The Politics of Personalized News Aggregation

Lin Hu* Anqi Li† Ilya Segal‡

This Draft: September 2020

Abstract

We study how personalized news aggregation for rational inattentive voters (NARI) affects policy polarization and public opinion. In a two-candidate electoral competition model, an attention-maximizing infomediary aggregates information about candidates' valence into news. Voters decide whether to consume news, trading off the expected gain from improved expressive voting against the attention cost. NARI generates policy polarization even if candidates are office-motivated. Personalized news serves extreme voters with skewed signals and makes them the disciplining entities of policy polarization. Analysis of disciplining voters' identities, preferences, and beliefs sheds light on the political consequences of recent regulatory proposals to tame tech giants.

Keywords: news aggregation for rational inattentive voters, electoral competition, policy polarization, public opinion

JEL codes: D72, D80, L10

*Research School of Finance, Actuarial Studies and Statistics, Australian National University, lin.hu@anu.edu.au.

†Department of Economics, Washington University in St. Louis, anqili@wustl.edu.

‡Department of Economics, Stanford University, isegal@stanford.edu.

1 Introduction

Recently, the idea that tech-enabled news personalization could affect political polarization has been put forward in the academia and popular press (Sunstein (2009); Pariser (2011); Gentzkow (2016); Obama (2017); Barbera (2020)). This paper studies how personalized news aggregation for rational inattentive voters affects policy polarization and public opinion in an electoral competition model.

Our premise is that rational demand for news aggregation in the digital era is driven by information processing costs. As more people get news online where the amount of available information (2.5 quintillion bytes) is vastly greater than what any individual can process in a lifetime, consumers must turn to infomediaries for news aggregation, personalized based on their individual data such as demographic and psychographic attributes, digital footprints, social network positions, etc.[1] In this paper, we abstract from the issue of information generation (e.g., original reporting), focusing instead on the role of infomediaries in aggregating available information into news that is easy to process and useful for the target audience.

We develop a model of news aggregation for rational inattentive consumers (hereafter NARI), in which an infomediary can flexibly aggregate source data into news using algorithm-driven systems. While flexibility is also assumed in the Rational Inattention model pioneered by Sims (1998) and Sims (2003) (hereafter RI), in that model consumers can aggregate information optimally themselves and so have no need for external aggregators. To model the demand for news aggregation, we assume that consumers can only choose whether to absorb the news offered to them but cannot digest news partially or

[1] News aggregators (e.g., aggregator sites, social media feeds) operate by sifting through myriad online sources and directing readers to stories they might find interesting. They have recently gained prominence as more people get news online, from social media and through mobile devices (Matsa and Lu (2016)). The top three popular news websites in 2019: Yahoo! News, Google News, and Huffington Post, are all aggregators. The role of social media feeds in the 2016 U.S. presidential election remains subject of hotly debate (Allcott and Gentzkow (2017)). See Athey and Mobius (2012), Athey, Mobius, and Pal (2017), and Chiou and Tucker (2017) for background readings and literature surveys.

selectively,[2] or aggregate information from sources themselves.[3] While this assumption is certainly stylized, it is, in our view, the simplest one that creates a role for news aggregators.[4]

If choosing to consume news, a consumer incurs an attention cost that is posterior separable (Caplin and Dean (2013)) while deriving utilities from improved decision-making. Consuming news is optimal if the expected utility gain exceeds the attention cost. As for the infomediary, we assume that its goal is to maximize the total amount of attention paid by consumers, interpreted as the advertising revenue generated from consumer eyeballs.[5] This stylized assumption captures the key trade-off faced by the infomediary, who uses useful and easy-to-process information to attract consumers' attention and to prevent them from tuning out. While we focus on the case of a monopolistic infomediary in order to capture the market power wielded by tech giants, we also investigate an extension to perfectly competitive infomediaries which, together with personalization, becomes equivalent to consumers optimally aggregating information themselves as in the standard RI model.

We embed the NARI model into an electoral competition game in which two candidates compete for office by choosing policies on a left-right spectrum. Voters vote expressively based on policies, as well as an uncertain valence state about candidates' fitness for office. News about the valence state is designed by an infomediary, who moves simultaneously with candidates. Voters make news consumption decisions before they observe policies and cast votes.

A consequence of NARI is that the infomediary gives binary recommendations as to

[2] Analyses of page activities (e.g., scrolling, viewport time) have established significant levels of user attention and engagement in online news reading (Lagun and Lalmas (2016); Mitchell, Stocking, and Matsa (2016)). Snippets (headlines plus excerpts) also contain substantial information despite that they do not always materialize into click-throughs (Dellarocas et al. (2016)).

[3] According to the studies in Footnote 1, aggregators reduce search costs and broaden readers' scopes compared to direct browsing and web-based searches.

[4] Strömberg (2015) first noted the difference between the RI model and standard media models, postulating that "in the RI model, information is assumed available, and voters choose what information to pay attention to given their cognitive constraints. In our (media) model, the media chooses what information is most profitable to make available to voters, while voters have no cognitive constraints." NARI lies somewhere in-between, in that voters must fully absorb the information given to them by paying costly attention.

[5] Section 2.2 details the business model of news aggregators.

which candidate one should vote for. Indeed, any information beyond voting recommendations would only raise attention cost without any corresponding benefit and would thus turn away news consumers whose participation constraints bind at the optimum. Furthermore, news consumers must strictly prefer to obey the recommendations given to them (hereafter *strict obedience*), because a consumer with a weakly preferred candidate that is independent of the voting recommendations would abstain from news consumption to save the attention cost.

An important implication of strict obedience is that local deviations from a symmetric policy profile wouldn't change voters' votes for either recommendation they may receive, suggesting that policy divergence could arise in equilibrium even if candidates are office-motivated. We define *policy polarization* as the maximal distance between candidates' positions among all symmetric perfect Bayesian equilibria. In the baseline model with three types of voters: left-wing, centrist, and right-wing, our main theorem shows that if strict obedience holds for every feasible symmetric policy profile (hereafter *uniform strict obedience*), then policy polarization is strictly positive and equals the *disciplining voter's policy latitude*.

A voter's *policy latitude* is the maximum policy such that even if a candidate deviates unilaterally from the symmetric policy profile to the voter's most preferred position, he still cannot *attract* the voter, i.e., win the voter's support even if news is unfavorable to his fitness. Policy latitude captures a voter's *resistance* to policy deviations, which decreases with his preference for the deviating candidate's policies and increases with his pessimism about the latter's fitness following unfavorable news; it is shown to be well-defined and strictly positive mainly due to uniform strict obedience.

A voter is *disciplining* if his policy latitude determines policy polarization. To pin down disciplining voters, we compare the case of *broadcast news*, in which the infomediary (e.g., commercial TV) must offer a single news signal to all voters, to the case of *personalized news*, in which the infomediary (e.g., Google News) can design different news signals for different voters. In the broadcast case, all voters receive the same voting recommendation, so a candidate's deviation is *profitable*, i.e., strictly increases his winning probability, if and only if it attracts a majority of voters. Under the usual assumptions, this is equivalent to attracting median voters, who are therefore disciplining. In the personalized case, the infomediary can provide conditionally independent signals to different types of voters (this assumption will be relaxed), so each type of voter is pivotal with a positive probability when voters' population distribution is sufficient dispersed. In that case, a deviation is shown to be profitable if and only if it attracts any type of voter, so voters with the smallest policy latitude are disciplining because they are the easiest to attract.

We examine factors that affect the identities and policy latitudes of disciplining voters. Our first comparative statics exercise concerns the transition from broadcast news to personalized news, the reverse of which is an integral part of Senator Elizabeth Warren's proposal to tame tech giants (Warren (2019)) and is a possible consequence of implementing the General Data Protection Regulation (GDPR) among EU citizens (The General Data Protection Regulation (2016)).[6] As argued above, candidates face more deviating opportunities in the personalized case than in the broadcast case, which makes policy polarization harder to sustain. However, there is a countervailing effect stemming from the skewness of news signals. In the broadcast case, the infomediary uses a symmetric signal to attract a broad and balanced audience. In contrast, the optimal personalized signals for extreme voters are skewed: to maximize the usefulness of news consumption, the recommendation to vote across party lines must be sufficiently strong and, in order to contain the attention cost, must also be sufficiently rare; most of the time, the recommendation is to vote along party lines, which by Bayes' rule can only shift the voter's belief moderately. Taken together, the optimal personalized signal exhibits both the *own-party bias* and *occasional big surprise* that have been documented in the empirical literature (Flaxman, Goel, and Rao (2016); Gentzkow (2016)).

[6] In early 2019, Senator Elizabeth Warren called for the big tech companies to be broken up and to meet the standard of nondiscriminatory dealing with customers. In 2016, GDPR was created to uphold the protection of personally identifiable information of EU citizens, mandating that data controllers must design information systems using the highest-possible privacy settings by default, and that no personal data may be processed unless it is done under the lawful bases specified by the regulation.

If extreme voters are disciplining in the personalized case, then the skewness of their personalized news signals is crucial for sustaining greater policy polarization than in the broadcast case. First, the occasional big surprise makes the base voters of a deviating candidate difficult to attract in the rare event where their news signal recommends the opposing candidate. Indeed, such a recommendation can be so strong that even the most attractive deviation to base voters cannot change their voting decisions. In that case, policy polarization is constrained by the deviation to attract opposition voters, who have an intrinsic preference against the deviating candidate's policies and therefore a big policy latitude. For these reasons, equilibrium could be more polarized than in the broadcast case, and we give conditions for this to be true when the attention cost is Shannon entropy-based.

Our second comparative statics exercise concerns introducing competition between infomediaries, which is advocated by the British government as a preferable way of regulating tech giants (The Digital Competition Expert Panel (2019)) and is mathematically equivalent to increasing voters' marginal attention cost in the monopolistic personalized case. Its policy polarization effect is negative, because increasing voters' marginal attention cost tempers their beliefs about candidates' fitness and, in turn, reduces their policy latitudes.

Our third comparative statics exercise concerns increasing mass polarization, modeled as a mean-preserving spread to voters' policy preferences (Fiorina and Abrams (2008); Gentzkow (2016)). Contrary to popular belief, its policy polarization effect is negative when news is personalized and voters' population distribution is sufficiently dispersed, in which case policy polarization is determined by the minimal policy latitude among all voters rather than median voters' policy latitude alone.

In Appendix A, we extend the baseline model to arbitrary finite types of voters and allow the infomediary to correlate personalized news signals across voters. We give a full characterization of policy polarization in this general setting and demonstrate, in particular, that correlation can only increase policy polarization. As an application of this result, we examine the consequence of large-scale adoptions of news aggregators such as Allsides.com, which have recently been built to battle the rising polarization through providing users with neutral viewpoints. Compared to the case of broadcast news, this innovation keeps the marginal news distribution unchanged while relaxing the assumption that news signals must be perfectly correlated among voters, so its policy polarization effect is negative.

Taken together, our results suggest that many hotly debated proposals to regulate tech giants could have unintended consequences for social welfare after their policy polarization effects have been taken into account. In Section 7, we discuss this issue in more detail, and lay out a plan for taking our theory to data.

1.1 Related literature

Rational inattention The literature on RI assumes the existence of a costly communication channel that aggregates source data into the optimal signal for decision-making. To create a role for news aggregators, we assume that communication channels are designed by an attention-maximizing infomediary, whereas voters must fully absorb the information given to them. Apart from this major departure from the RI paradigm, we otherwise follow the standard model of posterior-separable attention cost that nests Shannon entropy as a special case. Posterior separability (Caplin and Dean (2013); Caplin, Dean, and Leahy (2019)) has recently received attention from economists because of its axiomatic and revealed-preference foundations (Caplin and Dean (2015); Zhong (2017); Denti (2018); Tsakas (2019)), connections to sequential learning (Hébert and Woodford (2017); Morris and Strack (2017)), and validations by lab experiments (Ambuehl, Ockenfels, and Stewart (2019); Dean and Nelighz (2019)).

Matějka and Tabellini (2016) pioneer the study of electoral competition with rational inattentive voters. In their model, voters face normal uncertainties about candidates' policies (rather than valence), and the cost of information acquisition (in the form of variance reduction) can differ across candidates. The last assumption generates policy divergence between candidates by allowing them to target voters who pay different levels of attention and in turn exert different influences on the election outcome. Our model differs from Matějka and Tabellini (2016) in the source of uncertainty, the attention technology, and the mechanism that generates policy divergence between candidates.

Media bias The literature on media bias is thoroughly surveyed by Prat and Strömberg (2013), Strömberg (2015), and Anderson, Strömberg, and Waldfogel (2016). We add to the theoretical literature on demand-driven media bias, whose common explanations include limited information processing capacity, behavioral bias (Mullainathan and Shleifer (2005)), and reputational concern (Gentzkow and Shapiro (2006)). The idea that even rational consumers can exhibit a preference for biased news when constrained by information processing capacities dates back to Calvert (1985a) and is later expanded on by Suen (2004), Burke (2008), and Che and Mierendorff (2019). While these later models generate own-party biases (and implicitly occasional big surprises), they work with non-RI information aggregation technologies and do not examine the consequence of news bias for electoral competition. As we will explain in the discussion of Chan and Suen (2008), even seemingly miniature differences in the information aggregation technology can manifest into significant differences in election outcomes.[7]

In political science, *own-party bias*, or *party sorting*, refers to the positive correlation between a person's party affiliation and his propensity to support his own-party candidate. The past decade has witnessed a rise in party sorting but little change in voters' policy preferences (Fiorina and Abrams (2008); Gentzkow (2016)), a trend that could persist due to personalized news aggregation. Evidence for occasional big surprise and, more generally, Bayesian news consumers is surveyed by DellaVigna and Gentzkow (2010).[8] In the context of online news consumption, Flaxman, Goel, and Rao (2016) find that using news aggregators increases one's propensity to support his own-party candidate (i.e., own-party bias), as well as his opinion intensity when occasionally supporting his opposite-party

[7] For the same reason, one should not conflate RI with "rational ignorance," a term used by political scientists such as Prato and Wolton (2016) to refer to rigid information acquisition, i.e., pay a fixed cost for drawing a signal from a given probability distribution. Unless this probability distribution is carefully engineered, one couldn't obtain the same predictions as ours by working with rationally ignorant voters.

[8] In contrast, the non-Bayesian model of Mullainathan and Shleifer (2005) predicts a confirmatory bias but not any occasional big surprise.

candidate (i.e., occasional big surprise).[9]

Policy divergence in electoral competition models In most existing probabilistic voting models, voters' signals are taken to be continuously distributed, so even small changes in policy positions could affect voting decisions.[10] Under this assumption (and symmetry, as in our model), Calvert (1985b) famously establishes policy convergence between office-seeking candidates, and pioneers the use of policy preference for generating policy divergence between candidates. Uniform strict obedience stands in contrast to this assumption, though it is a natural consequence of NARI.

Electoral competition with flexible and profit-maximizing information aggregators Strömberg (2004) and Chan and Suen (2008) study electoral competition with flexible and profit-maximizing information aggregators, though their models and predictions differ completely from ours.[11] Strömberg (2004) assumes that the probability a government program is read by voters increases with its press coverage, and newspapers allocate limited spaces between multiple programs based on readers' revenue potentials. His research question concerns how newspaper reporting affects government budget allocation rather than platform convergence or divergence.

Chan and Suen (2008) examine a related research question to ours. In their model, voters care about whether the realization of a random state variable is above or below their personal thresholds, and media outlets partition state realizations using fixed threshold rules. A consequence of working with this information aggregation technology, rather than NARI, is that signal realizations are monotone in voters' thresholds (i.e., if a left-leaning voter is recommended to vote for candidate R, then a right-leaning voter must receive

[9] Another famous account for occasional big surprise is provided by Chiang and Knight (2011), who find that endorsements of presidential candidates that go against the bias of a newspaper are the most effective in shaping voters' voting decisions.

[10] See Duggan (2017) for a survey of probabilistic voting models.

[11] Assuming non-RI information aggregation technologies and fixed policy platforms, Duggan and Martinelli (2011) and Prat (2018) study political models in which the media's goal is to persuade voters with limited information processing capacities, and Perego and Yuksel (2018) examine how media entry affects the opinion disagreement among news consumers.

the same recommendation), so median voters are always disciplining despite a plurality of media. Starting from there, the analysis of Chan and Suen (2008) differs completely from ours. In particular, the authors generate policy divergence between candidates using the standard logic of the Calvert-Wittman model rather than uniform strict obedience.[12]

The remainder of this paper proceeds as follows: Section 2 introduces the baseline model; Sections 3 and 4 characterize equilibrium outcomes; Section 5 conducts comparative statics analyses; Section 6 investigates extensions of the baseline model; Section 7 concludes. Additional materials and mathematical proofs can be found in Appendices A-E.

2 Baseline model

2.1 Setup

Political players Two office-seeking candidates named L and R can adopt the policies in a compact interval $\mathcal{A} = [-\overline{a}, \overline{a}]$, where $\overline{a} > 0$. They face a unit mass of infinitesimal voters who are *left-wing* ($k = -1$), *centrist* ($k = 0$) or *right-wing* ($k = 1$). Each type $k \in \mathcal{K} = \{-1, 0, 1\}$ of voter has a population $q(k)$ and values a policy $a \in \mathcal{A}$ by $u(a, k) = -|t(k) - a|$. The population function $q : \mathcal{K} \to \mathbb{R}_{++}$ has support $\mathcal{K}$ and is symmetric around zero, whereas the bliss point function $t : \mathcal{K} \to int(\mathcal{A})$ is strictly increasing and satisfies $t(-k) = -t(k)$ for any $k \in \mathcal{K}$.

Voting Voting is *expressive*. For any given policy profile $\mathbf{a} = \langle a_L, a_R \rangle \in \mathcal{A}^2$, a type k voter earns the following utility difference from voting for candidate R rather than L:

$$v(\mathbf{a}, k) + \omega.$$

In the above expression,

$$v(\mathbf{a}, k) = u(a_R, k) - u(a_L, k)$$

[12] Specifically, Chan and Suen (2008) assume that voters privately observe an aggregate preference shock in addition to the voting recommendations given by media outlets, and that candidates have policy preferences. We do not make any of these assumptions.

captures the voter's differential valuation of candidates' policies, whereas ω is an uncertain state about which candidate is more fit for office (hereafter *valence state*, or simply *state*).[13] In the baseline model, ω takes the values in $\Omega = \{-1, 1\}$ with equal probability, so its prior mean equals zero.

News A *news signal* is a *finite* signal structure $\Pi : \Omega \to \Delta(\mathcal{Z})$, where each $\Pi(\cdot \mid \omega)$ specifies a probability distribution over a finite set $\mathcal{Z}$ of signal realizations conditional on the state realization being $\omega \in \Omega$.

News is provided by a *monopolistic* infomediary who is equipped with a *personalization technology* $\mathcal{S}$. $\mathcal{S}$ is a *partition* of voters' types, and each cell of it is called a *market segment*. The infomediary can distinguish between voters of different market segments but not those within the same market segment. Our focus will be on the coarsest and finest partitions named the *broadcast technology* $b = \{\mathcal{K}\}$ and *personalized technology* $p = \{\{k\} : k \in \mathcal{K}\}$, respectively: the former cannot distinguish between the various types of voters at all, whereas the latter can do so perfectly.

Under personalization technology $\mathcal{S} \in \{b, p\}$, the infomediary designs $|\mathcal{S}|$ news signals, one for each market segment. Within each market segment, voters decide whether to consume the signal that is offered to them. Consuming signal Π requires that voters fully absorb its information content. Doing so incurs an *attention cost* $\lambda \cdot I(\Pi)$, where $\lambda > 0$ is a scaling parameter called the *marginal attention cost*, and $I(\Pi)$ is the needed amount of *attention* for absorbing the information content of Π. After that, voters observe signal realizations, update beliefs about the state, and vote expressively. The infomediary's profit equals the total amount of attention paid by voters.

Game The game sequence is as follows.

[13] E.g., in the ongoing debate about the most effective measure to battle terrorism, $\omega = -1$ if the state favors using soft power such as diplomatic tactics, and $\omega = 1$ if the state favors using hard power such as military preemption. Candidate L and R are more experienced with using soft power and hard power, respectively, and whoever holding the experience that matches the true state enjoys an advantage over his opponent.

1. The infomediary designs news signal structures; voters observe the news signal structures offered to them and make consumption decisions accordingly.

2. Candidates choose policies without observing the moves in stage 1.

3. The state is realized.

4. Voters observe policies and news realizations and vote. The winner of the election is determined by simple majority rule with even tie-breaking.

The solution concept is *perfect Bayesian equilibrium* (hereafter *equilibrium*). Our goal is to characterize *all* equilibria in which candidates propose *symmetric* policy profiles of form $\langle -a, a \rangle$, $a \geq 0$ in stage 2 of the game.

2.2 Model discussions

Valence state We take the source data concerning candidates' fitness (e.g., original reporting) as given. The assumption of binary states eases exposition without affecting most qualitative predictions, and it will be relaxed in Appendix C.

News signal The infomediary aggregates source data into news content that shapes voter's belief about the valence state. The distribution of beliefs is determined by a signal structure $\Pi : \Omega \to \Delta(\mathcal{Z})$, which outputs signal realization $z \in \mathcal{Z}$ with probability

$$\pi_z = \sum_{\omega \in \Omega} \Pi(z \mid \omega) \cdot \frac{1}{2}.$$

Assume without loss of generality (w.l.o.g.) that $\pi_z > 0$ for any $z \in \mathcal{Z}$. Then

$$\mu_z = \sum_{\omega \in \Omega} \omega \cdot \Pi(z \mid \omega) / (2\pi_z)$$

is the posterior mean of the state conditional on the signal realization being $z \in \mathcal{Z}$, and it fully captures one's posterior belief after consuming the content indexed by z.

Attention cost We use a single parameter $\lambda > 0$ to capture factors that affect voters' (opportunity) cost of paying attention, e.g., distractions coming from the Internet and mobile devices, increased competition between firms for consumer eyeballs (Teixeira (2014); Dunaway (2016)), etc. We restrict λ to be constant across voters for now and will relax this assumption in Section 6.3.

The next assumption models attention as a scarce resource that reduces voter's uncertainty about the valence state before and after news consumption.

