WARWICK

ECONOMICS

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

Kirill Pogorelskiy & Matthew Shum

(This paper also appears as CAGE Working paper 427)

June 2019 No: 1199

(Revised August 2019)

Warwick Economics Research Papers

ISSN 2059-4283 (online)

ISSN 0083-7350 (print)

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

Kirill Pogorelskiy
University of Warwick†

Matthew Shum
Caltech‡

August 13, 2019

Abstract

More voters than ever get political news from their friends on social media platforms. Is this bad for democracy? Using context-neutral laboratory experiments, we find that biased (mis)information shared on social networks affects the quality of collective decisions relatively more than does segregation by political preferences on social media. Two features of subject behavior underlie this finding: 1) they share news signals selectively, revealing signals favorable to their candidates more often than unfavorable signals; 2) they naïvely take signals at face value and account for neither the selection in the shared signals nor the differential informativeness of news signals across different sources.

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: August 13, 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, Brian Knight, 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, U Rochester, 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, Andis Sofianos, and Alessandro Castagnetti 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).

How does social media affect voting behavior? A key channel concerns news and information sharing via the underlying social networks. 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. This approach overcomes a key challenge faced by field studies: the unobservability of voters' information sets arising from exposure to news across platforms and media sources. We study a voting game in which partisan players have the opportunity to share their news signals with other players on a small social network prior to voting. In this setting, we examine how the combination of media bias in individual signals and the limited communication possibilities in social networks distorts news sharing and influences electoral outcomes. By opting for a context-neutral rather than context-heavy political environment, we expect our experimental results to provide conservative estimates of the treatment effects, as well as an *objective* efficiency benchmark for the quality of collective choices.

We study elections with two candidates and two states of the world. While voters share common interests, in the sense that their payoffs are maximized when the elected candidate is the "correct" choice for the realized state, they are also partisans, and the electorate is split up into two groups supporting different candidates ex ante. 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 basic setting by introducing *media bias* in voters' private signals and a novel *social network* component that restricts signal sharing across groups. Specifically, prior to voting, voters decide whether or not to share their private signals with their neighbors on a social network, mimicking the news sharing protocol on social media platforms. We investigate, theoretically and experimentally, two network treatments: polarized network, in which each candidate's ex ante supporters are put together in a separate network component with no links in between (a variant of a "filter bubble") and complete network, in which everyone is connected to everyone else, relative to the benchmark of the empty network, which does not permit news sharing. We also vary the degree of media bias, which determines the accuracy of individual signals: from unbiased informative signals to moderately biased but still informative signals, with partisans more likely to receive signals favoring their candidate, to extremely biased uninformative signals (a variant of

[1] The June 2015 Pew Research Center report shows that 61% of Americans aged 18–33 get political news from Facebook in a given week, compared to 39% of Baby Boomers (journalism.org/2015/06/01/millennials-political-news).

"fake news").

The theoretical analysis of our model casts some doubt on the "filter bubble" story and highlights the positive role of moderate media bias. The filter bubble narrative suggests that "piercing" the bubble should be necessarily improving collective decisions by enabling more informative voting. We show that with media bias, the overall picture is more complicated, and the above conclusion may not hold. If voters are sufficiently sophisticated Bayesians, they should be able to extract information from verifiable signals shared with them, even when the signals are biased by partisan media sources and subject to filtering on social networks. Network filtering is helpful since it makes voters' incentives aligned to share all signal realizations, including those unfavorable to their ex ante preferred candidate. Media bias is helpful since biased signal realizations can align voter behind the "correct" candidate despite differences in partisan preferences. For instance, an endorsement of a left-wing candidate by a right-wing newspaper can be viewed as a very informative signal. When there is no network filtering, extracting information is more challenging under media bias, as it requires accounting for the differential informativeness of the signals across news sources. So while increased news sharing opportunities in larger networks may result in more informative voting, for the benefits to realize voters need to be sophisticated enough to be able to account for the signal sources in their Bayesian posteriors.

Our empirical results indicate that by and large, voter behavior is *not* sophisticated enough to be consistent with the complicated Bayesian equilibrium calculus. Rather, voters exhibit two behavioral features. First, they consistently engage in "selective" news sharing in social network treatments, whereby they share news that favors their partisan candidate 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: their news sharing behavior is at least somewhat sophisticated. Second, in their voting behavior, subjects more or less take signals at face value. As a result, in treatments where signals are informative albeit biased, voters do not fully account for the bias. In treatments where, by design, signals are completely uninformative, voters naïvely attribute information value to them, treating uninformative signals ("fake news") as real news, and reacting to them in their voting decisions. In sum, subjects' news sharing behavior indicates little, if any, regard for how the signals are subsequently acted upon – they behave as if they derive intrinsic (dis-)utility from sharing (un-)favorable signals; and subjects' voting behavior indicates failure to discount biased signal sources even when the bias makes signals completely uninformative.

The under-sophistication on the part of voters results in novel efficiency trade-offs between filtering and bias in social networks. We 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, as the news shared by sufficiently many voters is sufficiently accurate about the realized state of the world even when shared selectively. The complete network is more efficient than the polarized one, which filters voters according to partisanship. As news quality deteriorates under media bias, however, the complete social network becomes less efficient while the polarized social network remains similarly efficient. So we find that filtering does *not* reduce

efficiency under media bias, but the information benefits of news sharing in the complete network are *not* realized. These results are consistent with unrestricted social networks imposing high cognitive costs on voters in the presence of media bias, and indicate that sharing bad information on social networks is a bigger threat to information aggregation by voting than are filter bubbles created by social media.

The remainder of the paper is organized as follows. Subsection 1.1 provides a brief overview of the related literature. 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 presents additional detail of the experimental setup and data analyses. Appendix B contains proofs of the theoretical propositions. Appendix C contains experimental instructions.

1.1 Related literature

Methodologically, our paper is related to several theoretical and experimental studies of information aggregation by voting discussed below, which share some similarities with some of our experimental treatments but differ in several important ways. First, almost all the studies we are aware of did not allow for media bias in the voters' signals, which plays an important role in assessing the effects of social networks on voting. Second, previous studies (with a few exceptions) largely considered voters with homogeneous payoffs i.e. with no conflict in the preferences. Third, previous studies either did not allow for network-restricted deliberation (as in our polarized network), or modelled it as cheap talk. As a result, pertinent features of news sharing on social networks were not fully captured by the extant literature. With these caveats in mind, the two most closely related papers are Goeree and Yariv (2011) and Jackson and Tan (2013).

The benchmark case of the empty network and no bias has been studied in many papers on Condorcet jury voting, starting with Austen-Smith and Banks (1996). One difference in our treatments is the possibility of "no-signal" outcomes, which creates an endogenous group of "uninformed" voters, which can matter for decision making efficiency (see Austen-Smith and Feddersen (2005), Oliveros and Várdy (2015), and Elbittar et al. (2017)).

Our treatments with non-empty networks feature network-restricted deliberation among voters, where voters' signals are verifiable ("hard" evidence). A version of our treatment with no bias and complete network was analyzed theoretically in Schulte (2010) and Coughlan (2000). These papers, as well as Hagenbach, Koessler and Perez-Richet (2014) only focussed on full information revelation equilibria, which do not exist in our setting, so we provide a characterization of all semi-pooling equilibria. Jackson and Tan (2013) is the only other paper that studied such equilibria in a comparable deliberation game with many senders and receivers and hard evidence, but in their model voters were distinct from "experts" who obtain signals. In our model, in contrast, both roles

are combined. Our complete social network treatment shares some features with other theoretical and experimental papers studying communication in a voting setting.[2] 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.[3]

More broadly, our paper is related to the strands of the empirical literature in media economics studying media bias and social networks. These two issues have generally been treated separately in the existing literature, and a key contribution of this study is to consider their *joint* effects.

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) and Martin and Yurukoglu (2017), whereby television viewers who were exposed to the right-leaning Fox News stations tended to vote more for Republican candidates. Similarly, Enikolopov, Petrova and Zhuravskaya (2011) found that viewers exposed to an independent TV channel in Russia decreased their support for the government. Measuring media bias using observational data is challenging (e.g. Chiang and Knight (2011), Groeling (2013)), and evaluating the overall effects requires careful accounting for news consumption across different platforms (e.g. Kennedy and Prat (2017)). Using experimental data allows us to avoid these challenges: in our experiments we fix the media bias exogenously and make the subjects aware of the news-generating information structure.[4]

Social networks 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. At the same time, newsfeed filtering by social networks may dilute the Fox News effect due to increased political polarization. Bakshy, Messing and Adamic (2015) found that Facebook users who indicate their ideology in their profiles tend to maintain relationships with ideologically similar friends, and that content-sharing is well-aligned with ideology. However Gentzkow and Shapiro (2011) found relatively little online segregation by ideology, and Boxell, Gentzkow and Shapiro (2017) questioned the hypothesis that social media is a primary driver of increasing polarization in the U.S. Our experimental findings lend support to this interpretation of news sharing on social media, as our subjects by and large take signals at face value and fail to take into account the sources of their signals in their voting behavior. This partly mitigates concerns about social networks increasing polarization, but also means that when there is media bias, complete networks do not improve over polarized ones despite enabling access to a larger number of signals.

