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

Kirill Pogorelskiy
University of Warwick†

Matthew Shum
Caltech‡

February 21, 2019

Abstract

We study the relationship between news sharing on social networks 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 getting political news via social media may become more polarized in their voting behavior. Our results suggest that these concerns are warranted: subjects share news that is favorable to their party more often than the unfavorable news and take biased news at face value in their voting decisions, ignoring news sources. At the same time, the welfare implications of social media are driven by the quality of the shared news: with unbiased media, news sharing on social networks raises collective decision making efficiency, but efficiency deteriorates markedly in the presence of media bias despite the theoretical possibility of a moderate bias enabling more informative voting. Poor quality information, including uninformative ("fake") news, lowers efficiency more than do filter bubbles enabled by social media.

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

Keywords: news sharing, social networks, voting, media bias, fake news, polarization, filter bubble, lab experiments

* *First draft: May 22, 2017. This version: February 21, 2019.* We thank Marina Agranov, Simon Anderson, Dan Bernhardt, Micael Castanheira, Matt Chao, Khai X. Chiong, Ruben Enikolopov, Tim Feddersen, Guillaume Frechette, Jens Grosser, Marlene Guraieb, Ernan Haruvy, Asen Ivanov, Max Kwiek, Dimitri Landa, Aniol Llorente-Saguer, Matt Mitchell, Eugenia Nazrullaeva, Santiago Oliveros, Pietro Ortoleva, Tom Palfrey, Wolfgang Pesendorfer, Maria Petrova, Charlie Plott, Anja Prummer, Ronny Razin, Pedro Robalo, Alejandro Robinson, Daniel Sgroi, Jesse Shapiro, Erik Snowberg, Andis Sofianos, Francesco Squintani, Leeat Yariv, and audiences at QMUL, WBS, Political Economy Webinar, FSU, NYU, Oxford CESS (Nuffield), U Konstanz, Aix Marseille, Stanford GSB, GMU (ICES), USC, U York, 2016 North American ESA Meetings, U Zhejiang IO Conference, 15th Annual Media Economics Workshop in Barcelona, 2017 Lancaster GTC, 2018 Thurgau Experimental Economics Meeting, JHU "Influence and the Media" conference, BFI Chicago "Economics of Media and Communication Conference", Erasmus U Rotterdam Political Economy workshop, NICEP conference, Essex Experimental Economics workshop, and the Vienna Behavioral Public Economics workshop for useful discussions. We thank Chris Crabbe, Didi Egerton-Warburton, Mahnaz Nazneen, and Andis Sofianos for expert technical assistance. We thank Warwick Economics for financial support, and Umar Taj and Gallup Pakistan for administering our questions in their survey.

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

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

1 Introduction

Twitter, Facebook and other social media have become a primary source of news and a forum for political discourse for an increasing share of voters.[1] This may have dramatic consequences as social media platforms, faced with abundant shared information, filter content so that users only see news and advertising related to what they 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, inside their personal "filter bubble", reinforcing a limited 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).

Does 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, savvy towards social media, who overwhelmingly supported Remain, might have become convinced that the majority supported Remain and therefore turned out in smaller numbers than older voters, unfamiliar with social media, who were more likely to vote Leave. Similarly, in November 2016, an unexpected victory by Donald Trump in the U.S. Presidential election led to discussion of a "hidden majority" (beyond the awareness of the mainstream media) convinced by "fake news" who voted Trump into office.[2]

To date there has been little research formally analyzing and quantifying the impact of news sharing via social networks on collective decisions in the presence of media bias. This paper seeks to fill this gap. We construct an artificial, context-neutral social media environment in the laboratory, where we control the most important and subtle factors of the process: voters' preferences, social networks, and media bias. In this setting, we examine how the combination of media bias and the limited communication possibilities in social networks may distort news sharing and influence electoral outcomes.

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 welfare-maximizing choice depending on the realized state. There are two groups of voters who are "weak" partisans, i.e. voters in each group support a different partisan candidate ex ante but still prefer the elected candidate to be the "correct" choice for the state. Voters receive private signals about the state ("news") and vote for one of the two candidates, and the election is decided by a simple majority of votes. We extend this largely standard setting by introducing bias in voters' private signals and a novel social media component that restricts signal sharing across groups. Specifically, prior to voting, voters can share their private signals with their neighbors on the social network, mimicking the

[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).

[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.

news sharing protocol on social media platforms. If a voter decides not to share her signal, other voters in her network cannot fully infer the signal realization, since there is a genuine possibility that the voter did not receive any signal and hence could not have shared it.

In the experiment, we vary the treatment on two dimensions: *social networks*, which determine the size and type of news sharing audience, and *media bias*, which determines the accuracy of news items. 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 with no links in between) relative to the control of the empty network. We also vary the degree of media bias across treatments, from unbiased informative signals, to moderately biased informative signals where signals that are favorable to partisans are more likely, to extremely biased uninformative signals (a variant of "fake news").

One of our key findings is that, across social network and media bias treatments, subjects consistently engaged in "selective news sharing", whereby they share news that is favorable to their partisan candidate (in the sense of indicating that the candidate is more likely to be the correct choice for the state) more often than news that is unfavorable to their candidate. So we clearly reject a model of communication in which voters naïvely reveal all information that comes their way: they are at least somewhat sophisticated. But how sophisticated? We find that selective news sharing persists even in treatments where the probability of not receiving a signal is zero, which prevents the subjects from hiding unfavorable signals under the disguise of "no news". 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 interest 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.[3]

We also find that the effects of news sharing on voting depend crucially on the underlying quality of the news: Without media bias, social networks raise collective decision making efficiency. This positive effect is predicted by the theory – without media bias, the news shared by sufficiently many voters is sufficiently accurate about the realized state of the world. As news quality deteriorates under media bias, however, efficiency of the complete social network decreases. Theoretically, media bias need not lead to less efficient electoral outcomes, and may actually improve efficiency, if voters correctly account for the sources of the signals in their posteriors.[4] However, our results strongly refute this possibility: voters by and large take signals at face value and fail to discount signal sources. This happens even in our "fake news" treatments where all signals are uninformative by design: we find that even in this case, the majority of subjects react to their own signals as well as others' reported signals, treating uninformative signals as real news. This dimension of the failure of Bayesian updating differs from findings in previous studies, and suggests that bad information (e.g."fake" news) is a bigger threat to information aggregation by voting than are filter bubbles

[3] Selective news sharing has also been documented in the field, e.g. An et al. (2014), Garz et al. (2018).

[4] For instance, an endorsement of a left-wing candidate by a right-wing newspaper is a very informative signal.

enabled by social media. With good quality news, news sharing via social networks clearly aids information aggregation.

A key contribution of this study is to consider the *joint* effects of social networks and media bias, two issues that have generally been treated separately in the existing literature. Media bias at the source can directly affect electoral outcomes. This is the gist of the "Fox news effect" documented in DellaVigna and Kaplan (2007), whereby television viewers who were exposed to the right-leaning Fox News stations tended to vote more for Republican candidates.[5] Against such a backdrop, social media may amplify the Fox News effect if a biased news piece becomes more prominent as it is shared by a larger number of like-minded people.[6] Our results lend support to this interpretation of news sharing on social media, as our subjects by and large fail to take into account the sources of their signals in their voting behavior. At the same time, social networks have been shown to influence their users' exposure to ideologically misaligned news and opinion.[7] On Facebook, news is largely shared by "liking"; such semantics may accentuate the tendency of users to share only news that is congruent with their preferences, consistent with our finding of selective signal sharing in our experiments. However, our results also show clearly that the negative effects of social media only arise in conjunction with media bias: with unbiased news, news sharing on social networks actually leads to *better* electoral outcomes, by increasing the information that voters have at their disposal.

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 main results and their interpretation. As a robustness check, we supplement our lab experimental results with survey evidence from Pakistan in Section 5. The survey results are broadly consistent with our experimental findings. Section 6 concludes. Appendix A contains proofs of the theoretical propositions. Appendix B presents additional detail of the experimental setup and data analyses. 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

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

[6]In our experiments, media bias is exogenously given. For studies where media bias arises endogenously see, among others, Gentzkow and Shapiro (2006), Mullainathan and Shleifer (2005), Bernhardt et al. (2008), Duggan and Martinelli (2011), Anderson and McLaren (2012), and Piolatto and Schuett (2015). Measuring media bias using observational data is challenging; see e.g. Chiang and Knight (2011), Barberá et al. (2015), Groeling (2013), and Prat (2018), Kennedy and Prat (2017).

[7]One well-known study is Bakshy et al. (2015); see however evidence to the contrary in Gentzkow and Shapiro (2011) and Boxell et al. (2017).

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. Communication among voters, described below, is network-restricted according to exogenous social network configurations, which we describe below.[8]

To focus on information aggregation, we assume a voting environment with common interests, 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 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 for the state, translating into a lower expected payoff of C_b-partisans, and vice versa for an s_b signal. There are two different media sources that distribute the signals, and each group of supporters listens to a different source. We do not explicitly model media as strategic players but we vary the informativeness of the signals as discussed in detail below.

Timing and actions. We consider a game with 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 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.

[8] Thus, our setting features network-restricted *deliberation* among voters. Relevant theoretical work includes Coughlan (2000), Austen-Smith and Feddersen (2005, 2006), Gerardi and Yariv (2007), Dickson et al. (2008), Meirowitz (2007), and Schulte (2010). Battaglini (2017) considers public protests and finds that with precise enough signals, social media improves information aggregation. Empirical papers on deliberation and voting include Gole and Quinn (2014) and Iaryczower et al. (2018).

4. Voters vote for one of the candidates C_a or C_b.[9]

5. The candidate supported by the majority of votes is elected, and payoffs are realized. Ties are resolved by a random coin flip.

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] We study the following network treatments in our experiments:

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 signals, all players go to the voting stage.[11]

SN1: Polarized network. The social network consists of two separate and fully-connected components (which we will also refer to as "parties" below), 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 opposite ends of the spectrum, and communication is predominantly between like-minded people.[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.[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 verifiable and should be always believed.

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

[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. Martinelli and Palfrey (2017) survey experimental results in collective decision games.

[12] Several papers also studied heterogeneous voters (Maug and Yilmaz (2002), Mengel and Rivas (2017)) but did not allow for communication. Related theoretical studies with elements of communication between voters include Galeotti and Mattozzi (2011) and Le Quement and Yokeeswaran (2015).

[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), Bouton et al. (2017), and Breitmoser and Valasek (2018). Unlike those papers, we use verifiable signals rather than versions of cheap talk. Kawamura and Vlaseros (2017) studied information aggregation in the presence of public "expert" signals; in our setting public signals are generated endogenously via voters revealing their private signals.

2.2 Media bias treatments

Given the popular and policy debates on the bias and reliability of news circulated in real-world social networks, we study media bias treatments that vary the quality of the private news signals that subjects in our experiments receive:

MB0: No media bias. Signals are unbiased and informative.

MB1: Moderate partisan bias. Each voter receives the signal from the media source biased towards her ex ante preferred candidate. The interpretation is that biased media highlight good news for their candidate, and dampen good news for the opponent.[14] While "favorable" signals are more likely to be reported in both states than "unfavorable" signals, there is a difference in probabilities that depends on the realized state. Thus, signals are informative, and more so if they are "unfavorable".

MB2: Extreme partisan bias ("fake news"). As in MB1, favorable signals are more likely to be reported in both states. However the probability of getting a favorable signal is the *same* in both states – the signals are uninformative (not correlated with the realized state). This signal structure partially resembles "fake" news.