Assumption 1. *The needed amount of attention for consuming* $\Pi : \Omega \to \Delta(\mathcal{Z})$ *equals*

$$I(\Pi) = \sum_{z \in \mathcal{Z}} \pi_z \cdot h(\mu_z), \tag{1}$$

where the function $h : [-1, 1] \to \mathbb{R}_+$ *(i) is strictly convex and satisfies* $h(0) = 0$*, (ii) is continuous on* $[-1, 1]$ *and twice differentiable on* $(-1, 1)$*, and (iii) is symmetric around zero.*

Equation (1) coupled with Assumption 1(i) is equivalent to *weak posterior separability*, a notion proposed by Caplin and Dean (2013) to generalize Shannon's entropy as a measure of attention cost.[14] In the current setting, weak posterior separability stipulates that consuming the null signal requires no attention, and that more attention is needed for moving the posterior belief closer to the true state and as the news signal becomes more Blackwell informative. Assumption 1(ii) and (iii) impose regularities on our problem. Assumption 1(iii) is stronger than what we need for proving most results, stipulating that only the magnitude of the posterior mean could affect the attention cost, whereas its sign (which indicates the direction of belief updating) couldn't.[15]

[14]Various foundations for posterior separability have been proposed since Shannon (1948), where a voter ask a series of yes-or-no questions about the performance state at a constant marginal cost. The more questions the voter asks, the more precise his posterior belief is about the performance state. According to Shannon (1948), the minimal average number of questions that needs to be asked in order to implement a signal structure equals approximately the mutual information between the source data and output signal. More recently, Hébert and Woodford (2017) and Morris and Strack (2017) study optimal stopping problems where a decision maker consults one of the many sources sequentially and incurs a (time and belief-dependent) flow cost until the process is randomly terminated. These authors provide general conditions under which the expected total cost is posterior separable in the continuous-time limit.

[15]For proving most results, all we need is that the function $h(\cdot, k)$ (parameterized by voter's type k)

Assumption 1 is satisfied by many commonly used attention functions. While most upcoming analysis won't make functional form assumptions about h, it is still useful to keep in mind two special cases $h(\mu) = \mu^2$ and $h(\mu) = H((1+\mu)/2)$ (H denotes the binary entropy function), in which $I(\Pi)$ equals the reductions in the variance and Shannon entropy of the valence state, respectively, through consuming Π.

Expressive voting with costly news consumption We share the same view as Prat and Strömberg (2013) that instrumental voting is an important motive for consuming political news. Our voters trade off the gain from improved expressive voting against the cost of news consumption, just like their counterparts in Strömberg (2004) and Matějka and Tabellini (2016).[16] In Section 6.2, we discuss the consequence of introducing alternative motives for news consumption (e.g., entertainment) to our model.

Business model of news aggregators A major revenue source for news aggregators is displaying advertising.[17] Ad revenue is increasing in the attention paid by news consumers, because the more information content a consumer absorbs, the longer he stays on the platform (for reasons discussed in Footnote 14) and so is exposed to more ads.[18] Apart from the fixed operating cost we do not explicitly model, the main expenditure incurred by the infomediary stems from its arrangements between sources, which range from no payment (e.g., Google News), a lump-sum payment (e.g., Yahoo! News), to revenue sharing (e.g., Facebook News Feed). To capture these institutional details, we set the infomediary's gross profit equal to the total amount of attention paid by voters, taking comfort in the fact that applying any strictly increasing transformation to this gross profit function wouldn't affect

is symmetric between symmetric types of voters, i.e., $h(\mu, k) = h(-\mu, -k)$ for any μ and k. The only exceptions are Theorem 1 and its corollaries, but even they would remain unchanged if the departure from Assumption 1(iii) is small.

[16] In Chan and Suen (2008) and Perego and Yuksel (2018), voters consume news to improve expressive voting decisions, subject to their bandwidth constraints.

[17] Most news aggregators display ads on their platforms. An exception is Google News, which displays zero ads and instead refers readers to source webpages as well as the main Google search engine where Google ads are displayed.

[18] Click here for tactics deployed by Facebook such as playing mid-roll ads when users are already in the "lean-back" reading/watching mode.

qualitative prediction.

Personalization technology We model the personalization technology as a partition of voters' types and use its fineness to capture the degree of personalization. The transition from broadcast technology to personalized technology (hereafter $b \to p$) is interesting to study for three reasons. First, if one interprets the broadcast technology as being adopted by traditional media outlets, say, commercial TV, then studying $b \to p$ yields insights into how modernization of media affects political outcomes.[19] Second, if one is interested in the consequence of banning personalization (as part of the recent regulatory proposals to tame tech giants), then studying the reverse transition $p \to b$ would be a natural starting point. Third, once we understand the two basic technologies, we can immediately extend the analysis to other voter partitions, such as those that divide the same type of voters into multiple subgroups, and those that are finer than the broadcast technology yet coarser than the personalized technology. Interested readers can consult Sections 6.1 and 6.3 for further details.

Game sequence Our game sequence captures some realities while keeping our analysis as simple as possible. Here we comment on two main assumptions, which are both stylized. Our first assumption says that policies only become observable to voters at the voting stage. We do not explicitly model the activities (e.g., political advertising, canvassing, debates) that make this happen but note that they typically take place right before the election day.[20] Given this, it is reasonable to assume that the design of news signals (which aggregates an evolving state of affairs into breaking news) and the news consumption decisions are made independently of candidates' policies.

Our second assumption says that candidates do not observe news signal structures when

[19] Commercial TV also generates most revenues from advertising, and its programs are curated to attract viewers with the highest revenue potentials (Hamilton (2004)). In our view, one of the biggest differences between commercial TV and news aggregators lies in the ability to provide personalized content.

[20] Many countries including Canada, France, Japan, and U.K. limit the lengths of political campaigns. Even in the U.S. where such regulation is absent, the majority of election ads are aired in weeks closest to the election day because their persuasive effects are believed to be short-lived (Gerber et al. (2011)).

crafting policies. Here we are motivated by the reality that the algorithms behind modern news aggregators represent trade secrets and cannot be easily reverse-engineered by third parties. According to computer scientists who work on algorithm audit, the easiest way to peek into these algorithms is to survey users (Eslami et al. (2015)), which leads us to believe that voters (at least partially) observe the news signals offered to them.

Under the above assumptions, allowing candidates to observe (signals of) the valence state before they move wouldn't affect the symmetric policy profiles that can arise in equilibria, so assuming that candidates move before the valence state is realized is less restrictive than it seems.

3 Optimal news signals

In this section, we fix any symmetric policy profile $\mathbf{a} = \langle -a, a \rangle$ with $a \geq 0$ and solve for the news signals that maximize the infomediary's profit (hereafter *optimal news signals*). We formalize the infomediary's problem in Section 3.1 and characterize its solutions in Section 3.2. All results of Section 3.2 are proven for arbitrary finite types of voters with general policy preferences in Appendix E.1.

To facilitate analysis, we say that candidate L and left-wing voters form the *left-wing party*, whereas candidate R and right-wing voters form the *right-wing party*. Under this definition, voters share party affiliations with the candidate whose policy position they most prefer.

3.1 Infomediary's problem

Under personalization technology $\mathcal{S} \in \{b, p\}$, any optimal news signal for market segment $s \in \mathcal{S}$ solves

$$\max_{\Pi} I(\Pi) \cdot \mathcal{D}(\Pi; \mathbf{a}, s) \tag{s}$$

where $\mathcal{D}(\Pi; \mathbf{a}, s)$ denotes the demand for news signal Π in market segment s under policy profile $\mathbf{a}$. To figure out the demand for news, note that extreme voters consume news in order to be convinced to vote across party lines. After news consumption, a voter strictly prefers candidate R to L if $v(\mathbf{a}, k) + \mu_z > 0$, and he strictly prefers candidate L to R if $v(\mathbf{a}, k) + \mu_z < 0$. Ex ante, the expected gain from consuming Π is

$$V(\Pi; \mathbf{a}, k) = \sum_{z \in \mathcal{Z}} \pi_z \cdot \nu(\mu_z; \mathbf{a}, k)$$

where

$$\nu(\mu_z; \mathbf{a}, k) = \begin{cases} [v(\mathbf{a}, k) + \mu_z]^+ & \text{if } k \leq 0, \\ -[v(\mathbf{a}, k) + \mu_z]^- & \text{if } k > 0. \end{cases}$$

Therefore,

$$\mathcal{D}(\Pi; \mathbf{a}, s) = \sum_{k \in \mathcal{K}: V(\Pi; \mathbf{a}, k) \geq \lambda \cdot I(\Pi)} q(k, s)$$

where $q(k, s)$ denotes the population of type k voters in market segment s.

If a solution to Problem (s) has zero demand, then it will be regarded the same as a degenerate signal. This rules out uninteresting situations in which the infomediary deters news consumption using nondegenerate signals.

3.2 Main features

Binary voting recommendations and strict obedience We first demonstrate that any optimal news signal has at most two realizations and, if binary, prescribes voting recommendations its consumers strictly prefer to obey.

Formally, we say that a signal realization z *endorses* candidate R (resp. L) and *disapproves* candidate L (resp. R) if $\mu_z > 0$ (resp. $\mu_z < 0$). For binary signals, we write $\mathcal{Z} = \{L, R\}$. Since Bayes' plausibility mandates that the expected posterior mean of the state must equal the prior mean, i.e.,

$$\sum_{z \in \mathcal{Z}} \pi_z \cdot \mu_z = 0, \tag{BP}$$

it is without loss to assume that $\mu_L < 0 < \mu_R$. In this way, we can interpret each signal realization $z \in \{L, R\}$ as an endorsement for candidate z and a disapproval of candidate $-z$.[21] In addition, we can define *strict obedience* as follows.

Definition 1. *A binary signal with posterior means $\langle \mu_L, \mu_R \rangle$ induces* strict obedience from its consumers *if the latter strictly prefer the endorsed candidate to the disapproved candidate under both signal realizations, i.e.,*

$$v(\mathbf{a}, k) + \mu_L < 0 < v(\mathbf{a}, k) + \mu_R. \quad \text{(SOB)}$$

The next lemma formalizes the claim made at the beginning of this section.

Lemma 1. *Under Assumption 1, the following hold for any symmetric policy profile $\langle -a, a \rangle$ with $a \geq 0$.*

(i) Any optimal broadcast signal has at most two realizations.

(ii) Any optimal personalized signal for any voter has at most two realizations.

(iii) Any optimal signal, if binary, induces strict obedience from its consumers.

Lemma 1 is proven differently for the cases of broadcast and personalized news. In the personalized case, the result follows from the fact that individual voters make binary voting decisions. Given this, any information beyond voting recommendations would only raise the attention cost without any corresponding benefit and would thus turn away voters whose participation constraints bind at the optimum. For these voters, maximizing attention is equivalent to maximizing the usefulness of news consumption at the maximal attention level.

The broadcast case is proven by aggregating voters with binding participation constraints into a representative voter. Under the assumptions that voters' preferences exhibiting increasing differences between policies and types and all voters share the same

[21] Since the state is binary, it is without loss to identify any binary signal with the profile $\langle \mu_L, \mu_R \rangle$ of posterior means: $\Pi(z = R \mid \omega = 1) = \frac{-\mu_L(1+\mu_R)}{\mu_R - \mu_L}$ and $\Pi(z = R \mid \omega = -1) = \frac{-\mu_L(1-\mu_R)}{\mu_R - \mu_L}$.

marginal attention cost,[22] only extreme voters' participation constraints can be binding, whereas centrist voters' participation constraint is slack.[23] Then using the concavification method developed by Aumann and Maschler (1995) and Kamenica and Gentzkow (2011), we demonstrate that the optimal personalized signal for the representative voter has at most two realizations. The analysis exploits the assumption of binary states, which will be relaxed in Section 6.2.

Strict obedience (SOB) is a consequence of rational and flexible information aggregation. Indeed, if a consumer of a binary news signal has a (weakly) preferred candidate that is independent of the voting recommendations, then he would abstain from news consumption (because doing so saves attention cost without affecting expressive voting utility), a contradiction.

The next assumption imposes regularities on our problem.

Assumption 2. *The following hold for any symmetric policy profile* $\langle -a, a \rangle$ *with* $a \geq 0$.

(i) The optimal news signal consumed by any voter under any personalization technology is nondegenerate and, hence, is binary and satisfies (SOB).

(ii) The posterior means of the state induced by the news signal in Part (i) takes values in $(-1, 1)$.

Assumption 2 is *sufficient but not necessary* for conducting the upcoming analysis. Assumption 2(i) is satisfied if voters' marginal attention cost isn't too high and their policy preferences aren't too extreme; it implies that (SOB) holds for every feasible symmetric policy profile and will hereafter be referred to as *uniform strict obedience*. Assumption 2(ii) holds if voters' marginal attention cost isn't too low, and it is imposed so that we can obtain strict comparative statics results. Figure 1 reduces Assumption 2 to model primitives in the case of entropy attention cost. In Section 6.2, we discuss the consequence of relaxing this assumption in great detail.

[22] Increasing difference, formally defined as $v(a, a', k)$ being increasing in k for any $a' > a$, means that voters more prefer right-leaning policies to left-leaning ones as they become more pro-right.

[23] In Section 6.3, we allow the marginal attention cost to differ across voters and reconcile our findings with the stylized fact that extreme voters tend to be more attentive to politics than centrist voters.

Assumptions 1 and 2 together guarantee the uniqueness of optimal news signals.

Lemma 2. *Under Assumptions 1 and 2, the optimal news signal consumed by any voter is unique for any pair of personalization technology and symmetric policy profile.*

In what follows, we will use $\Pi^{\mathcal{S}}(a,k)$ to denote the optimal news signal consumed by type k voters under personalization technology $\mathcal{S}$ and policy profile $\langle -a, a\rangle$. For each $\Pi^{\mathcal{S}}(a,k)$, we will use $\mu_z^{\mathcal{S}}(a,k)$ to denote the posterior mean of the state conditional on the signal realization being $z \in \{L, R\}$, and $\pi^{\mathcal{S}}(a,k)$ to denote the probability that the signal realization is R. In the broadcast case, we will suppress the notation of k and write $\Pi^b(a)$, $\mu_z^b(a)$, and $\pi^b(a)$ instead.

Skewness The next theorem concerns the skewness of optimal news signals.

Theorem 1. *Under Assumptions 1 and 2, the following hold for any symmetric policy profile $\langle -a, a\rangle$ with $a > 0$.*

(i) The optimal broadcast signal is symmetric, in that it endorses each candidate with equal probability, and the endorsements shift voters' beliefs by the same magnitude, i.e., $\pi^b(a) = 1/2$, $|\mu_L^b(a)| = \mu_R^b(a)$.

(ii) In the personalized case, the optimal signals are symmetric between symmetric types of voters, i.e., $|\mu_L^p(a,-k)| = \mu_R^p(a,k)$ for any $k \in \mathcal{K}$. Moreover,

(a) the optimal signal for median voters is symmetric, i.e., $\pi^p(a,0) = 1/2$ and $|\mu_L^p(a,0)| = \mu_R^p(a,0)$;

(b) the optimal signal for any extreme voter is skewed, in that it endorses the voter's own-party candidate more often than his opposite-party candidate, although the endorsement for the opposite-party candidate shifts the voter's belief more significantly than the endorsement for his own-party candidate, i.e.,

- $\pi^p(a,k) < 1/2$ *and* $|\mu_L^p(a,k)| < \mu_R^p(a,k)$ *if* $k < 0$;
- $\pi^p(a,k) > 1/2$ *and* $|\mu_L^p(a,k)| > \mu_R^p(a,k)$ *if* $k > 0$.

(iii) The optimal broadcast signal attracts less attention from any voter than his optimal personalized signal, i.e., $I\left(\Pi^b(a)\right) < I\left(\Pi^p(a,k)\right)$ *for any* $k \in \mathcal{K}$.

Theorem 1(i) holds because the representative voter's preference is symmetric in a symmetric environment. To develop intuition for Theorem 1(ii), recall that an extreme voter consumes news in order to be convinced to vote across the party line. Since the corresponding signal realization must move the posterior mean of the state far away from the prior mean, it is costly to process and, in order to contain the attention cost, must occur with a small probability. Hereafter we shall refer to this signal realization as an *occasional big surprise*. Most of the time, the recommendation is to vote for the own-party candidate, which by Bayes' plausibility can only shift the voter's belief moderately. We will refer to this signal realization as an *own-party bias*, and note that it constitutes the flip side of occasional big surprise.

We finally turn to Theorem 1(iii). As demonstrated earlier, the optimal broadcast signal is designed for a representative voter with a symmetric preference. Yet the news consumption decision is made by extreme voters who prefer skewed signals to symmetric ones. This conflict of interests limits the amount of attention that the optimal broadcast signal can attract from any voter compared to his optimal personalized signal.

4 Equilibrium policies

This section endogenizes candidates' policy positions. Under personalization technology $\mathcal{S} \in \{b, p\}$, a news profile $\widetilde{\boldsymbol{\mu}}$ and a symmetric policy profile $\langle -a, a\rangle$ with $a \geq 0$ can arise in a perfect Bayesian equilibrium if

- $\widetilde{\boldsymbol{\mu}}$ is a $|\mathcal{S}|$-dimensional random variable, where the marginal probability distribution of each dimension $s \in \mathcal{S}$ solves Problem (s), taking $\langle -a, a\rangle$ as given;
- a maximizes candidate R's winning probability, taking candidate L's policy $-a$, $\widetilde{\boldsymbol{\mu}}$, voters' news consumption decisions, and their expressive voting strategies (as functions of *actual* policies and news realizations) as given.

Our goal is to characterize *all* perfect Bayesian equilibria of this form. To this end, we develop the needed concepts in Section 4.1 and present the main characterization theorem in Section 4.2.

Before proceeding, note that the analysis so far has pinned down the marginal news distribution for each market segment but leaves the joint news distribution across market segments unspecified. This is because under the assumption of expressive voting, a voter cares only about his marginal news distribution but not the joint news distribution, despite that the latter clearly affects candidates' strategic reasoning. In Appendix A, we consider *all* joint news distributions that are *consistent* with the marginal news distributions solved in Section 3. Here, we restrict attention to news signals that are *conditionally independent* across market segments for any given state realization. The implication of this restriction will soon become clear.

4.1 Key concepts

All concepts of this section are defined for a given pair of personalization technology $\mathcal{S} \in \{b, p\}$ and population function q. We first describe how a unilateral deviation from a symmetric policy profile $\langle -a, a \rangle$ with $a \geq 0$ can affect expressive voting decisions. Due to symmetry, it suffices to consider candidate R's deviation to a'.

Definition 2. *A unilateral deviation of candidate R from a symmetric policy profile $\langle -a, a \rangle$ with $a \geq 0$ to a'* attracts *type k voters if it wins the latter's support even when their news signal endorses candidate L, i.e.,*

$$v\left(-a, a', k\right) + \mu_L^{\mathcal{S}}(a, k) > 0.$$

It repels *type k voters if it loses the latter's support even when their news signal endorses candidate R, i.e.,*

$$v\left(-a, a', k\right) + \mu_R^{\mathcal{S}}(a, k) < 0.^{24}$$

[24] Replacing strict inequalities with weak inequalities won't affect the upcoming analysis.

Note that if a' attracts (resp. repels) a voter, then it makes him vote for (resp. against) candidate R unconditionally. If it neither attracts or repels a voter, then it has no effect on his voting decisions. Also note that $a' = t(k)$ is the most attractive deviation to type k voters.

We next describe equilibrium outcomes. In what follows, we will use $\mathcal{E}^{\mathcal{S},q}$ to denote the set of nonnegative policy as such that the symmetric policy profile $\langle -a, a\rangle$ can arise in equilibrium. We are interested in the *policy polarization* $a^{\mathcal{S},q} = \max \mathcal{E}^{\mathcal{S},q}$, defined as the maximal symmetric equilibrium policy, and whether all policies between zero and policy polarization can arise in equilibrium. The following concepts facilitate analysis.

Definition 3. $\phi^{\mathcal{S}}(-a, a', k) = v(-a, a', k) + \mu_L^{\mathcal{S}}(a, k)$ *is type k voters'* susceptibility *to candidate R's deviation from a symmetric policy profile $\langle -a, a\rangle$ with $a \geq 0$ to a' following unfavorable news to R's fitness, and $-\phi^{\mathcal{S}}(-a, a', k)$ is type k voters'* resistance *to a' in this situation.*

Definition 4. *The k-*proof set $\Xi^{\mathcal{S}}(k)$ *is the set of nonnegative policy as such that no unilateral deviation of candidate R from $\langle -a, a\rangle$ attracts type k voters, i.e., $\phi^{\mathcal{S}}(-a, a', k) \leq 0$ $\forall a'$. Since $a' = t(k)$ is the most attractive deviation to type k voters,*

$$\Xi^{\mathcal{S}}(k) = \left\{a \in [0, \overline{a}] : \phi^{\mathcal{S}}(-a, t(k), k) \leq 0\right\}.$$

Type k voters' policy latitude $\xi^{\mathcal{S}}(k)$ *is the maximum of the k-proof set, i.e., $\xi^{\mathcal{S}}(k) = \max \Xi^{\mathcal{S}}(k)$. Type k voters are* disciplining *if their policy latitude determines policy polarization, i.e., $a^{\mathcal{S},q} = \xi^{\mathcal{S}}(k)$.*

Note that a voter's policy latitude decreases (resp. increases) with his susceptibility (resp. resistance) to candidate R's policy deviations.

4.2 Main characterization

The next theorem gives a full characterization of the equilibrium policy set.

Theorem 2. *Assume Assumptions 1 and 2. For any pair of personalization technology $\mathcal{S} \in \{b, p\}$ and population function q, policy polarization is strictly positive, and all policies between zero and policy polarization can arise in equilibrium, i.e., $a^{\mathcal{S},q} > 0$ and $\mathcal{E}^{\mathcal{S},q} = \left[0, a^{\mathcal{S},q}\right]$. Moreover, disciplining voters always exist, and their identities are as follows.*

(i) In the broadcast case, median voters are always disciplining, i.e., $a^{b,q} = \xi^{b}(0)\ \forall q$.

(ii) In the personalized case, median voters are disciplining if they constitute a majority of the population. Otherwise voters with the smallest policy latitude are disciplining, i.e.,

$$a^{p,q} = \begin{cases} \xi^{p}(0) & \text{if } q(0) > 1/2, \\ \min\limits_{k \in \mathcal{K}} \xi^{p}(k) & \text{otherwise.} \end{cases}$$

In what follows, we will sketch the proof for Theorem 2 in Section 4.2.1 and examine the determinants of disciplining voters in Section 4.2.2. Results of Section 4.2.1 are restated and proven for arbitrary finite types of voters holding general policy preferences in Appendices A and E.2, respectively. Results of Section 4.2.2 are proven for the distance utility function in Appendix E.4.

