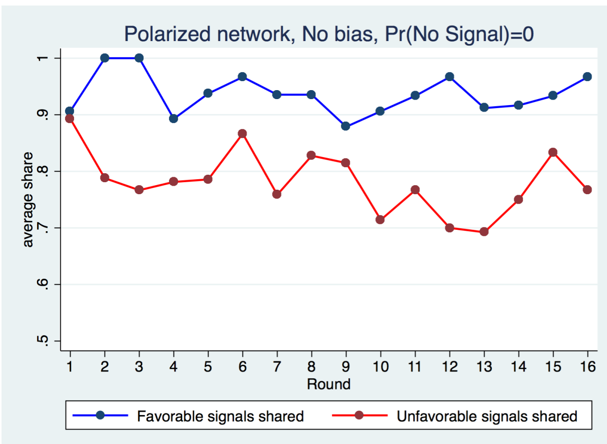
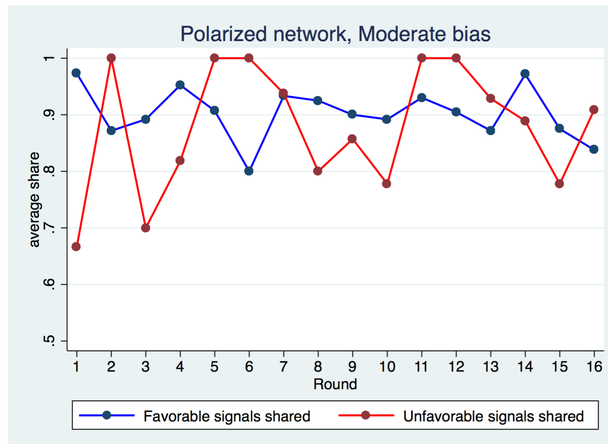


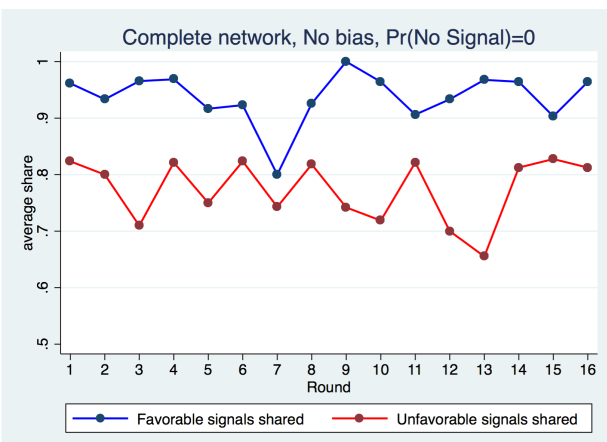
Figure 6: Average signal sharing rates across rounds: No bias and Extreme bias



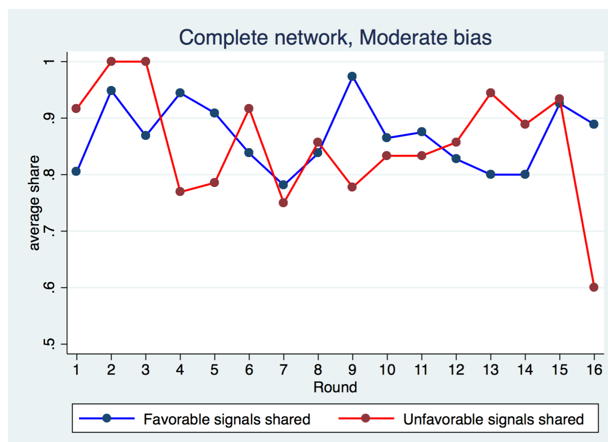
(a) SN1, MB0, r=0



(b) SN1, MB2



(c) SN2, MB0, r=0



(d) SN2, MB2

Figure 7: Average signal sharing rates across rounds: No bias with $r = 0$ and Moderate bias