In all MB treatments, there is no news (empty signal) with probability r, with $r = 0.2$ for most of our treatments. The positive probability of an empty signal is crucial – if each voter receives a signal with probability one, then her decision not to reveal a signal will be fully "recovered" by other voters as an attempt to suppress information unfavorable to her candidate. In order to see whether voters take these strategic considerations into account, we experimentally check this case below, by setting $r = 0$ for two sessions.

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

Table 1: Signal Structure

	No bias, MB0 possible signals			**Moderate bias, MB1** possible signals			**Extreme bias, MB2** possible signals		
State	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	$q_j^j(1-r)$	$(1-q_j^j)(1-r)$	r	$q(1-r)$	$(1-q)(1-r)$	r
	0.56	0.24	0.2	0.72	0.08	0.2	0.56	0.24	0.2
$-j$	$(1-q)(1-r)$	$q(1-r)$	r	$(1-q_j^{-j})(1-r)$	$q_j^{-j}(1-r)$	r	$q(1-r)$	$(1-q)(1-r)$	r
	0.24	0.56	0.2	0.48	0.32	0.2	0.56	0.24	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, $j, -j \equiv \{a, b\} \setminus j$, for each MB treatment. First line: formal expressions using our notation. Second line: actual values computed using experimental parameters $q = 0.7, r = 0.2, q_j^j = 0.9, q_j^{-j} = 0.4$.

[14]We do not consider a possibility of negative campaigning *in addition* to the good news bias, but it is straightforward to re-arrange the signals to get a model of negative campaigning *instead* of the positive one. See also Morton et al. (2018) who investigate effects of biased signals on voting efficiency using a different experimental design.

2.3 Experimental procedures

Our experiment combines three social network treatments and three media bias treatments, and 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, and further 6 sessions in May 2018. Sessions lasted between 50 and 80 minutes. In total, 590 subjects participated.[15]

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. We paid two random rounds from each network treatment, with GBP payoffs for each player specified in Eq (1). 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, subjects' member IDs and the composition of the party groups changed randomly every round.[16] More details of the experimental setup, including a screenshot of subject computer screen and summary statistics for the sessions and subject payoffs, are contained in Appendix B.

3 Equilibrium Predictions

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 focus on deriving equilibrium predictions specific to the parameters of the experiment, which are taken as a primitive by all formal statements.[17]

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 piv-

[15] All our analyses except the subject summary characteristics in Table 9 use the data from 530 subjects. We did not use the data from the 60 subjects who had participated in the two 2016 sessions that we later had to discard, as described in Table 8 in Appendix B.

[16] As specified in the experimental instructions in Appendix C, we intended to keep subject "party" affiliation fixed for all rounds, to make subjects more attached to their party identity. However, a software parameter issue discovered during the 2018 sessions resulted in subjects' party being randomly re-assigned every round. Since in every round and every interface screen each subject's current party identity was always correctly and clearly indicated, we do not consider this deviation from our original design an issue for our analyses.

[17] The baseline case of SN0+MB0 has been studied in many papers on jury voting, starting with 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; see Austen-Smith and Feddersen (2005) and Oliveros and Várdy (2015). A version of our SN2+MB0 treatment was analyzed theoretically in Schulte (2010) and Coughlan (2000). These papers, as well as 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 MB1 treatments appear to be new in the literature.

otality.[18] Then i prefers to have C_a elected if and only if this maximizes her expected utility:

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

For the utility specification in (1), inequality (2) holds for C_a-partisans if and only if $p_i > t_a$ and for C_b-partisans if and only if $p_i > t_b$, where

$$t_a = 7/34 \approx 0.206, \qquad t_b = 1 - 7/34 \approx 0.794. \quad (3)$$

A *sincere Bayesian voting strategy* in which players vote for the candidate who maximizes their expected utility, is as follows: C_a-partisans vote for C_a whenever $p_i > t_a$, C_b-partisans vote for C_a whenever $p_i > t_b$. This voting strategy is always a best response at the voting stage and, moreover, a unique best response if players are pivotal with positive probability.[19] With these preliminaries at hand, we 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.

Before going over predictions for specific treatments in detail, we showcase an important trade-off between the network structure and information aggregation by voting which emerges from the equilibrium analysis. Fixing the total electorate size, a complete network enables exposure to a larger number of signals relative to the polarized network: this can improve everyone's welfare. However, full information revelation fails under a complete network because when the audience for messages has heterogeneous preferences, partisans have incentives to hide unfavorable signals – this is not the case in the homogeneous polarized network. Thus in theory both social networks, by permitting news sharing, should result in higher welfare than the empty network, but *having all voters grouped in a polarized network by party affiliation can be welfare-improving relative to a complete network*. Moreover, this may be more pronounced under moderate media bias as unfavorable signals in those treatments are more informative.

3.1 Equilibria without informative communication

We begin with the empty network [SN0] treatments (third column in Table 2), where no communication is permitted among voters. Each voter can only condition on her own signal and the equilibrium play to deduce the optimal voting strategy.

With unbiased media [MB0], signals are equally informative about either state. For each C_j-partisan, $j \in \{a, b\}$, conditional probabilities of private signals are specified in Table 1. Conditional on observing signals s_a and s_b, respectively, with our parameters and the common prior of 1/2, player i's posterior is $p_i(\theta = a|s_a) = 1/(1+(1-q)/q) = 0.7$ and $p_i(\theta = a|s_b) = 1/(1+q/(1-q)) = 0.3$. If all players used sincere Bayesian voting strategies, then a unique best response for a C_a-partisan

[18] 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.

[19] 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.

Table 2: Brief Summary of Equilibrium Predictions for each Treatment

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

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. 'Informative voting' indicates whether voters respond to their information in equilibrium. No informative voting means that each voter votes for her partisan candidate, regardless of their information. Partial informative voting means that voters vote according to their information under some combinations of news signals shared with them.

would be to vote for C_a after either signal, as $p_i(\theta = a|\cdot) \geq 0.3 > t_a \approx 0.206$, and for a C_b-partisan to vote for C_b after either signal, as $p_i(\theta = a|\cdot) \leq 0.7 < t_b \approx 0.794$. Hence we obtain an equilibrium in which voters ignore their signals, so voting is uninformative and does not aggregate information.[20]

Proposition 1. ***[Unbiased media and empty social network]*** *Under MB0+SN0, voting is not informative: In the unique sincere voting equilibrium, all players vote for their partisan candidate, regardless of their signal, and the election results in a tie.*

Under moderately biased media [MB1], signals are weakly informative (as specified in Table 1), allowing partisans to update from the prior of 1/2. This leads to

Proposition 2. ***[Moderate media bias and empty network]*** *Under MB1+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 helpful for information aggregation in the case of no communication* – sincere voting becomes informative, unlike the no bias case. This happens because under the 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 her partisan candidate is the wrong choice for the state.[21]

[20] Full equilibrium derivation requires conditioning on pivotality in addition to own signal, but this does not change the result: there is no equilibrium in which everyone with a non-empty signal votes according to their signal (Lemma 1 in Appendix A). 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.

[21] In this way, moderate bias may play a similarly beneficial role for information aggregation as correlation neglect (e.g. Levy and Razin (2015)). This also resembles the classical result in Calvert (1985) for a single decision maker faced with the choice between a neutral and a biased expert (which however arises from a different mechanism).

In treatments with extreme media bias [MB2] (the bottom row in Table 2), signals are completely uninformative: conditional on getting any signal, voters are more likely to receive signals supporting their partisan candidates, *regardless of the realized state*. Hence posterior beliefs never move away from the prior ($= 1/2$), which rationalizes any signalling strategy. However, the unique sincere Bayesian equilibrium requires everyone to vote (non-informatively) for their partisan candidate, as in Proposition 1.

3.2 Equilibria with informative communication via social networks

In social networks, voters have opportunities to exchange signals with others during the communication stage prior to voting. In the complete network [SN2] treatments (the last column in Table 2), voters can share their signals with *all* the other voters or with no one. Since both states are equally likely and revealed signals are "hard" evidence, the posterior belief of the state depends on the difference between the number of reported signals favoring each candidate, and, in the moderate bias case, on the sources of the signals. Non-reported signals also affects beliefs: if some voter 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.

In a complete network, since signals are shared with everyone including the supporters of the other candidate, revealing an unfavorable signal may push these supporters to vote against one's preferred partisan candidate. Thus there is no equilibrium in which all voters always share their non-empty signals during the communication stage and use sincere Bayesian voting strategies at the voting stage.[22] Hence we characterize all *semi-pooling equilibria*, in which all players fully reveal favorable signals and, with some positive probability, hide unfavorable signals pretending to be uninformed. In this case, players must form consistent (and sufficiently pessimistic) beliefs about the state given the reported empty signals, to the extent that those won't affect their posterior much. At the voting stage, all players vote according to sincere Bayesian strategies based on their equilibrium posteriors.[23]

Proposition 3. ***[Unbiased media and complete network]*** *Under MB0+SN2, there is no full information revelation equilibrium. 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^*. Each such equilibrium is characterized by fixing any $\nu^* \in (.922, 1]$.*

In the polarized network [SN1] treatments (the fourth column in Table 2), voters can only share signals with members of their own party. In this case, they cannot directly affect beliefs in the other group – their signals can only be pivotal for the posteriors held in their own group – so that it becomes incentive compatible for voters to reveal even unfavorable signals.

[22] See Lemmata 2 and 3 in Appendices A.1 and A.2. Similar logic arises in two-member committees with conflicts in members' preferences (Li et al., 2001).

[23] We provide the details of posterior calculations in the proofs of respective propositions in Appendix A.

Proposition 4. ***[Unbiased media and polarized network]*** *Under MB0+SN1, there is a full information revelation equilibrium, in which all voters with non-empty signals reveal them at the communication stage and believe with probability one that non-revealing voters are uninformed.*

Finally, for moderately biased media [MB1] treatments, Appendix A contains Propositions 5 and 6 that are qualitatively similar to Propositions 3 and 4 above.

4 Experimental Findings

In this section we discuss the main results from our experiments. In interpreting the results we focus on key theoretical predictions from the previous section, that (i) polarized networks may be more efficient than complete networks; (ii) voters may share news selectively, passing along favorable signals and suppressing unfavorable ones. In addition, (iii) we also want to test whether voters behave in line with sincere Bayesian voting strategies, conditional on all available information and the equilibrium behavior. We examine these issues in turn.

4.1 Collective decision-making efficiency

The most natural measure of efficiency in our informationally scarce environment is a *Bayesian efficiency* benchmark adapted to partisan subgroups.[24] Each round, in each subgroup we check if the majority decision within subgroup agrees with the choice a benevolent Bayesian social planner would have made for this subgroup, had she observed voters' private signals realized in that round in both subgroups (for the empty and complete network treatments) or in only one partisan subgroup (for the polarized network treatments).[25] This binary score is averaged across all subgroups and rounds for each treatment, and reported in Table 3. In the discussion, we focus on the more conservative statistics reported in the "Last 12" columns of Table 3, which exclude the first four rounds (to look at more experienced decisions).

When signals are informative, as in the "No media bias" and "Moderate media bias" treatments in the top two panels of Table 3, social networks improve efficiency, as they provide the opportunities for sharing informative news. For instance, under unbiased media, efficiency in the polarized network is higher by $0.877 - 0.727 = 15$ percentage points ($p = 0.000$) relative to the empty

[24] An alternative measure is "group success" rates (how often the elected candidate matches the realized state) but it does not consider the possibility that due to sampling and media bias, the realized signals may not reflect the true state sufficiently for voters to reach the correct decision. The Bayesian efficiency criterion corrects for this by conditioning on the realized signals. For completeness, group success rates are reported in Table 10 in Appendix B.