4.2.1 Proof sketch

Broadcast case Since all voters receive the same voting recommendation in this case, a deviation of candidate R is *profitable*, i.e., strictly increases his winning probability, if and only it attracts a majority of voters. Under the usual assumption that voters' policy preferences exhibit increasing differences, this is equivalent to attracting median voters, so a symmetric policy profile $\langle -a, a \rangle$ with $a \geq 0$ can arise in equilibrium if and only if a belongs to the 0-proof set:

$$\mathcal{E}^{b,q} = \Xi^{b}(0)\ \ \forall q.$$

Personalized case In this case, median voters remain disciplining if they constitute a majority of the population. Otherwise no type of voter alone forms a majority coalition, and a deviation is profitable if it attracts *any* type of voter, *holding other things constant.* The reason is pivotality: since the infomediary can now offer conditionally independent signals to different types of voters, the above deviation strictly increases candidate R's winning probability when the remaining voters disagree about which candidate to vote for.

The above argument leaves open the question of whether attracting some voters would cause the repulsion of others. Fortunately, this concern is ruled out by the next lemma.

Lemma 3. *Assume Assumptions 1 and 2. In the case where* $\mathcal{S} = p$ *and* $q(0) \leq 1/2$*, a symmetric policy profile* $\langle -a, a \rangle$ *with* $a \geq 0$ *can arise in equilibrium if and only if no unilateral deviation of candidate* R *from it to* $a' \in [-a, a)$ *attracts any voter whose bliss point lies inside* $[-a, a]$*.*

The proof of Lemma 3 examines two kinds of (global) deviations: (1) $a' \notin [-a, a]$ and (2) $a' \in [-a, a)$. By committing the first kind of deviation, candidate R may indeed attract, say, right-wing voters. But this success must cause the repulsion of left-wing voters and, hence, cannot increase candidate R's winning probability. Meanwhile, the second kind of deviation repels no voter, and it attracts no voter whose bliss point lies outside $[-a, a]$. If, in addition, it attracts no voter whose bliss point lies inside $[-a, a]$, then the original policy profile $\langle -a, a \rangle$ can be sustained in equilibrium.[25]

By Lemma 3, a policy profile $\langle -a, a \rangle$ with $a \in [0, t(1))$ can arise in equilibrium if and only a belongs to the 0-proof set, and a policy profile $\langle -a, a \rangle$ with $a \in [t(1), \overline{a}]$ can arise in equilibrium if and only a belongs to the k-proof set for all $k \in \mathcal{K}$:

$$\mathcal{E}^{p,q} = \left([0, t(1)) \bigcap \Xi^p(0) \right) \bigcup \left([t(1), \overline{a}] \cap \bigcap_{k \in \mathcal{K}} \Xi^p(k) \right).$$

Equilibrium policy set It remains to prove existence of disciplining voters with a positive policy latitude, and to demonstrate that all policies between zero and disciplining voters' policy latitude can arise in equilibrium. For starters, notice that the k-proof set must con-

[25] Note that replacing win-maximizing candidates with vote-maximizing ones or those with mixed motives won't alter the above arguments and, hence, Lemma 3 or Theorem 2.

tain $|t(k)|$. For $k \geq 0$, this is true because of uniform strict obedience: at $\langle -|t(k)|, |t(k)|\rangle$, candidate R is already adopting type k voters' most preferred position and so cannot attract them by committing any deviation, i.e., $v(-|t(k)|, |t(k)|, k) + \mu_L^{\mathcal{S}}(|t(k)|, k) < 0$. After more work, we can show that all policies between $|t(k)|$ and type k voters' policy latitude belong to the k-proof set. We summarize these results in the next lemma.

Lemma 4. *The following hold under Assumptions 1 and 2.*

(i) $\xi^b(0) > 0$ *and* $\Xi^{\mathcal{S}}(0) = \left[0, \xi^b(0)\right]$.

(ii) $\xi^p(k) > |t(k)|$ *and* $[t(k), \xi^p(k)] \subseteq \Xi^p(k)\ \forall k \in \mathcal{K}$.

Based on Lemmas 3 and 4, we develop an algorithm for computing the equilibrium policy set in the appendix. For the case where $\mathcal{S} = p$ and $q(0) \leq 1/2$, straightforward algebra shows that $\mathcal{E}^{p,q} = [0, \xi^p(0)] \cup \emptyset = [0, \min_{k\in\mathcal{K}} \xi^p(k)]$ if $\xi^p(0) < t(1)$ and $\mathcal{E}^{p,q} = [0, t(1)) \cup [t(1), \min_{k\in\mathcal{K}} \xi^p(k)] = [0, \min_{k\in\mathcal{K}} \xi^p(k)]$ if $\xi^p(0) \geq t(1)$, thus completing the proof of Theorem 2.[26]

Remark Personalized news aggregation allows candidates to benefit from attracting extreme voters in addition to median voters. Voters with the smallest policy latitude are the most susceptible to policy deviations and hence the easiest to attract, so the deviation to their bliss point constrains policy polarization.[27] While this effect makes polarization harder to sustain compared to the broadcast case, there is another effect stemming from changes in marginal news distributions and, hence, policy latitudes. We now turn to this second effect, and will come back to the first effect in Section 6.1.

[26] With arbitrary finite types of voters, policy polarization is shown to equal the minimal policy latitude across all voter coalitions that can influence the election outcome (see Appendix A). If members of the disciplining coalition have moderate types, then policy polarization wouldn't exceed all voters' bliss points. For now, the reader can think of the disciplining voter as representing the disciplining coalition of the general model. According to Barber and McCarty (2015), there is good evidence that candidates adopt more extreme positions than the representative voters of their constituencies.

[27] For similar reasons, there exist situations in which deviations to extreme voters' bliss points are profitable to candidate R, whereas the deviation to median voters' bliss point isn't. Detailed constructions are available upon request.

4.2.2 Disciplining voter

The next lemma decomposes a voter's policy latitude into (i) the negative of his bliss point and (ii) the magnitude of his belief about candidate R's fitness following unfavorable news.[28]

Lemma 5. *Let everything be as in Theorem 2. Then the following happen when $\overline{a}$ is sufficiently large.*

(i) $\xi^b(0) = |\mu_L^b| \coloneqq \mu_L^b(t(1))$.

(ii) $\forall k \in \mathcal{K}$, $\xi^p(k) = -t(k) + |\mu_L^p(k)|$, *where* $\mu_L^p(k) \coloneqq \mu_L^p(|t(k)|, k)$.

As a voter's bliss point increases, he more prefers candidate R's policies, which reduces his policy latitude. At the same time, he seeks bigger occasional surprises from news consumption and so becomes more pessimistic about candidate R's fitness following unfavorable news, which increases his policy latitude. This tension between policy preference and belief pins down the disciplining voter.

To illustrate, consider candidate R's choice between attracting his base (right-wing voters) and attracting his opposition (left-wing voters). While right-wing voters most prefer candidate R's policies, they are the most pessimistic about his fitness following unfavorable news. In contrast, left-wing voters are the most optimistic about candidate R's fitness following unfavorable news, yet they have an inherent preference against his policies.[29] In the case where $-t(-1) + |\mu_L^p(-1)| < -t(1) + |\mu_L^p(1)|$, left-wing voters have a smaller policy latitude and hence are easier to attract than right-wing voters. Using symmetry $\mu_L^p(-k) = -\mu_R^p(k)$ and $t(-k) = -t(k)$ to simplify, we obtain

$$|\mu_L^p(1)| - \mu_R^p(1) > 2t(1), \tag{$*$}$$

[28] Note that in the broadcast case, the relevant voter is the representative voter with zero bliss point.

[29] This trade-off, while delicate, seems an inevitable consequence of personalization. In related but distinct settings, e.g., personalized campaign, it has been shown that advances in micro-targeting technologies could introduce the trade-off between mobilizing base voters and persuading swing voters to the equilibrium reasoning about policy platforms (Herrera, Levine, and Martinelli (2008); Prummer (2020)). According to Hersh (2015), such a trade-off creates identification problems that can nevertheless be solved by conducting large-scale interviews among campaign staff and volunteers. We discuss how methods akin to that of Hersh (2015) could alleviate our identification problem in Section 7.

which says that extreme voters' signals must be sufficiently skewed that the beliefs induced by the occasional big surprise and own-party bias differ by a significant amount. The next example reduces Condition $(*)$ to model primitives in the case of entropy attention cost.[30]

Example 1. When attention cost is Shannon entropy-based, Condition $(*)$ holds when the marginal attention cost λ is sufficiently high and extreme voters' policy preferences (parameterized by $t(1)$) are sufficiently extreme (as depicted in Figure 1).

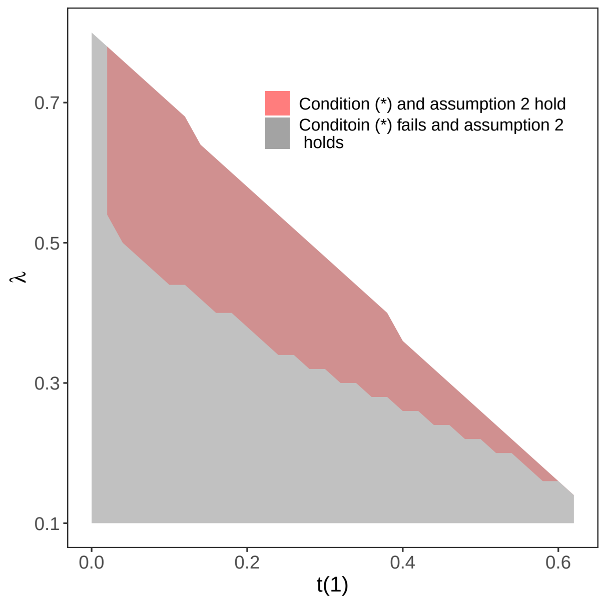

Figure 1: Condition $(*)$: $u(a,k) = -|t(k) - a|$ and entropy attention cost.

As λ increases and so paying attention becomes more costly, the infomediary makes news signals less Blackwell informative to prevent voters from tuning out (Lemma 12 of Appendix E.3 formalizes this claim). In the process of doing so, the infomediary is reluctant to reduce $|\mu_L^p(1)|$, the occasional big surprise that makes news consumption useful for right-wing voters. Instead, she reduces $\mu_R^p(1)$ significantly, which relaxes Condition $(*)$.

Meanwhile, as $t(1)$ increases, right-wing voters more prefer candidate R's policies, so to convince them to vote for candidate L requires a stronger belief than before, i.e., $|\mu_L^p(1)|$ must increase. Also since news consumption becomes less useful to them, $\mu_R^p(1)$ must

[30] In an earlier version of this paper, we did Example 1 for alternative cost function hs and obtained qualitatively similar results to those presented above. The material is available upon request.

decrease to prevent them from tuning out. Thus $|\mu_L^p(1)| - \mu_R^p(1)$ increases, which relaxes Condition $(*)$ when $t(1)$ is sufficiently large. $\diamondsuit$

5 Comparative statics

This section examines how equilibrium policies depend on the personalization technology $\mathcal{S}$, voters' marginal attention cost λ and their population distribution q. Since all policies between zero and policy polarization can arise in equilibrium, it is without loss to focus on the comparative statics of policy polarization. As policy polarization increases, the equilibrium policy set increases in the strong set order.

5.1 Personalization technology

The next proposition prescribes the sufficient and necessary condition for personalized news aggregation to increase policy polarization.

Proposition 1. *Fix any population function q, assume Assumptions 1 and 2, and let $\overline{a}$ be large. Then policy polarization is strictly higher in the personalized case than in the broadcast case if and only if in the personalized case, either (i) median voters are disciplining, or (ii) extreme voters are disciplining and have a bigger policy latitude than the median voters in the broadcast case, i.e.,*

$$\xi^b(0) < \min\{\xi^p(1), \xi^p(-1)\}. \qquad (**)$$

Condition $(**)$ automatically holds if $\xi^b(0) \leq t(1)$. If $\xi^b(0) > t(1)$, then Condition $(**)$ is equivalent to:

(i) $|\mu_L^p(1)| - |\mu_L^b| > t(1)$ if right-wing voters are disciplining in the personalized case;

(ii) $|t(-1)| > |\mu_L^b| - |\mu_L^p(-1)|$ if left-wing voters are disciplining in the personalized case.

Proposition 1 follows immediately from Theorem 2 and Lemma 5. Proposition 1(i) exploits the fact that median voters' personalized signal is more Blackwell informative than the optimal broadcast signal, so their policy latitude increases as news aggregation becomes personalized. The past decade has witnessed a rise in the "apathy and indifference" that centrist voters display to candidates' policy positions (see, e.g., Barber and McCarty (2015)). Proposition 1(i) suggests that this trend could persist due to personalized news aggregation.

Proposition 1(ii) shows that if extreme voters are disciplining in the personalized case, then the skewness of their personalized signals is crucial for sustaining greater policy polarization than in the broadcast case. The role of skewness differs according to which type of extreme voter is disciplining. If right-wing voters are disciplining, then the only explanation for why they could have a bigger policy latitude than the median voters in the broadcast case must be the occasional big surprise of their personalized signal: in order to satisfy Condition ($**$), right-wing voters must be significantly more pessimistic about candidate R's fitness following unfavorable news than the median voters in the broadcast case, i.e., $|\mu_L^p(1)| - |\mu_L^b| > t(1)$.

In the case where opposition voters are disciplining, extreme voters' personalized signals must be sufficiently skewed in order to satisfy Condition ($*$). Thus when contemplating deviations from the equilibrium policy profile, candidate R wouldn't target his base, because doing so is either needless (in the likely event where he already captures his base) or futile (in the unlikely event where the base is convinced of his unfitness). Instead, he appeals to his opposition, which is itself challenging if the latter has a strong preference against his policies, i.e., $|t(-1)| > |\mu_L^b| - |\mu_L^p(-1)|$. Note the role of skewness in the above argument, which is crucial yet indirect.

The next example reduces Condition ($**$) to model primitives in the case of entropy attention cost.

Example 1 (Continued)**.** Here we focus on the case where base voters are disciplining, i.e., Condition ($*$) fails, so Condition ($**$) becomes $|\mu_L^p(1)| - |\mu_L^b| > t(1)$ (Appendix D covers the case where opposition voters are disciplining). In Section 4.2.2, we already explained why large values of marginal attention cost λ and policy preference parameter $t(1)$ can lead right-wing voters to seek big surprise $|\mu_L^p(1)|$s from news consumption. In the current context, this suggests that increasing λ and $t(1)$ could (and indeed) relax Condition ($**$) (as depicted in Figure 2). Below we explain why this intuition remains true after we introduce $|\mu_L^b|$ into the analysis.

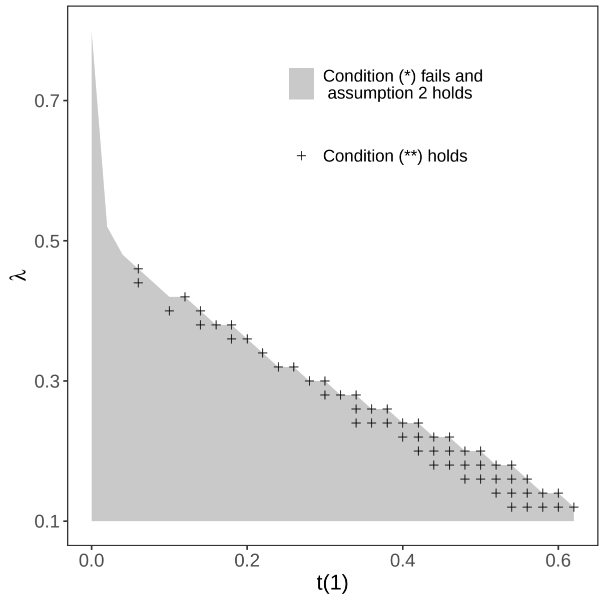

Figure 2: Condition (∗∗): $u(a,k) = -|t(k) - a|$, entropy attention cost, Condition (∗) fails.

First, recall that as λ increases, the infomediary makes news signals less Blackwell informative to prevent voters from tuning out. In the personalized case, she can reduce $|\mu_L^p(1)|$ and $\mu_R^p(1)$ differently to make news consumption still useful for right-wing voters (as discussed in Section 4.2.2). Such flexibility is absent in the broadcast case, where $|\mu_L^b|$ and μ_R^b must be reduced by the same magnitude. Thus $|\mu_L^b|$ decreases faster than $|\mu_L^p(1)|$, which relaxes Condition (∗∗).

Next, notice that as $t(1)$ increases, extreme voters find news consumption less useful, so the broadcast signal must become less Blackwell informative to prevent them from tuning out, i.e., $|\mu_L^b|$ must decrease. Meanwhile, $|\mu_L^p(1)|$ must increase, because to convince right-wing voters to vote for candidate L requires a bigger occasional surprise than before. Thus $|\mu_L^b(1)| - |\mu_L^b|$ increases, which relaxes Condition (∗∗) when $t(1)$ is sufficiently large. $\diamondsuit$

5.2 Marginal attention cost

The next proposition shows that policy polarization is decreasing in voters' marginal attention cost.

Proposition 2. *Let $\lambda'' > \lambda' > 0$ be two marginal attention costs such that the corresponding environments satisfy Assumption 1 and 2. For each $\lambda \in \{\lambda', \lambda''\}$, write the policy polarization under personalization technology $\mathcal{S}$ and population function q as $a^{\mathcal{S},q}(\lambda)$. Then $a^{\mathcal{S},q}(\lambda'') < a^{\mathcal{S},q}(\lambda')$.*

Proposition 2 holds because increasing the marginal attention cost makes optimal news signals less Blackwell informative and, in turn, reduces voters' policy latitudes. In Appendix B, we allow multiple infomediaries to compete for voters using signals that maximize their expected utilities rather than attention. We find that introducing competition between infomediaries–which is advocated by the British government as a preferable way of regulating tech giants to Warren's ban of personalization (The Digital Competition Expert Panel (2019))–is mathematically equivalent to increasing the marginal attention cost in the monopolistic personalized case, so its policy polarization effect is negative by Proposition 2. Intuitively, optimal news signals overfeed voters with information about candidate valence compared to competitive signals. Competition corrects this overfeeding problem and reduces policy polarization.

5.3 Population distribution

Recently, a growing body of the literature has been devoted to the understanding of voter polarization, also termed *mass polarization*. Notably, Fiorina and Abrams (2008) define mass polarization as a bimodal distribution of voters' policy preferences on a liberal-conservative scale, and Gentzkow (2016) advocates using the average ideological distance between Democrats and Republicans to measure mass polarization. Inspired by these authors, we define increasing mass polarization as a mean-preserving spread of voters' policy preferences and examine its consequence on policy polarization. Note that our exercise is purely conceptual, since evidence on increasing mass polarization is mixed at best (as argued forcefully by the above authors). Our goal is to call the reader's attention to the following message: with personalized news aggregation, increasing mass polarization may decrease rather than increase policy polarization.

Proposition 3. *Under Assumptions 1 and 2,* $a^{p,q} \geq a^{p,q'}$ *for any two population functions* q *and* q' *such that* $q(0) > q'(0)$*, and* $a^{p,q} > a^{p,q'}$ *if and only if* $q(0) > 1/2 \geq q'(0)$ *and* $\min\{\xi^p(-1), \xi^p(1)\} < \xi^p(0)$.

Proposition 3 is immediate from Theorem 2: as we keep redistributing voters' population from the center to the margin, candidates would eventually benefit from attracting extreme voters in addition to median voters. If extreme voters have smaller policy latitudes than median voters (as in Example 1), then a reduction in policy polarization would ensue. While caution should be exercised when extrapolating Proposition 3 to general environments, its warning message nevertheless warrants attention. Appendix A.4.2 proves a similar result for the case of general voters and quadratic attention cost.

6 Extensions

6.1 Joint news distribution

In Appendix A, we relax the assumption that news signals are conditionally independent across market segments and instead consider all joint news distributions that are consistent with the marginal distributions solved in Section 3. We also extend the analysis to arbitrary finite types of voters holding general policy preferences.

Our analysis leverages a new concept called *influential coalition*. Loosely speaking, a coalition of voters is influential if attracting all its members, holding other things constant, strictly increases the deviating candidate's winning probability. In the broadcast case, all voters consume the same signal, so a coalition of voters is influential if and only if it is a majority coalition. In the personalized case, non-majority coalitions can be influential, due to the imperfect correlation between the signals consumed by different voters. The next table compiles the influential coalitions in the baseline model.

	$\mathcal{S} = b$	$\mathcal{S} = p$
$q(0) > 1/2$	majority coalitions	majority coalitions
$q(0) < 1/2$	majority coalitions	nonempty coalitions

Table 1: influential coalitions under any symmetric policy profile $\langle -a, a \rangle$ with $a \geq 0$: baseline model.

As it turns out, influential coalitions depend on the joint news distribution only through the *news configuration* $\boldsymbol{\chi}$—a matrix that compiles all voting recommendation profiles that occur with positive probabilities. Let $\mathcal{C}$s denote the influential coalitions formed under a given pair $\langle \boldsymbol{\chi}, q \rangle$ of news configuration and voter population distribution. Theorem 3 of Appendix A.3 shows that under certain regularity conditions, the set of nonnegative policy as such that the symmetric policy profile $\langle -a, a \rangle$ can arise in equilibrium under personalization technology $\mathcal{S}$, news configuration $\boldsymbol{\chi}$ and population function q is

$$\left[0, \min_{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}, q \rangle} \xi^{\mathcal{S}}(\mathcal{C})\right],$$

where $\xi^{\mathcal{S}}(\mathcal{C})$ is the policy latitude of $\mathcal{C}$ and depends *only* on marginal news distributions. The messages are twofold. First, in general, policy polarization is disciplined by the influential coalition with the smallest policy latitude. Second, factors that enrich influential coalitions through the news configuration $\boldsymbol{\chi}$ or the population distribution q reduce policy polarization, holding marginal news distributions constant.

Implications Theorem 3 sheds light on the role of joint news distribution in determining policy polarization. Consider first the transition from broadcast news and personalized news. As demonstrated earlier, personalized news aggregation enriches influential coalitions and changes marginal news distributions compared to the case of broadcast news. Since the first effect reduces policy polarization, the reader can safely attribute the increase in policy polarization as shown in Proposition 1 to the second effect (Proposition 5 of Appendix A.4.1 formalizes this claim).