[2]See in particular, Guarnaschelli, McKelvey and Palfrey (2000), Austen-Smith and Feddersen (2006), Gerardi and Yariv (2007), Meirowitz (2007), Goeree and Yariv (2011), Le Quement and Marcin (2016), Buechel and Mechtenberg (2016), Bouton, Llorente-Saguer and Malherbe (2017), and Breitmoser and Valasek (2018). See Martinelli and Palfrey (2017) for a comprehensive survey of experimental results in collective decision games.

[3]Several papers also studied heterogeneous voters partitioned into groups like in our polarized network treatment (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), Le Quement and Yokeeswaran (2015), and Battaglini (2017). Empirical papers on deliberation and voting include Iaryczower, Shi and Shum (2018).

[4]Media outlets in our paper are not modelled as strategic players but rather as a device delivering news signals to consumers. This contrasts with several theoretical studies where media bias arises *endogenously* (e.g. Mullainathan and Shleifer (2005), Gentzkow and Shapiro (2006), Bernhardt, Krasa and Polborn (2008), Duggan and Martinelli (2011), Anderson and McLaren (2012), and Piolatto and Schuett (2015)).

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. Selective news sharing has also been documented in the field (e.g. An, Quercia and Crowcroft (2014), Garz, Sörensen and Stone (2018)) and in the lab (e.g. in sender-receiver disclosure experiments, Jin, Luca and Martin (2017) report that senders disclose more favorable information more often), suggesting robustness of the finding across contexts. However, our results also show clearly that the negative effects of social media on collective outcomes 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.

2 The Model and Experimental Design

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

Preferences. We model voters as *weak partisans* who are biased toward a particular candidate but still care about choosing the correct candidate for the state, i.e. are responsive to information and can be persuaded to change their mind. 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 $s \neq s_\emptyset$ decides whether or not to share s with her neighbors on the network. She cannot selectively choose who to share the signal with: either the signal is shared with everyone in her network or with no one. She also cannot lie about her signal if she decides to share it (no cheap talk). After everyone's signal sharing choice is carried out, voters observe signals shared with them, if any.
4. Voters vote for one of the candidates C_a or C_b.[5]
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.[6] We study the following network treatments in our experiments:

SN0: Empty network. 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.

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.

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.

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

[6] Homophily in social networks has received a lot of attention since McPherson, Smith-Lovin and Cook (2001). Recent studies include, e.g. Golub and Jackson (2012), Baccara and Yariv (2013), Halberstam and Knight (2016).

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. The idea is to capture the setting in which a voter comes across a news article on the Internet, and decides whether or not to Facebook "like" it. If "liked", the article is publicly shared with her Facebook friends, however if not "liked", she can always plausibly pretend she has not been aware of that article in the first place (perhaps to avoid questions about why she did not share it).

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.[7] 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 share her signal would be "unraveled" 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.

[7] 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, Piovesan and Tyran (2019) who investigate effects of low-accuracy (but still unbiased) signals on voting efficiency using a very different experimental design.

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 $q = 0.7, r = 0.2, q_j^j = 0.9, q_j^{-j} = 0.4$.

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.[8]

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, and counter-balancing the order of the treatments. 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. With two to three ten-person electorates deciding simultaneously and independently each round, this design mitigates repeated play considerations. 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 A.

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 B. 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.

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

[8] All our analyses except the subject summary characteristics in Table 8 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 7 in Appendix A.

including her payoff type, her private signal, equilibrium messages received from others, and pivotality.[9] Then i prefers to have C_a elected if and only if this maximizes her expected utility:

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

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

$$t_a = 7/34 \approx 0.206, \qquad t_b = 1 - 7/34 \approx 0.794. \tag{3}$$

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.[10] With these preliminaries at hand, we characterize type-symmetric Bayesian equilibria for each combination of media bias and social network treatments.[11] To aid the reader, we have summarized the main empirical implications for each treatment combination in Table 2.

Equilibrium analysis (described in more detail below, with formal statements in Appendix B) reveals two main trade-offs between the network structure and informative voting. First, moderate media bias can improve informative voting compared to no bias. The reason is that in the presence of moderate bias, rational voters put different weights on favorable and unfavorable signals in their posteriors depending on the partisan bias of the signal source. As a result, the posteriors are updated asymmetrically: a favorable signal coming from the source biased towards own party does not move the posterior far from the prior, but an unfavorable signal coming from the same source moves it quite a lot from the prior. This improves incentives to vote in line with one's information compared to the no bias case. Second, having all voters grouped in a polarized network by party affiliation can be welfare-improving relative to a complete network. The reason is that in a complete network full information revelation fails – when the audience for messages has heterogeneous preferences, partisans have incentives to hide unfavorable signals – which is not the case in the homogeneous polarized network. Of course, it is an empirical question (explored in Section 4 below) whether the subjects are sophisticated enough to follow the equilibrium logic.

3.1 Equilibria without informative communication

We begin with the empty network treatments (third column in Table 2), where no communication is permitted among voters. Each voter can condition only on her own signal and the equilibrium

[9] 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.

[10] 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.

[11] There are other equilibria, e.g. those in which everyone always votes for the same candidate so 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.

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
		(Result 1)	(Result 2)	(Result 2)
Moderate (MB1)	News sharing	n/a	All	Selective
	Informative voting	Partial	Partial	Partial
		(Result 1)	(Result 2)	(Result 2)
Extreme (MB2)	News sharing	Irrelevant (anything goes)		
	Informative voting	No		
		(Result 3)		

Notes: All predictions are derived using experimental parameters. "Results" refer to statements in Subections 3.1 and 3.2. The corresponding formal statements are provided in Propositions 1 – 6 in Appendix B. '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.

play in determining how to vote.

With unbiased media, signals are symmetrically informative ex ante about the realized state: with parameters specified in Table 1, player i's posterior conditional on observing s_a is $p_i(\theta = a|s_a) = 1/(1+(1-q)/q) = 0.7$ and conditional on observing s_b it is $p_i(\theta = a|s_b) = 1/(1+q/(1-q)) = 0.3$. So both posteriors are equidistant from the prior of 1/2 and stay within the interval between belief thresholds, $[t_a, t_b]$, in which the unique best response for each voter is to vote for their ex ante preferred candidate. So an equilibrium obtains in which voters base their vote on their prior rather than on their signals.[12]

With moderately biased media, however, signals become *asymmetrically* informative ex ante about the realized state. For instance, for a C_a-partisan, the posterior about the "favorable" state $\theta = a$ conditional on observing a "favorable" signal s_a is $p_i(\theta = a|s_a) = 1/\left(1 + \left(1 - q_a^b\right)/q_a^a\right) = 0.6$, but conditional on an "unfavorable" signal s_b it is $p_i(\theta = a|s_b) = 1/\left(1 + q_a^b/\left(1 - q_a^a\right)\right) = 0.2 < t_a$. So an unfavorable signal provides *more* information than a favorable signal, as it moves the posterior farther from 1/2 and out of the interval between belief thresholds, $[t_a, t_b]$, making it a best response for each partisan to vote according to their signal, i.e. an unfavorable signal strongly indicates to this voter that her partisan candidate is the wrong choice for the state.[13] Accordingly, *moderate bias might be helpful for information aggregation* – sincere voting becomes informative, unlike in

[12] When all other players use sincere Bayesian voting strategies, a unique best response for a C_a-partisan is to vote for C_a after either signal, as $p_i(\theta = a|\cdot) \geq 0.3 > t_a$, and for a C_b-partisan to vote for C_b after either signal, as $p_i(\theta = a|\cdot) \leq 0.7 < t_b$. Conditioning on pivotality in addition to own signal does not alter best responses in our setting.

[13] When all other players use sincere Bayesian voting strategies, a best response for a C_j-partisan i is to vote for C_j after signal s_j, and vote for C_{-j} after signal s_{-j}, as now $p_i(\theta = j|s_{-j}) < t_j$. Hence 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).

the no bias case:

Result 1. *In the empty network treatments, equilibrium voting is at best partially informative. With unbiased media, all players vote for their partisan candidate regardless of their signal. With moderately biased media, players vote according to their signal (and vote for their partisan candidate if they get no signal) so voting becomes partially informative.*

This result summarizes the equilibrium predictions of Propositions 1 and 4 in Appendix B.