[25] Each partisan subgroup $C_j, j \in \{a, b\}$ is assigned a binary indicator of 1 whenever the majority in C_j votes for candidate C_a if and only if the Bayesian posterior p_j (based on all realized signals in both C_a and C_b for the empty and complete networks, and based on all realized signals in C_j only for the polarized network) exceeds the critical threshold t_j defined in (3). The modified efficiency measure for the polarized networks accounts for signal sharing constraints: the standard measure that uses signals from both subgroups mechanically implies a lower efficiency of the polarized network relative to the complete one in treatments with informative signals. For the sake of completeness, Table 11 in Appendix B uses the standard efficiency measure for all treatments. Both measures coincide for extreme bias treatments where signals are uninformative.

Table 3: Bayesian Efficiency by Treatment

Network			**No media bias**			
	N		[**All**]	N		[**Last 12**]
Empty	176	.716	(.021)	132	.727	(.024)
Polarized	288	.866	(.015)	216	.877	(.016)
Complete	272	.936	(.010)	204	.939	(.012)
Filter bubble effect		$-.069^{***}$	(.018)		$-.061^{***}$	(.020)
			Moderate Media Bias			
	N		[**All**]	N		[**Last 12**]
Empty	160	.700	(.021)	120	.708	(.025)
Polarized	144	.872	(.018)	108	.875	(.021)
Complete	144	.868	(.019)	108	.857	(.023)
Filter bubble effect		.004	(.026)		.019	(.031)
Media bias effect		$-.068^{***}$	(.022)		$-.082^{***}$	(.026)
			Extreme Media Bias			
	N		[**All**]	N		[**Last 12**]
Empty	112	.915	(.018)	84	.905	(.022)
Polarized	144	.833	(.023)	108	.843	(.026)
Complete	256	.713	(.017)	192	.721	(.018)
Filter bubble effect		$.120^{***}$	(.029)		$.121^{***}$	(.032)
Media bias effect		$-.223^{***}$	(.020)		$-.217^{***}$	(.022)

Notes: N is the number of group decision observations. Standard errors are in parentheses. Averages and standard errors based on all data are reported in columns labeled "All", those that exclude the first four rounds in "Last 12". No-bias non-empty network treatments with no signal probability $r = 0.2$ are pooled together with those in which $r = 0$. The filter bubble effect is the difference in efficiency between the polarized and complete network, for each bias treatment. The media bias effect is the difference in efficiency between the complete network under the respective bias treatment and under no bias. Significance codes: $^{***} < 0.01$, $^{**} < 0.05$, $^{*} < 0.1$

network, and by 21.1 p. p. ($p = 0.000$) in the complete network.[26]

However, the "Extreme media bias" results, in the bottom panel of Table 3, provide an interesting counterpoint. Under extreme media bias, signals are uninformative, so a Bayesian decision-maker would discard all signals and simply vote for the partisan candidate in each subgroup. Hence the efficiency measure here essentially tallies how often subjects voted for their partisan candidate. We see clearly that social networks lead to *lower* efficiency, with both complete and polarized networks being less efficient than the empty network: apparently, sharing of uninformative signals leads subjects to vote *against* their partisan candidate. The loss in efficiency from social networks in this setting does not imply a lack of information aggregation as, by design, there is no information to be aggregated in this treatment. Rather, it indicates that subjects are treating uninformative signals as informative, and acting upon them. Such behavior is consistent with the "fake news" interpretation of the extreme bias treatments.[27]

The efficiency comparison between polarized and complete networks (fixing media bias) represents the *filter bubble effect*. We find that while filter bubble effects are negative under no bias, implying that the polarized networks are less efficient than the complete ones (87.73% vs. 93.87%, $p = 0.002$), this is no longer the case under moderate bias (87.50% vs. 85.65%, $p = 0.551$) or

[26] If we use the standard efficiency measure for polarized network, it is still more efficient than the empty network, by $0.780 - 0.727 = 5.3$ percentage points ($p = 0.090$).

[27] In Table 11 in Appendix B we report the linear regression results (clustering standard errors at the treatment level) that confirm our efficiency findings and mitigate concerns about multiple hypothesis testing.

extreme bias (84.26% vs. 72.14%, $p = 0.000$).

Contrasting filter bubble effects with *media bias effects* (i.e. the efficiency comparisons between complete networks under the media bias and complete networks under no bias) shows that the latter have a larger impact on efficiency: while the negative filter bubble effect averages to -6.1 p.p., the negative media bias effect averages to at least -8.2 p.p. (under moderate bias) and the negative "fake" news effect (under extreme bias) is -21.7 p.p.

In summary, news sharing via social networks per se, does not appear to harmfully impact collective decisions; rather the culprit is media bias. When signals are informative, social networks allow for better information aggregation, leading to more efficient collective decisions compared to the empty network.

Result 1 (Efficiency)**.** *Social networks improve efficiency relative to the empty network when news signals are informative. However, with completely uninformative signals, social networks lower efficiency as voters act based upon uninformative signals. Under moderate bias, both complete and polarized networks have similar efficiency, even though the complete network can potentially reveal a larger number of informative signals. Overall, media bias treatments decrease efficiency relatively more than do filter bubbles.*

4.2 News-sharing patterns

We will now look at the news-sharing patterns, focusing on the selective news sharing prediction from the semi-pooling equilibria (cf. Proposition 3), whereby voters should communicate favorable signals but suppress unfavorable ones. To examine this prediction, we look at three different subgroups of signals: First, we look at the average signal sharing rate for all non-empty signals, clustered by subject (reported in Table 4 under "All signals sent"). Second, we look at the average signal sharing rates for signals favorable to the subject's party ("Fav. signals sent"), i.e. the ones that match the subject's preferred party color. Third, we look at average signal sharing rates for signals unfavorable to the subject's party ("Unfav. signals sent").

Table 4: 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%	Moderate	90	.897	(.015)	90	.923	(.016)	87	.814	(.031)
Complete	–	–	90	.881	(.016)	90	.913	(.016)	88	.780	(.034)
Polarized	–	Extreme	90	.827	(.021)	90	.891	(.019)	89	.678	(.039)
Complete	–	–	120	.825	(.016)	120	.888	(.018)	119	.673	(.034)

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

In the polarized network, all subjects in the same subnetwork have fully aligned interests, so in theory, there should be full signal revelation (Propositions 4 and 6). Table 4 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 we see from the "Fav. signals sent" and "Unfav. signals sent" columns of Table 4 that *subjects are selective in the signals they share*: in all treatments, favorable signals are shared 89–94% of the time but unfavorable signals are shared only 68–81% of the time. These differences are significant (in all cases, $p \leq 0.002$).[28]

At first glance, such selective news sharing behavior is consistent with the predictions of the semi-pooling equilibria, in which favorable signals are always revealed and unfavorable signals are hidden with a positive probability (Propositions 3 and 5). However, the sharing rates of unfavorable signals – 67.8% or higher in all treatments – far outstrip the maximum of about 8% predicted in the semi-pooling equilibria (Proposition 3).[29]

Instead, a simpler explanation for our selective news sharing finding is a behavioral inclination – that subjects derive intrinsic utility from sharing favorable signals, and disutility from sharing unfavorable signals. To test for this possibility, in rows 3–4 of Table 4 we present the unbiased treatments in which the probability of obtaining an empty signal is zero: in these cases, subjects who don't share a signal are not able to pretend to be uninformed. Interestingly, in those treatments, average signal sharing rates are virtually the same as in the treatments discussed above, where the probability of an empty signal was set to 20%. This suggests that subjects' selective news sharing is not guided by backward-induction reasoning about how their (non-)reports will affect others' posterior beliefs and votes, as in the semi-pooling equilibrium. Rather, there is something inherently distasteful about sharing an unfavorable signal which trumps the possibility that their non-report will be "unraveled."[30]

Result 2 (Selective news sharing)**.** *While subjects share the majority of their signals with those connected to them in a network in all treatments, full signal revelation is rejected in all network treatments. There is selective news sharing: unfavorable signals are communicated less often than favorable ones. However, sharing rates of unfavorable signals are too high to be consistent with signalling strategies in semi-pooling equilibria.*

4.3 Voting patterns

We turn to the analysis of individual behavior at the voting stage.

Sincere Bayesian voting by subject. First, we check whether subjects take into account their signal as well as signals revealed by others in their vote in a way consistent with the theoretical

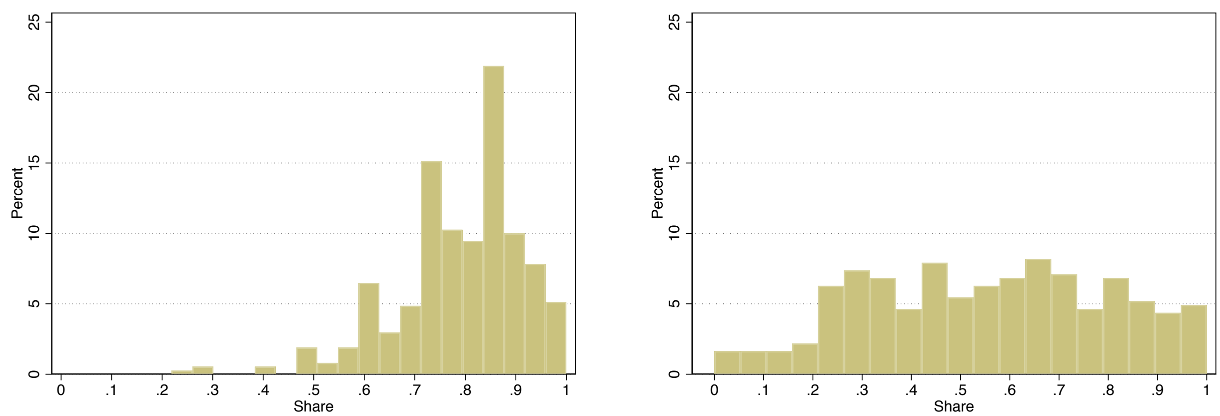
[28] The dynamics of signal sharing across experimental rounds are illustrated in Figures 6–7 in Appendix B. Except for the first few initial rounds (expiring around round 4, hence our focus on the last 12 rounds in Table 3), the average sharing rates do not vary much over time. Unfavorable signals in treatments with media bias are naturally less likely to occur so those sharing rates exhibit a higher variance over time than the sharing rates of favorable signals.

[29] For the moderate bias case, the maximum sharing rate of unfavorable signals is about 14%, still way lower than the observed rates (Proposition 5).

[30] In sender-receiver disclosure experiments, Jin et al. (2017) report that senders disclose more favorable information more often, which suggests robustness of the selective news sharing result across contexts.

predictions, by computing for each treatment the theoretical Bayesian posterior as a function of shared and private signals in each group-round.[31]

To estimate how often subjects followed sincere Bayesian voting 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, obtaining a binary indicator of consistent voting. Our subject-level measure is the fraction of voting decisions that are fully consistent with a sincere Bayesian strategy. The distribution of full consistency across all treatments is illustrated in Figure 1(a). In the second approach, illustrated in Figure 1(b), we only looked at those cases in which the Bayesian posterior counselled a player to vote against their partisan candidate, and computed how often they voted accordingly. This can be considered a more stringent measure of sincere Bayesian voting as it involves voting in line with the information from the signals even when this information contradicts one's partisan bias.[32]



(a) $N = 370$ subjects

(b) $N = 366$ subjects

Figure 1: Do subjects vote consistently with sincere Bayesian voting? Panel (a): using all votes; Panel (b): using only votes against partisan candidate. y-axis: percent of subjects; x-axis: share of each subject's votes (out of 32 decisions per subject) consistent with sincere Bayesian voting strategies.