Consider next the case of personalized news. In Appendix A.4.1, we show that influential coalitions are the *richest for any given voter population distribution* when news signals are conditionally independent across voters, and they are the *richest across all scenarios* if, in addition, voters' population distribution is uniform across types. Two implications are immediate.

- Relaxing the assumption of conditional independence can *only* increase policy polarization, whereas dividing the same type of votes into multiple subgroups and providing them with imperfectly correlated signals can *only* decrease policy polarization.
- The term $\min_{k\in\mathcal{K}} \xi^p(k)$ (as in Case $q(0) \leq 1/2$ of Theorem 2(ii)) constitutes the *exact lower bound* for the policy polarization that can be attained in the personalized case. As long as this lower bound stays positive, changing the environment (e.g., enriching voters' type space, dividing the same type of voters into multiple subgroups) wouldn't render policy polarization trivial.

Recently, news aggregators such as Allsides.com have been built to battle the rising polarization through providing users with unbiased viewpoints. In our model, providing extreme voters with unbiased viewpoints is mathematically equivalent to keeping their marginal news distributions the same as in the broadcast case, but relaxing the assumption that news signals must be perfectly correlated across voters. At the opposite extreme where news signals are conditionally independent across voters, right-wing voters have the smallest policy latitude $-t(1) + |\mu_L^b|$ and are therefore disciplining. In that case, policy polarization is smaller than its counterpart $|\mu_L^b|$ under $\mathcal{S} = b$, and it is smaller than its counterpart under $\mathcal{S} = p$ if either centrist voters or right-wing voters are disciplining in the latter case (so policy polarization equals $|\mu_L^p(0)|$ and $-t(1) + |\mu_L^p(1)|$, respectively).

6.2 Beyond uniform strict obedience

The analysis so far has assumed uniform strict obedience, i.e., the news signal consumed by any voter satisfies (SOB) for any given pair of personalization technology and feasible policy profile. Three assumptions together guarantee that this is the case: (i) the state is binary, (ii) optimal news signals are nondegenerate, and (iii) voters face binary decision problems. We now relax these assumptions one by one, showing that uniform strict obedience is sufficient but not necessary for generating policy divergence between candidates.[31]

General state distribution Appendix C extends the model presented in Appendix A (featuring arbitrary finite types of voters with general policy preferences) to general state distributions. The findings are threefold. First, any optimal personalized signal that makes its consumers' participation constraint binding has at most two realizations (for the same reason given in the baseline model). Second, any optimal broadcast signal that induces consumption from all voters and makes some voters' participation constraints binding has at most three realizations. Finally, policy polarization equals zero in the new case of three signal realizations.[32]

To develop intuition for the broadcast case, we aggregate voters with binding participation constraints into a representative voter (as we did in the baseline model). If the marginal attention cost is constant across voters (as we assumed in the baseline model), then only the most extreme types of voters can have binding participation constraints.[33] Together, these voters make three decisions in total: LL, LR and RR, where the first and second letters stand for the voting decisions made by the most left-leaning type and the most right-leaning type, respectively. The optimal broadcast signal maximizes the attention of the representative voter acting on behalf of these voters and so prescribes at most three

[31] Finding the sufficient and necessary condition for NARI to generate policy divergence between candidates is a challenging yet fascinating topic for future research.

[32] We require all voters to consume news in order to maintain consistency between the baseline model. We will relax this assumption later in this section.

[33] We assume a constant marginal attention cost across voters to keep consistency between the baseline model. Generically, at most two types of voters can have binding participation constraints in a symmetric environment.

voting recommendation profiles. For the case where all three decision profiles LL, LR and RR are recommended with positive probabilities, it can be shown that the posterior mean of the state given the recommendation profile LR must equal zero. As argued in the footnote, this implies that $\langle 0, 0 \rangle$ is the only symmetric policy profile that can arise in equilibrium.[34]

Exclusion from news consumption The next example shows that policy polarization could still be positive even if extreme voters are excluded from news consumption, suggesting that abstentions from news consumption due to high marginal attention costs or mistaken exclusions of some voters from news consumption due to model misspecifications wouldn't necessarily render policy polarization trivial.

Example 2. Let everything be as in the baseline model except that extreme voters are excluded from news consumption. As shown in Appendix E.4, all symmetric policy profile $\langle -a, a \rangle$s with $a \in \left[0, \underline{\xi}\right]$ can arise in equilibrium, where $\underline{\xi} = \min\left\{t\left(1\right), \xi^{\mathcal{S}}\left(0\right)\right\}$. ♢

Beyond binary decision problems So far we've focused on the role of news consumption in improving expressive voting decisions. While other motives for news consumption (e.g., entertainment) are certainly important and should definitely be incorporated into future researches, we also believe that limited attention is ubiquitous and, as pointed out by Downs (1957), "constitutes a basic step towards understanding politics." By now, it is well known that if voters face finite decision problems, e.g., categorical thinking, then the outcome of NARI is a finite signal (Matějka and McKay (2015)). Indeed, the same conclusion can be drawn for some decision problems in which both the state space and decision space are infinite (Jung et al. (2019)). Below we give an example in which policy polarization is positive, despite that extreme voters observe more than two signal realizations and do not always have preferences between candidates. Providing microfoundations for these signals, while desirable, is beyond the scope of the current paper.

[34] For any symmetric policy profile $\langle -a, a \rangle$ with $a > 0$, the deviation to $a' = 0$ weakly increases candidate R's winning probability when the voting recommendation profile is either LL or RR (Lemma 3), and it strictly increases candidate R's winning probability when the voting recommendation profile is LR. In contrast, no unilateral deviation from $\langle 0, 0 \rangle$ increases candidate R's winning probability when the voting recommendation profile is LL or RR (Lemma 3) or LR.

Example 3. In the baseline model, suppose extreme voters' personalized signals have three realizations L, M and R, and they strictly prefer candidate L to R given signal realizaton L, are indifferent between the two candidates given signal realization M, and strictly prefer candidate R to L given signal realization R. As shown in Appendix E.4, the policy profile $\langle -t(1), t(1)\rangle$ can arise in equilibrium if median voters' policy latitude exceeds $t(1)$. $\diamondsuit$

6.3 Other extensions

Skewness effect vs. level effect Compared to the optimal broadcast signal, optimal personalized signals for extreme voters are skewed and attract a greater amount of attention from its consumers. Hereafter we shall name these effects as the *skewness effect* and *level effect*, respectively. The level effect increases policy polarization because of the trade-off between policy and valence. The next example shows that the skewness effect alone can increase policy polarization.

Example 1 (Continued)**.** In the current example, only extreme voters can be disciplining in the personalized case. Fix these voters' attention to the broadcast level $I\left(\Pi^b(a)\right)$ and solve for their most preferred signals. Then compute the minimal policy latitude among these voters and compare the result to the policy polarization in the broadcast case. If the former exceeds the latter, than we say that skewness effect alone could increase policy polarization.[35] As depicted in Figure 3, this is indeed the case for some parameter values. $\diamondsuit$

Heterogeneous marginal attention cost A consequence of assuming a constant marginal attention cost across voters is that median voters pay more attention than extreme voters in the personalized case. Allowing the marginal attention cost to differ across voters helps reconcile our findings with the stylized fact that extreme voters tend to be more attentive to politics than centrist voters (see, e.g., Barber and McCarty (2015)). The only qualitative

[35] This is certainly not the only way to single out the skewness effect. An alternative approach, which generates qualitatively similar results, is to fix voters' bandwidths upfront.

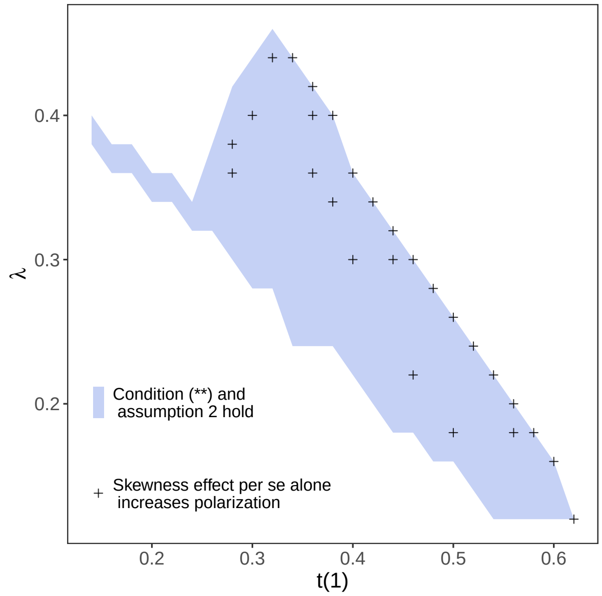

Figure 3: Skewness effect: $u(a,k) = -|t(k) - a|$, entropy attention cost.

difference this change might *but not necessarily* cause is to make median voters' participation constraint binding and extreme voters' participation constraints slack in the broadcast case. In that case, the broadcast signal coincides with median voters' personalized signal, so personalized news aggregation could only decrease policy polarization.

Alternative personalization technologies Our analysis can be generalized beyond the two basic personalization technologies. In Section 6.1, we already discussed the consequence of dividing the same type of voters into multiple subgroups. For voter partitions that are finer than the broadcast technology yet coarser than the personalized technology, we can first aggregate, for each market segment, voters with binding participation constraints into a representative voter, and then solve for the optimal personalized signals for representative voters. The aggregation problem is the most straightforward if all market segments consist of consecutive types of voters and do not simultaneously contain left-leaning and right-leaning voters, because in that case, at most one type of voter's participation constraint is binding within each market segment.

7 Concluding remarks

Tech-enabled personalization is ubiquitous and seems to maximize social surplus by best serving consumers' needs. In our opinion, a caveat to this argument is that tech giants now operate major news aggregators and could therefore affect political decisions and outcomes. The current paper studies the policy polarization effect of personalized news aggregation for rational inattentive voters. We find that after taking this effect into account, the welfare consequences of many hotly debated regulatory proposals to tame tech giants become less clear-cut. For example, while both banning news personalization and introducing competition between infomediaries clearly make voters better off and the infomediary worse off, holding policies constant, they also affect policy polarization and, in turn, players' equilibrium utilities in a delicate manner. For this reason, we suggest that caution be exercised and our characterization for policy polarization be considered when evaluating the overall impacts of these proposals.

Our characterization for equilibrium policies holds as long as voters' signals are binary and satisfy (SOB). While NARI provides a foundation for such signal structures, we don't deny the existence of other foundations and are open to applying our theory to other settings. In a companion paper Li, Hu, and Segal (2020), we build on the tools developed in the current paper to study the effects of personalized news aggregation on electoral accountability and selection. One could also imagine IO applications of our theory, e.g., studying the effects of personalized product recommendation on product differentiation and consumer welfare.

So far, we have restricted news consumers to single-homing, i.e., obtain news from a single infomediary. Given the monopoly power wielded by tech giants, we believe we have taken a useful first step towards studying news aggregation for rational inattentive voters. At the same time, it is tempting to believe in the high-level idea that all media outlets are, in some sense, information aggregators. Formalizing this idea requires major theoretical innovations which, at the end of the day, allow us to establish the multi-homing counterpart of the NARI model.

At least three challenges must be tackled before we can take our theory to data. First and foremost, it is imperative to test the model of posterior-separable attention cost in the context of political decision-making, building on the tools developed by Ambuehl, Ockenfels, and Stewart (2019) and Dean and Nelighz (2019) for testing posterior separability in the lab. Second, it is important to identify exogenous shocks to major news aggregators, which, in our opinion, stem mainly from the experiments conducted by tech companies, or from unexpected regulatory interventions such as Spain's shutdown of Google News in 2014 (Athey, Mobius, and Pal (2017)). Finally, our theory predicts that with personzalized news aggregation, the identity of the disciplining voter would depend endogenously on voters' policy latitudes. In the context of personalized campaign, Hersh (2015) pioneers the use of large-scale interviews among campaign staff and volunteers to deal with a similar identification problem, whereby advances in micro-targeting technologies make the target voter group difficult to observe to analysts (see Footnote 29 for further discussions). We believe methods akin to that developed by Hersh (2015) could alleviate our identification problem, and we hope someone, maybe us, will put this idea to practice in the future.

A General model

This appendix has two purposes. The first purpose is to extend the baseline model to general voters with a finite set $\mathcal{K} = \{-K, \cdots, 0, \cdots, K\}$ of types, a population function $q : \mathcal{K} \to \mathbb{R}_{++}$, and a utility function $u : \mathcal{A} \times \mathcal{K} \to \mathbb{R}$. Throughout this appendix, K can be any positive integer, q has support $\mathcal{K}$ and is symmetric around zero, and u satisfies the following assumption.[36]

Assumption 3. *The following hold for any $k \in \mathcal{K}$.*

continuity and concavity $u(\cdot, k)$ *is continuous and concave.*

[36]Assumption 3 holds for many standard utility functions in the literature, e.g., $-|t(k) - a|$ and $-(t(k) - a)^2$.

symmetry $u(a,k) = u(-a,-k)$ *for any* $a \in \mathcal{A}$.

inverted V-shape $u(\cdot,k)$ *is strictly increasing on* $[-\overline{a}, t(k)]$ *and is strictly decreasing on* $[t(k), \overline{a}]$*, where the bliss point function* $t : \mathcal{K} \to int(\mathcal{A})$ *is strictly increasing and satisfies* $t(k) = -t(-k)$ *for any* $k \in \mathcal{K}$.

In addition,

increasing differences $u(a,k) - u(a',k)$ *is increasing in* k *for any* $a > a'$. *For any* $a > 0$, $u(a,k) - u(-a,k)$ *is strictly positive if* $k > 0$*, equals zero if* $k = 0$ *and is strictly negative if* $k < 0$.

The second purpose of this appendix is to relax the assumption that news is conditionally independent across market segments. To this end, we develop new concepts in Appendix A.1 and conduct equilibrium analyses in Appendices A.2-A.4. Omitted proofs can be found in Appendix E.2.

A.1 Key concepts

Joint news distribution A *joint news distribution* is a tuple $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^- \rangle$ of *news configuration* $\boldsymbol{\chi}$ and *probability vectors* $\mathbf{b}^+$ and $\mathbf{b}^-$. The news configuration $\boldsymbol{\chi}$ is a matrix of $|\mathcal{K}|$ rows. Each column of $\boldsymbol{\chi}$ constitutes a profile of the voting recommendations to type $-K, \cdots, K$ voters that occurs with a strictly positive probability. Each entry of $\boldsymbol{\chi}$ is either 0 or 1, where 0 means that candidate R is disapproved and 1 means that he is endorsed. For example, the news configuration is

$$\boldsymbol{\chi}^* = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}$$

if $\mathcal{S} = b$, and it is

$$\boldsymbol{\chi}^{**} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 1 & \cdots & 0 & \cdots & 1 \\ 0 & 0 & 1 & \cdots & 0 & 1 & \cdots & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 1 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 1 & \cdots & 1 \end{bmatrix}}_{2^{|\mathcal{K}|} \text{ columns}}$$

if $\mathcal{S} = p$ and news signals are conditionally independent across market segments. The vectors $\mathbf{b}^{+}$ and $\mathbf{b}^{-}$ compile the probabilities that columns of $\boldsymbol{\chi}$ occur in states $\omega = 1$ and $\omega = -1$, respectively. By definition, all elements of $\mathbf{b}^{+}$ or $\mathbf{b}^{-}$ are strictly positive and add up to one.

We restrict attention to *symmetric* joint news distributions. To formally define symmetry, let $\mathbf{x}$ be a generic voting recommendation profile, $\mathbf{1}$ be the $|\mathcal{K}|$-vector of all ones, and

$$\mathbf{P} = \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix}$$

be a $|\mathcal{K}| \times |\mathcal{K}|$ permutation matrix. Define the *symmetry operator* Σ as

$$\Sigma \circ \mathbf{x} = \mathbf{P}\left(\mathbf{1} - \mathbf{x}\right),$$

so $\mathbf{x}$ recommends candidate $z \in \{L, R\}$ to type k voters if and only if $\Sigma \circ \mathbf{x}$ recommends candidate $-z$ to type $-k$ voters.

Definition 5. *A news configuration* $\boldsymbol{\chi}$ *is* symmetric *if for any* $m \in \{1, \cdots, \text{columns}(\boldsymbol{\chi})\}$*, there exists* $n \in \{1, \cdots, \text{columns}(\boldsymbol{\chi})\}$ *such that* $\Sigma \circ [\boldsymbol{\chi}]_m = [\boldsymbol{\chi}]_n$*. A joint news distribution* $\langle \boldsymbol{\chi}, \mathbf{b}^{+}, \mathbf{b}^{-} \rangle$ *is* symmetric *if* $\boldsymbol{\chi}$ *is symmetric and* $[\mathbf{b}^{+}]_m = [\mathbf{b}^{-}]_n$ *for any* $m, n \in \{1, \cdots, \text{columns}(\boldsymbol{\chi})\}$ *such that* $\Sigma \circ [\boldsymbol{\chi}]_m = [\boldsymbol{\chi}]_n$*.*[37]

[37] $[\cdot]_m$ denotes both the m^{th} entry of a column vector and the m^{th} column of a matrix. columns$(\boldsymbol{\chi})$ denotes the number of columns of $\boldsymbol{\chi}$.

In words, a joint news distribution is symmetric if the probability that a voting recommendation profile $\mathbf{x}$ occurs in state $\omega = 1$ equals the probability that $\Sigma \circ \mathbf{x}$ occurs in state $\omega = -1$. We consider symmetric joint news distributions that are *consistent* with the marginal news distributions solved in Section 3. In Footnote 21, we solved for the probabilities that the news signal consumed by type k voters endorses candidate R in states $\omega = 1$ and $\omega = -1$, respectively. Compiling these probabilities across type $-K, \cdots, K$ voters for any given personalization technology $\mathcal{S} \in \{b, p\}$ and symmetric policy profile $\langle -a, a \rangle$ with $a \geq 0$ yield two $|\mathcal{K}|$-vectors $\boldsymbol{\pi}^{\mathcal{S},+}(a)$ and $\boldsymbol{\pi}^{\mathcal{S},-}(a)$.

Definition 6. *A news configuration $\boldsymbol{\chi}$ is $\langle \mathcal{S}, a \rangle$-*consistent *for some $\mathcal{S} \in \{b, p\}$ and $a \geq 0$ if there exist probability vectors $\mathbf{b}^+$ and $\mathbf{b}^-$ such that the joint news distribution $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^- \rangle$ is $\langle \mathcal{S}, a \rangle$-*consistent*, i.e.,*

$$\boldsymbol{\chi}\mathbf{b}^+ = \boldsymbol{\pi}^{\mathcal{S},+}(a) \;\; \textit{and} \; \boldsymbol{\chi}\mathbf{b}^- = \boldsymbol{\pi}^{\mathcal{S},-}(a).$$

*$\boldsymbol{\chi}$ is $\mathcal{S}$-*consistent *if it is $\langle \mathcal{S}, a \rangle$-consistent for all $a \geq 0$.*

Note that $\boldsymbol{\chi}^*$ is b-consistent and indeed the only $\langle b, a \rangle$-consistent news configuration for any given $a \geq 0$. $\boldsymbol{\chi}^{**}$ is p-consistent, but it is in general not uniquely p-consistent.

Susceptibility to policy deviations The next definition extends Definitions 2 and 3 to sets of voters.

Definition 7. *Under personalization technology $\mathcal{S}$, a unilateral deviation of candidate R from a symmetric policy profile $\langle -a, a \rangle$ with $a \geq 0$ to a'* attracts *a set $\mathcal{D} \subseteq \mathcal{K}$ of voters if it attracts all members of $\mathcal{D}$, i.e., $\phi^{\mathcal{S}}(-a, a', k) > 0 \; \forall k \in \mathcal{D}$. This is equivalent to $\phi^{\mathcal{S}}(-a, a', \mathcal{D}) > 0$, where*

$$\phi^{\mathcal{S}}\left(-a, a', \mathcal{D}\right) = \min_{k \in \mathcal{D}} \phi^{\mathcal{S}}\left(-a, a', k\right)$$

*is the $\mathcal{D}$-*susceptibility *to a' following unfavorable news to candidate R's fitness.*

Influential coalition The next concept is integral to the upcoming analysis.

Definition 8. *Fix any personalization technology* $\mathcal{S} \in \{b, p\}$*, symmetric policy profile* $\langle -a, a \rangle$ *with* $a \geq 0$*, and population function* q*, and let the default be the strictly obedient outcome induced by any joint news distribution* $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^- \rangle$ *that is* $\langle \mathcal{S}, a \rangle$*-consistent. Then a set* $\mathcal{C} \subseteq \mathcal{K}$ *of voters constitutes an* R-influential coalition*, or* influential coalition *for short, if attracting* $\mathcal{C}$ *while holding other things constant strictly increases candidate* R*'s winning probability compared to the default.*

By definition, majority coalitions are influential, and supersets of influential coalitions are influential. In the broadcast case, all voters consume the same news signal, so a coalition of voters is influential if and only if it is a majority coalition. In the personalized case, non-majority coalitions can be influential due to the imperfect correlation between voters' news signals (see Table 1 for an illustration). As the next lemma demonstrates, influential coalitions depend on the joint news distribution only through the news configuration, and they are independent of candidates' policy positions if the news configuration is $\mathcal{S}$-consistent.

Lemma 6. *Let everything be as in Definition 8. Then influential coalitions depend on the joint news distribution* $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^- \rangle$ *only through the news configuration* $\boldsymbol{\chi}$*, and they are independent of the policy profile* $\langle -a, a \rangle$ *if* $\boldsymbol{\chi}$ *is* $\mathcal{S}$*-consistent.*

A.2 Main lemma

The next lemma gives a full characterization of the symmetric policy profiles that can arise in equilibrium, hence extending Lemma 3 of Section 4.2.1 to general voters and joint news distributions.

Lemma 7. *Fix any pair of personalization technology* $\mathcal{S} \in \{b, p\}$ *and population function* q*, and assume Assumptions 1-3. Then the following are equivalent.*

(i) A symmetric policy profile $\langle -a, a \rangle$ *with* $a \geq 0$ *can arise in an equilibrium with a joint news distribution* $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^- \rangle$ *that is* $\langle \mathcal{S}, a \rangle$*-consistent.*

(ii) No unilateral deviation of candidate R from $\langle -a, a\rangle$ to $a' \in [-a, a)$ attracts any influential coalition formed under $\langle \boldsymbol{\chi}, q\rangle$ whose members have ideological bliss points in $[-a, a]$.