3.2 Equilibria with informative communication via social networks

Next, we consider social networks, in which voters have opportunities to exchange signals with others during the communication stage prior to voting. In our social network treatments (the last two columns in Table 2), voters decide whether or not to share their signals with all of their neighbors on the network, that is *all* other voters (under complete network) or *like-minded* partisans *only* (under polarized network).

Since signals are "hard" evidence, sharing informative signals on social networks should improve voting outcomes relative to the empty network. However, to properly interpret the extra signals received via social networks requires some sophistication on the part of voters: the Bayesian posterior depends not only on the reported signals but also on the non-reported ones: 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 fact, in the complete network, since signals are shared with everyone including the supporters of the other candidate, voters have incentives to hide signals that are unfavorable to their partisan candidate.[14] In the polarized network, voters can share signals only with members of their own party. In this case, they cannot directly affect beliefs in the other party – their signals can only be pivotal for the posteriors held in their own party – so that it becomes incentive compatible for voters to share even unfavorable signals, resulting in more informative voting.

Result 2. *In polarized networks, there is a fully revealing equilibrium in which all players with non-empty signals reveal them at the communication stage and believe with probability one that non-revealing players are uninformed. So both favorable and unfavorable signals are revealed with probability one. In contrast, in complete networks, there is no fully revealing equilibrium. Rather, there are semi-pooling equilibria in which players engage in selective news sharing, revealing favorable signals with probability one, but hiding unfavorable signals with high probability (> 0.92 under no media bias, > 0.86 under moderate media bias).*

This result summarizes the equilibrium predictions of Propositions 2 and 3 (no bias) and Propositions 5 and 6 (moderate bias) in Appendix B. For moderately biased media treatments, as discussed in Subsection 3.1, favorable and unfavorable signals enter the Bayesian posteriors with different weights that depend on signal sources. This creates important trade-offs between voter

[14] Similar logic arises in two-member committees with conflict in members' preferences (Li, Rosen and Suen, 2001).

sophistication and signal information *when media bias and social networks co-exist*: Complete networks may provide more information than polarized ones but require more voter sophistication (accounting for the *sources* of the reported and the strategically hidden signals). Polarized networks may provide less information than complete ones but also require less voter sophistication (all signals should be shared and come from the same source). In all cases, moderate media bias requires voters to treat favorable and unfavorable signals differently in their posteriors.

A different kind of voter sophistication is required in treatments with extreme media bias (the bottom row in Table 2). In those treatments, voters are more likely to receive signals favoring their partisan candidates but signals are uncorrelated with the state, no matter the source. The Bayesian posterior never moves away from the prior of 1/2 and in the unique sincere Bayesian equilibrium everyone votes (non-informatively) for their partisan candidate.

Result 3. *In the extreme bias treatments, signals are biased but completely uninformative. All players vote for their partisan candidate regardless of their own signal and/or any reported signals.*

Overall, Results 1 – 3 indicate that the joint effect of media bias and social networks on voting outcomes substantively depends on how sophisticated the majority of the voters are in sharing news, interpreting news received from others, and accounting for their information in their voting decisions. How these factors interact is an empirical question we turn to next.

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) voters may share news selectively, passing along favorable signals and suppressing unfavorable ones; (ii) voters behave in line with sincere Bayesian voting strategies, conditional on all available information and the equilibrium behavior. In addition, (iii) we compute decision-making efficiency for each treatment and check whether polarized networks may be more efficient than complete networks, and whether media bias may lead to more informative voting. We examine these issues in turn.

4.1 News-sharing patterns

We first look at the news-sharing patterns, focusing on the selective news sharing prediction from the semi-pooling equilibria (cf. Result 2), whereby voters should communicate favorable signals but suppress unfavorable ones. Subjects' signal-sharing activity is summarized in Table 3.

In the polarized network, all subjects in the same subnetwork have fully aligned interests, so in theory, there should be full signal revelation (Propositions 3 and 6 in Appendix B). Table 3 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" and "Unfav. signals" columns of Table 3 that *subjects are selective in the signals they share*: in *all* treatments,

Table 3: Signal Sharing Patterns by Treatment

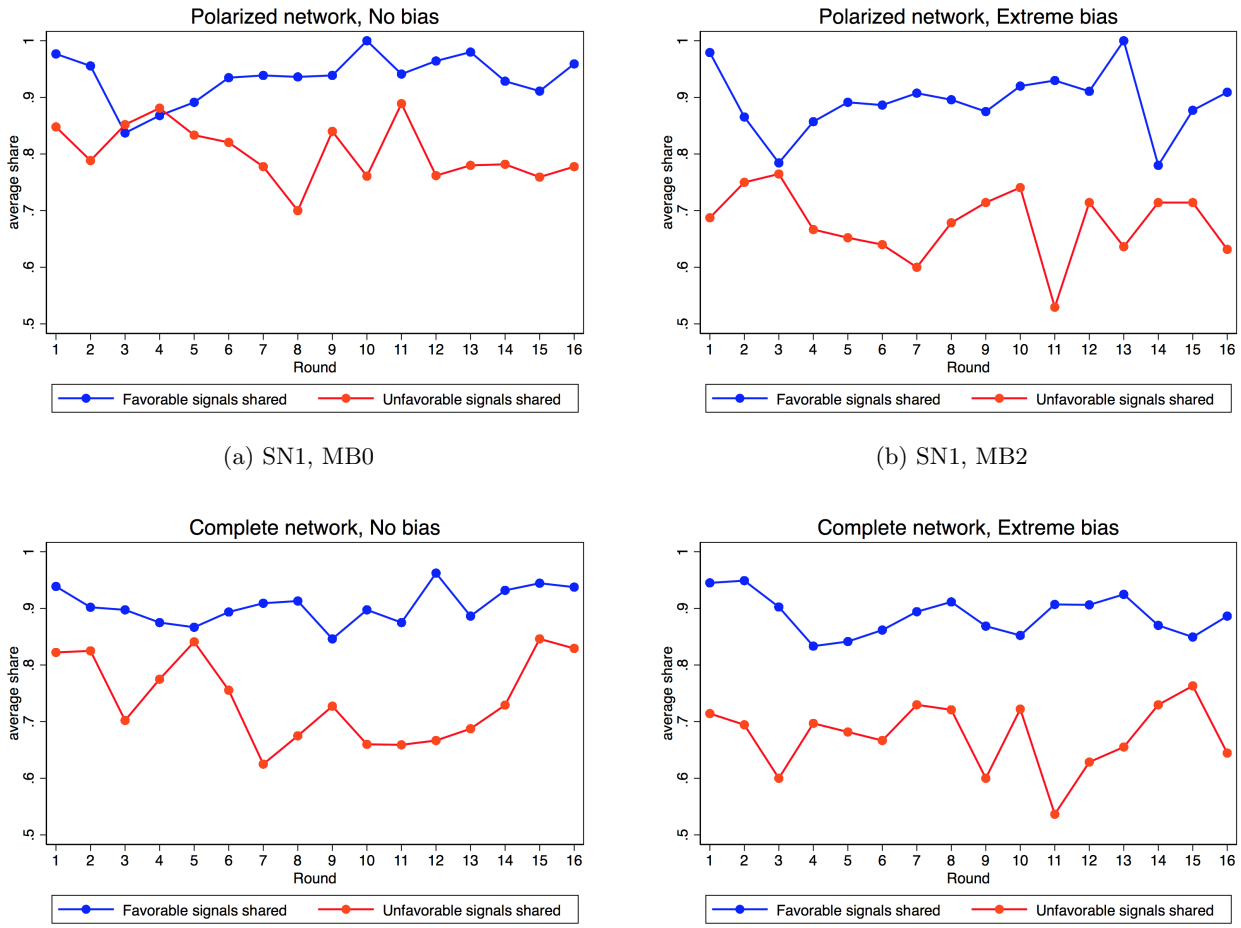
Network	Pr(No sig.)	Bias	N	All signals		N	Fav. signals		N	Unfav. signals		Fav − Unfav
Polarized	20%	No	120	.869	(.019)	120	.936	(.017)	120	.803	(.029)	.133***
Complete	–	–	110	.825	(.021)	110	.906	(.021)	110	.738	(.032)	.168***
Polarized	0%	–	60	.861	(.026)	60	.936	(.023)	60	.782	(.042)	.154***
Complete	–	–	60	.854	(.025)	60	.938	(.026)	60	.774	(.042)	.164***
Polarized	20%	Moderate	90	.897	(.015)	90	.923	(.016)	87	.814	(.031)	.109***
Complete	–	–	90	.881	(.016)	90	.913	(.016)	88	.780	(.034)	.133***
Polarized	–	Extreme	90	.827	(.021)	90	.891	(.019)	89	.678	(.039)	.213***
Complete	–	–	120	.825	(.016)	120	.888	(.018)	119	.673	(.034)	.215***

Notes: N = number of subjects, standard errors in parentheses are clustered by subject. All rates are conditional on getting a non-empty signal. Significance codes: *** < 0.01, ** < 0.05, * < 0.1

favorable signals are shared 89–94% of the time but unfavorable signals are shared only 68–81% of the time. These differences (in the last column) are significant (in all cases, $p \leq 0.002$) and robust to subsamples, e.g. using only the first halves of all sessions. The dynamics of signal sharing across experimental rounds (within network treatment condition) are illustrated in Figures 1–2. Except for the first few initial rounds (expiring around round 4), the average sharing rates do not vary much over time, thus providing no evidence of learning. Unfavorable signals in treatments with media bias are naturally less likely to occur so sharing rates of those exhibit a higher variance over time than the sharing rates of favorable signals.