From Figure 1(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 1 reveals that this seemingly high consistency is partly due to sincere Bayesian strategy prescribing voting for one's partisan candidate. If we look exclusively at those cases in which the sincere Bayesian strategy prescribes voting against one's partisan candidate (in panel (b)), the consistency rate is markedly lower, and more than half the subjects fail to vote against their partisan candidate when doing so is prescribed by the Bayesian strategy. Nevertheless, there is a non-negligible proportion of subjects using sincere Bayesian strategies.[33]

[31]The formulae for these posteriors are provided in the proofs of Propositions 1–6 in Appendix A.

[32]Figure 1(b) by construction excludes the extreme bias treatments, since in those the Bayesian posterior never counsels a player to vote against their partisan candidate. To ensure comparability, we also excluded the extreme bias treatments from Figure 1(a). Figure 8 in Appendix B has the extreme bias treatments included.

[33]Moreover, it is precisely those subjects who score highly in panel (b) who also score highly in panel (a), as

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 partisan candidate.[34]

Sincere Bayesian voting by treatment. Next, we check how consistency with sincere Bayesian strategies varies by treatment. We classified each subject's vote decision in each round as "correct" if it was consistent with the sincere Bayesian strategy given the voter's theoretical posterior (as in Figure 1 (a)). We also investigated consistency with sincere Bayesian strategy given the voter's empirical posterior estimated from the data.[35] In Table 5, we report the frequency of subjects' votes which are consistent with sincere Bayesian voting, using both the theoretical (columns 4-5) and the empirical posteriors (columns 6-7). Thus columns 4–5 of Table 5 display individual subject Bayesian consistency scores from Figure 1(a) averaged across each treatment combination. In addition, Table 5 displays consistency with extreme bias treatments (not included in Figure 1), for which a sincere Bayesian strategy prescribes always voting for the default choice.

Table 5: Sincere Bayesian Voting by Treatment

Network	**Bias**	N	**Consistent vote share (theoretical posterior)**		**Consistent vote share (empirical posterior)**	
Empty	No	110	.645	(.016)	.645	(.016)
Polarized	–	180	.807	(.011)	.779	(.010)
Complete	–	170	.861	(.010)	.870	(.011)
Empty	Moderate	100	.835	(.015)	.770	(.016)
Polarized	–	90	.810	(.014)	.740	(.017)
Complete	–	90	.742	(.014)	.787	(.012)
Empty	Extreme	70	.777	(.018)	.777	(.018)
Polarized	–	90	.738	(.018)	.728	(.019)
Complete	–	120	.673	(.014)	.669	(.014)

Notes: N is the number of subjects, standard errors in parentheses are clustered by subject. No bias, non-empty network treatments with $r = 0$ and $r = 0.2$ are pooled together.

Both of our consistency measures (based on empirical and theoretical posteriors) tell the same story under the no bias and extreme bias treatments. The average consistency scores follow the Bayesian efficiency patterns we observed earlier in Table 3. Without bias, the complete network is *more consistent* than the polarized than the empty one. Under extreme bias (rows 7-9), when signals are uninformative by design, the complete network is *less consistent* than the polarized than the empty one (although the latter difference is not significant). Thus while sincere Bayesian strategies simply dictate voting for one's partisan candidate regardless of any signals, subjects

illustrated in Figure 9 in Appendix B, which depicts consistency with sincere Bayesian strategies only for those subjects who score more than 60% correct in Figure 1(b).

[34]We also looked at subjects' signal sharing patterns conditional on their Bayesian voting types. Consistently with results in Table 4, we difference in the probability of sharing favorable vs. unfavorable signals depends very little on the bias and network structure. More-consistent subjects exhibit a slightly bigger gap in selective news sharing than the less-consistent ones.

[35]We estimated the empirical frequency of state a conditional on voter's own signal (s_a, s_b, or $s_\emptyset$), and for network treatments, conditional on the difference in the number of s_a and s_b signals revealed by others (ignoring signal sources and empty signals). Since the difference in the number of signals can vary between -9 and 9 (in complete networks), to improve estimation we collapsed values of -9 and -8 into a single "< -7" category, and values of 8 and 9 into a single "> 7" category. In the complete network treatment data, the number of revealed signals is between -6 and 6 for about 94.4% of the time (no bias) to 98.89% of the time (moderate bias) to 99.76% of the time (extreme bias).

Table 6: Voting for partisan candidate, complete network

		No bias	**Extreme**	**Moderate**
Fav. Signal	Yes	$.093^{***}$	$.108^{***}$	$.164^{***}$
		(.024)	(.025)	(.028)
	No	$-.110^{***}$	$-.214^{***}$	$-.169^{***}$
		(.024)	(.029)	(.040)
# Fav, co-partisans		$.073^{***}$	$.077^{***}$	$.074^{***}$
		(.007)	(.008)	(.012)
# Unfav, co-partisans		$-.063^{***}$	$-.095^{***}$	$-.092^{***}$
		(.008)	(.012)	(.017)
# Fav, anti-partisans		$.071^{***}$	$.086^{***}$	$.055^{***}$
		(.007)	(.013)	(.017)
# Unfav, anti-partisans		$-.053^{***}$	$-.071^{***}$	$-.091^{***}$
		(.007)	(.009)	(.010)
Realized state		✓	✓	✓
N		2,550	2,400	1,350
% correctly classified		87.69	76.46	82.67

Notes: N is the number of individual decisions (first round observations are dropped). The numbers represent average marginal effects on Prob (Vote for Partisan Candidate).

actually vote *against* their partisan candidate with non-negligible probability. Essentially, they treat fake news as real news.

Under moderate bias (rows 4-6), the observed voting behavior in the empty and polarized networks appears more consistent than in the complete network using the theoretical posteriors, while consistency is highest for complete network using the empirical posteriors. This difference arises from accounting for signal sources in the theoretical posterior under the moderate bias and non-empty networks, which we investigate more below.

Result 3 (Voting according to sincere Bayesian strategies)**.** *While a significant proportion of subjects vote consistently with equilibrium sincere Bayesian voting strategies, the majority use voting strategies that are biased towards their partisan candidate relative to the available information. At the same time, subjects largely take uninformative signals at face value. Without media bias, the complete network has the highest average consistency with sincere Bayesian strategies. With media bias, the complete network has the lowest average consistency.*

Do voters account for the sources of the signals? Given limited consistency with sincere Bayesian voting, we estimate a binary logit model of individual vote where the dependent variable is set to one when an individual votes for her partisan candidate, and zero otherwise. We include voter's private signal, the number of favorable and unfavorable signals (with respect to the voter's partisanship) shared by other voters, and their interactions, controlling for the realized state (which is unobserverd by voters but is correlated with realized signals in all but extreme bias treatments). The average marginal effects for the complete network case are reported in Table 6.

Table 6 makes two important observations. First, an extra favorable signal, regardless of source, significantly increases the likelihood of voting for the partisan candidate, while an extra unfavorable signal significantly decreases this likelihood, in all media bias treatments. Second, there are no significant differences in the average marginal effects of signals of the same kind (favorable or unfavorable) revealed by co-partisans vs. anti-partisans.[36] Under moderate bias, favorable signals are more likely in both states. Since co-partisans and anti-partisans receive signals biased in the opposite directions, favorable signals from co-partisans are much less informative than favorable signals from anti-partisans (similarly, unfavorable signals from co-partisans are more informative than unfavorable signals from anti-partisans). A fully rational Bayesian updating would imply a larger positive marginal effect in the "# Fav, anti-partisans" row than in the "# Fav, co-partisans" row, and a smaller negative marginal effect in the "# Unfav, co-partisans" row than in the "# Unfav, anti-partisans" row. Since those coefficients are not significantly different, it appears that voters do not account for the signal sources. We obtain

Result 4 (Determinants of individual votes)**.** *While subjects account for the kind of signals revealed by others (favorable to them vs. unfavorable to them) in their voting decisions, they do not account for the signal sources (i.e. whether signals are revealed by co-partisans or anti-partisans).*

5 Further Evidence: Survey Results from Pakistan

As a robustness check to see how our results – obtained in a laboratory setting – translate to a real-world setting, in this Section we present survey evidence collected for us by Gallup Pakistan. The goal was to check whether, like in our main findings, subjects selectively share information that is favorable to their party more often than the unfavorable information, and how likely they are to revise their prior beliefs in the presence of unfavorable information.

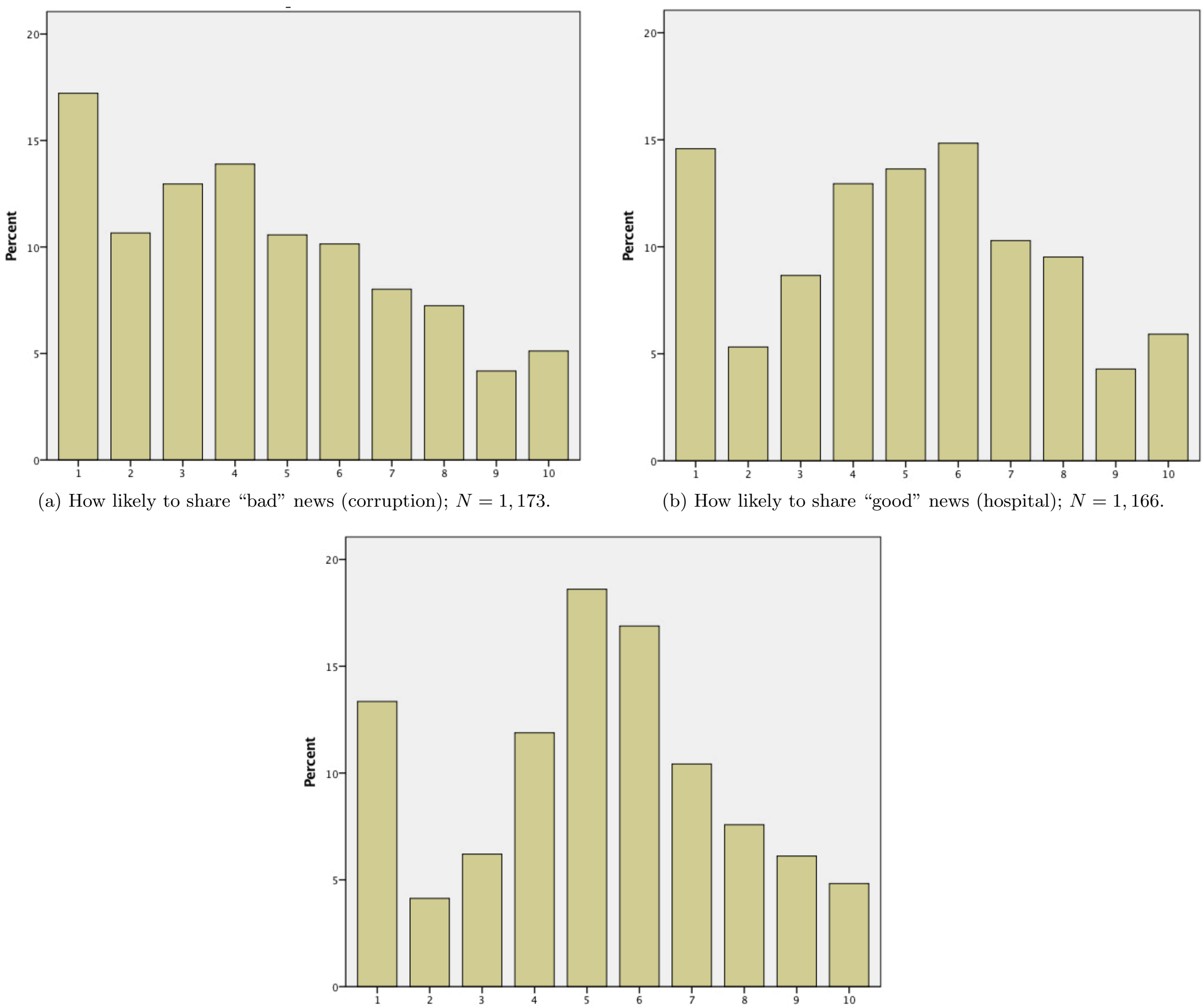
We introduced five questions on social media into a standard questionnaire administered by Gallup interviewers to a panel of the Pakistani population, representative with respect to the province level and urban/rural split, during January 16th–20th, 2017. In the survey questions we asked how likely a respondent was to share a "favorable news" (that their favorite political candidate was a major force behind building a new hospital), to share an "unfavorable news" (that their favorite political candidate was accused of corruption), and to revise their opinion after an "unfavorable news" item 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.[37]

Figures 2–3 present the survey responses graphically.[38] Comparing Figures 2(a) and 2(b), we see 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

[36] For revealed fav. signals, Wald $p = .828$ for no bias, $p = .674$ for moderate bias, and $p = .879$ for extreme bias; for revealed unfav. signals, Wald $p = .913$ for no bias, $p = .788$ for moderate bias, and $p = .341$ for extreme bias.