In what follows, we will use $\mathcal{E}^{\mathcal{S},\boldsymbol{\chi},q}$ denote the set of the nonnegative policy as such that the symmetric policy profile $\langle -a, a\rangle$ can arise in equilibrium under personalization technology $\mathcal{S}$, news configuration $\boldsymbol{\chi}$ and population function q. Lemma 7 prescribes a two-step procedure for computing this set.

Step 1. Compute the influential coalitions formed under $\langle \boldsymbol{\chi}, q\rangle$. For each $a \geq 0$, check if any unilateral deviation of candidate R from $\langle -a, a\rangle$ to $a' \in [-a, a)$ attracts any influential coalition whose members have bliss points in $[-a, a]$. If the answer is negative, then add a to the temporary output set $\widetilde{\mathcal{E}}^{\mathcal{S},\boldsymbol{\chi},q}$.

Step 2. For each $a \in \widetilde{\mathcal{E}}^{\mathcal{S},\boldsymbol{\chi},q}$, check if $\boldsymbol{\chi}$ is $\langle \mathcal{S}, a\rangle$-consistent. If the answer is negative, then remove a from $\widetilde{\mathcal{E}}^{\mathcal{S},\boldsymbol{\chi},q}$.

The above procedure outputs $\mathcal{E}^{\mathcal{S},\boldsymbol{\chi},q}$. If $\boldsymbol{\chi}$ is $\mathcal{S}$-consistent, then no element of $\widetilde{\mathcal{E}}^{\mathcal{S},\boldsymbol{\chi},q}$ needs to be removed in Step 2, i.e., $\widetilde{\mathcal{E}}^{\mathcal{S},\boldsymbol{\chi},q} = \mathcal{E}^{\mathcal{S},\boldsymbol{\chi},q}$.

A.3 Main theorem

This appendix gives characterizations of the equilibrium policy set $\mathcal{E}^{\mathcal{S},\boldsymbol{\chi},q}$ for $\mathcal{S}$-consistent news configuration $\boldsymbol{\chi}$s. As before, we define the maximum of $\mathcal{E}^{\mathcal{S},\boldsymbol{\chi},q}$ as the policy polarization and denote it by $a^{\mathcal{S},\boldsymbol{\chi},q}$. The next definition generalizes Definition 4 to sets of voters.

Definition 9. *For any personalization technology $\mathcal{S} \in \{b, p\}$ and any set $\mathcal{D} \subseteq \mathcal{K}$ of voters, define the $\mathcal{D}$-*proof set* $\Xi^{\mathcal{S}}(\mathcal{D})$ as the set of nonnegative policy as such that no unilateral deviations of candidate R from the symmetric policy profile $\langle -a, a\rangle$ attract $\mathcal{D}$, i.e.,*

$$\Xi^{\mathcal{D}}(\mathcal{D}) = \left\{ a \geq 0 : \max_{a' \in \mathcal{A}} \phi^{\mathcal{S}}\left(-a, a', \mathcal{D}\right) \leq 0 \right\}.$$

Also define $\mathcal{D}$'s policy latitude $\xi^{\mathcal{S}}(\mathcal{D})$ *as the maximum of the $\mathcal{D}$-proof set, i.e.,*

$$\xi^{\mathcal{S}}(\mathcal{D}) = \max \Xi^{\mathcal{S}}(\mathcal{D}).$$

$\mathcal{D}$ *is* disciplining *under personalization technology $\mathcal{S}$, news configuration $\boldsymbol{\chi}$ and population distribution q if its policy latitude determines the policy polarization, i.e., $a^{\mathcal{S},\chi,q} = \xi^{\mathcal{S}}(\mathcal{D})$.*

We next state the needed assumptions for conducting the upcoming analysis. In addition to Assumptions 1-3, we require that the susceptibility function $\phi^{\mathcal{S}}(-a, a', k)$ must be increasing in a in a local region.[38]

Assumption 4. *$\phi^{\mathcal{S}}(-a, a', k)$ is increasing in a on $[|t(k)|, \overline{a}]$ for any $\mathcal{S} \in \{b, p\}$, $k \in \mathcal{K}$ and $a' \in \mathcal{A}$.*

We now state our main theorem, which extends Theorem 2 of Section 4.2 to general voters and joint news distributions.

Theorem 3. *Fix any personalization technology $\mathcal{S} \in \{b, p\}$, $\mathcal{S}$-consistent news configuration $\boldsymbol{\chi}$, and population function q, and let $\mathcal{C}$s denote the influential coalition formed under $\langle \boldsymbol{\chi}, q \rangle$. Under Assumptions 1-4, $\mathcal{E}^{\mathcal{S},\boldsymbol{\chi},q} = \left[0, a^{\mathcal{S},\boldsymbol{\chi},q}\right]$, where $a^{\mathcal{S},\boldsymbol{\chi},q} = \min_{\mathcal{C}s \text{ formed under } \langle \boldsymbol{\chi}, q \rangle} \xi^{\mathcal{S}}(\mathcal{C}) > 0$.*

Theorem 3 conveys two messages. First, for general voters and joint news distributions, policy polarization is disciplined by the influential coalition with the smallest policy latitude. Second, marginal news distributions affect policy polarization through policy latitudes, whereas the joint news distribution does so through the news configuration, holding marginal news distributions constant.

A.4 Comparative statics

This appendix examines the comparative statics of the equilibrium policy set $\mathcal{E}^{\mathcal{S},\boldsymbol{\chi},q}$ for $\mathcal{S}$-consistent $\boldsymbol{\chi}$s. Since $\mathcal{E}^{\mathcal{S},\boldsymbol{\chi},q} = \left[0, a^{\mathcal{S},\boldsymbol{\chi},q}\right]$, it is without loss to focus on the comparative

[38] Recall that $\phi^{\mathcal{S}}(-a, a', k) := v(-a, a', k) + \mu_L^{\mathcal{S}}(a, k)$. Since $v(-a, a', k)$ is increasing in a on $[|t(k)|, \overline{a}]$ by Assumption 3 **inverted V-shape**, Assumption 4 holds if $\mu_L^{\mathcal{S}}(a, k)$ doesn't vary significantly with a on $[|t(k)|, \overline{a}]$. As demonstrated by Lemma 11 of Appendix E.2.1, this is the case if $\mathcal{S} = b$ or if $\mathcal{S} = p$ and either $u(a, k) = -|t(k) - a|$ or $h(\mu) = \mu^2$.

statics of $a^{\mathcal{S},\boldsymbol{\chi},q}$.

A.4.1 Influential coalitions

This appendix examines factors that affect policy polarization through influential coalitions. The analysis exploits a simple observation, namely enriching the news configuration enriches influential coalitions and, by Theorem 3, reduces policy polarization. Formally, we say that $\boldsymbol{\chi}$ is *richer than* $\boldsymbol{\chi}'$ and write $\boldsymbol{\chi} \succeq \boldsymbol{\chi}'$ if $\boldsymbol{\chi}$ prescribes more voting recommendation profiles than $\boldsymbol{\chi}'$.

Definition 10. *$\boldsymbol{\chi} \succeq \boldsymbol{\chi}'$ if every column of $\boldsymbol{\chi}'$ is a column of $\boldsymbol{\chi}$.*

Lemma 8. *$\{\mathcal{C}s$ formed under $\langle \boldsymbol{\chi}', q\rangle\} \subseteq \{\mathcal{C}s$ formed under $\langle \boldsymbol{\chi}, q\rangle\}$ for any personalization technology $\mathcal{S} \in \{b, p\}$, any $\mathcal{S}$-consistent news configurations $\boldsymbol{\chi}$ and $\boldsymbol{\chi}'$ such that $\boldsymbol{\chi} \succeq \boldsymbol{\chi}'$, and any population function q.*

We examine two implications of Lemma 8. We first consider the case of personalized news, showing that holding marginal news distributions constant, policy polarization is minimized for any given voter population distribution when news signals are conditionally independent across voters, and it is minimized across all scenarios if, in addition, voters' population distribution is uniform across types.

Proposition 4. *Under Assumptions 1-4, $\min_{k \in \mathcal{K}} \xi^p(k) = a^{p,\boldsymbol{\chi}^{**},uniform} \leq a^{p,\boldsymbol{\chi}^{**},q} \leq a^{p,\boldsymbol{\chi},q}$ for any p-consistent news configuration $\boldsymbol{\chi}$ and any population function q.*

Proof. To establish the second inequality, note that $\boldsymbol{\chi}^{**} \succeq \boldsymbol{\chi}$ for any p-consistent $\boldsymbol{\chi}$. To establish the first equality and inequality, note that under $\boldsymbol{\chi}^{**}$ and uniform population distribution, each type of voter is influential, and the collection of influential coalitions $2^{\mathcal{K}} - \{\emptyset\}$ is the richest across all scenarios. Combining these observations with Theorem 3 and Lemma 8 gives the desired result. □

Consider next the transition from broadcast news to personalized news. Since this transition enriches the news configuration, its policy polarization effect is negative, holding other things constant. Given this, the reader can safely attribute the increasing policy polarization shown in Proposition 1 to changes in marginal news distributions only.

Proposition 5. *$\{\mathcal{C}s$ formed under $\langle \boldsymbol{\chi}^*, q \rangle\} \subseteq \{\mathcal{C}s$ formed under $\langle \boldsymbol{\chi}, q \rangle\}$ for any p-consistent news configuration $\boldsymbol{\chi}$ and any population function q.*

Proof. $\forall \boldsymbol{\chi}$ and q as above, $\{\mathcal{C}s$ formed under $\langle \boldsymbol{\chi}, q \rangle\} \supseteq \{$majority coalitions$\}$
$= \{\mathcal{C}s$ formed under $\langle \boldsymbol{\chi}^*, q \rangle\}$ □

A.4.2 Mass polarization

This appendix continues to investigate the policy polarization effect of increasing mass polarization. The next definition is inspired by Fiorina and Abrams (2008) and Gentzkow (2016).

Definition 11. The mass is more polarized *under q' than q if q has second-order stochastic dominance over q' (write $q \succeq_{SOSD} q'$), i.e., $\sum_{k=m}^{K} q(k) \leq \sum_{k=m}^{K} q'(k)\ \forall m = 1, \cdots, K$.*

The analysis assumes quadratic attention cost.

Assumption 5. *$h(\mu) = \mu^2$.*

The next proposition proves a similar result to Proposition 3 for general voters and p-consistent news configurations.

Proposition 6. *Under Assumptions 1-3 and 5, $a^{p,\boldsymbol{\chi},q} \geq a^{p,\boldsymbol{\chi},q'}$ for any p-consistent news configuration $\boldsymbol{\chi}$ and any two population functions q and q' such that $q \succeq_{SOSD} q'$.*

B Competitive infomediaries

In the environment laid out in Appendix A, suppose $m(k) \geq 2$ infomediaries compete for the attention of type k voters for each $k \in \mathcal{K}$. A *market segment* is a pair (k, i), where $k \in \mathcal{K}$ represents the type of the voters being served, and $i \in \{1, \cdots, m(k)\}$ represents the serving infomediary. The population of the voters in market segment (k, i) is $\rho(k, i)$, where $\rho(k,i) > 0$ and $\sum_{i=1}^{m(k)} \rho(k,i) = q(k)$. The functions m and ρ are symmetric, i.e., $m(k) = m(-k)$ and $\rho(k,i) = \rho(-k,i)$ for any $k \in \mathcal{K}$ and $i = 1, \cdots, m(k)$, and they are taken as given throughout this appendix.

For any given symmetric policy profile $\mathbf{a} = \langle -a, a \rangle$ with $a \geq 0$, the news signal for market segment (k,i) maximizes the expected utilities of the voters therein (as in the standard RI model):

$$\max_{\Pi} V(\Pi; \mathbf{a}, k) - \lambda \cdot I(\Pi).$$

Across market segments, we consider all joint news distributions that are symmetric and consistent with the marginal news distributions that solve the above problem (hereafter *c-consistency*). As in Appendix A.1, we can represent a joint news distribution by its matrix form and define the *c*-consistency of the news configuration. This exercise is omitted for brevity.

We examine the policy polarization effect of introducing competition between infomediaries. To facilitate comparison between the case of monopolistic personalized news, we redefine *p*-consistency by first forming market segments using functions m and ρ and then restricting the same type of voters to receiving the same voting recommendation. By Lemma 6, equilibrium policies are fully determined by (1) $\mathcal{S} \in \{c, p\}$, which pins down marginal news distributions, (2) the news configuration $\boldsymbol{\chi}$, and (3) the functions m and ρ. Hereafter we shall use $\mathcal{E}^{\mathcal{S},\boldsymbol{\chi},m,\rho}$ to denote the equilibrium policy set and $a^{\mathcal{S},\boldsymbol{\chi},m,\rho}$ to denote the policy polarization.

The next proposition prescribes sufficient conditions for competition to reduce policy polarization.

Proposition 7. *Fix any functions m and ρ as above, and assume Assumptions 1-4 for $\mathcal{S} \in \{c, p\}$. Then $\mathcal{E}^{c,\boldsymbol{\chi},m,\rho} = [0, a^{c,\boldsymbol{\chi},m,\rho}] \subsetneq \mathcal{E}^{p,\boldsymbol{\chi}',m,\rho} = \left[0, a^{p,\boldsymbol{\chi}',m,\rho}\right]$ for any c-consistent news configuration $\boldsymbol{\chi}$ and any p-consistent news configuration $\boldsymbol{\chi}'$ such that $\boldsymbol{\chi} \succeq \boldsymbol{\chi}'$.*

Proof. See Appendix E.3. □

Two forces are acting in the same direction. First, competitive news signals maximize voters' expected utilities rather than their attention and are therefore less Blackwell informative than monopolistic personalized news signals. As infomediaries stop overfeeding voters with information about the valence state, voters become more susceptible to policy deviations, so a reduction in policy polarization ensues. Second, introducing competition between infomediaries reduces the correlation between the signals consumed by the same type of voters, whose policy polarization effect is negative by Proposition 4.

C General state distribution

This appendix extends the analysis so far to general state distributions. In the environment laid out in Appendix A, suppose the valence state is a distributed on $\mathbb{R}$ according to a c.d.f. G that is absolute continuous and symmetric around zero.[39] A news signal is a mapping $\Pi : \mathbb{R} \to \Delta(\mathcal{Z})$, where each $\Pi(\cdot \mid \omega)$ specifies a probability distribution over a finite set $\mathcal{Z}$ of signal realizations when the state realization is $\omega \in \mathbb{R}$. Under signal structure Π,

$$\pi_z = \int_{\omega \in \mathbb{R}} \Pi(z \mid \omega) \, dG(\omega)$$

is the probability that the signal realization is $z \in \mathcal{Z}$. Assume w.l.o.g. that $\pi_z > 0$ for any $z \in \mathcal{Z}$. Then

$$\mu_z = \int_{\omega \in \mathbb{R}} \omega \Pi(z \mid \omega) \, dG(\omega) / \pi_z$$

is the posterior mean of the state conditional on the signal realization being $z \in \mathcal{Z}$.

The next assumption is adapted from Matějka and McKay (2015).

Assumption 6. *The needed amount of attention for consuming* $\Pi : \mathbb{R} \to \Delta(\mathcal{Z})$ *is*

$$I(\Pi) = H(G) - \mathbb{E}_{\Pi}[H(G(\cdot \mid z))]$$

where $H(G)$ *is the entropy of the valence state, and* $H(G(\cdot \mid z))$ *is the conditional entropy*

[39] Results below also hold for discrete Gs. ω is the differential valence between the two candidates expressed in terms of voters' utilities and, hence, is a real-valued random variable.

of the valence state given signal realization z.

In what follows, we will fix any symmetric policy profile $\langle -a, a\rangle$ with $a > 0$ and use $\Pi^{\mathcal{S}}(a,k)$ to denote any optimal news signal consumed by type k voters under personalization technology $\mathcal{S} \in \{b,p\}$. In the case of $\mathcal{S} = b$, we will drop the notation k and simply write $\Pi^b(a)$. For each $\Pi^{\mathcal{S}}(a,k)$, we will use $\mathcal{Z}^{\mathcal{S}}(a,k)$ to denote the set of signal realizations, and $\mu_z^{\mathcal{S}}(a,k)$ to denote the posterior mean of the state conditional on the signal realization being $z \in \mathcal{Z}^{\mathcal{S}}(a,k)$. The next proposition extends Lemma 1 and Theorem 1 to general state distributions.

Proposition 8. *Fix any symmetric policy profile $\langle -a, a\rangle$ with $a > 0$, and assume Assumptions 3 and 6. Then,*

(i) any optimal personalized signal $\Pi^p(a,k)$ that is nondegenerate and makes type k voters' participation constraint binding must satisfy $|\mathcal{Z}^p(a,k)| = 2$, (SOB) *and the skewness properties stated in Theorem 1(ii);*

(ii) any optimal broadcast signal $\Pi^b(a)$ that is nondegenerate, induces consumption from all voters and makes some voter's participation constraint binding must satisfy $|\mathcal{Z}^b(a)| \in \{2,3\}$:

(a) if $|\mathcal{Z}^b(a)| = 2$, then $\Pi^b(a)$ satisfies (SOB) *and the skewness properties stated in Theorem 1(i);*

(b) if $|\mathcal{Z}^b(a)| = 3$, then we can write $\mathcal{Z}^b(a) = \{LL, LR, RR\}$, where $\mu^b_{LL}(a) < 0$, $\mu^b_{LR}(a) = 0$, and $\mu^b_{RR}(a) = |\mu^b_{LL}(a)| > 0$. Also for any $k \in \mathcal{K}$, we must have $v(\mathbf{a},k)+\mu^b_{LL}(a) < 0$, $\text{sgn}\left(v(\mathbf{a},k)+\mu^b_{LR}(a)\right) = \text{sgn}(k)$, and $v(\mathbf{a},k)+\mu^b_{RR}(a) > 0$.

Proof. See Appendix E.3. □

D Completing Example 1

This appendix covers the case where opposition voters are disciplining, i.e., Condition $(*)$ holds, so Condition $(**)$ becomes $|t(-1)| > |\mu^b_L| - |\mu^p_L(-1)|$. As depicted in Figure 4, the

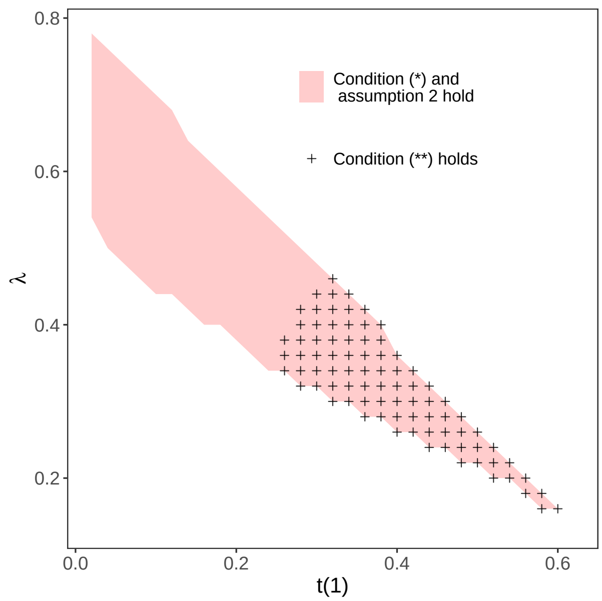

Figure 4: Condition $(**)$: $u(a,k) = -|t(k) - a|$, entropy attention cost, Condition $(*)$ holds.

last condition holds when extreme voters' policy preference parameter $t(1)$ takes sufficiently large values. As $t(1)$ increases, left-wing voters seek a bigger occasional surprise from news consumption than before, so $\mu_R^p(-1)$ must increase. Also since they find news consumption less useful, $|\mu_L^p(-1)|$ must decrease to prevent them from tuning out. Meanwhile in the broadcast case, $|\mu_L^b|$ $(= \mu_R^b)$ must decrease to prevent extreme voters from tuning out. Thus for sufficiently large $t(1)$s, the right-hand side of Condition $(**)$ is small whereas the left-hand side of it is big, which explains the pattern depicted in Figure 4.

E Mathematical proofs

E.1 Omitted proofs from Section 3

This appendix proves the results of Section 3 in the general environment laid out in Appendix A. The analysis takes an arbitrary symmetric policy profile $\mathbf{a} = \langle -a, a \rangle$ with $a \geq 0$ as given. Since the state is binary, it is without loss to identify a signal realization z with the posterior mean μ_z of the state it induces. A signal structure is then a profile $\langle \mu_z, \pi_z \rangle_{z \in \mathcal{Z}}$ of posterior mean μ_zs and their probability π_zs, and it must satisfy plausibility

(BP): $\sum_{z\in\mathcal{Z}} \pi_z \mu_z = 0$.

As discussed in Footnote 21, it is without loss to identify a binary signal structure with $\langle \mu_L, \mu_R \rangle$. Type k voters' expected gain from consuming $\langle \mu_L, \mu_R \rangle$ equals

$$V\left(\langle \mu_L, \mu_R \rangle; \mathbf{a}, k\right) = \begin{cases} \pi_R \left[v\left(\mathbf{a}, k\right) + \mu_R\right]^+ & \text{if } k \leq 0 \\ -\pi_L \left[v\left(\mathbf{a}, k\right) + \mu_L\right]^- & \text{if } k > 0 \end{cases} \tag{2}$$

where

$$\pi_L = \frac{\mu_R}{\mu_R - \mu_L} \text{ and } \pi_R = -\frac{\mu_L}{\mu_R - \mu_L}. \tag{3}$$

Proof of Lemmas 1 and 2 We prove Lemmas 1 and 2 together in three steps.