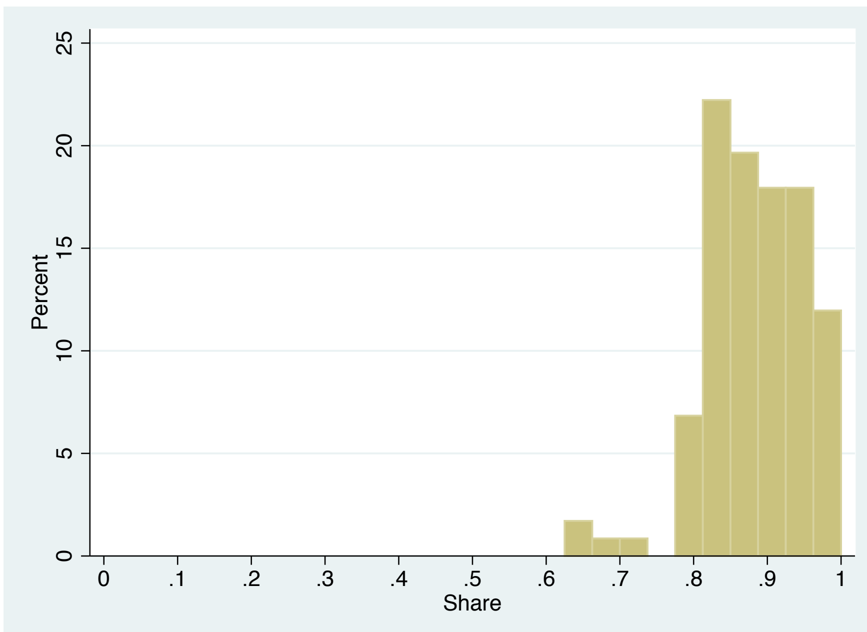


Figure 8: Distribution of average individual vote shares consistent with sincere Bayesian strategy, based on those subjects who had more than 60% correct decisions for votes against own favorite candidate, all treatments except MB1, 117 subjects.

Appendix C Instructions (MB0 or MB2, sequence SN1-SN2)

Welcome. This is an experiment in decision making, and for your participation you will be paid in cash at the end of the experiment. Different participants may earn different amounts. What you earn depends partly on your decisions, partly on decisions of others, and partly on chance.

The entire experiment will take place through computers. It is important that you do not talk with other participants and only use the computer software as instructed. We kindly ask you to turn off your phones and other mobile devices. If you violate the rules, we may ask you to leave the experiment.

This experiment will take between 50 and 80 minutes. If for any reason you are unable to stay for the entire duration of the experiment, please tell us now. In this experiment all interactions between participants are via the computers. You will interact anonymously and your decisions will be recorded together with your randomly assigned subject ID number. Neither your name, nor names of other participants will be ever made public and will not be used for any research purpose, only for payments records.

We will now start with a brief instruction period. If you have any questions during this time, raise your hand and your question will be answered so everyone can hear. If any difficulties arise after the experiment has begun, raise your hand, and an experimenter will come and assist you.

The experiment today has two parts. Each part consists of several procedurally identical rounds. At the end of the experiment, four rounds total (two from each part) will be randomly selected as paid rounds. The rounds that are not selected will not be paid out. Your total payoff consists of the points you will earn in the selected rounds, converted to pounds, plus your show-up fee of £8. The conversion rate of points to pounds is as follows: 100 points make £1.5. Every participant will be paid out in private so that no other participant can see how much you have earned. There will be a practice round followed by two parts, with 16 rounds in each part. Let me now show you what will occur every round.

[Turn on the projector]

Once the experiment begins, you will be randomly assigned a type: Green or Blue. Each type is equally likely. You will keep this type for all rounds. Every round, you will be randomly placed into one of three 10-person groups, with 5 Green types and 5 Blue types in each group. All three groups are totally independent and all interactions and payoffs in one group do not affect interactions and payoffs in any of the other groups. Hence from now on I will focus what?s going on within one 10-person group only. This is the screen you will see at the beginning of every round in part 1.