[37] The exact questions are available in Table 13 in Appendix B. The questionnaire was administered in an Urdu translation.

[38] The actual response frequencies for each category are in Table 13, and the main demographic characteristics of the survey respondents are in Table 12 in Appendix B.



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

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

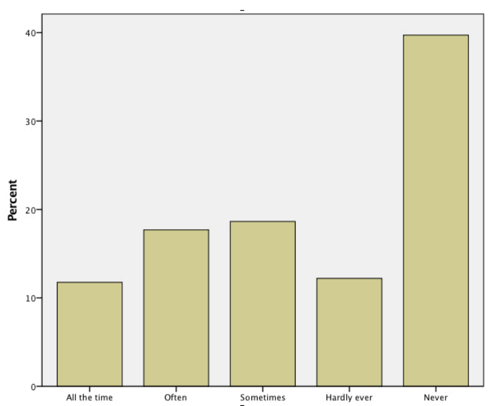
(c) How likely to revise opinion if Facebook friend shared "bad" news; $N = 1,161$.

Figure 2: 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.

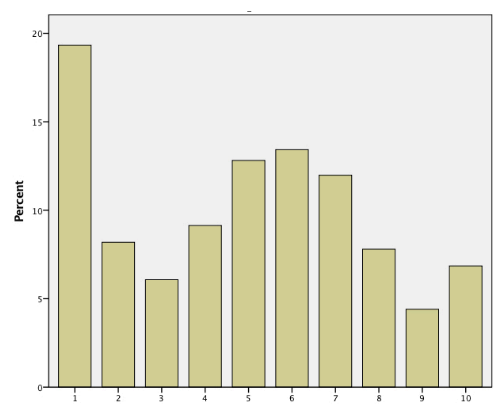
social media users. Furthermore, about 45.8% of social media users are "likely" to revise their opinion after a Facebook friend shared a "bad" news article. Both of these findings are consistent with the experimental results described earlier.

From Figure 3(a), we see that about 48.1% of respondents regularly get the news about politics and government from social media. In the full sample (including those who do not use social media), according to Figure 3(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 – 4.72). The histogram of trustworthiness features a peak at the lower tail, which consists primarily of those subjects who answered "Never" in Figure 3(a).[39]

[39]Moreover, this histogram differs from results from a similar question asked to subjects from our lab experiments; see Figure 5 in Appendix B.



(a) Get news about politics & gov't from social media



(b) How trustworthy is information from social media? Scale 1=not at all, 10=extremely trustworthy

Figure 3: Distribution of survey responses about social media as a source of news (left panel) and degree of trust (right panel); $N = 1,803$.

6 Concluding Remarks

The proliferation of news and information filtering 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 information filtering and targeting makes it less likely that voters will 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 for the idea that by filtering out unfavorable content social media may lead to polarization in voting behavior. Voters publicly share signals favorable to their party more often than signals unfavorable to their party, take uninformative signals at face value, and ignore signal sources. This suggests that improving the quality of information shared on social networks is of first order policy importance, especially in light of recent Facebook announcements regarding the change in the newsfeed algorithm that would prioritize "organic" news and posts shared by users over that of advertisers and publishers.[40]

Our results that voters selectively share signals favorable to their party are different from findings in the "confirmation bias" and "information avoidance" literatures[41] – 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 I only share bad news with you, you may drop me as a Facebook friend). In ongoing work,

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

[41] 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.

we are exploring how to enrich our experimental setting to allow for these types of effects.

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.

Breitmoser, Y. and Valasek, J. (2018). A Rationale for Unanimity in Committees. Working paper.

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.

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.

Galeotti, A. and Mattozzi, A. (2011). “Personal Influence”: Social Context and Political Competition. *American Economic Journal: Microeconomics*, 3(1):307–327.

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.

Gole, T. and Quinn, S. (2014). Committees and status quo bias: Structural evidence from a randomized field experiment. *University of Oxford Department of Economics Discussion Paper*, 733.

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.

Iaryczower, M., Shi, X., and Shum, M. (2018). Can Words Get in the Way? The Effect of Deliberation in Collective Decision-Making. *Journal of Political Economy*, 126(2).

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

Kawamura, K. and Vlaseros, V. (2017). Expert information and majority decisions. *Journal of Pubic Economics*, 147:77–88.

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.

Li, H., Rosen, S., and Suen, W. (2001). Conflicts and Common Interests in Committees. *American Economic Review*, 91(5):1478–1497.

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.

Morton, R. B., Piovesan, M., and Tyran, J.-R. (2018). The dark side of the vote: Biased voters, social information, and information aggregation through majority voting. *In Press, Games and Economic Behavior.*

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. (2018). Media power. *Journal of Political Economy*, 126(4):1747–1783.

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: $\mu : \{C_a, C_b\} \times \{s_a, s_b, s_\emptyset\} \to \{s_a, s_b, s_\emptyset\}$, $\mu(\cdot, s_\emptyset) = s_\emptyset$, 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 equilibrium where all voters vote informatively.*

Proof. Since there is no abstention and each voter is independently uninformed with probability $r = 0.2$, a fully informative equilibrium is not 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 the 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)

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

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 posterior calculations become more complicated and require taking the expectation over each 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 conclusion continues to hold, and there is no informative voting. □

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

Proof. Let $k := \#s_a - \#s_b$, $k \in \{-n, \dots, n\}$ 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

$$p_i(\theta = a|k_{-i}) = \frac{\overbrace{((1-r)q)^{\#_{-i}(s_a)}((1-r)(1-q))^{\#_{-i}(s_b)}r^{n-1-\#_{-i}(s_a)-\#_{-i}(s_b)}}^{\text{denote as } X}}{X + ((1-r)(1-q))^{\#_{-i}(s_a)}((1-r)q)^{\#_{-i}(s_b)}r^{n-1-\#_{-i}(s_a)-\#_{-i}(s_b)}}$$
$$= \frac{1}{1+q^{-k_{-i}}(1-q)^{k_{-i}}} = \frac{1}{1+\left(\frac{1-q}{q}\right)^{k_{-i}}} \tag{4}$$

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.[42] □

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

$$\mu_{-j}(\tilde{s}^j_\emptyset, \nu) \equiv \Pr(s = s_{-j}|\tilde{s}^j_\emptyset) = \frac{\frac{1}{2}(1-r)(1-q)\nu + \frac{1}{2}(1-r)q\nu}{\frac{1}{2}(1-r)(1-q)\nu + \frac{1}{2}(1-r)q\nu + r} = \frac{\frac{1}{2}(1-r)\nu}{\frac{1}{2}(1-r)\nu + r} \tag{5}$$

Of course, if they receive a non-empty signal, they believe it, since signals are hard evidence. Note also that $1 - \mu_{-j}(\tilde{s}^j_\emptyset, \nu) \equiv \Pr(s = s_\emptyset|\tilde{s}^j_\emptyset)$.[43] 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,

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

[43] 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.

being the expectation of a binomial random variable, is

$$h_{-j}(\#_{-i}(\tilde{s}_\emptyset^j)) = \sum_{\ell=0}^{\#_{-i}(\tilde{s}_\emptyset^j)} \ell \binom{\#_{-i}(\tilde{s}_\emptyset^j)}{\ell} (\mu)^\ell (1-\mu)^{\#_{-i}(\tilde{s}_\emptyset^j)-\ell} \equiv \mu \cdot \#_{-i}(\tilde{s}_\emptyset^j) \tag{6}$$

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

$$\begin{aligned}
&p_i(\theta = a | k_{-i}, \pi_{-i}(\tilde{s}_\emptyset)) = \\
&= \frac{\overbrace{((1-r)q)^{\#_{-i}(s_a)+h_a(\#_{-i}(\tilde{s}_\emptyset^b))}((1-r)(1-q))^{\#_{-i}(s_b)+h_b(\#_{-i}(\tilde{s}_\emptyset^a))} r^{n-1-h_b(\#_{-i}(\tilde{s}_\emptyset^a))-h_a(\#_{-i}(\tilde{s}_\emptyset^b))}}^{\text{denote as } X}}{X + ((1-r)(1-q))^{\#_{-i}(s_a)+h_a(\#_{-i}(\tilde{s}_\emptyset^b))}((1-r)q)^{\#_{-i}(s_b)+h_b(\#_{-i}(\tilde{s}_\emptyset^a))} r^{n-1-h_b(\#_{-i}(\tilde{s}_\emptyset^a))-h_a(\#_{-i}(\tilde{s}_\emptyset^b))}} \\
&= \frac{1}{1 + \left(\frac{1-q}{q}\right)^{\#_{-i}(s_a)+h_a(\#_{-i}(\tilde{s}_\emptyset^b))-\#_{-i}(s_b)-h_b(\#_{-i}(\tilde{s}_\emptyset^a))}} = \frac{1}{1 + \left(\frac{1-q}{q}\right)^{k_{-i}+\pi_{-i}(\tilde{s}_\emptyset)}} \\
&= \frac{1}{1 + \left(\frac{1-q}{q}\right)^{k_{-i}+\mu k_{-i}(\tilde{s}_\emptyset)}}
\end{aligned} \tag{7}$$

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 (5). 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:

$$p_{-i}(\theta = a|k_{-i} - \mathbb{1}_{\{s_i = s_b\}}, \pi'(\tilde{s}_\emptyset)) < p_{-i}(\theta = a|k_{-i}, x_a(\tilde{s}_\emptyset)) \tag{8}$$

and

$$p_{-i}(\theta = a|k_{-i} + \mathbb{1}_{\{s_i = s_a\}}, \pi'(\tilde{s}_\emptyset)) > p_{-i}(\theta = a|k_{-i}, x_b(\tilde{s}_\emptyset)) \tag{9}$$

(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 (8), 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 (9), 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' and $k'(\tilde{s}_\emptyset)$ (note: the latter number determines $\pi'(\tilde{s}_\emptyset)$ for a fixed μ), 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 (7), 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 values of k_{-i} and $k_{-i}(\tilde{s}_\emptyset)$, there are four possibilities, and the corresponding critical values can be expressed via the following equations:

$$\mu^*_{i)1)}(s_b, C_a) = \frac{\ln\left(\frac{1}{t_b} - 1\right) - (k_{-i} - \mathbb{1}_{\{s_i = s_b\}}) \ln \frac{1-q}{q}}{k_{-i}(\tilde{s}_\emptyset) \ln \frac{1-q}{q}} \tag{10}$$

$$\mu^*_{i)2)}(s_b, C_a) = \frac{\ln\left(\frac{1}{t_b} - 1\right) - k_{-i} \ln \frac{1-q}{q}}{(k_{-i}(\tilde{s}_\emptyset) - \mathbb{1}_{\{s_i = s_b\}}) \ln \frac{1-q}{q}} \tag{11}$$

$$\mu^*_{ii)1)}(s_a, C_b) = \frac{\ln\left(\frac{1}{t_a} - 1\right) - (k_{-i} + \mathbb{1}_{\{s_i = s_a\}}) \ln \frac{1-q}{q}}{k_{-i}(\tilde{s}_\emptyset) \ln \frac{1-q}{q}} \tag{12}$$

$$\mu^*_{ii)2)}(s_a, C_b) = \frac{\ln\left(\frac{1}{t_a} - 1\right) - k_{-i} \ln \frac{1-q}{q}}{(k_{-i}(\tilde{s}_\emptyset) + \mathbb{1}_{\{s_i = s_a\}}) \ln \frac{1-q}{q}} \tag{13}$$