Step 1. Show that the optimal personalized signal for any type k voter is unique and has at most two realizations. An optimal personalized signal for type k voter solves

$$\max_{\mathcal{Z}, \Pi:\Omega\rightarrow\Delta(\mathcal{Z})} I\left(\Pi\right) \text{ subject to } V\left(\Pi; \mathbf{a}, k\right) \geq \lambda I\left(\Pi\right). \tag{4}$$

Let $\gamma \geq 0$ denote a Lagrange multiplier associated with voter k's participation constraint, and write the complementary slackness constraints as

$$\gamma \geq 0, V\left(\Pi; \mathbf{a}, k\right) \geq \lambda I\left(\Pi\right), \text{and } \gamma\left[V\left(\Pi; \mathbf{a}, k\right) - \lambda I\left(\Pi\right)\right] = 0. \tag{5}$$

If $\gamma = 0$, then the solution to (4) is the true state and so is unique and binary. If $\gamma > 0$, then reformulate (4) as

$$\max_{\mathcal{Z}, \langle \mu_z, \mu_z \rangle_{z\in\mathcal{Z}}, \gamma\geq 0} V\left(\Pi; \mathbf{a}, k\right) - \lambda\left(\gamma\right) I\left(\Pi\right) \text{ subject to (BP) and (5)} \tag{6}$$

where $\lambda\left(\gamma\right) := \lambda - 1/\gamma$. Note that the maximand of (6) equals

$$\sum_{z\in\mathcal{Z}} \pi_z \left[\nu\left(\mu_z; \mathbf{a}, k\right) - \lambda\left(\gamma\right) h\left(\mu_z\right)\right]$$

and so is posterior separable, and moreover the solution to (6) is the true state if $\lambda(\gamma) \leq 0$. If $\lambda(\gamma) > 0$, then $\nu(\mu; \mathbf{a}, k) - \lambda(\gamma) h(\mu)$ is the maximum of two strictly concave functions of μ: (i) $-\lambda(\gamma) h(\mu)$, and (ii) $v(\mathbf{a}, k) + \mu - \lambda(\gamma) h(\mu)$ if $k \leq 0$ and $-v(\mathbf{a}, k) - \mu - \lambda(\gamma) h(\mu)$ if $k > 0$. Since (i) and (ii) single-cross at $\mu = -v(\mathbf{a}, k)$, their maximum is M-shaped, so applying the concavification method developed by Kamenica and Gentzkow (2011) yields a unique solution with at most two signal realizations. Given this, we can restrict $\mathcal{Z}$ to $\mathcal{Z} : |\mathcal{Z}| \leq 2$ in (4) and prove that a solution to (4) exists.

To prove that the solution to (4) is unique, it suffices to show that γ is unique in the case where $\gamma > 0$ and $\lambda(\gamma) > 0$. If the contrary were true, then take two distinct Lagrange multipliers $\gamma_1 > \gamma_2 > 0$ for the voter's participation constraint. For each $i = 1, 2$, write $\lambda_i = \lambda - 1/\gamma_i$ and Π_i for the unique solution to (6). From strict optimality, i.e., the voter strictly prefers Π_i to Π_{-i} when the attention cost parameter in (6) is given by λ_i, we deduce that

$$\lambda_1 (I(\Pi_1) - I(\Pi_2)) > V(\Pi_1; \mathbf{a}, k) - V(\Pi_2; \mathbf{a}, k) > \lambda_2 (I(\Pi_1) - I(\Pi_2)).$$

Simplifying using $\lambda_1 > \lambda_2$ yields $I(\Pi_1) > I(\Pi_2)$, so Π_1 and Π_2 cannot both be the solutions to (4), a contradiction.

Step 2. Show that any optimal broadcast signal has at most two realizations. An optimal broadcast signal solves

$$\max_{\mathcal{Z}, \Pi:\Omega \to \Delta(\mathcal{Z})} I(\Pi) \sum_{k \in \mathcal{K}: V(\Pi; \mathbf{a}, k) \geq \lambda \cdot I(\Pi)} q_k. \tag{7}$$

For any solution to (7), let $\mathcal{BS}$ denote the set of voters with either binding or slack participation constraints, and note that $\mathcal{BS} \neq \emptyset$. Since $v(\mathbf{a}, k)$ is increasing in k by Assumption 3 **increasing differences**, it follows from the definition of $V(\cdot)$ that $V(\Pi; \mathbf{a}, k)$ is increasing in k on $\{k \in \mathcal{K} : k \leq 0\}$ and is decreasing in k on $\{k \in \mathcal{K} : k \geq 0\}$ for any Π. Given this, we can write $\mathcal{BS}$ as $\{k \in \mathcal{K} : k_1 \leq k \leq k_2\}$, where $k_1 \leq 0 \leq k_2$.

To ease exposition, we shall treat, w.l.o.g., adjacent types of voters with binding participation constraints as a single entity. Given this, only k_1 and k_2 can have binding participation constraints. Let $\gamma_1 \geq 0$ and $\gamma_2 \geq 0$ denote the Lagrange multipliers associated with their participation constraints, and let $\mathcal{B} \subseteq \{k_1, k_2\}$ denote the set of voters with binding participation constraints. Write the complementary slackness constraints as

$$\gamma_i \geq 0, V(\Pi; \mathbf{a}, k_i) \geq \lambda I(\Pi), \text{and } \gamma_i [V(\Pi; \mathbf{a}, k_i) - \lambda I(\Pi)] = 0 \tag{8}$$

and reformulate (7) as

$$\max_{\substack{\mathcal{Z}, \Pi: \Omega \to \Delta(\mathcal{Z}) \\ k_1 \leq 0 \leq k_2, \gamma_1, \gamma_2 \geq 0}} I(\Pi) \sum_{k=k_1}^{k_2} q_k + \sum_{i=1}^{2} \gamma_i [V(\Pi; \mathbf{a}, k_i) - \lambda I(\Pi)] \text{ subject to (8).} \tag{9}$$

Consider three cases. First, if $\mathcal{B} = \emptyset$, then the solution to (9) is the true state. Second, if $|\mathcal{B}| = 1$, then the solution to (9) is the optimal personalized signal for the voter in $\mathcal{B}$. Finally, if $|\mathcal{B}| = 2$, then (9) becomes

$$\max_{\substack{\mathcal{Z}, \langle \pi_z, \mu_z \rangle_{z \in \mathcal{Z}} \\ k_1 \leq 0 \leq k_2, \gamma_1, \gamma_2 \geq 0}} \sum_{z \in \mathcal{Z}} \pi_z f(\mu_z) \text{ subject to (BP) and (8)} \tag{10}$$

where

$$f(\mu_z) \coloneqq \sum_{i=1,2} \frac{\gamma_i}{\gamma_1 + \gamma_2} \nu(\mu_z; \mathbf{a}, k_i) - \left(\lambda - \frac{\sum_{k=k_1}^{k_2} q_k}{\gamma_1 + \gamma_2} \right) h(\mu_z).$$

Write $\widetilde{\gamma}_i = \frac{\gamma_i}{\gamma_1 + \gamma_2}$ for $i = 1, 2$, Q for $\sum_{k=k_1}^{k_2} q_k$, and $\widetilde{\lambda}$ for $\lambda - \frac{Q}{\gamma_1 + \gamma_2}$. Note that the solution to (10) is the true state if $\widetilde{\lambda} \leq 0$. If $\widetilde{\lambda} > 0$, then $f(\mu_z)$ is the maximum of three strictly concave functions of μ: (i) $\widetilde{\gamma}_1 [v(\mathbf{a}, k_1) + \mu] - \widetilde{\lambda} h(\mu)$, (ii) $-\widetilde{\lambda} h(\mu)$, and (iii) $-\widetilde{\gamma}_2 [v(\mathbf{a}, k_2) + \mu] - \widetilde{\lambda} h(\mu)$, where (i) and (ii) intersect at $\mu = -v(\mathbf{a}, k_1) > 0$, and (ii) and (iii) intersect at $\mu = -v(\mathbf{a}, k_2) < 0$. Let f^+ denote the concave closure of f, and note that $\mu_1 \coloneqq \inf\{\mu : f^+(\mu) > f(\mu)\}$ and $\mu_2 \coloneqq \sup\{\mu : f^+(\mu) > f(\mu)\}$ exist and satisfy $\mu_1 < 0 < \mu_2$. If $f^+(0)$ cannot be expressed as a convex combination of $f^+(\mu_1)$ and $f^+(\mu_2)$, then (10) has a unique degenerate solution. If $f^+(0)$ can be expressed as a convex combination of $f^+(\mu_1)$ and $f^+(\mu_2)$, then among all solution(s) to (10), the binary signal

$\langle \mu_1, \mu_2 \rangle$ is the most Blackwell informative and, hence, constitutes the unique solution to the original attention maximization problem. Taken together, we obtain that in all cases, we can restrict $\mathcal{Z}$ to $\mathcal{Z} : |\mathcal{Z}| \leq 2$ in (7) and prove that a solution (7) exists. This completes the proof.

Step 3. Show that in the broadcast case, if it is optimal to induce news consumption from all voters using a binary signal, then the optimal signal is unique and symmetric.

By assumption, we have $k_1 = -K$ and $k_2 = K$, so $\mathcal{B} \subseteq \{-K, K\}$. If $\mathcal{B} = \emptyset$, then the solution to (9) is the true state and so is unique and symmetric. The case where $|\mathcal{B}| = 1$ is impossible, because in that case, the solution to (9) is the optimal personalized signal for the voter in $\mathcal{B}$, which violates the participation constraint of the voter in $\{-K, K\} - \mathcal{B}$. Finally, if $\mathcal{B} = \{-K, K\}$, then take any optimal broadcast signal $\langle \mu_L, \mu_R \rangle$, and notice that $V(\langle \mu_L, \mu_R \rangle; \mathbf{a}, -K) = \lambda I(\langle \mu_L, \mu_R \rangle) = V(\langle \mu_L, \mu_R \rangle; \mathbf{a}, K) > 0$. Simplifying using (2) and $v(\mathbf{a}, K) = -v(\mathbf{a}, -K)$ yields $\mu_L + \mu_R = 0$, so $\langle \mu_L, \mu_R \rangle$ is symmetric, and μ_L is a solution to

$$\max_{\mu \in [-1,0]} h(\mu) \text{ s.t. } \frac{1}{2} [v(\mathbf{a}, -K) - \mu]^+ \geq \lambda h(\mu). \tag{11}$$

Since h is strictly convex and strictly decreasing on $[-1, 0]$, (11) either admits no solution (in which case the optimal signal isn't binary to begin with) or admits a unique solution. □

Proof of Theorem 1 We focus on the proof of Part (ii), which concerns the skewness of personalized signals. Part (i) concerning the symmetry of the broadcast signal has already been shown in the proof of Lemmas 1 and 2. Part (iii) that compares the needed amounts of attention for consuming the broadcast signal and personalized signals requires no more proof than the verbal argument in the main text.

We only prove Part (ii) for an arbitrary left-wing voter $k < 0$, for whom we write $v(\mathbf{a}, k) = v$ and note that $v < 0$. Let $\langle \mu_L, \mu_R \rangle$ denote the voter's optimal personalized signal, which by Assumption 2 must satisfy

$$V\left(\langle\mu_L,\mu_R\rangle;\mathbf{a},k\right)=-\frac{\mu_L}{\mu_R-\mu_L}\left[v+\mu_R\right]^+\geq\lambda I\left(\langle\mu_L,\mu_R\rangle\right)>0$$

and, hence, $v+\mu_R>0$. We wish to demonstrate that $\mu_L+\mu_R>0$. Notice first that $\mu_L+\mu_R\geq 0$, because if the contrary were true, i.e., $\mu_L+\mu_R<0$, then the voter would strictly prefer $\langle-\mu_R,-\mu_L\rangle$ to $\langle\mu_L,\mu_R\rangle$:

$$\begin{aligned}V\left(\langle-\mu_R,-\mu_L\rangle;\mathbf{a},k\right)=\frac{\mu_R}{\mu_R-\mu_L}\left(v-\mu_L\right)&\\ >-\frac{\mu_L}{\mu_R-\mu_L}\left(v+\mu_R\right)=V\left(\langle\mu_L,\mu_R\rangle;\mathbf{a},k\right).&\end{aligned}$$

To show that $\mu_L+\mu_R\neq 0$, take the Lagrange multiplier $\gamma>0$ associated with the voter's participation constraint as given, and reduce (6) to

$$\max_{\substack{\langle\mu_L,\mu_R\rangle\in\\ [-1,0]\times[-v,1]}}\frac{-\mu_L}{\mu_R-\mu_L}\left[v+\mu_R\right]^+-\left(\lambda-1/\gamma\right)\left[\frac{\mu_R}{\mu_R-\mu_L}h\left(\mu_L\right)+\frac{-\mu_L}{\mu_R-\mu_L}h\left(\mu_R\right)\right].$$

Note that $\lambda-1/\gamma>0$ must hold in order for this problem to admit an interior solution (as required by Assumption 2), and that any interior solution must satisfy the following first-order conditions:

$$\begin{aligned}v+\mu_R&=\left(\lambda-1/\gamma\right)\left[\Delta h-h'\left(\mu_L\right)\Delta\mu\right]\\ \text{and }-\left(v+\mu_L\right)&=\left(\lambda-1/\gamma\right)\left[h'\left(\mu_R\right)\Delta\mu-\Delta h\right]\end{aligned}$$

where $\Delta h\coloneqq h\left(\mu_R\right)-h\left(\mu_L\right)$ and $\Delta\mu\coloneqq\mu_R-\mu_L$. If $\mu_L+\mu_R=0$, then $\Delta h=0$ and $h'\left(\mu_R\right)=-h'\left(\mu_L\right)$ by Assumption 1, so the right-hand sides of the first-order conditions are the same. Meanwhile, the left-hand sides differ, which leads to a contradiction. □

E.2 Omitted proofs from Appendix A

E.2.1 Useful lemmas and their proofs

Proof of Lemma 6 Fix any personalization technology $\mathcal{S}$, symmetric policy profile $\langle -a, a\rangle$ with $a \geq 0$, $\langle \mathcal{S}, a\rangle$-consistent news distribution $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^-\rangle$, and population function q. Let $\mathbf{q}$ denote the $|\mathcal{K}|$-column vector that compiles the populations of voters $-K, \cdots, K$. Let the default be the strictly obedient outcome induced by the joint news distribution.

Define two matrix operations. First, for any $\mathcal{C} \subseteq \mathcal{K}$, let $\boldsymbol{\chi}_{\mathcal{C}}$ be the resulting matrix from replacing every row $k \in \mathcal{C}$ of $\boldsymbol{\chi}$ with a row of all ones. Second, for any matrix $\mathbf{A}$, let $\widehat{\mathbf{A}}$ be the resulting matrix from rounding the entries of $\mathbf{A}$, i.e., replacing those entries above 1/2 with 1 and those below 1/2 with zero. By definition, the row vector $\widehat{\mathbf{q}^\top \boldsymbol{\chi}}$ compiles candidate R's default winning probabilities across the voting recommendation profiles that occur with strictly positive probabilities, and $(\widehat{\mathbf{q}^\top \boldsymbol{\chi}}\mathbf{b}^+ + \widehat{\mathbf{q}^\top \boldsymbol{\chi}}\mathbf{b}^-)/2$ is candidate R's default winning probability in expectation. After candidate R commits a unilateral deviation from $\langle -a, a\rangle$ that attracts a set $\mathcal{C} \subseteq \mathcal{K}$ of voters without affecting anything else, his winning probability vector becomes $\widehat{\mathbf{q}^\top \boldsymbol{\chi}_{\mathcal{C}}}$, and his expected winning probability becomes $(\widehat{\mathbf{q}^\top \boldsymbol{\chi}_{\mathcal{C}}}\mathbf{b}^+ + \widehat{\mathbf{q}^\top \boldsymbol{\chi}_{\mathcal{C}}}\mathbf{b}^-)/2$. Since $\widehat{\mathbf{q}^\top \boldsymbol{\chi}_{\mathcal{C}}} \geq \widehat{\mathbf{q}^\top \boldsymbol{\chi}}$, the deviation strictly increases candidate R's winning probability in expectation if and only if it does so under some voting recommendation profile, i.e., $(\widehat{\mathbf{q}^\top \boldsymbol{\chi}_{\mathcal{C}}}\mathbf{b}^+ + \widehat{\mathbf{q}^\top \boldsymbol{\chi}_{\mathcal{C}}}\mathbf{b}^-)/2 > (\widehat{\mathbf{q}^\top \boldsymbol{\chi}}\mathbf{b}^+ + \widehat{\mathbf{q}^\top \boldsymbol{\chi}}\mathbf{b}^-)/2$ if and only if $\widehat{\mathbf{q}^\top \boldsymbol{\chi}_{\mathcal{C}}} \neq \widehat{\mathbf{q}^\top \boldsymbol{\chi}}$. The last condition is equivalent to $\mathcal{C}$ being an influential coalition, and it depends on $\mathcal{S}$, $\langle -a, a\rangle$, $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^-\rangle$, and q only through $\langle \boldsymbol{\chi}, q\rangle$. □

Proof of Lemma 7 Fix any personalization technology $\mathcal{S}$, symmetric policy profile $\langle -a, a\rangle$ with $a \geq 0$, $\langle \mathcal{S}, a\rangle$-consistent news distribution $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^-\rangle$ and population function q. Let $\mathcal{C}$s denote the influential coalitions formed under $\langle \boldsymbol{\chi}, q\rangle$. Consider any unilateral deviation of candidate R from $\langle -a, a\rangle$ to a'. Below we demonstrate that a' is unprofitable if and only if (i) $a' \notin [-a, a]$, or (ii) $a' \in [-a, a)$ and it doesn't attract any influential coalition whose members have ideological bliss points in $[-a, a]$.

Step 1. Show that no $a' > a$ strictly increases candidate R's winning probability. Fix any $a' > a$, and note that no type $k \leq 0$ voter finds a' attractive by Assumption 3:

$$\begin{aligned} v\left(-a, a', k\right) + \mu_L^{\mathcal{S}}(a, k) &< v(-a, a, k) + \mu_L^{\mathcal{S}}(a, k) && (\textbf{inverted V-shape}) \\ &\leq v(-a, a, 0) + \mu_L^{\mathcal{S}}(a, k) && (\textbf{increasing differences}) \\ &= 0 + \mu_L^{\mathcal{S}}(a, k) && (\textbf{symmetry}) \\ &< 0. \end{aligned}$$

Given this, as well as the symmetry of the joint news distribution, it suffices to show that if a' attracts any type $k > 0$ voter, then it must repel type $-k$ voters, i.e.,

$$v\left(-a, a', k\right) + \mu_L^{\mathcal{S}}(a, k) > 0 \Longrightarrow v\left(-a, a', -k\right) + \mu_R^{\mathcal{S}}(a, -k) < 0.$$

The argument below exploits the symmetry of marginal news distributions, i.e., $\mu_R^{\mathcal{S}}(a, -k) = -\mu_L^{\mathcal{S}}(a, k)$, which together with Assumption 3 **symmetry** implies

$$\begin{aligned} v\left(-a, a', -k\right) + \mu_R^{\mathcal{S}}(a, -k) &= u\left(a', -k\right) - u(-a, -k) + \mu_R^{\mathcal{S}}(a, -k) \\ &= u\left(-a', k\right) - u(a, k) - \mu_L^{\mathcal{S}}(a, k). \end{aligned}$$

Thus, if $v(-a, a', k) + \mu_L^{\mathcal{S}}(a, k) = u(a', k) - u(-a, k) + \mu_L^{\mathcal{S}}(a, k) > 0$, then

$$\begin{aligned} v\left(-a, a', -k\right) + \mu_R^{\mathcal{S}}(a, -k) &= u\left(-a', k\right) - u(a, k) - \mu_L^{\mathcal{S}}(a, k) \\ &< u\left(a', k\right) + u\left(-a', k\right) - [u(a, k) + u(-a, k)] \leq 0 \end{aligned}$$

where the last inequality follows from Assumption 3 **concavity**.

Step 2. Show that no $a' < -a$ strictly increases candidate R's winning probability. The proof closely parallels that offered in Step 1. First, note that no $a' < a$ attracts any type

$k \geq 0$ voter by Assumption 3 **inverted V-shape**:

$$v\left(-a, a', k\right) + \mu_L^p(a, k) < v(-a, -a, k) + \mu_L^p(a, k) = 0 + \mu_L^p(a, k) < 0.$$

Second, if any a' as above attracts any type $k < 0$ voter, then it must repel type $-k$ voters for the same reason given in Step 1. Combining these observations gives the desired result.

Step 3. Show that no $a' \in [-a, a)$ repels any voter. Fix any $a' \in [-a, a)$. From Assumption 3 **inverted V-shape** and (SOB), it follows that if $t(k) \leq a'$, then

$$v\left(-a, a', k\right) + \mu_R^{\mathcal{S}}(a, k) > v(-a, a, k) + \mu_R^{\mathcal{S}}(a, k) > 0,$$

and if $t(k) > a'$, then

$$v\left(-a, a', k\right) + \mu_R^{\mathcal{S}}(a, k) \geq v(-a, -a, k) + \mu_R^{\mathcal{S}}(a, k) = 0 + \mu_R^{\mathcal{S}}(a, k) > 0.$$

Combining these observations yields $v(-a, a', k) + \mu_R^{\mathcal{S}}(a, k) > 0$ for any k.

Step 4. Show that no $a' \in [-a, a)$ attracts any voter whose bliss point lies outside $[-a, a]$. Fix any $a' \in [-a, a)$. From Assumption 3 **inverted V-shape** and (SOB), we deduce that if $t(k) \leq -a$, then

$$v\left(-a, a', k\right) + \mu_L^{\mathcal{S}}(a, k) \leq v(-a, -a, k) + \mu_L^{\mathcal{S}}(a, k) = 0 + \mu_L^{\mathcal{S}}(a, k) < 0,$$

and if $t(k) > a$, then

$$v\left(-a, a', k\right) + \mu_L^{\mathcal{S}}(a, k) < v(-a, a, k) + \mu_L^{\mathcal{S}}(a, k) < 0.$$

Combining these observations yields $v(-a, a', k) + \mu_L^{\mathcal{S}}(a, k) < 0$ for any k.

Combining Steps 1-4 shows a deviation a' from $\langle -a, a \rangle$ strictly increases candidate R's winning probability if and only if it belongs to $[-a, a]$ and attracts an influential coalition whose members have ideological bliss points in $[-a, a]$. Ruling out such deviations leads us to sustain $\langle -a, a\rangle$ in an equilibrium. $\square$

The next lemma gives characterizations of the $\mathcal{D}$-proof set for any $\mathcal{D} \subseteq \mathcal{K}$.