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 (Result 2). However, the sharing rates of unfavorable signals – 67.8% or higher in all treatments – far outstrip the maximum of about 8% (about 14% under the moderate bias) predicted in the semi-pooling equilibria.

Could selective news sharing be a best response by sophisticated senders to "naïve" receivers who take reported signals at face value? This possibility seems strained: since each sender is also a receiver, a sophisticated sender should not, at the same time, be a naïve receiver. As an indirect test of sender sophistication, however, in rows 3–4 of Table 3 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 cannot hide under the guise of being 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% (see also Figures 2(a) and 2(c)). This suggests that senders are not sophisticated enough as to be guided by backward-induction reasoning about how their (non-)reports will affect others' posterior beliefs and votes, as in the semi-pooling equilibrium. Rather, signals are shared (or not) with seemingly little regard for how the signals will be interpreted or acted upon, and there appears to be a behavioral disinclination towards sharing unfavorable signals.



(a) SN1, MB0

(b) SN1, MB2

(c) SN2, MB0

(d) SN2, MB2

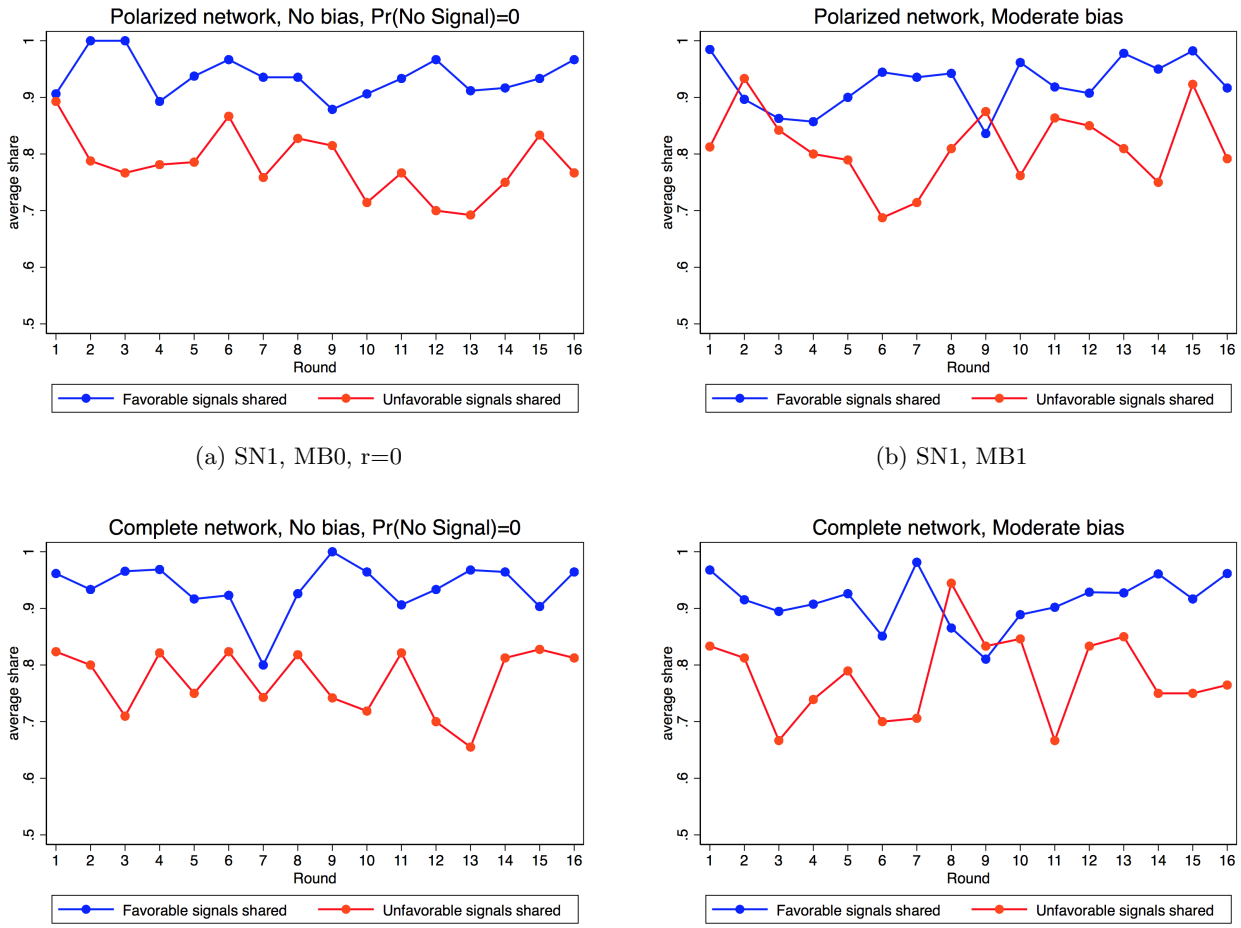
Figure 1: Average signal sharing rates across rounds: No bias and Extreme bias.

4.2 Assessing Bayesian voting

In this section we consider more formally whether subjects account for their information in their vote, in a way that is consistent with the theoretical predictions.

Sincere Bayesian voting by subject. To estimate how often subjects followed sincere Bayesian voting strategies, we start by computing, for each subject, the fraction of her votes that is consistent with the sincere Bayesian strategy. This histogram of this measure, across all subjects, is presented in Figure 3(a). Here, we see that a substantial proportion of subjects use voting strategies that seem consistent with sincere Bayesian voting 60% of the time or more.

This rather rosy picture, however, is incomplete as illustrated in Figure 3(b), which plots the histogram only for the subsample in which the sincere Bayesian strategy counsels voting *against* a subject's partisan candidate (i.e. each point in this histogram tabulates the frequency that a subject voted in line with her information when it contradicts her partisan bias.) A comparison of Figures 3(a) and 3(b) suggests that the high consistency in Figure 3(a) is somewhat spurious, as it