[show the initial Green screen]

At the top of the screen there is information about the current round and your type. This screen is for a Green type, and the screen for a Blue type looks very similar, as I'll show in a moment.

[show the initial Blue screen]

Let me start by explaining the left-hand side of the screen. There are two possible states represented by two wheels [MB2: at the top]: a Blue wheel, which has a larger area covered with Blue, and a Green wheel, which has a larger area covered with Green.

In each round the computer randomly selects one of these wheels with equal chances. We will call the selected wheel the "chosen" wheel. The identity of the chosen wheel is not revealed to anyone. However, each participant has a chance at getting a hint (that we call a signal) about the chosen wheel.

[MB0 treatment]

Here is how signals work: Once you indicate that you are ready to receive your signal by clicking on the OK button, the computer will cover the chosen wheel so that you won't be able to see its colors, and let you spin it. Here is how this will look like.

[show the screen with spin button]

Once you click the Spin button, the covered chosen wheel will be spun randomly, and after the rotation stops, there will be one of the two possible outcomes: Your attempt to receive a signal will either be successful or not. If it is successful, you will see a colored strip (Green or Blue) through a slit in the cover of the chosen wheel.

[show the screen with successful finished spin]

Notice that as a result of a random spin, the Blue color is more likely to come up on the Blue wheel, and the Green color is more likely to come up on the Green wheel, however there is also a possibility that the Blue color comes up on the Green wheel and that the Green color comes up on the Blue wheel. If your attempt to receive a signal is not successful, you will see a text saying "No signal".

[show the no-signal screen]

There is a 20% chance that you will get a "No signal" outcome, regardless of which wheel was chosen.

[MB2 treatment]

Here is how signals work: Once you indicate that you are ready to receive your signal by clicking on the OK button, the computer will take the wheel displayed directly below the chosen wheel (this wheel is called Bluish if you are a Blue type, and Greenish if you are a Green type), cover it so that you won't be able to see its colors, and let you spin this covered wheel. Here is how this will look like.

[show the screen with spin button]

Once you click the Spin button, the covered wheel (Bluish or Greenish, depending on your type) will be spun randomly, and after the rotation stops, there will be one of the two possible outcomes: Your attempt to receive a signal will either be successful or not. If it is successful, you will see a colored strip (Green or Blue) through a slit in the cover of the covered wheel.

[show the screen with finished spin]

Notice that as a result of a random spin, the Blue color is more likely to come up on the Bluish wheel, and the Green color is more likely to come up on the Greenish wheel, however there is also a possibility that the Blue color comes up on the Greenish wheel and that the Green color comes up on the Bluish wheel, and these likelihoods differ depending on which wheel has been chosen. If your attempt to receive a signal is not successful, you will see a text saying "No signal". There is a 20% chance that you will get a "No signal", regardless of which wheel is chosen.

[Continue for both MB treatments + SN]

Now that you have received a signal (or no signal), let's look at the right hand side of the screen. At the top of the screen there is a graph showing positions and connections of all participants in your group, including yourself. If you did not receive a signal you are not required to do anything at this stage, just have to wait for others. If you did receive a signal, you now have a choice of whether or not to send your signal to all those participants connected to you in the network.

[For SN1 (polarized)]

Notice that all Blue types are connected to one another and all Green types are connected to one another, so if you are a Blue type you can send your signal to four other Blue types, and if you are a Green type you can send your signal to four other Green types. Correspondingly you cannot receive signals from subjects of the other type, and their rows in the table are grayed out.

[For SN2 (complete)] Notice that everyone is connected to everyone else, so if you are a Blue type you can send your signal to four other Blue types and five Green types, and if you are a Green type you can send your signal to four other Green types and five Blue types. Correspondingly you can receive signals from subjects of either type.

[Continue for both SN1+SN2]

Once everyone has decided about sending or not sending their signals, the table on the right will be updated showing the signals sent.