Lemma 9. *Let everything be as in Theorem 3. Then for any $k \in \{0, \cdots, K\}$ and any $\mathcal{D} \subseteq \{-k, \cdots, k\}$ such that $\mathcal{D} \cap \{-k, k\} \neq \emptyset$, we must have $\xi^{\mathcal{S}}(\mathcal{D}) > t(k)$ and $[t(k), \overline{a}] \cap \Xi^{\mathcal{S}}(\mathcal{D}) = \left[t(k), \xi^{\mathcal{S}}(\mathcal{D})\right]$.*

Proof. Let k and $\mathcal{D}$ be as above. Recall the definition of the $\mathcal{D}$-proof set:

$$\Xi^{\mathcal{S}}(\mathcal{D}) := \left\{a \geq 0 : \max_{a' \in \mathcal{A}} \phi^{\mathcal{S}}\left(-a, a', \mathcal{D}\right) \leq 0\right\}$$

where $\phi^{\mathcal{S}}(-a, a', \mathcal{D}) := \min_{k' \in \mathcal{D}} \phi^{\mathcal{S}}\left(-a, a', k'\right)$ is the $\mathcal{D}$-susceptibility to candidate R's deviation from $\langle -a, a\rangle$ to a'. By Assumption 3 **inverted V-shape**, we can restrict attention to $a' \in [\min t(\mathcal{D}), \max t(\mathcal{D})] := \widetilde{\mathcal{D}}$, i.e.,

$$\Xi^{\mathcal{S}}(\mathcal{D}) = \left\{a \geq 0 : \max_{a' \in \widetilde{\mathcal{D}}} \phi^{\mathcal{S}}\left(-a, a', \mathcal{D}\right) \leq 0\right\}$$

where $t(\mathcal{D})$ denotes the image of $\mathcal{D}$ under the mapping t.

Fix the policy profile to be $\langle -t(k), t(k)\rangle$, and take any $a' \in \widetilde{\mathcal{D}}$. From Assumption 3 and (SOB), we deduce that a' doesn't attract type k voters:

$$\begin{aligned}
\phi^{\mathcal{S}}\left(-t(k), a', k\right) &:= v\left(-t(k), a', k\right) + \mu_L^{\mathcal{S}}(t(k), k) \\
&\leq v(-t(k), t(k), k) + \mu_L^{\mathcal{S}}(t(k), k) && (\textbf{inverted V-shape}) \\
&< 0, && \text{(SOB)}
\end{aligned}$$

and it doesn't attract type $-k$ voters, either:

$$\begin{aligned}
&\phi^{\mathcal{S}}\left(-t\left(k\right),a',-k\right)\\
&\coloneqq v\left(-t\left(k\right),a',-k\right)+\mu_{L}^{\mathcal{S}}\left(t\left(k\right),-k\right)\\
&\leq v\left(-t\left(k\right),t\left(-k\right),-k\right)+\mu_{L}^{\mathcal{S}}\left(t\left(k\right),-k\right) && (\textbf{inverted V-shape})\\
&=0+\mu_{L}^{\mathcal{S}}\left(t\left(k\right),-k\right) && (\textbf{symmetry})\\
&<0.
\end{aligned}$$

Thus $\phi^{\mathcal{S}}\left(-t\left(k\right),a',\mathcal{D}\right)\coloneqq\min_{k'\in\mathcal{D}}\phi^{\mathcal{S}}\left(-t\left(k\right),a',k'\right)<0$, and taking maximum over a' yields $\max_{a'\in\widetilde{\mathcal{D}}}\phi^{\mathcal{S}}\left(-t\left(k\right),a',\mathcal{D}\right)<0$. Meanwhile, Assumption 4 implies that $\phi^{\mathcal{S}}\left(-a,a',\mathcal{D}\right)$ is increasing in a on $\left[t\left(k\right),\overline{a}\right]$ for any a'. Taking maximum over a' yields

$$\begin{aligned}
\max_{a'\in\widetilde{\mathcal{D}}}\phi^{\mathcal{S}}\left(-a_1,a',\mathcal{D}\right)&=\phi^{\mathcal{S}}\left(-a_1,\arg\max_{a'\in\widetilde{\mathcal{D}}}\phi^{\mathcal{S}}\left(-a_1,a',\mathcal{D}\right),\mathcal{D}\right)\\
&<\phi^{\mathcal{S}}\left(-a_2,\arg\max_{a'\in\widetilde{\mathcal{D}}}\phi^{\mathcal{S}}\left(-a_1,a',\mathcal{D}\right),\mathcal{D}\right)\\
&\leq\max_{a'\in\widetilde{\mathcal{D}}}\phi^{\mathcal{S}}\left(-a_2,a',\mathcal{D}\right)\ \forall a_2>a_1\geq t\left(k\right),
\end{aligned}$$

so $\max_{a'\in\widetilde{\mathcal{D}}}\phi^{\mathcal{S}}\left(-a,a',\mathcal{D}\right)$ is increasing in a on $\left[t\left(k\right),\overline{a}\right]$. Taken together, we obtain that $\mathcal{D}$'s policy latitude exceeds $t\left(k\right)$:

$$\xi^{\mathcal{S}}\left(\mathcal{D}\right)\coloneqq\max\Xi^{\mathcal{S}}\left(\mathcal{D}\right)=\max\left\{a\geq0:\max_{a'\in\widetilde{\mathcal{D}}}\phi^{\mathcal{S}}\left(-a,a',\mathcal{D}\right)\leq0\right\}>t\left(k\right)$$

and that all policies in $\left[t\left(k\right),\xi^{\mathcal{S}}\left(\mathcal{D}\right)\right]$ belong to the $\mathcal{D}$-proof set:

$$\left[t\left(k\right),\overline{a}\right]\cap\Xi^{\mathcal{S}}\left(\mathcal{D}\right)=\left\{a\geq t\left(k\right):\max_{a'\in\widetilde{\mathcal{D}}}\phi^{\mathcal{S}}\left(-a,a',\mathcal{D}\right)\leq0\right\}=\left[t\left(k\right),\xi^{\mathcal{S}}\left(\mathcal{D}\right)\right].$$

□

The next lemma characterizes the susceptibility function for the case of personalized news and quadratic attention cost.

Lemma 10. *Under Assumptions 1, 2(i), 3, and 5, the following hold for any $a \geq 0$ and $a' \in [-a, a]$.*

(i) $\phi^p(-a, a', k)$ *is decreasing in* k *on* $\{k \in \mathcal{K} : k \leq 0\}$ *and is increasing in* k *on* $\{k \in \mathcal{K} : k \geq 0\}$.

(ii) $\phi^p(-a, a', k) \leq \phi^p(-a, a', -k)$ *for any* $k > 0$.

Proof. Let a and a' be as above. Tedious algebra (available upon request) reduces Assumption 2(i) to $2\lambda > 1$ and $4\lambda v(-\overline{a}, \overline{a}, K) < 1$, under which

$$\mu_L^p(a, k) = \begin{cases} -2v(-a, a, k) - 1/(2\lambda) & \text{if } k \leq 0, \\ -1/(2\lambda) & \text{if } k > 0. \end{cases} \tag{12}$$

Also recall that $\phi^p(-a, a', k) := v(-a, a', k) + \mu_L^p(a, k)$.

Part (i): If $k \leq 0$, then

$$\begin{aligned} \phi^p\left(-a, a', k\right) &= v\left(-a, a', k\right) - 2v(-a, a, k) - \frac{1}{2\lambda} \\ &= u\left(a', k\right) - u(-a, k) - 2\left[u(a, k) - u(-a, k)\right] - \frac{1}{2\lambda} \\ &= u\left(a', k\right) + u(-a, k) - 2u(a, k) - \frac{1}{2\lambda} \\ &= -\left[v\left(a', a, k\right) + v(-a, a, k)\right] - \frac{1}{2\lambda} \end{aligned}$$

where the last line is decreasing in k by Assumption 3 **increasing differences**. If $k > 0$, then $\phi^p(-a, a', k) = v(-a, a', k) - 1/(2\lambda)$ and so is increasing in k.

Part (ii): Under Assumption 3, the following holds for any $k > 0$:

$$\begin{aligned}
&\phi^p\left(-a, a', k\right) - \phi^p\left(-a, a', -k\right) \\
&= v\left(-a, a', k\right) - \frac{1}{2\lambda} - \left[v\left(-a, a', -k\right) - 2v\left(-a, a, -k\right) - \frac{1}{2\lambda}\right] \\
&= v\left(-a, a', k\right) - v\left(a, -a', k\right) - 2v\left(-a, a, k\right) && (\textbf{symmetry}) \\
&= \left[u\left(-a, k\right) - u\left(-a', k\right)\right] - \left[u\left(a, k\right) - u\left(a', k\right)\right] \\
&= v\left(a', a, -k\right) - v\left(a', a, k\right) && (\textbf{symmetry}) \\
&\leq 0. && (\textbf{increasing differences})
\end{aligned}$$

□

Lemma 11. *Under Assumptions 1-3, Assumption 4 holds if $\mathcal{S} = b$ or if $\mathcal{S} = p$ and either $u(a, k) = -|t(k) - a|$ or $h(\mu) = \mu^2$.*

Proof. We wish to verify that $\phi^{\mathcal{S}}(-a, a', k) := v(-a, a', k) + \mu_L^{\mathcal{S}}(a, k)$ is increasing in a on $[|t(k)|, \overline{a}]$ for any $k \in \mathcal{K}$ and $a' \in \mathcal{A}$ in the above scenarios. Since $v(-a, a', k)$ is strictly increasing in a on $[|t(k)|, \overline{a}]$ by Assumption 3 **inverted V-shape**, it suffices to show that $\mu_L^{\mathcal{S}}(a, k)$ is nondecreasing in a on $[|t(k)|, \overline{a}]$.

Case 1. $\mathcal{S} = b$ Recall that $\mu_L^b(a)$ is the unique solution to

$$\max_{\mu \in [-1,0]} h(\mu) \text{ s.t. } \frac{1}{2}\left[v(-a, a, -K) - \mu\right]^+ \geq \lambda h(\mu)$$

where h is strictly convex and strictly decreasing on $[-1, 0]$. Meanwhile, the following holds for any $a' > a \geq 0$ under Assumption 3:

$$\begin{aligned}
& v\left(-a', a', -K\right) - v\left(-a, a, -K\right) \\
&= u\left(a', -K\right) - u\left(-a', -K\right) - u\left(a, -K\right) + u\left(-a, -K\right) \\
&= u\left(a', -K\right) - u\left(a, -K\right) - \left[u\left(a', K\right) - u\left(a, K\right)\right] && (\textbf{symmetry}) \\
&= v\left(a, a', -K\right) - v\left(a, a', K\right) \\
&\leq 0, && (\textbf{increasing differences})
\end{aligned}$$

so $v\left(-a, a, -K\right)$ is decreasing in a. Combining these observations gives the desired result.

Case 2. $\mathcal{S} = p$ **and** $u\left(a, k\right) = -|t\left(k\right) - a|$ In this case, $v\left(-a, a, k\right)$ is invariant with a on $[|t\left(k\right)|, \overline{a}]$, so $\mu_L^p\left(a, k\right) \equiv \mu_L^p\left(|t\left(k\right)|, k\right)$ on $[|t\left(k\right)|, \overline{a}]$.

Case 3. $\mathcal{S} = p$ **and** $h\left(\mu\right) = \mu^2$ In this case, a careful inspection of (12) (i.e., the expression for $\mu_L^p\left(a, k\right)$) gives the desired result. □

E.2.2 Proofs of theorems and propositions

Proof of Theorem 3 Fix any personalization technology $\mathcal{S} \in \{b, p\}$, $\mathcal{S}$-consistent news configuration $\boldsymbol{\chi}$, and population function q. Let $\mathcal{C}$ denote a generic influential coalition formed under $\langle \boldsymbol{\chi}, q \rangle$. For each $k = 0, \cdots, K-1$, define

$$A\left(k\right) = \begin{cases} [t\left(k\right), t\left(k+1\right)) \cap \displaystyle\bigcap_{\mathcal{C} \subseteq \{-k, \cdots, k\}} \Xi^{\mathcal{S}}\left(\mathcal{C}\right) & \text{if } \exists \mathcal{C} \subseteq \{-k, \cdots, k\}, \\ [t\left(k\right), t\left(k+1\right)) & \text{otherwise.} \end{cases}$$

Also define

$$A\left(K\right) = [t\left(K\right), \overline{a}] \cap \bigcap_{\mathcal{C}} \Xi^{\mathcal{S}}\left(\mathcal{C}\right).$$

Lemma 7 shows that

$$\mathcal{E}^{\mathcal{S},\boldsymbol{\mathcal{X}},q} = \bigcup_{k=0}^{K} A(k).$$

Below we prove by induction that $\cup_{k=0}^{K} A(k) = \left[0, \min_{\mathcal{C}} \xi^{\mathcal{S}}(\mathcal{C})\right]$.

Step 0. Letting $k = 0$ in the statement of Lemma 9 yields $\Xi^{\mathcal{S}}(\{0\}) = \left[0, \xi^{\mathcal{S}}(\{0\})\right]$, so

$$A(0) = \begin{cases} \left[0, \xi^{\mathcal{S}}(\{0\})\right] & \text{if } \{0\} \text{ is influential and } \xi^{\mathcal{S}}(\{0\}) < t(1), \\ [0, t(1)) & \text{otherwise.} \end{cases}$$

For the first case, note that $A(k) \subseteq [t(k), t(k+1)] \cap \Xi^{\mathcal{S}}(\{0\}) = \emptyset$ for any $k \geq 1$, and $\min_{\mathcal{C}} \xi^{\mathcal{S}}(\mathcal{C}) = \xi^{\mathcal{S}}(\{0\})$ because $\xi^{\mathcal{S}}(\mathcal{C}) > t(1)$ for any $\mathcal{C} \neq \{0\}$ by Lemma 9. Taken together, we obtain $\cup_{k=0}^{K} A(k) = \left[0, \min_{\mathcal{C}} \xi^{\mathcal{S}}(\mathcal{C})\right]$ and terminate the procedure. In the second case, we proceed to the next step.

Step m. The output of Step $m-1$ is $\cup_{k=0}^{m-1} A(k) = [0, t(m))$. Then from Lemma 9, which shows that $[t(m), \overline{a}] \cap \Xi^{\mathcal{S}}(\mathcal{C}) = \left[t(m), \xi^{\mathcal{S}}(\mathcal{C})\right]$ for any $\mathcal{C} \subseteq \{-m, \cdots, m\}$ such that $\mathcal{C} \cap \{-m, m\} \neq \emptyset$, it follows that

$$\cup_{k=0}^{m} A(k) = \begin{cases} \left[0, \min_{\mathcal{C} \subseteq \{-m, \cdots, m\}} \xi^{\mathcal{S}}(\mathcal{C})\right] & \text{if } \min_{\mathcal{C} \subseteq \{-m, \cdots, m\}} \xi^{\mathcal{S}}(\mathcal{C}) < t(m+1), \\ [0, t(m+1)) & \text{otherwise.} \end{cases}$$

For the first case, note that $A(k) \subseteq [t(k), t(k+1)] \cap \cap_{\mathcal{C} \subseteq \{-m, \cdots, m\}} \xi^{\mathcal{S}}(\mathcal{C}) = \emptyset$ for any $k \geq m+1$, and $\min_{\mathcal{C}} \xi^{\mathcal{S}}(\mathcal{C}) = \min_{\mathcal{C} \subseteq \{-m, \cdots, m\}} \xi^{\mathcal{S}}(\mathcal{C})$ because $\xi^{\mathcal{S}}(\mathcal{C}') > t(m+1)$ for any $\mathcal{C}' \nsubseteq \{-m, \cdots, m\}$ by Lemma 9. Taken together, we obtain $\cup_{k=0}^{K} A(k) = \left[0, \min_{\mathcal{C}} \xi^{\mathcal{S}}(\mathcal{C})\right]$ and terminate the procedure. In the second case, we proceed to the next step.

The above procedure terminates in at most $K+1$ steps, and the output is always $\cup_{k=0}^{K} A(k) = \left[0, \min_{\mathcal{C}} \xi^{\mathcal{S}}(\mathcal{C})\right]$. □

Proof of Proposition 6 We wish to demonstrate that $\min_{\mathcal{C}s \text{ formed under } \langle \boldsymbol{\chi}, q \rangle} \xi^p(\mathcal{C}) \geq \min_{\mathcal{C}s \text{ formed under } \langle \boldsymbol{\chi}, q' \rangle} \xi^p(\mathcal{C})$ holds for any p-consistent $\boldsymbol{\chi}$ and any two population functions q and q' such that $q \succeq_{SOSD} q'$. The proof below exploits the following consequences of Lemma 10: for any $a \geq 0$ and $a' \in [-a, a]$, (i) $\phi^p(-a, a', -K) = \max_{k \in \mathcal{K}} \phi^p\left(-a, a', k\right)$, and (ii) $\phi^p(-a, a', k)$ is decreasing in k on $\{k : k \leq 0\}$ and is increasing in k on $\{k : k \geq 0\}$.

Step 1. Show that $\xi^p(\mathcal{D}) > t(K)$ for any $\mathcal{D} \subseteq \mathcal{K}$. Fix any $a' \in [t(-K), t(K)]$ and any $\mathcal{D} \subseteq \mathcal{K}$, and notice two things. First,

$$
\begin{aligned}
& \phi^p\left(-t(K), a', \mathcal{D}\right) \\
& \coloneqq \min_{k \in \mathcal{D}} \phi^p\left(-t(K), a', k\right) \\
& \leq \max_{k \in \mathcal{D}} \phi^p\left(-t(K), a', k\right) \\
& \leq \phi^p\left(-t(K), a', -K\right) && \text{(Lemma 10)} \\
& \leq \phi^p\left(-t(K), t(-K), -K\right) && \text{(Assumption 3 } \textbf{inverted V-shape}\text{)} \\
& \coloneqq v\left(-t(K), t(-K), K\right) + \mu_L^p\left(t(K), -K\right) \\
& = 0 + \mu_L^p\left(t(K), -K\right) && \text{(Assumption 3 } \textbf{symmetry}\text{)} \\
& < 0.
\end{aligned}
$$

Second, since $\phi^p(-a, a', k)$ is increasing in a on $[t(K), \overline{a}]$ for any $k \in \mathcal{D}$ by Lemma 11, $\phi^p(-a, a', \mathcal{D}) \coloneqq \min_{k \in \mathcal{D}} \phi^p\left(-a, a', k\right)$ is increasing in a on $[t(K), \overline{a}]$, too. Taken together, we obtain

$$
\begin{aligned}
\xi^p(\mathcal{D}) &\coloneqq \max\left\{a \geq 0 : \max_{a' \in [t(-K), t(K)]} \phi^p\left(-a, a', \mathcal{D}\right) \leq 0\right\} \\
&= \max\left\{a \geq t(K) : \max_{a' \in [t(-K), t(K)]} \phi^p\left(-a, a', \mathcal{D}\right) \leq 0\right\}.
\end{aligned}
$$

Step 2. There are three kinds of influential coalitions: (a) $\max \mathcal{C} \leq 0$, (b) $\min \mathcal{C} \geq 0$, and (c) $\min \mathcal{C} < 0 < \max \mathcal{C}$. Consider case (a), and notice two things. First, the following are equivalent for any $a \geq t(K)$ and any $a' \in [-a, a]$ by Lemma 10: (i) $\phi^p(-a, a', \mathcal{C}) \leq 0$, (ii) $\phi^p(-a, a', \max \mathcal{C}) \leq 0$, and (iii) $\phi^p(-a, a', \{k : k \leq \max \mathcal{C}\}) \leq 0$. Second, since $\mathcal{C}$ is influential and $\mathcal{C} \subseteq \{k : k \leq \max \mathcal{C}\}$, $\{k : k \leq \max \mathcal{C}\}$ is influential, too. Combining these observations yields

$$\min_{\substack{\mathcal{C}s \text{ formed under } \langle \boldsymbol{\chi}, q \rangle \\ \text{s.t. } \max \mathcal{C} \leq 0}} \xi^p(\mathcal{C}) = \min_{\substack{\mathcal{C}s \text{ formed under } \langle \boldsymbol{\chi}, q \rangle \\ \text{s.t. } \mathcal{C} = \{k : k \leq \alpha\}, \alpha \leq 0}} \xi^p(\mathcal{C}). \tag{13}$$

A close inspection of (13) reveals two things. First,

$$\xi^p(\{k : k \leq \alpha\}) = \max \left\{ a \geq t(K) : \max_{a' \in [-t(K), t(K)]} \phi^p\left(-a, a', \{k : k \leq \alpha\}\right) \leq 0 \right\}$$

is increasing in α on $\{\alpha : \alpha \leq 0\}$ by Lemma 10. Second, every set $\{k : k \leq \alpha\}$ with $\alpha < 0$ is more likely to be influential under q' than under q because $q \succeq_{SOSD} q'$. Thus

$$\min_{\substack{\mathcal{C}s \text{ formed under } \langle \boldsymbol{\chi}, q \rangle \\ \text{s.t.} \mathcal{C} = \{k : k \leq \alpha\}, \alpha \leq 0}} \xi^p(\mathcal{C}) \geq \min_{\substack{\mathcal{C}s \text{ formed under } \langle \boldsymbol{\chi}, q' \rangle \\ \text{s.t. } \mathcal{C} = \{k : k \leq \alpha\}, \alpha \leq 0}} \xi^p(\mathcal{C}),$$

which proves the desired result for case (a). The proofs for cases (b) and (c) are similar and are therefore omitted for brevity. □

E.3 Omitted proofs from Appendices B and C

Lemma 12. *Fix any personalization technology $\mathcal{S} \in \{b, p\}$ and any symmetric policy profile $\langle -a, a \rangle$ with $a \geq 0$. Assume Assumption 1. Then for any two marginal attention costs $\lambda'' > \lambda' > 0$ such that the corresponding environments satisfy Assumption 2, the optimal news signal consumed by any type k voter is more Blackwell informative when the marginal attention cost equals λ' than when it equals λ''.*

Proof. In the broadcast case, the optimal news signal is symmetric and so becomes less Blackwell informative as λ increases. For the personalized case, we only prove the result for an arbitrary type $k \leq 0$ voter, for whom we write $v(\mathbf{a}, k) = v$ and note that $v \leq 0$.