(a) SN1, MB0, r=0

(b) SN1, MB1

(c) SN2, MB0, r=0

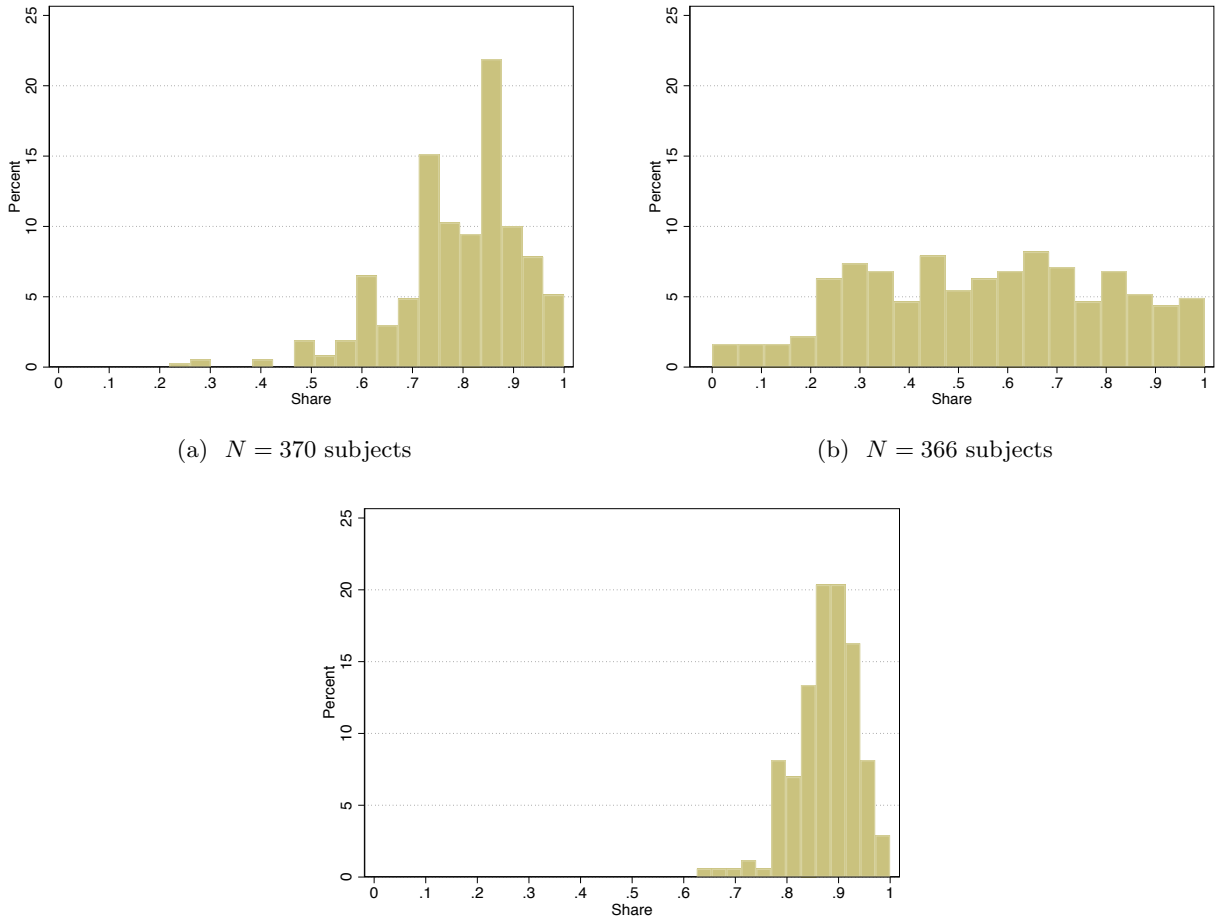
(d) SN2, MB1

Figure 2: Average signal sharing rates across rounds: No bias with $r = 0$ and Moderate bias.

arises when the Bayesian strategy coincides with voting for one's partisan candidate.[15] 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. Moreover, it is precisely those subjects who score highly in panel (b) who also score highly in panel (a), as illustrated in Figure 3(c), which depicts consistency with sincere Bayesian strategies only for those subjects who score more than 60% correct in Figure 3(b).

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 their theoretical posterior (as in Figure

[15]Figure 3(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 3(a). Figure 8 in Appendix A has the extreme bias treatments included.



(a) $N = 370$ subjects

(b) $N = 366$ subjects

(c) $N = 172$ subjects

Figure 3: Do subjects vote consistently with sincere Bayesian voting? Panel (a): using all votes; Panel (b): using only votes against partisan candidate; Panel (c): using all votes by only those subjects who scored more than 60% in Panel (b). 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.

3(a)). We also investigated consistency with the sincere Bayesian strategy using the subject's empirical posterior estimated from the data.[16] In Table 4, we report the frequency of subjects' votes which are consistent with sincere Bayesian voting, using both the theoretical (columns 4–5) and the empirical (columns 6–7) posteriors. Thus columns 4–5 of Table 4 display individual subject Bayesian consistency scores from Figure 3(a) averaged across each treatment combination. In addition, Table 4 displays consistency with extreme bias treatments (not included in Figure 3), for which the sincere strategy prescribes always voting for the default partisan choice.

[16] 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 4: 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 = number of subjects, std 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. 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 actually vote *against* their partisan candidate with non-negligible probability, i.e. they are influenced by the signals from others' in their network. 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 (while ignoring signal sources and empty signals in the empirical posterior), which we investigate more below.

Do voters account for the sources of the signals? Given limited consistency with sincere Bayesian voting, we next examine how the signals received by voters affect their voting decisions. We estimate a binary logit model of individual vote where the dependent variable is an indicator for whether a subject votes for her partisan candidate. 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 5.

Table 5 contains 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

Table 5: 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 = number of individual decisions (first round observations are dropped). The numbers represent average marginal effects on Prob (Vote for Partisan Candidate).

or unfavorable) revealed by co-partisans vs. anti-partisans.[17] A rational Bayesian would recognize that signals which are not aligned with partisanship – favorable signals from anti-partisans, and unfavorable signals from co-partisans – are *very* informative (under moderate media bias), and should therefore carry a larger weight in the voting decision. At the same time, under extreme bias, all signals are uninformative, and should not affect voting decisions. However, the estimated coefficients indicate that subjects take all signals more or less at face value; their voting decisions reflect signal labels rather than signal informativeness. Complete social networks in the presence of media bias make it very hard for the voters to correctly interpret their information.

4.3 Collective decision-making efficiency

The individual voter behavior discussed and analyzed above underlies the overall collective decision efficiency across the various treatments, which we turn to now. For assessing efficiency, we use a Bayesian efficiency measure, which conditions on the realized signals in the experiments.[18] Each round, in each subgroup we check if the majority decision within the 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). This way the measure accounts for signal sharing constraints in the polarized network treatments, enabling a fair efficiency comparison between polarized and complete networks; otherwise, polarized networks,

[17] For revealed favorable signals, Wald $p = .828$ (no bias), $p = .674$ (moderate bias), and $p = .879$ (extreme bias); for revealed unfavorable signals, Wald $p = .913$ (no bias), $p = .788$ (moderate bias), and $p = .341$ (extreme bias).

[18] This measure differs from *group success rates* – how often the elected candidate matches the realized state – which 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. Group success rates are reported in Table 9 in Appendix A and results are qualitatively similar.

which prevent signals from being shared across subgroups, would mechanically yield lower efficiency relative to complete networks, in which signals can be shared with all voters.[19] This binary measure, averaged across all subgroups and rounds for each treatment, is reported in Table 6. In the discussion, we focus on the more conservative statistics reported in the "Last 12" columns of Table 6, which exclude the first four rounds, to look at more experienced decisions.

Table 6: 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)
Polarized − Complete		−.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)
Polarized − Complete		.004	(.026)		.019	(.031)
Complete w/bias − Complete w/o bias		−.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)
Polarized − Complete		.120***	(.029)		.121***	(.032)
Complete w/bias − Complete w/o bias		−.223***	(.020)		−.217***	(.022)

Notes: N = 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 $r = 0$. "Polarized − Complete" is the difference in efficiency between the polarized and complete network, for each bias treatment (a measure of the "filter bubble" effect). "Complete w/bias − Complete w/o bias" is the difference in efficiency between the complete network under the respective bias treatment and under no bias (a measure of the "media bias" effect). Significance codes: *** < 0.01, ** < 0.05, * < 0.1

When signals are informative, as in the "No media bias" and "Moderate media bias" treatments in the top two panels of Table 6, social networks improve efficiency relative to the empty network, 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 network, and by 21.1 p. p. ($p = 0.000$) in the complete network. However, the benefits of complete networks diminish under moderate media bias: despite allowing access to a larger number of signals, the complete network is no longer more efficient than the polarized one. As discussed in Subsection 4.2, this is because under media bias, the extra information revealed in the complete networks requires some voter sophistication to interpret correctly, yet voters take signals at face value and fail to account for the sources. Overall, the Bayesian efficiency patterns largely follow the average consistency with sincere Bayesian voting reported in Table 4.

[19] Of course, using the standard efficiency measure is perfectly fine to compare the empty network with a non-empty one. All of our efficiency comparisons are robust to the modified efficiency measure, as Table 10 in Appendix A demonstrates.

These numbers also allow us to compare the effects of segregation on social networks with those arising from media bias. The efficiency comparison between polarized and complete networks (fixing media bias) estimates the effect of restricted news sharing on voting (a "filter bubble" effect). We find that while these 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 extreme bias (84.26% vs. 72.14%, $p = 0.000$). Importantly, the *efficiency of polarized networks does not vary much across treatments* so these results are largely driven by changes in the efficiency of complete networks.

On the other hand, the *media bias effect* can be estimated by the efficiency comparison between complete networks with media bias and complete networks without bias. These effects are always negative (the complete network has the highest efficiency without media bias), and affect the efficiency relatively more than does partisan-restricted news sharing imposed by the polarized network: While the network restrictions reduce efficiency by *at most* −6.1 p.p., the media bias reduces efficiency by *at least* −8.2 p.p. (under moderate bias), and the negative "fake" news effect (under extreme bias) reaches −21.7 p.p. These results are consistent with information benefits of complete networks being offset by additional cognitive costs they impose on voters in treatments with media bias, where signals should not be taken at face value.