[show screen with updated table]

[Continue for all, including SN0]

Now you will have to make a guess about the chosen wheel by clicking on the respective button – Green or Blue. Once all individual guesses have been made, your group guess will be determined as follows: if more subjects submitted a guess of Green than a guess of Blue, the group guess will be Green. If more subjects submitted a guess of Blue than a guess of Green, the group guess will be Blue. If there is an equal number of guesses for Blue and Green, the group guess about the chosen wheel will be decided by a coin flip, and with equal chances will be either Blue or Green. Your potential payoff from the round, if it is selected as a paid round, will depend on three things: (i) your type, (ii) the identity of the chosen wheel, (iii) and the group guess.

If you are a Blue type:

- if the chosen wheel is Blue, and the group guess is Blue, you get 150 points.

  [show screen with blue round outcome]

- If the chosen wheel is Green and the group guess is Green, you get 50 points.

  [show screen with green round outcome]

- In all other cases (if the group guess does not match the chosen wheel color), you get 15 points.

  [show screen with mismatched outcome]

If you are a Green type:

- If the chosen wheel is Green and the group guess is Green, you get 150 points.

  [show screen with green round outcome]

- if the chosen wheel is Blue, and the group guess is Blue, you get 50 points.

  [show screen with blue round outcome]

- In all other cases (if the group guess does not match the chosen wheel color), you get 15 points.

  [show screen with mismatched outcome]

In other words, you get 15 points if the group guess does not match the chosen wheel color, 50 points if the group guess is correct and the chosen wheel color is not the same as your type, and 150 points if the group guess is correct and the chosen wheel color is the same as your type.

This process will be repeated for 16 rounds. In every round you will be randomly assigned to one of the three groups, the computer will randomly choose a new wheel in each group, and so on. Then will move to part 2, and I will explain the details once part 1 is over.

After part 2, at the end of the experiment you will be also asked to answer a short questionnaire which will not affect your payoff. Remember that two rounds from each part (four total) will be randomly selected as paid rounds, and every round in either part, including the very first, and the very last, has the same chance of being selected as a paid round. Are there any questions about the procedure?

[wait for response]

We will now start with one practice round. The practice round will not be paid. Is everyone ready?

[wait for response, start multi-stage server]

Now follow my instructions very carefully and do not do anything until I ask you to do so. Please locate the 'Client-Multistage' icon on your desktop, and double click on it. The program will ask you to type in your name. Don't do this. Instead, please type in the number of your computer station.

[wait for subjects to connect to server]

We will now start the practice round. Do not hit any keys or click the mouse button until you are told to do so. Please pull up the dividers between your cubicles.

[start first practice match]

This is the end of the practice round. Are there any questions?

[wait for response]

Now let's start part 1 of the actual experiment. If there are any problems from this point on, raise your hand and an experimenter will come and assist you.

[start part 1, turn off slides]

This was the last round of part 1. Before we move on to part 2, let me explain the differences from the previous part.

[SN1-SN2]

Once you are ready to receive the signal

[show the spin screen]

and click on the spin button

[show the finished spin screen]

Notice that now everyone is connected to everyone else, so if you are a Blue type you can send your signal to four other Blue types and five Green types, and if you are a Green type you can send your signal to four other Green types and five Blue types. Correspondingly you can receive signals from subjects of either type.

[are there any questions?]

[start part 2]

This was the last round. A small window showing your payoff in pounds has popped up. Your total payoff will be the amount shown under "Total payoff", rounded up to the nearest 50p plus the £8 show-up fee. So it will be the amount you see on the screen plus £8.

All payoffs are recorded in the system so you don't need to write down anything. Now I need you all to click on the OK button in the small window so that you can proceed to filling out a short questionnaire while I am preparing your payments. After you finish the questionnaire, please remain seated and do not close any windows. You will be paid in a booth at the front row one at a time, you will be called in by your station number. Please bring all your things with you when you go to the booth as you will then leave the experiment. Please refrain from discussing this experiment while you are waiting to receive payment so that privacy regarding individual choices and payoffs may be maintained. Thanks very much for your participation.