The critical values of ν, denoted ν^* are obtained by reversing (5):

$$\nu = \frac{2\mu r}{(1-\mu)(1-r)}. \tag{14}$$

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 (5). Since ν^* is incentive-compatible for all possible communication outcomes, sincere Bayesian voting based on each player's equilibrium posterior remains a best response even conditional on vote pivotality. Voting is informative and information gets aggregated. □

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:

$$\begin{aligned} p_i(\theta = a|k_{-i}(C_j)) &= \frac{\overbrace{((1-r)q)^{\#_{-i}(s_a(C_j))}((1-r)(1-q))^{\#_{-i}(s_b(C_j))} r^{\frac{n}{2}-1-\#_{-i}(s_a(C_j))-\#_{-i}(s_b(C_j))}}^{\text{denote as } X}}{X + ((1-r)(1-q))^{\#_{-i}(s_a(C_j))}((1-r)q)^{\#_{-i}(s_b(C_j))} r^{\frac{n}{2}-1-\#_{-i}(s_a(C_j))-\#_{-i}(s_b(C_j))}} \\ &= \frac{1}{1 + q^{-k_{-i}(C_j)}(1-q)^{k_{-i}(C_j)}} = \frac{1}{1 + \left(\frac{1-q}{q}\right)^{k_{-i}(C_j)}} \end{aligned} \tag{15}$$

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 $p_{-j}(\theta = a|k(C_j))$:

$$\begin{aligned}
p_{-j}(\theta = a|k(C_j)) &= \frac{p_i(\theta = a|k(C_j))}{1 + \left(\frac{1-q}{q}\right)^{\mu(k_{-i}(C_{-j})|\theta=a)}} + \frac{1 - p_i(\theta = a|k(C_j))}{1 + \left(\frac{1-q}{q}\right)^{\mu(k_{-i}(C_{-j})|\theta=b)}} \\
&= \frac{p_i(\theta = a|k(C_j))}{1 + \left(\frac{1-q}{q}\right)^{\mu(k_{-i}(C_{-j})|\theta=a)}} + \frac{1 - p_i(\theta = a|k(C_j))}{1 + \left(\frac{1-q}{q}\right)^{-\mu(k_{-i}(C_{-j})|\theta=a)}} \\
&= p_i(\theta = a|k(C_j)) \left(\frac{1}{1 + \left(\frac{1-q}{q}\right)^{\mu(k_{-i}(C_{-j})|\theta=a)}} - \frac{1}{1 + \left(\frac{1-q}{q}\right)^{-\mu(k_{-i}(C_{-j})|\theta=a)}} \right) \\
&\quad + \frac{1}{1 + \left(\frac{1-q}{q}\right)^{-\mu(k_{-i}(C_{-j})|\theta=a)}} \\
&\approx 0.5901 \cdot p_i(\theta = a|k(C_j)) + 0.2049 \qquad (16)
\end{aligned}$$

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

$$p_j(\theta = a|k(C_j)) = \frac{1}{1 + \left(\frac{1-q}{q}\right)^{k(C_j)}} \qquad (17)$$

□

A.2 The case of moderate media bias

Proof of Proposition 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

$$p_i(\theta = j|s_j) = \frac{(1-r)q_j^j}{(1-r)q_j^j + (1-r)(1-q_j^{-j})} = \frac{1}{1+\left(\frac{1-q_j^{-j}}{q_j^j}\right)} = 0.6 \tag{18}$$

$$p_i(\theta = j|s_{-j}) = \frac{(1-r)(1-q_j^j)}{(1-r)(1-q_j^j) + (1-r)q_j^{-j}} = \frac{1}{1+\left(\frac{q_j^{-j}}{1-q_j^j}\right)} = 0.2 \tag{19}$$

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 $p_i(\theta = j|s_{-j}) < t_j$ (see (3)). 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

$$\begin{aligned}
p_i(\theta = a|k_{-i}(\text{f}), k_{-i}(\text{uf})) &= \frac{\overbrace{(q_a^a)^{\#_{-i}(s_a(C_a))}(q_b^a)^{\#_{-i}(s_a(C_b))}(1-q_a^a)^{\#_{-i}(s_b(C_a))}(1-q_b^a)^{\#_{-i}(s_b(C_b))}}^{\text{denote as } Z}}{Z + (1-q_a^b)^{\#_{-i}(s_a(C_a))}(1-q_b^b)^{\#_{-i}(s_a(C_b))}(q_a^b)^{\#_{-i}(s_b(C_a))}(q_b^b)^{\#_{-i}(s_b(C_b))}} \\
&= \frac{1}{1+\left(\frac{1-q_a^b}{q_a^a}\right)^{\#_{-i}(s_a(C_a))}\left(\frac{1-q_b^b}{q_b^a}\right)^{\#_{-i}(s_a(C_b))}\left(\frac{q_a^b}{1-q_a^a}\right)^{\#_{-i}(s_b(C_a))}\left(\frac{q_b^b}{1-q_b^a}\right)^{\#_{-i}(s_b(C_b))}} \\
&= \frac{1}{1+\left(\frac{2}{3}\right)^{\#_{-i}(s_a(C_a))}\left(\frac{1}{4}\right)^{\#_{-i}(s_a(C_b))}\left(\frac{1}{4}\right)^{-\#_{-i}(s_b(C_a))}\left(\frac{2}{3}\right)^{-\#_{-i}(s_b(C_b))}} \\
&= \frac{1}{1+\left(\frac{2}{3}\right)^{k_{-i}(\text{f})}\left(\frac{1}{4}\right)^{k_{-i}(\text{uf})}}
\end{aligned} \tag{20}$$

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 7 (by revealing her signal, i pushes the others' posterior over one of the two critical thresholds, t_a and t_b).

Table 7: 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.

Similarly to the no bias case, Table 7 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 MB1+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.*[44] *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 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

$$\mu_{-j} \equiv \Pr(s = s_{-j} | \tilde{s}^j_\emptyset) = \frac{\frac{1}{2}(1-r)(1-q^j_j)\nu + \frac{1}{2}(1-r)q^{-j}_j \nu}{\frac{1}{2}(1-r)(1-q^j_j)\nu + \frac{1}{2}(1-r)q^{-j}_j \nu + r}$$
$$= \frac{\frac{1}{2}(1-r)\nu(1-q^j_j + q^{-j}_j)}{\frac{1}{2}(1-r)\nu(1-q^j_j + q^{-j}_j) + r} \tag{21}$$

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 (6)). 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, $k_{-i}(\tilde{s}_\emptyset) \in \{-5\mathbb{1}_{i\in C_b}, -4, \ldots, 4, 5\mathbb{1}_{i \in C_a}\}$.[45] 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,

$$k_{-i}(\text{f}) - k_{-i}(\text{uf}) = k_{-i}(\tilde{s}_\emptyset) + \mathbb{1}_{i\in C_b} - \mathbb{1}_{i\in C_a} \tag{22}$$

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-

[44] See Eq (23), 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.

[45] 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.

revealed) by others becomes

$$
\begin{aligned}
&p_i(\theta = a|k_{-i}(\text{f}), k_{-i}(\text{uf}), \pi_{-i}(\tilde{s}_\emptyset)) = \\
&= \frac{\overbrace{(q_a^a)^{\#_{-i}(s_a(C_a))}(q_b^a)^{\#_{-i}(s_a(C_b))+h_a(\#_{-i}(\tilde{s}_\emptyset^b))}(1-q_a^a)^{\#_{-i}(s_b(C_a))+h_b(\#_{-i}(\tilde{s}_\emptyset^a))}(1-q_b^a)^{\#_{-i}(s_b(C_b))}}^{\text{denote as } Z}}{Z + (1-q_a^b)^{\#_{-i}(s_a(C_a))}(1-q_b^b)^{\#_{-i}(s_a(C_b))+h_a(\#_{-i}(\tilde{s}_\emptyset^b))}(q_a^b)^{\#_{-i}(s_b(C_a))+h_b(\#_{-i}(\tilde{s}_\emptyset^a))}(q_b^b)^{\#_{-i}(s_b(C_b))}} \\
&= \frac{1}{1 + \left(\frac{1-q_a^b}{q_a^a}\right)^{\#_{-i}(s_a(C_a))}\left(\frac{1-q_b^b}{q_b^a}\right)^{\#_{-i}(s_a(C_b))+h_a(\#_{-i}(\tilde{s}_\emptyset^b))}\left(\frac{q_a^b}{1-q_a^a}\right)^{\#_{-i}(s_b(C_a))+h_b(\#_{-i}(\tilde{s}_\emptyset^a))}\left(\frac{q_b^b}{1-q_b^a}\right)^{\#_{-i}(s_b(C_b))}} \\
&= \frac{1}{1 + \left(\frac{2}{3}\right)^{\#_{-i}(s_a(C_a))}\left(\frac{1}{4}\right)^{\#_{-i}(s_a(C_b))+h_a(\#_{-i}(\tilde{s}_\emptyset^b))}\left(\frac{1}{4}\right)^{-\#_{-i}(s_b(C_a))-h_b(\#_{-i}(\tilde{s}_\emptyset^a))}\left(\frac{2}{3}\right)^{-\#_{-i}(s_b(C_b))}} \\
&= \frac{1}{1 + \left(\frac{2}{3}\right)^{k_{-i}(\text{f})}\left(\frac{1}{4}\right)^{k_{-i}(\text{uf})+\pi_{-i}(\tilde{s}_\emptyset)}} = \frac{1}{1 + \left(\frac{2}{3}\right)^{k_{-i}(\text{f})}\left(\frac{1}{4}\right)^{k_{-i}(\text{uf})+\mu k_{-i}(\tilde{s}_\emptyset)}} \qquad (23)
\end{aligned}
$$

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 (23):

$$p_i(\theta = a|k_{-i}(\text{f}), k_{-i}(\text{uf}), \pi_{-i}(\tilde{s}_\emptyset)) = \frac{1}{1 + \left(\frac{2}{3}\right)^{k_{-i}(\text{f})}\left(\frac{1}{4}\right)^{k_{-i}(\text{uf})+\mu k_{-i}(\tilde{s}_\emptyset)}} \qquad (24)$$

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 from (21). 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:

$$p_{-i}(\theta = a|k'(\text{f}), k_a(s_b), \pi'(\tilde{s}_\emptyset)) < p_{-i}(\theta = a|k'(\text{f}), k'(\text{uf}), x_a(\tilde{s}_\emptyset)) \qquad (25)$$

and

$$p_{-i}(\theta = a|k'(\text{f}), k_b(s_a), \pi'(\tilde{s}_\emptyset)) > p_{-i}(\theta = a|k'(\text{f}), k'(\text{uf}), x_b(\tilde{s}_\emptyset)) \qquad (26)$$

(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 (25), 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 (26), 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 (23), 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:

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

$$\mu^*_{i)2)}(s_b, C_a) = \frac{\ln\left(\frac{1}{t_b} - 1\right) - k_{-i}(\text{f}) \ln\frac{2}{3} - k_{-i}(\text{uf}) \ln\frac{1}{4}}{(k_{-i}(\tilde{s}_\emptyset) - \mathbb{1}_{\{s_i = s_b\}}) \ln\frac{1}{4}} \tag{28}$$

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

$$\mu^*_{ii)2)}(s_a, C_b) = \frac{\ln\left(\frac{1}{t_a} - 1\right) - k_{-i}(\text{f}) \ln\frac{2}{3} - k_{-i}(\text{uf}) \ln\frac{1}{4}}{(k_{-i}(\tilde{s}_\emptyset) + \mathbb{1}_{\{s_i = s_a\}}) \ln\frac{1}{4}} \tag{30}$$

The critical values of ν, denoted ν^* are obtained by reversing (21):

$$\nu = \frac{2\mu r}{(1-\mu)(1-r)(1-q_j^j + q_j^{-j})}, \tag{31}$$

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 $k_{-i}(\text{f})$, $k_{-i}(\text{uf})$, 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 $(.862, 1]$.