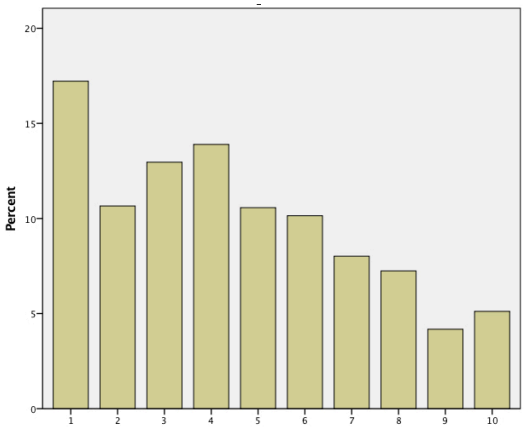
The "Extreme media bias" results, in the bottom panel of Table 6, strengthen these interpretations. Under extreme media bias, signals are uninformative, so a Bayesian decision-maker would discard all signals and simply vote for the partisan candidate in each subgroup. Hence the efficiency measure here 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.[20]

In summary, news sharing via social networks per se, does not appear to harmfully impact collective decisions despite pronounced lack of voter sophistication; rather the culprit is media bias. When media bias is present in social networks, more voter sophistication is required to extract the information from the shared news. When signals are informative, social networks allow for better information aggregation, leading to more efficient collective decisions compared to the empty network.

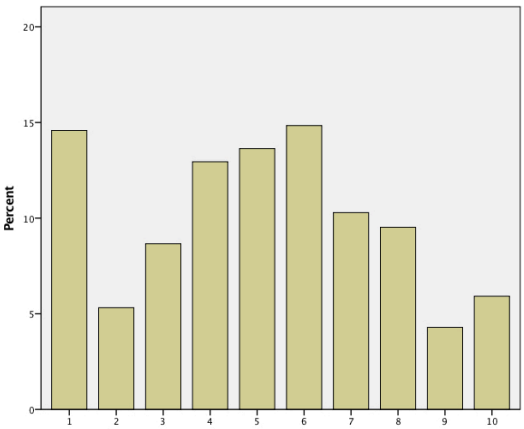
[20] In Table 10 in Appendix A we report the linear regression results (clustering standard errors at the treatment level) that confirm our efficiency findings and mitigate concerns about multiple hypothesis testing.

5 Corroborating Evidence from the Field: a Pakistani Survey

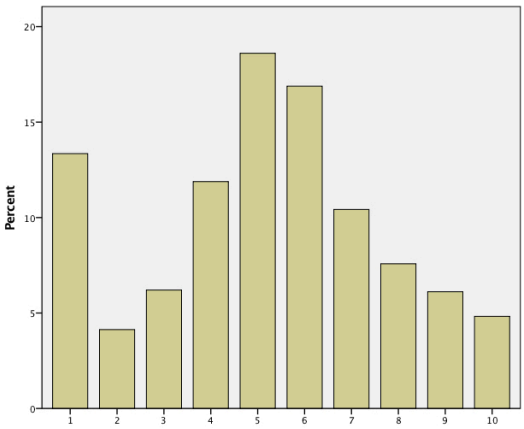
Our experimental treatments focused on stylized and extreme cases of social networks (empty, polarized, and complete) and media bias in voters' private signals (no bias, moderate, and extreme). While we expect our results to provide lower bounds on voting behavior in a more realistic setting, a host of additional factors could be relevant for social media studies in the field. 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 that we collected via 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 beliefs in the presence of unfavorable information.



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



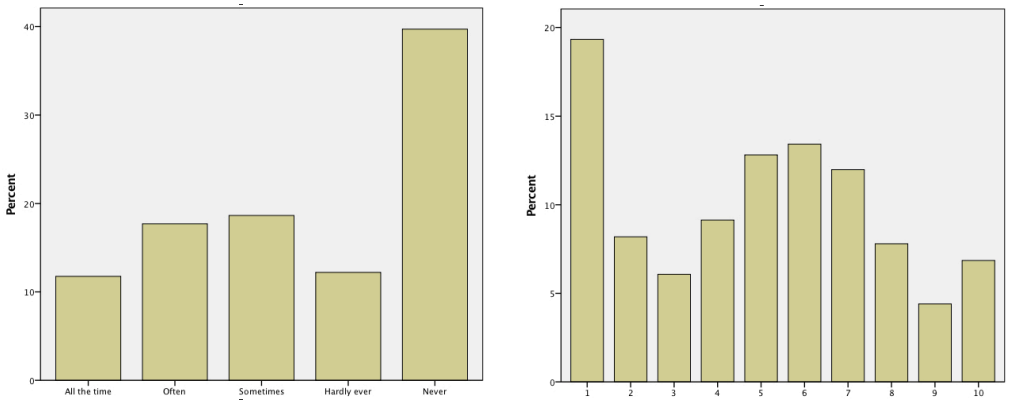
(b) How likely to share "favorable" news (hospital); $N = 1,166$.



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

Figure 4: 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.

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



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

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.

Figures 4–5 present the survey responses graphically.[21] Comparing Figures 4(a) and 4(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 "unfavorable" news about the favorite candidate is "likely" to be shared by about 34.7% of social media users. In contrast, "favorable" news about the favorite candidate is "likely" to be shared by about 44.9% of social media users. Furthermore, about 45.8% of social media users are "likely" to revise their opinion after a Facebook friend shared a "unfavorable" news article. Both of these findings are consistent with the experimental results described earlier.

Almost half (48.1%) of respondents regularly get the news about politics and government from social media (see Figure 5(a)). These numbers look strikingly similar to the figures for American voters from the September 2017 Pew Research Center report on news use across social media platforms.[22] According to Figure 5(b), in the full sample including those who do not use social media, about 44.5% view the information from social media as "trustworthy", and the average media trustworthiness on the 10-point scale is 4.89, very similar to the average media trust reported in the experiments (4.72). The histogram of trustworthiness features a peak at the lower tail, which consists primarily of those subjects who don't use social media (answered "Never" in Figure 5(a)).

[21] The exact survey questions along with response frequencies are available in Table 12 in Appendix A. The questionnaire was administered in an Urdu translation. The main demographic characteristics of the survey respondents are in Table 11 in Appendix A.

[22] See journalism.org/2017/09/07/news-use-across-social-media-platforms-2017.

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 joint effects of social networks and media bias on voter behavior.

Overall, our results show that the quality of information shared via social networks affects collective decisions relatively more than do the partisan-based information sharing restrictions. While such restrictions limit the amount of information available to voters – reducing efficiency when there is no media bias – the unrestricted networks impose high cognitive costs on voters whenever media bias is present. Consequently, voters ignore signal sources and take their information largely at face value. In line with majority of voters having limited sophistication, they display a behavioral bias towards selective news sharing, and publicly share signals favorable to their party more often than signals unfavorable to their party in all of our network treatments. These findings suggest that improving the quality of information shared on social networks is of first order policy importance.

Our results that voters selectively share signals favorable to their party are different from findings in the "confirmation bias" and "information avoidance" literatures[23] – 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 with you the news you don't like, you may unfriend me on Facebook) or selective recall (e.g. see a recent study by Zimmermann (2019)). In ongoing work, we are exploring how to enrich our experimental setting to allow for these types of effects.

References

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

An, Jisun, Daniele Quercia, and Jon Crowcroft. 2014. "Partisan Sharing: Facebook Evidence and Societal Consequences." *COSN '14*, 13–24. New York, NY, USA:ACM.

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

Austen-Smith, David, and Timothy J Feddersen. 2005. "Deliberation and Voting Rules." In *Social Choice and Strategic Decisions*. 269–316. Springer.

[23] See Golman, Hagmann and Loewenstein (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.

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

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

Bakshy, Eytan, Solomon Messing, and Lada A Adamic. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science*, 348(6239): 1130–1132.

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

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

Bouton, Laurent, Aniol Llorente-Saguer, and Frédéric Malherbe. 2017. "Unanimous rules in the laboratory." *Games and Economic Behavior*, 102: 179–198.

Boxell, Levi, Matthew Gentzkow, and Jesse M Shapiro. 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, Yves, and Justin Valasek. 2018. "A Rationale for Unanimity in Committees." Working paper.

Buechel, Berno, and Lydia Mechtenberg. 2016. "The Swing Voter's Curse in Social Networks." Working paper.

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

Chiang, Chun-Fang, and Brian Knight. 2011. "Media Bias and Influence: Evidence from Newspaper Endorsements." *Review of Economic Studies*, 78(3): 795–820.

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

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

Duggan, John, and Cesar Martinelli. 2011. "A Spatial Theory of Media Slant and Voter Choice." *Review of Economic Studies*, 78: 640–666.

Elbittar, Alexander, Andrei Gomberg, César Martinelli, and Thomas R Palfrey. 2017. "Ignorance and bias in collective decisions." *Forthcoming, Journal of Economic Behavior & Organization.*

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and political persuasion: Evidence from Russia." *American Economic Review*, 101(7): 3253–3285.

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

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

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

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

Garz, Marcel, Jil Sörensen, and Daniel F Stone. 2018. "Congeniality and News Demand: Evidence from Facebook." Working paper.

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

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

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

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

Golman, Russel, David Hagmann, and George Loewenstein. 2017. "Information Avoidance." *Journal of Economic Literature*, 55(1): 96–135.

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

Groeling, Tim. 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, Serena, Richard D McKelvey, and Thomas R Palfrey. 2000. "An Experimental Study of Jury Decision Rules." *American Political Science Review*, 94(2): 407–423.

Hagenbach, Jeanne, Frédéric Koessler, and Eduardo Perez-Richet. 2014. "Certifiable Pre-play Communication: Full Disclosure." *Econometrica*, 82(3): 1093–1131.

Halberstam, Yosh, and Brian Knight. 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, Matias, Xiaoxia Shi, and Matthew Shum. 2018. "Can Words Get in the Way? The Effect of Deliberation in Collective Decision-Making." *Journal of Political Economy*, 126(2): 688–734.

Jackson, Matthew O, and Xu Tan. 2013. "Deliberation, disclosure of information, and voting." *Journal of Economic Theory*, 148: 2–30.

Jin, Ginger Zhe, Michael Luca, and Daniel Martin. 2017. "Is no News (Perceived as) Bad News? An Experimental Investigation of Information Disclosure." Working paper.

Kawamura, Kohei, and Vasileios Vlaseros. 2017. "Expert information and majority decisions." *Journal of Pubic Economics*, 147: 77–88.

Kennedy, Patrick, and Andrea Prat. 2017. "Where Do People Get Their News?" Working paper.

Le Quement, Mark Thordal, and Isabel Marcin. 2016. "Communication and voting in heterogeneous committees: An experimental study." Working paper.

Le Quement, Mark Thordal, and Venuga Yokeeswaran. 2015. "Subgroup deliberation and voting." *Social Choice and Welfare*, 155–186.

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

Li, Hao, Sherwin Rosen, and Wing Suen. 2001. "Conflicts and Common Interests in Committees." *American Economic Review*, 91(5): 1478–1497.

Martinelli, César, and Thomas R Palfrey. 2017. "Communication and Information in Games of Collective Decision: A Survey of Experimental Results." Working paper.

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

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

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology*, 27: 415–444.

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

Mengel, Friederike, and Javier Rivas. 2017. "Common value elections with private information and informative priors: Theory and experiments." *Games and Economic Behavior*, 104: 190–221.

Morton, Rebecca B, Marco Piovesan, and Jean-Robert Tyran. 2019. "The dark side of the vote: Biased voters, social information, and information aggregation through majority voting." *Games and Economic Behavior*, 113: 461–481.

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

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

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

Piolatto, Amedeo, and Florian Schuett. 2015. "Media competition and electoral politics." *Journal of Public Economics*, 130: 80–93.

Schulte, Elisabeth. 2010. "Information aggregation and preference heterogeneity in committees." *Theory and Decision*, 69: 97–118.

Zimmermann, Florian. 2019. "The Dynamics of Motivated Beliefs." *Forthcoming, American Economic Review.*

For Online Publication

Appendix A Additional details and results

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 6. 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 6). 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". As specified in the experimental instructions in Appendix C, we intended to keep subject "party"

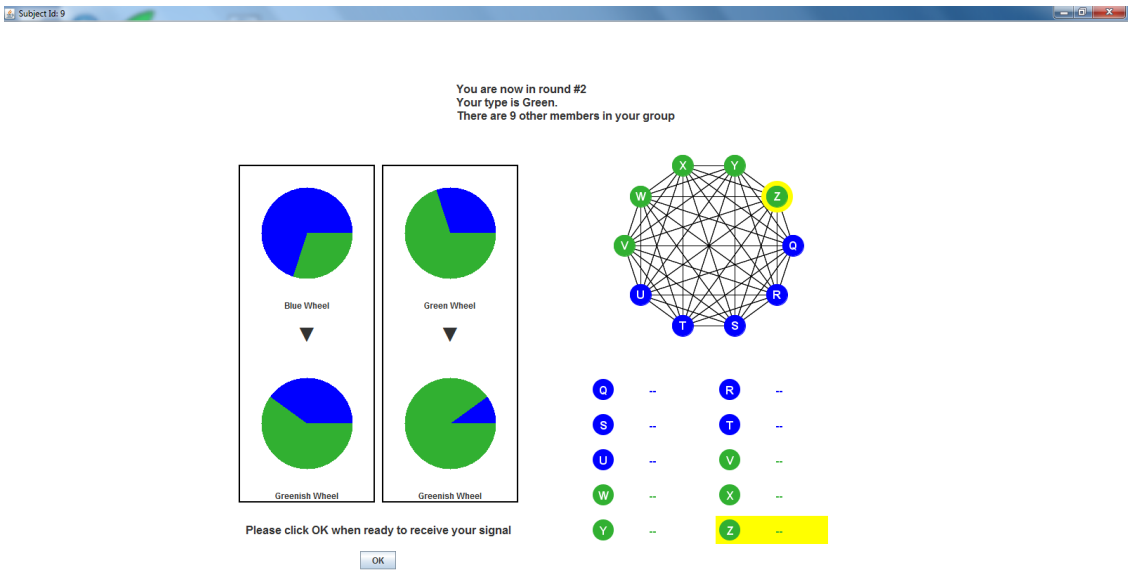


Figure 6: User Interface (MB1+SN2 Treatment, initial screen for a Green partisan).

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 as Figure 6 illustrates, we do not consider this deviation from our original design an issue for our analyses.

Table 7: 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 8: 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 9: 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 = 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 10: Regressing Bayesian Efficiency on Treatment Dummies and Controls

		(1)		(2)	
Constant		.726***	(.034)	.723***	(.033)
Network	Polarized	.053*	(.031)	.150***	(.030)
	Complete	.212***	(.027)	.212***	(.027)
Bias	Moderate	−.019	(.030)	−.019	(.030)
	Extreme	.178***	(.026)	.178***	(.026)
Network × Bias	Polarized × Moderate	.049	(.035)	.017	(.036)
	Polarized × Extreme	−.115**	(.046)	−.212***	(.046)
	Complete × Moderate	−.063*	(.033)	−.063*	(.033)
	Complete × Extreme	−.395***	(.035)	−.395***	(.035)
Round		✓		✓	
N		12,720		12,720	
# clusters		42		42	

Notes: OLS with standard errors clustered at half-session (treatment level, more conservative) in parentheses, using the last 12 rounds. Dependent variable: (1) Bayesian efficiency based on signals realized in all groups, (2) Bayesian efficiency based on signals realized in subgroups only (for polarized network treatments) and in all groups (for all other treatments). Network is relative to empty network, Bias is relative to no bias. Significance codes: *** < 0.01, ** < 0.05, * < 0.1

Table 11: Summary characteristics of the Gallup Pakistan survey respondents

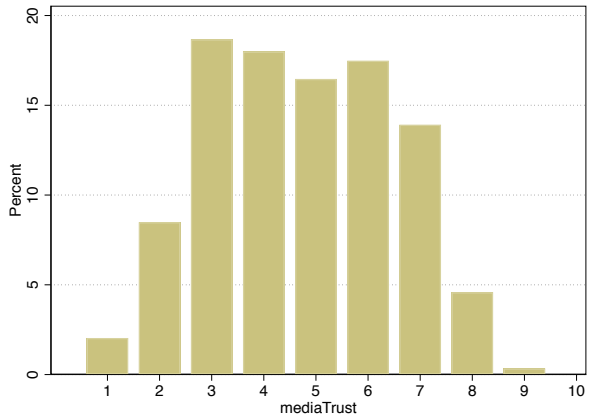
Variable	Label	Values	Percent
ur	Location type		
		Rural	69.99
		Urban	30.01
*d*1	Gender		
		Male	50.14
		Female	49.86
*d*2_1	Age (years)		
		< 29	34.00
		$30 - 50$	53.13
		$51 - 65$	12.76
		> 65	0.11
*d*3	Education		
		Illiterate	11.37
		Literate but no formal	9.76
		Up to Primary	14.37
		Middle school	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 12: 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 7: Lab subject responses ($N = 589$). Scale: 1=not at all trustworthy, 10=extremely trustworthy.