An important difference for belief updating in Proposition 5 compared to Proposition 3 is that in the moderate bias case, the posterior beliefs depend not only on the number of revealed signals but also on their sources: 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 carry less weight in the posterior than revealed unfavorable signals. □

Proposition 6. ***[Moderate media bias and polarized network]*** *Given our experimental parameters, under MB1+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.*[46] *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,

[46]The posteriors are given by Eq (32) – (33), supplemented by i's signal.

with some modifications regarding the expressions for the posteriors. Namely, (15) becomes

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

$$p_i(\theta = a|\#_{-i}(s_a(C_b)), \#_{-i}(s_b(C_b))) = \frac{1}{1 + \left(\frac{q_b^b}{1-q_b^a}\right)^{\#_{-i}(s_b(C_b))} \left(\frac{1-q_b^b}{q_b^a}\right)^{\#_{-i}(s_a(C_b))}} \tag{33}$$

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

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

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

and in state $\theta = b$:

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

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

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 $\theta = a$:

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

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

and in state $\theta = b$:

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

$$\mu_{C_b}(\#(s_b(C_a))|\theta = b) = \sum_{\alpha=0}^{\frac{n}{2}} \sum_{\beta=0}^{\frac{n}{2}-\alpha} \frac{\beta(\frac{n}{2})!}{\alpha!\beta!(\frac{n}{2}-\alpha-\beta)!}((1-r)(1-q_a^b))^{\alpha}((1-r)q_a^b)^{\beta} r^{\frac{n}{2}-\alpha-\beta} \tag{41}$$

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

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

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 $i \in C_a$,

$$\begin{aligned} p_b^{C_a}(\theta = a) &= \frac{p_i(\theta = a|\#(s_a(C_a)), \#(s_b(C_a)))}{1 + \left(\frac{q_b^b}{1-q_b^a}\right)^{\mu_{C_a}(\#(s_b(C_b))|\theta=a)} \left(\frac{1-q_b^b}{q_b^a}\right)^{\mu_{C_a}(\#(s_a(C_b))|\theta=a)}} \\ &+ \frac{1 - p_i(\theta = a|\#(s_a(C_a)), \#(s_b(C_a)))}{1 + \left(\frac{q_b^b}{1-q_b^a}\right)^{\mu_{C_a}(\#(s_b(C_b))|\theta=b)} \left(\frac{1-q_b^b}{q_b^a}\right)^{\mu_{C_a}(\#(s_a(C_b))|\theta=b)}} \\ &\approx 0.488 \cdot p_i(\theta = a|\#(s_a(C_a)), \#(s_b(C_a))) + 0.288 \end{aligned} \tag{42}$$

and for $i \in C_b$,

$$\begin{aligned} p_a^{C_b}(\theta = a) &= \frac{p_i(\theta = a|\#(s_a(C_a)), \#(s_b(C_a)))}{1 + \left(\frac{q_a^b}{1-q_a^a}\right)^{\mu_{C_b}(\#(s_b(C_a))|\theta=a)} \left(\frac{1-q_a^b}{q_a^a}\right)^{\mu_{C_b}(\#(s_a(C_a))|\theta=a)}} \\ &+ \frac{1 - p_i(\theta = a|\#(s_a(C_a)), \#(s_b(C_a)))}{1 + \left(\frac{q_a^b}{1-q_a^a}\right)^{\mu_{C_b}(\#(s_b(C_a))|\theta=b)} \left(\frac{1-q_a^b}{q_a^a}\right)^{\mu_{C_b}(\#(s_a(C_a))|\theta=b)}} \\ &\approx 0.488 \cdot p_i(\theta = a|\#(s_a(C_a)), \#(s_b(C_a))) + 0.224 \end{aligned} \tag{43}$$

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. $\square$

Appendix B Additional details

Experimental setup and procedures. 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 4. 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 4). 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".

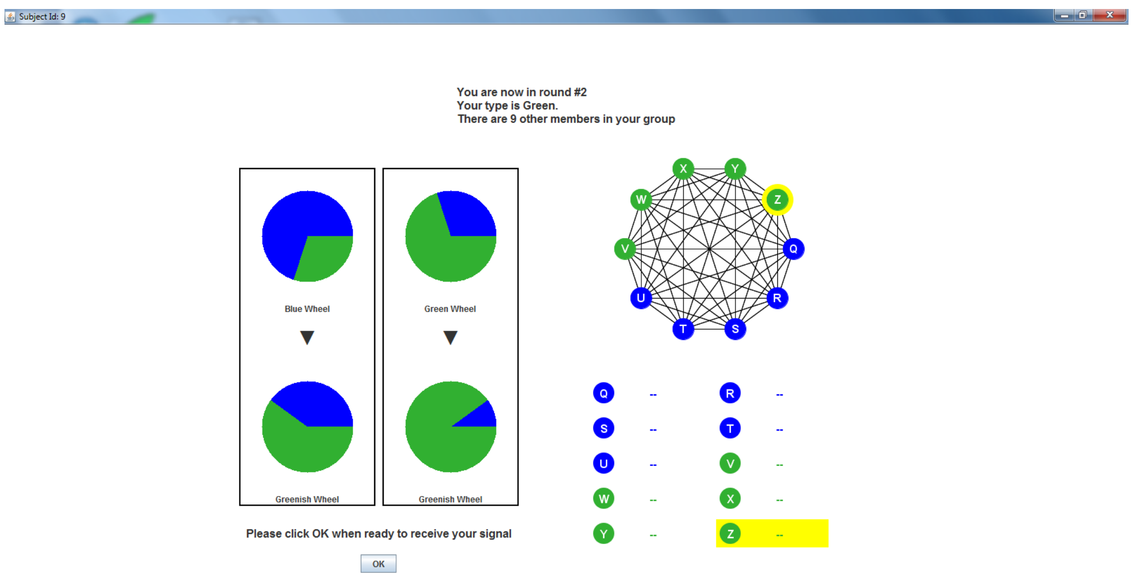


Figure 4: User Interface (MB1+SN2 Treatment, initial screen for a Green partisan).

Table 8: 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	20%	Complete	Polarized	30	11.50
17	10/18/16, 15:30	No	0%	Polarized	Complete	30	14.00
18	5/17/18, 09:00	Moderate	20%	None	Complete	20	12.13
19	5/17/18, 11:00	Moderate	–	Complete	None	30	11.50
20	5/17/18, 13:30	Moderate	–	None	Polarized	30	12.33
21	5/17/18, 15:30	Moderate	–	Polarized	None	20	11.50
22	5/21/18, 10:00	Moderate	–	Polarized	Complete	20	12.13
23	5/21/18, 13:00	Moderate	–	Complete	Polarized	20	12.75

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 1 and 4 we could not have varied the network treatment due to minor software issues with the polarized network display and decided to keep the complete network treatment throughout those sessions. In sessions 15 and 16 a typo in the parameter file produced incorrect signal accuracies for green partisans. We decided not to use the data from those sessions in the analysis and repeated the respective treatments in sessions 22 and 23.

Table 9: Summary of lab participant characteristics

Trait	N	Mean	Std. Dev.
Age, years	589	21.99	4.49
Male, %	590	39.32	49.58
English native, %	589	49.58	50.04
Previous experiments, #	589	6.32	8.68
Deception experiments, #	573	1.62	2.77
More quantitative, %	589	68.42	46.52
Media trust	589	4.72	1.77

Notes: We had 590 subjects total. One subject accidentally quit the software before finishing the post-treatment questionnaire, so for most characteristics $N = 589$. All characteristics are self-reported. "Deception experiments" is the average number of experiments involving deception (as perceived by the subject) 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 10: Group Success Rates By Treatment

Network	No media bias					
	N	[**All**]		N	[**Last 12**]	
Empty	176	.744	(.029)	132	.769	(.032)
Polarized	288	.816	(.020)	216	.806	(.023)
Complete	272	.862	(.020)	204	.873	(.023)
	Moderate Media Bias					
	N	[**All**]		N	[**Last 12**]	
Empty	160	.663	(.031)	120	.667	(.036)
Polarized	144	.767	(.030)	108	.773	(.035)
Complete	144	.809	(.031)	108	.815	(.035)
	Extreme Media Bias					
	N	[**All**]		N	[**Last 12**]	
Empty	112	.415	(.038)	84	.429	(.046)
Polarized	144	.493	(.035)	108	.482	(.041)
Complete	256	.502	(.030)	192	.477	(.034)

Notes: The group decision success is coded as 1 if the elected candidate matches the realized state, 1/2 if there is a tie, and 0 in all remaining cases, and the group success rate is computed by averaging group decision success across all group decisions in a given treatment. N is the number of group decision observations. Standard errors are in parentheses. Averages and standard errors based on all data are reported in columns labeled "All", those that exclude the first four rounds in "Last 12". No-bias non-empty network treatments with no signal probability $r = 0.2$ are pooled together with those in which $r = 0$.

Table 11: Regressing Bayesian Efficiency on Treatment Dummies and Controls

Constant		.726***
		(.034)
Network	Polarized	.053*
		(.031)
	Complete	.212***
		(.027)
Bias	Moderate	−.019
		(.030)
	Extreme	.177***
		(.026)
Network × Bias	Polarized × Moderate	.049
		(.035)
	Polarized × Extreme	−.115**
		(.046)
	Complete × Moderate	−.063*
		(.033)
	Complete × Extreme	−.395***
		(.035)
Round		✓
N		12,720
# clusters		42

Notes: Dependent variable: Bayesian efficiency. Linear regression with standard errors clustered at half-session (treatment level, more conservative) in parentheses. Network is relative to empty network, Bias is relative to no bias. Significance codes: *** < 0.01, ** < 0.05, * < 0.1

Table 12: Summary characteristics of the Gallup Pakistan survey respondents

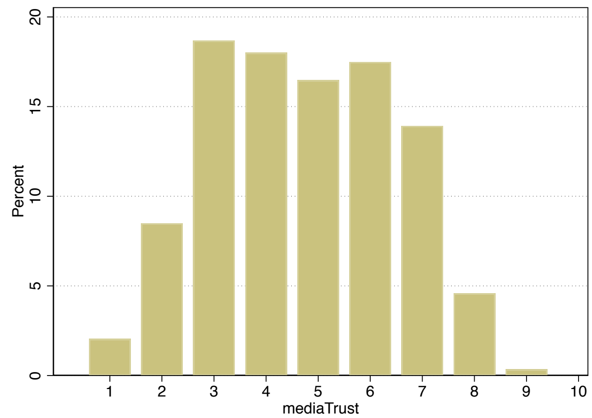
Variable	Label	Values	Percent
ur	Location type		
		Rural	69.99
		Urban	30.01
*d*1	Gender		
		Male	50.14
		Female	49.86
*d*2_1	Age (years)		
		< 29	34.00
		$30 - 50$	53.13
		$51 - 65$	12.76
		> 65	0.11
*d*3	Education		
		Illiterate	11.37
		Literate but no formal	9.76
		Up to Primary	14.37
		Middle school	16.53
		Matric	19.69
		Intermediate	10.65
		Graduate	13.98
		Postgraduate	2.66
		Professional/Doctor	0.33
		No response	0.67
*d*4_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
*d*5	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
*d*6	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 13: Responses to questions about social media

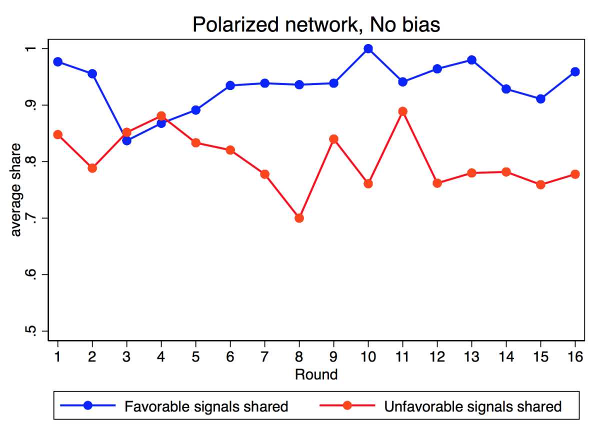
Question	Options	Percent
How often do you get the news about politics and government from social media (e.g., Facebook, Twitter, and the like)?	All the time	11.76
	Often	17.69
	Sometimes	18.64
	Hardly ever	12.20
	Never	39.71
Consider a hypothetical situation in which you have come across a news article 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 $N = 1,803$.

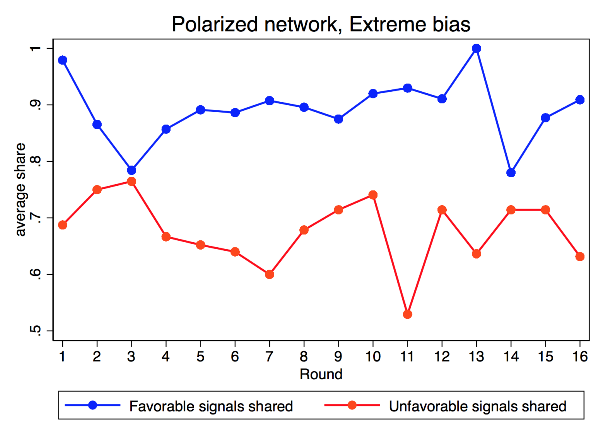


(a) How trustworthy is information from social media

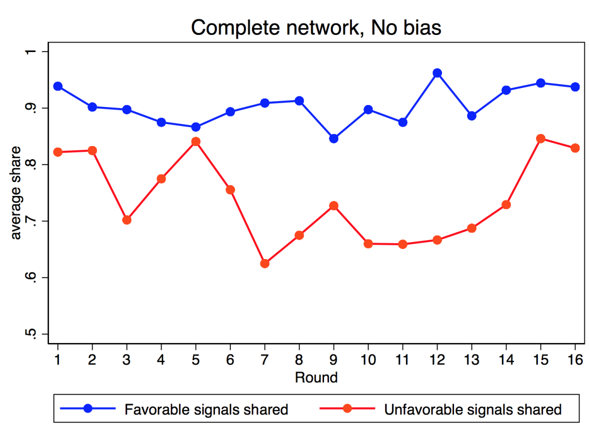
Figure 5: Lab subject responses ($N = 589$). Scale: 1=not at all trustworthy, 10=extremely trustworthy.



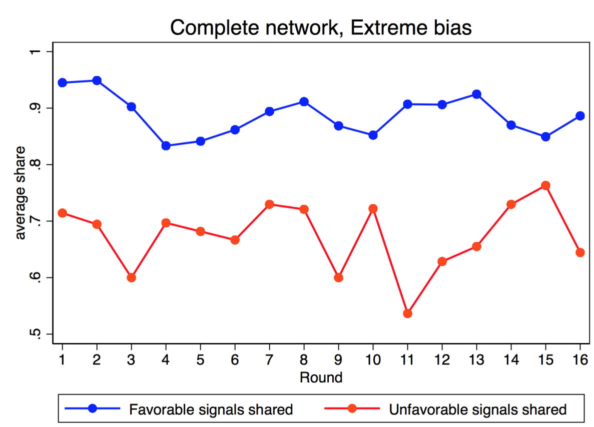
(a) SN1, MB0



(b) SN1, MB2

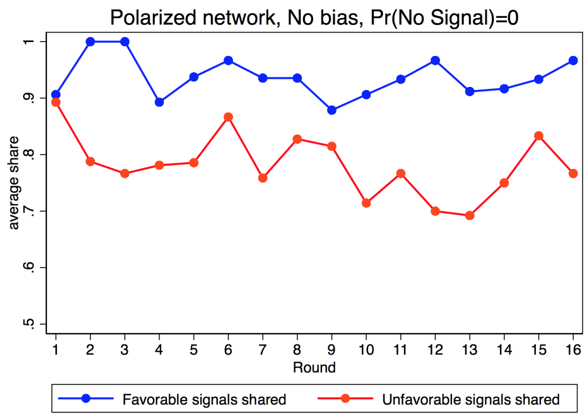


(c) SN2, MB0

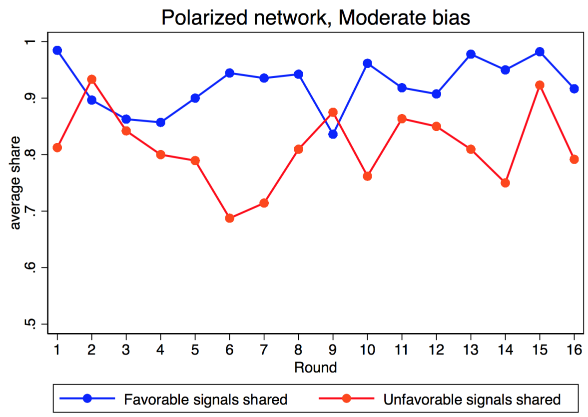


(d) SN2, MB2

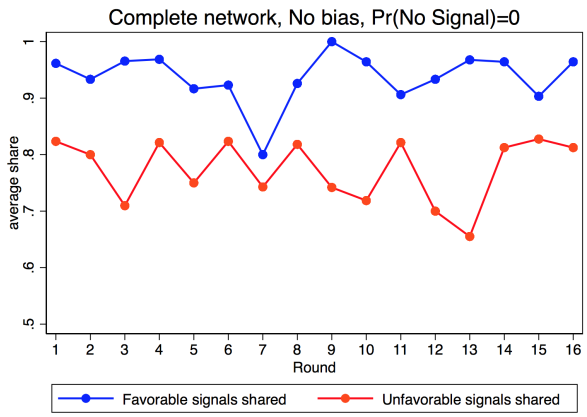
Figure 6: Average signal sharing rates across rounds: No bias and Extreme bias



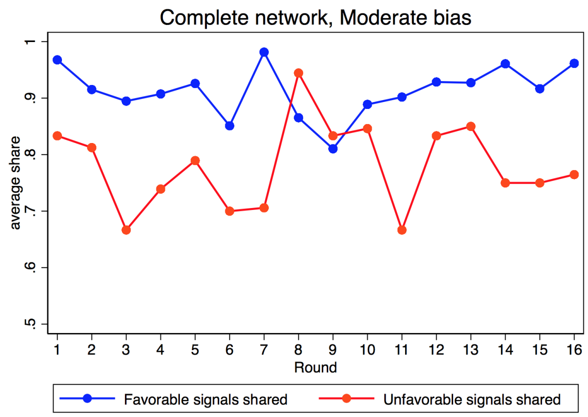
(a) SN1, MB0, r=0



(b) SN1, MB1



(c) SN2, MB0, r=0



(d) SN2, MB1

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

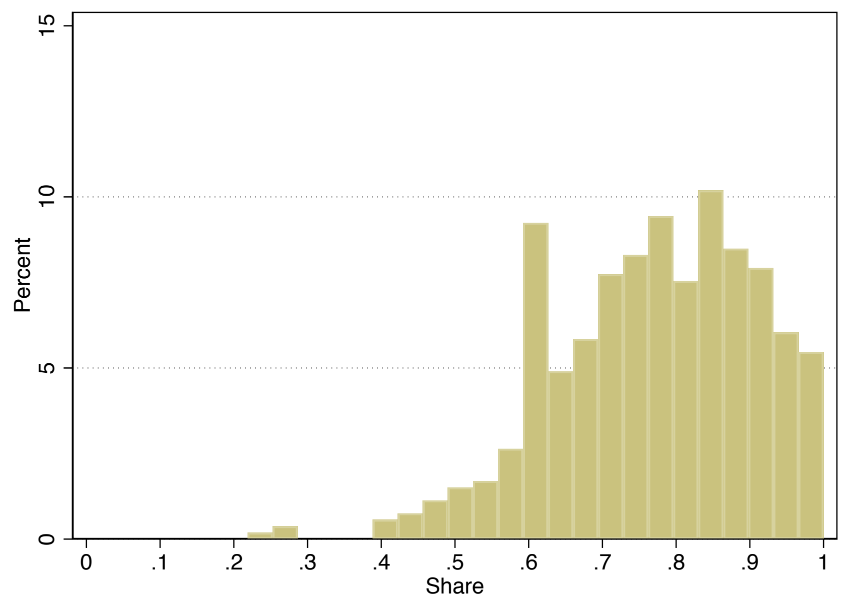


Figure 8: Are subjects consistent with sincere Bayesian voting? Using all votes and all treatments *including MB2*, 530 subjects. y-axis: percent of subjects; x-axis: share of each subject's votes (out of 32 decisions per subject) consistent with sincere Bayesian voting strategies.

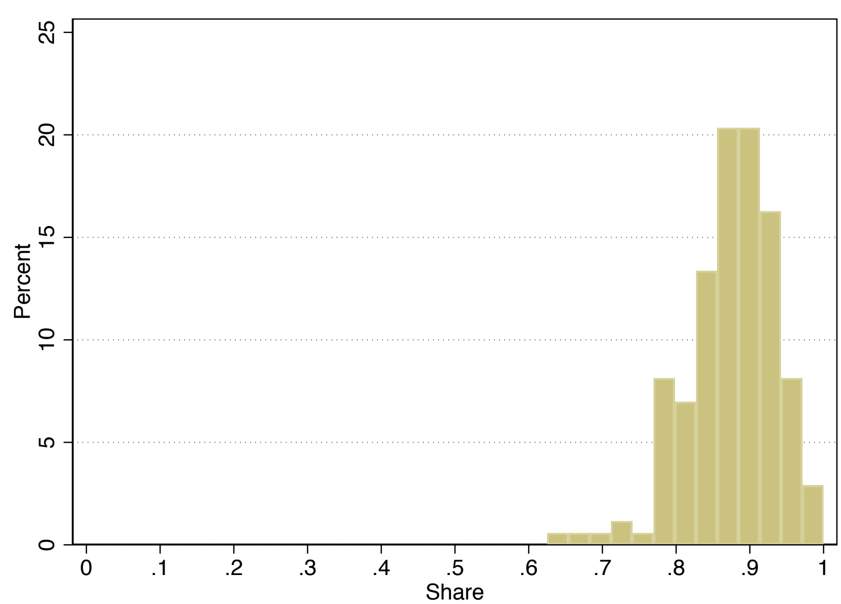


Figure 9: Distribution of average individual vote shares consistent with sincere Bayesian strategy, based on those subjects who had more than 60% correct decisions for votes against own favorite candidate, all treatments except MB2, 172 subjects.

Appendix C Instructions (MB0 or MB1, 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 on 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 [MB1: 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.

[MB1 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.