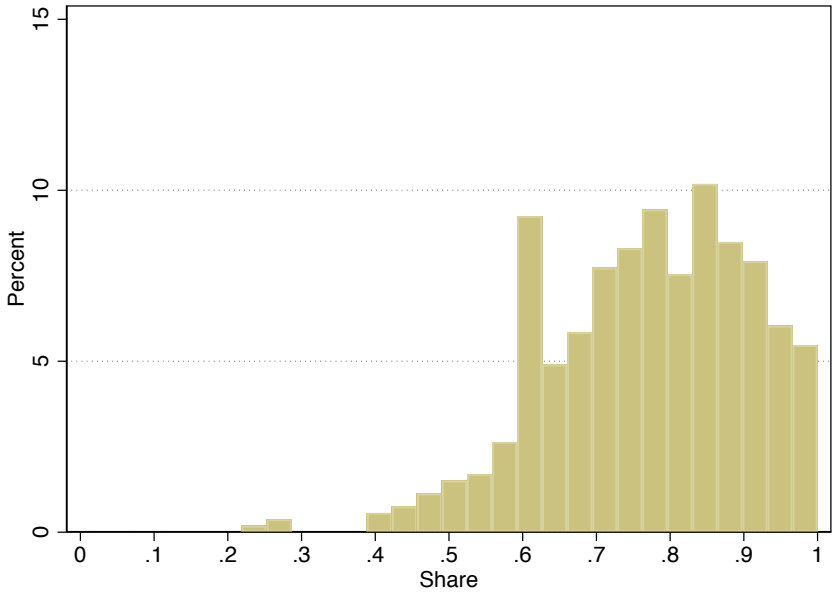


Figure 8: Are subjects consistent with sincere Bayesian voting? Using all votes and all treatments *including MB2*, 530 subjects. y-axis: percent of subjects; x-axis: share of each subject's votes (out of 32 decisions per subject) consistent with sincere Bayesian voting strategies.

Appendix B 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.

B.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)

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

Hence even though state b is more likely, i still prefers to vote for C_a, i.e. not to vote informatively. Now, if some voters may be uninformed, the 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. □

Proposition 1. ***[Unbiased media and empty 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 (since partisan groups have equal sizes).*

Proof of Proposition 1. Follows from Lemma 1 (see also the discussion in Subsection 3.1). □

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

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

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.[24] □

Proposition 2. ***[Unbiased media and complete network]*** *Under MB0+SN2, there is no full information revelation equilibrium. However, there is a range of semi-pooling equilibria, in which all C_j-partisans, $j \in \{a, b\}$, with favorable signals s_j reveal them truthfully at the communication stage, and hide the unfavorable signals s_{-j}, with a commonly known equilibrium probability ν^*. At the voting stage, each player i has a potentially different posterior p_i, which depends on the difference between the number of s_a and s_b signals.*[25] *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 $\nu^* \in (.922, 1]$.*

Proof of Proposition 2. 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

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

[25] See Eq (7) where k_{-i}, the difference in the number of revealed s_a and s_b signals is supplemented by i's private signal.

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 $1-\mu_{-j}(\tilde{s}^j_\emptyset, \nu) \equiv \Pr(s = s_\emptyset|\tilde{s}^j_\emptyset)$.[26] Due to symmetry, $\mu_a(\tilde{s}^b_\emptyset, \cdot) = \mu_b(\tilde{s}^a_\emptyset, \cdot)$, thus we can omit the subscript and simply write $\mu = \mu(\nu)$, where μ is an increasing function of ν.

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

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

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

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

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

[26] 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.

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. □

Proposition 3. ***[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. At the voting stage, each C_j-partisan*

has a common posterior p_j, which depends on the difference between the number of s_a and s_b signals revealed by C_j-partisans.[27] *Given these posteriors, C_a-partisans vote for C_a as long as $p_a > t_a$, and otherwise vote C_b. C_b-partisans vote for C_b as long as $p_b < t_b$, and otherwise vote C_a.*

Proof of Proposition 3. 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 $p_{-j}(\theta = a|k(C_j))$:

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

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,

[27] See Eq (17).

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

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

□

B.2 The case of moderate media bias

Proposition 4. ***[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.*

Proof of Proposition 4. Suppose i is C_j-partisan. With parameters from Table 1, player i's posterior conditional on observing signals s_j and s_{-j} is given by

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

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

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 for C_j. If everyone else is voting sincerely and informatively (with the exception of the uninformed[28]), i's vote is pivotal when the difference between the number of s_a and s_b signals amongst others, denoted k_{-i}, takes values in $\{-1, 0, 1\}$. This leaves the following five cases to consider: i) $k_{-i} = 1$ (or -1) due to an extra signal s_j from a C_j-partisan; ii) $k_{-i} = 1$ (or -1) due to an extra signal s_j from a C_{-j}-partisan; iii) $k_{-i} = 1$ (or -1) due to an extra signal s_{-j} from a C_j-partisan; iv) $k_{-i} = 1$ (or -1) due to an extra signal s_{-j} from a C_{-j}-partisan; or v) $k_{-i} = 0$ and 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, then 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.*

[28] i's vote may or may not be pivotal for the same value of k_{-i}, depending on the partisanship and the realized number of the uninformed voters, assumed to vote for their default choice. However, the empty signals are equally likely in both partisan groups and are uncorrelated with the state, so in expectation, the uninformed voters do not affect the pivotal calculations.

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

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

The transition from line 2 to line 3 follows by substituting $q_a^a = q_b^b = 0.9$ and $q_b^a = q_a^b = 0.4$ from our parameters. Denote $k_a(s_a) := k_{-i}(\text{f}) + \mathbb{1}_{s_a}$, $k_a(s_b) := k_{-i}(\text{uf}) - \mathbb{1}_{s_b}$ the reported difference in the number of s_a and s_b signals with i's signal included, if i is C_a-partisan; and $k_b(s_a) := k_{-i}(\text{uf}) + \mathbb{1}_{s_a}$, $k_b(s_b) := k_{-i}(\text{f}) - \mathbb{1}_{s_b}$ the reported difference in the number of s_a and s_b signals with i's signal included, if i is C_b-partisan. Denote $\boldsymbol{k} := (k_a(s_a), k_a(s_b), k_b(s_a), k_b(s_b))$ the vector of decision-relevant revealed signal differences. i's message is pivotal for different values of $k_{-i}(\text{f}), k_{-i}(\text{uf})$, as described in Table 13 (by revealing her signal, i pushes the others' posterior over one of the two critical thresholds, t_a and t_b).

Similarly to the no bias case, Table 13 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.*[29] *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 $\nu^* \in (0.862, 1]$.*

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, $\tilde{s}_\emptyset^j$, reported by a C_j-partisan, all other players believe that this signal is actually an unfavorable signal to C_j with probability

$$\mu_{-j} \equiv \Pr(s = s_{-j} | \tilde{s}_\emptyset^j) = \frac{\frac{1}{2}(1-r)(1-q_j^j)\nu + \frac{1}{2}(1-r)q_j^{-j}\nu}{\frac{1}{2}(1-r)(1-q_j^j)\nu + \frac{1}{2}(1-r)q_j^{-j}\nu + r} = \frac{\frac{1}{2}(1-r)\nu(1-q_j^j+q_j^{-j})}{\frac{1}{2}(1-r)\nu(1-q_j^j+q_j^{-j}) + r} \qquad (21)$$

Due to symmetry, $\mu_a(\tilde{s}_\emptyset^b) = \mu_b(\tilde{s}_\emptyset^a)$, thus we can omit the subscript and simply write μ. Fix player i and denote $\#_{-i}(\tilde{s}_\emptyset^j)$ 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 $\#_{-i}(\tilde{s}_\emptyset^j)$ reported empty signals is $h_{-j}(\#_{-i}(\tilde{s}_\emptyset^j)) =$

[29] 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.

Table 13: 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.

$\mu \cdot \#_{-i}(\tilde{s}^j_\emptyset)$ (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, $k_{-i}(\tilde{s}_\emptyset) \in \{-5\mathbb{1}_{i\in C_b}, -4, \ldots, 4, 5\mathbb{1}_{i\in C_a}\}$.[30] 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 expected difference in the number of hidden 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-revealed) by others becomes

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

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. Note that the expression for player i's posterior, obtained in (23), 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:

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

[30] 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.

and

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

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

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

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

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

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

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

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

It is straightforward to show that any $\nu \geq \nu^*$ is also incentive compatible. Thus we obtain a series of critical values ν^* that depend on i's partisanship, her signal, and different combinations of $k_{-i}(\text{f})$, $k_{-i}(\text{uf})$, 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 2 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]*** *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.*[31] *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.*

Proof of Proposition 6. The analysis mirrors the case of polarized network and no bias, with some modifications regarding the expressions for the posteriors. Namely, (15) becomes

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

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

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

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

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

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

and in state $\theta = b$:

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

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

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

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

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

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

and in state $\theta = b$:

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

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

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

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

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 $i \in C_a$,

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

and for $i \in C_b$,

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

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

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.