Step 1. Suppose the infomediary maximizes the voter's expected utility as in the standard RI model. For the same reason given in the proof of Lemma 1, the solution to the infomediary's problem is unique and has at most two signal realizations. In the case where the solution is binary and induces interior beliefs, the first-order conditions of utility maximization are

$$v + \mu_R = \lambda \left[\Delta h - h'(\mu_L)\Delta\mu\right] \tag{14}$$

$$\text{and} - (v + \mu_L) = \lambda \left[h'(\mu_R)\Delta\mu - \Delta h\right] \tag{15}$$

where $\Delta h \coloneqq h(\mu_R) - h(\mu_L)$ and $\Delta\mu \coloneqq \mu_R - \mu_L$. Summing up (14) and (15) yields

$$h'(\mu_R) - h'(\mu_L) = 1/\lambda, \tag{16}$$

and using (16) when differentiating (15) with respect to λ yields

$$\begin{aligned}\frac{d\mu_L}{d\lambda} =& \Delta h - h'(\mu_R)\Delta\mu + \lambda \left[\begin{array}{c} h'(\mu_R)\frac{d\mu_R}{d\lambda} - h'(\mu_L)\frac{d\mu_L}{d\lambda} - h''(\mu_R)\frac{d\mu_R}{d\lambda}\Delta\mu \\ -h'(\mu_R)\frac{d\mu_R}{d\lambda} + h'(\mu_R)\frac{d\mu_L}{d\lambda}\end{array}\right] \\ =& \Delta h - h'(\mu_R)\Delta\mu - \lambda h''(\mu_R)\frac{d\mu_R}{d\lambda}\Delta\mu + \frac{d\mu_L}{d\lambda}.\end{aligned}$$

Therefore,

$$\frac{d\mu_R}{d\lambda} = \frac{\Delta h - h'(\mu_R)\Delta\mu}{\lambda h''(\mu_R)\Delta\mu} < 0, \tag{17}$$

where the inequality holds because h is symmetric around zero, $h' > 0$ on $[0, 1]$, $h'' > 0$, and $\Delta\mu > 0$.[40] Meanwhile, differentiating (16) with respect to λ yields

$$h''(\mu_L)\frac{d\mu_L}{d\lambda} = h''(\mu_R)\frac{d\mu_R}{d\lambda} + \frac{1}{\lambda^2},$$

[40] For readers interested in the algebra, note that $\Delta h - h'(\mu_R)\Delta\mu = h(\mu_R) - h(|\mu_L|) - h'(\mu_R)\Delta\mu < h'(\mu_R)(\mu_R - |\mu_L| - \Delta\mu) = -2h'(\mu_R)|\mu_L| < 0$ if $\mu_R > |\mu_L|$, and $h(\mu_R) - h(|\mu_L|) - h'(\mu_R)\Delta\mu \leq -h'(\mu_R)\Delta < 0$ if $\mu_R \leq |\mu_L|$.

and simplifying using (17) yields

$$\frac{d\mu_L}{d\lambda} = \frac{\Delta h - h'(\mu_L)\,\Delta\mu}{\lambda h''(\mu_L)\,\Delta\mu} > 0. \tag{18}$$

Together, (17) and (18) imply that the utility-maximizing signal (hereafter denoted by $\Pi^c(a,k,\lambda)$) becomes less Blackwell informative as λ increases.

Step 2. Write type k voter's optimal personalized signal as $\Pi^p(a,k,\lambda)$, and recall that $\Pi^p(a,k,\lambda) = \Pi^c(a,k,\lambda - 1/\gamma(\gamma))$ for some $0 < \gamma(\lambda) < 1/\lambda$. Write $\beta(\lambda) = \lambda - 1/\gamma(\lambda)$. If we can show that $\beta(\lambda'') > \beta(\lambda')$ for any $\lambda'' > \lambda' > 0$ that satisfy Assumption 2, then $\Pi^p(a,k,\lambda')$ is more Blackwell informative than $\Pi^p(a,k,\lambda'')$ by Step 1.

Suppose to the contrary that $\beta(\lambda'') \leq \beta(\lambda')$ for some $\lambda'' > \lambda' > 0$ as above. Write $\Pi^p(a,k,\lambda') = \Pi'$ and $\Pi^p(a,k,\lambda'') = \Pi''$, and note that Π' and Π'' are nondegenerate and induce interior beliefs by assumption. From Step 1, we deduce that if $\beta(\lambda'') < \beta(\lambda')$, then Π'' is more Blackwell informative than Π', so in particular $I(\Pi'') > I(\Pi') > 0$. Then from $V(\Pi';\mathbf{a},k) = \lambda' I(\Pi')$ and $V(\Pi'';\mathbf{a},k) = \lambda'' I(\Pi'')$, it follows that $V(\Pi'';\mathbf{a},k) - \lambda' I(\Pi'') > 0 = V(\Pi';\mathbf{a},k) - \lambda' I(\Pi')$, which together with $I(\Pi'') > I(\Pi')$ implies that Π' is not optimal when $\lambda = \lambda'$, a contradiction. Meanwhile, if $\beta(\lambda') = \beta(\lambda'')$, then $\Pi' = \Pi''$ and, hence, $V(\Pi';\mathbf{a},k) = \lambda' I(\Pi') < \lambda'' I(\Pi'') = V(\Pi'';\mathbf{a},k)$, a contradiction. □

Proof of Proposition 7 Since the optimal personalized signal for any voter is more Blackwell informative than his competitive signal (Lemma 12), the following holds for any $a \geq 0$, a', and $\mathcal{D} \subseteq \mathcal{K}$:

$$\begin{aligned}\phi^c\left(-a,a',\mathcal{D}\right) &:= \min_{k\in\mathcal{D}} \phi^c\left(-a,a',k\right) := \min_{k\in\mathcal{D}} v\left(-a,a',k\right) + \mu_L^c(a,k) \\ &> \min_{k\in\mathcal{D}} v\left(-a,a',k\right) + \mu_L^p(a,k) := \phi^p\left(-a,a',\mathcal{D}\right).\end{aligned}$$

Thus letting $\mathcal{S} = c$ in the proof of Lemma 9 yields $\xi^c(\mathcal{D}) < \xi^p(\mathcal{D})$, where $\xi^c(\mathcal{D})$ denotes $\mathcal{D}$'s policy latitude in the competitive case. Then for any c-consistent $\boldsymbol{\chi}$ and p-consistent

$\boldsymbol{\chi}'$ such that $\boldsymbol{\chi} \succeq \boldsymbol{\chi}'$, we must have

$$\begin{aligned}
\mathcal{E}^{c,\boldsymbol{\chi},\rho} &= \left[0, \min_{\mathcal{C}s \text{ formed under } \langle \boldsymbol{\chi},\rho\rangle} \xi^{c}\left(\mathcal{C}\right)\right] && \text{(Theorem 3; } \boldsymbol{\chi} \text{ is } c\text{-consistent)} \\
&\subseteq \left[0, \min_{\mathcal{C}s \text{ formed under } \langle \boldsymbol{\chi}',\rho\rangle} \xi^{c}\left(\mathcal{C}\right)\right] && \text{(Proposition 4; } \boldsymbol{\chi} \succeq \boldsymbol{\chi}'\text{)} \\
&\subsetneq \left[0, \min_{\mathcal{C}s \text{ formed under } \langle \boldsymbol{\chi}',\rho\rangle} \xi^{p}\left(\mathcal{C}\right)\right] && \left(\xi^{c}\left(\mathcal{C}\right) < \xi^{p}\left(\mathcal{C}\right)\right) \\
&= \mathcal{E}^{p,\boldsymbol{\chi}',\rho}, && \text{(Theorem 3; } \boldsymbol{\chi}' \text{ is } p\text{-consistent)}
\end{aligned}$$

hence completing the proof. □

Proof of Proposition 8 Fix any $\mathbf{a} = \langle -a, a\rangle$ with $a > 0$. As demonstrated in the proof of Lemma 1, it is without loss to strengthen Assumption 3 **increasing differences** to $v(\mathbf{a}, k)$ being strictly increasing in k, because if $v(\mathbf{a}, k) = v(\mathbf{a}, k+1)$ for some k, then we can treat type k and $k+1$ voters as a single entity.

Part (i): Any optimal personalized signal that is nondegenerate and makes its consumers' participation constraint binding (let $\gamma(k) > 0$ denote the corresponding Lagrange multiplier) solves

$$\max_{\mathcal{Z},\Pi:\mathbb{R}\to\Delta(\mathcal{Z})} V\left(\Pi; \mathbf{a}, k\right) - \left(\lambda - 1/\gamma\left(k\right)\right) I\left(\Pi\right) \tag{19}$$

where $\lambda(k) := \lambda - 1/\gamma(k) > 0$ must hold in order to satisfy $\gamma(k) > 0$. By Matějka and McKay (2015), any nondegenerate solution to (19) must be binary and, hence, satisfy (SOB). Take any such solution, and let $\mathcal{L}(k)$ denote the likelihood that the voter votes for R rather than L. Below we demonstrate that $\mathcal{L}(k) < 1$ if $k < 0$, $\mathcal{L}(k) = 1$ if $k = 0$, and $\mathcal{L}(k) > 1$ if $k > 0$, which together with Bayes' plausibility implies the skewness properties stated in Theorem 1(ii).

By Matějka and McKay (2015), the probability that voter k votes for R in state ω equals

$$\frac{\mathcal{L}\left(k\right)\exp\left(\frac{v(\mathbf{a},k)+\omega}{\lambda(k)}\right)}{\mathcal{L}\left(k\right)\exp\left(\frac{v(\mathbf{a},k)+\omega}{\lambda(k)}\right)+1}. \tag{20}$$

Thus for any given $\lambda(k) > 0$, (19) becomes

$$\max_{\mathcal{L}\in[0,+\infty)} \mathbb{E}_G \left[(v(\mathbf{a},k)+\omega) \frac{\mathcal{L}\exp\left(\frac{v(\mathbf{a},k)+\omega}{\lambda(k)}\right)}{\mathcal{L}\exp\left(\frac{v(\mathbf{a},k)+\omega}{\lambda(k)}\right)+1} \right] - \lambda(k) I(\mathcal{L})$$

where $I(\mathcal{L})$ denotes the mutual information of the valence state and the voting decision made based on $\mathcal{L}$ and (20). Since the maximand has strict increasing differences in $(\mathcal{L}, v(\mathbf{a},k))$, $\mathcal{L}(k)$ must be increasing in k. If, in addition, $\mathcal{L}(k) = 1$ if and only if $k = 0$ (equivalently, $v(\mathbf{a},k) = 0$), then we are done.

To verify that $\mathcal{L}(k) = 1$ if and only if $k = 0$, we write $\widetilde{v}$ for $v(\mathbf{a},k)/\lambda(k)$, $\widetilde{\omega}$ for $\omega/\lambda(k)$, and $\mathcal{L}$ for $\mathcal{L}(k)$. Since G is symmetric around zero, $\mathcal{L} = 1$ if and only if

$$\int_0^\infty \frac{\exp(\widetilde{\omega}+\widetilde{v})}{\exp(\widetilde{\omega}+\widetilde{v})+1} + \frac{\exp(-\widetilde{\omega}+\widetilde{v})}{\exp(-\widetilde{\omega}+\widetilde{v})+1} dG(\omega) = \frac{1}{2}. \tag{21}$$

A close inspection of (21) reveals that its the left-hand side is strictly increasing in $\widetilde{v}$, so we only need to verify (21) for $\widetilde{v} = 0$, which is clearly true.

Part (ii): Take any optimal broadcast signal that induces consumption from all voters, and let $\mathcal{B} \neq \emptyset$ denote the set of voters with binding participation constraints. Since $v(\mathbf{a},k)$ is strictly increasing in k, we must have $\mathcal{B} \subseteq \{-K, K\}$ and, indeed, $\mathcal{B} = \{-K, K\}$, because if $\mathcal{B} \subsetneq \{-K, K\}$, then the signal we begin with is the optimal personalized signal for the voter in $\mathcal{B}$ and so violates the participation constraint of the voter in $\{-K, K\} - \mathcal{B}$.

For each $k \in \mathcal{B}$, let $\gamma(k) > 0$ denote the Lagrange multiplier associated with voter k's participation constraint. As demonstrated in the proof of Lemma 1, we can reformulate the infomediary's problem as

$$\max_{\mathcal{Z},\Pi:\mathbb{R}\to\Delta(\mathcal{Z})} \sum_{k\in\mathcal{B}} \frac{\gamma(k)}{\sum_{k\in\mathcal{B}}\gamma(K)} V(\Pi;\mathbf{a},k) - \left(\lambda - \frac{1}{\sum_{k\in\mathcal{B}}\gamma(k)}\right) I(\Pi) \tag{22}$$

where $\lambda - \frac{1}{\sum_{k\in\mathcal{B}}\gamma(k)} > 0$ must hold in order to satisfy $\gamma(k) > 0$ for all $k \in \mathcal{B}$. A careful inspection of (22) reduces it to the same kind of the optimal information acquisition problem studied by Matějka and McKay (2015), whereby a representative voter makes three decisions LL, LR and RR on behalf of the voters in $\mathcal{B}$ (the first and second letters stand for the voting recommendations to type $-K$ and K voters, respectively) and pays an information acquisition cost that is proportional to the mutual information of the valence state and voting decision. By Matějka and McKay (2015), any solution $\Pi : \Omega \to \Delta(\mathcal{Z})$ to (22) must satisfy $\mathcal{Z} \subseteq \{LL, LR, RR\}$ and make the voters in $\mathcal{B}$ obey the voting recommendations given to them (hereafter *obedience*).

Case 1. $|\mathcal{Z}| = 2$ In this case, $\mathcal{Z}$ must equal $\{LL, RR\}$, and Π must induce strict obedience from its consumers, i.e., $v(\mathbf{a}, -K) + \mu_{LL} > 0$ and $v(\mathbf{a}, K) + \mu_{RR} < 0$. To show that Π is symmetric, i.e., $\Pi(LL \mid \omega) = \Pi(RR \mid -\omega)$ a.e., suppose the contrary is true, and consider a new signal structure $\Pi' : \Omega \to \Delta(\mathcal{Z})$ where $\Pi'(LL \mid \omega) = \Pi(RR \mid -\omega)$ for all ω. For each $z \in \mathcal{Z}$, write π'_z for $\int \Pi'(z \mid \omega)\, dG(\omega)$ and μ'_z for $\int \omega \Pi'(z \mid \omega)\, dG(\omega) / \pi'_z$. By construction, we have $\pi'_{LL} = \pi_{RR}$, $\pi'_{RR} = \pi_{LL}$, $\mu'_{LL} = -\mu_{RR}$, $\mu'_{RR} = -\mu_{LL}$, and $I(\Pi) = I(\Pi')$. Thus

$$\begin{aligned}
V\left(\Pi'; \mathbf{a}, -K\right) &= \pi'_{RR}\left[v(\mathbf{a}, -K) + \mu'_{RR}\right] \\
&= \pi_{LL}\left[-v(\mathbf{a}, K) - \mu_{LL}\right] \\
&= V(\Pi; \mathbf{a}, K) \\
&= \lambda I(\Pi) && (K \in \mathcal{B}) \\
&= V(\Pi; \mathbf{a}, -K), && (-K \in \mathcal{B})
\end{aligned}$$

and $V(\Pi'; \mathbf{a}, K) = V(\Pi; \mathbf{a}, K)$ can be shown analogously. So compared to Π (or Π'), the signal structure $\frac{1}{2}\Pi + \frac{1}{2}\Pi'$ generates the same utility of consumption to the representative voter in (22) but incurs a strictly lower attention cost because $I(\Pi)$ is strictly convex in its argument (see, e.g., Theorem 2.7.4. of Cover and Thomas (2006)). Thus Π isn't a solution to (22), a contradiction.

Case 2. $\mathcal{Z} = \{LL, LR, RR\}$ In this case, the voters in $\mathcal{B}$ must strictly prefer to obey the voting recommendations prescribed by LL and RR, and weakly prefer to obey the voting recommendations prescribed by LR, i.e., $v(\mathbf{a}, k) + \mu_{LL} < 0 < v(\mathbf{a}, k) + \mu_{RR}$ for all $k \in \mathcal{B}$, and $v(\mathbf{a}, -K) + \mu_{LR} \leq 0 \leq v(\mathbf{a}, K) + \mu_{LR}$. To show that Π is symmetric, i.e., $\Pi(LL \mid \omega) = \Pi(RR \mid -\omega)$, $\Pi(LR \mid \omega) = \Pi(LR \mid -\omega)$, and $\Pi(RR \mid \omega) = \Pi(LL \mid -\omega)$ a.e., suppose the contrary is true, and consider a new signal structure $\Pi' : \Omega \to \Delta(\mathcal{Z})$ where $\Pi'(LL \mid \omega) = \Pi(RR \mid -\omega)$, $\Pi'(LR \mid \omega) = \Pi(LR \mid -\omega)$, and $\Pi'(RR \mid \omega) = \Pi(LL \mid -\omega)$. By construction, we have $\pi'_{LL} = \pi_{RR}$, $\pi'_{LR} = \pi_{LR}$, $\pi'_{RR} = \pi_{LL}$, $\mu'_{LL} = -\mu_{RR}$, $\mu'_{LR} = -\mu_{LR}$, and $\mu'_{RR} = -\mu_{LL}$. Combining this with obedience yields

$$\begin{aligned}
v(\mathbf{a}, -K) + \mu'_{LL} &= -v(\mathbf{a}, K) - \mu_{RR} < 0 \\
v(\mathbf{a}, -K) + \mu'_{LR} &= -v(\mathbf{a}, K) - \mu_{LR} \leq 0 \\
v(\mathbf{a}, -K) + \mu'_{RR} &= -v(\mathbf{a}, K) - \mu_{LL} > 0 \\
v(\mathbf{a}, K) + \mu'_{LL} &= -v(\mathbf{a}, -K) - \mu_{RR} < 0 \\
v(\mathbf{a}, K) + \mu'_{LR} &= -v(\mathbf{a}, -K) - \mu_{LR} \geq 0 \\
\text{and } v(\mathbf{a}, K) + \mu'_{RR} &= -v(\mathbf{a}, -K) - \mu_{LL} > 0,
\end{aligned}$$

so

$$\begin{aligned}
V\left(\Pi'; \mathbf{a}, -K\right) &= \pi'_{RR}\left[v(\mathbf{a}, -K) + \mu'_{RR}\right] \\
&= \pi_{LL}\left[-v(\mathbf{a}, K) - \mu_{LL}\right] = V(\Pi; \mathbf{a}, K) = \lambda I(\Pi) = V(\Pi; \mathbf{a}, -K),
\end{aligned}$$

and $V(\Pi'; \mathbf{a}, K) = V(\Pi'; \mathbf{a}, K)$ can be shown analogously. The remainder of the proof follows that for case 1 and is therefore omitted for brevity. □

E.4 Omitted proofs from Sections 4-6

Throughout this appendix, suppose $\mathcal{K} = \{-1, 0, 1\}$ and $u(a,k) = -|t(k) - a|$, so in particular

$$v(-a, a, k) = \begin{cases} -2a & \text{if } t(k) < -a, \\ 2t(k) & \text{if } -a \leq t(k) \leq a, \\ 2a & \text{if } t(k) > a. \end{cases}$$

In the proof of Lemmas 1 and 2, letting $v(-a, a, k) \equiv v(-|t(k)|, |t(k)|, k)$ on $[|t(k)|, \overline{a}]$ yields $\mu_L^b(a) \equiv \mu_L^b(t(1)) := \mu_L^b$ on $[t(1), \overline{a}]$ and $\mu_L^p(a, k) \equiv \mu_L^p(|t(k)|, k) := \mu_L^p(k)$ on $[|t(k)|, \overline{a}]$ for any k.

Proof of Lemma 5 Part (i): When proving Lemma 11, we already demonstrated that $\mu_L^b(a)$ is nondecreasing in a, so $\phi^b(-a, 0, 0) = a + \mu_L^b(a)$ is strictly increasing in a. Then from $\phi^b(-a, 0, 0)\Big|_{a=0} = \mu_L^b(0) < 0$, it follows that $\xi^b(0) := \max\left\{0 \leq a \leq \overline{a} : \phi^b(-a, 0, 0) \leq 0\right\}$ is the unique root of $\phi^b(-a, 0, 0)$ when $\overline{a}$ is sufficiently large. If, in addition, $\xi^b(0) \geq t(1)$, then solving $\phi^b(-a, 0, 0) = 0$ yields $\xi^b(0) = -\mu_L^b\left(\xi^b(0)\right) = -\mu_L^b$.

Part (ii): For $k = 1$, notice that $\phi^p(-a, t(1), 1) = a + t(1) + \mu_L^p(1)$ when $a \geq t(1)$, and $\phi^p(-a, t(1), 1)\Big|_{a=t(1)} = v(-t(1), t(1), 1) + \mu_L^p(t(1), 1) < 0$ by (SOB). Therefore, $\xi^p(1) := \max\{0 \leq a \leq \overline{a} : \phi^p(-a, t(1), 1) \leq 0\}$ is the unique root of $\phi^p(-a, t(1), 1)$, which equals $-\left[t(1) + \mu_L^p(1)\right]$ when $\overline{a}$ is sufficiently large. The proofs for $k = 0$ and $k = -1$ are similar and are therefore omitted for brevity. $\square$

Completing Example 2 Let everything be as in the baseline model, except that extreme voters abstain from news consumption and vote along party lines when they are indifferent between the candidates. Take any symmetric policy profile $\langle -a, a \rangle$ with $a \in \left[0, \underline{\xi}\right]$, where $\underline{\xi} := \min\left\{t(1), \xi^{\mathcal{S}}(0)\right\}$. Note that any unilateral deviation of candidate R from $\langle -a, a \rangle$ doesn't attract median voters, and it cannot increase the total number of votes that extreme voters cast to candidate R. Combining these observations gives the desired result. $\square$

Completing Example 3 Fix the policy profile to $\langle -t(1), t(1) \rangle$. For each $k \in \{-1, 1\}$, write $\mu_z(k)$ for the posterior mean of the state conditional on type k voters' signal realization being $z \in \{L, M, R\}$, where $v(-t(1), t(1), k) + \mu_L(k) < v(-t(1), t(1), k) + \mu_M(k) = 0 < v(-t(1), t(1), k) + \mu_R(k)$ by assumption. Combining this with Bayes' plausibility yields $\mu_L(k) < 0 < \mu_R(k)$.

Consider any unilateral deviation of candidate R from $\langle -t(1), t(1) \rangle$ to a'. Clearly, no $a' \notin [-t(1), t(1)]$ constitutes a profitable deviation, and no $a' \in [-t(1), t(1)]$ attracts median voters whose policy latitude is assumed to be greater than $t(1)$. It remains to show that no $a' \in [-t(1), t(1))$ affects extreme voters' voting decisions. To avoid repetition, we only prove this result for $k = -1$ and omit the proof for $k = 1$. By Assumption 3, the following holds for $z = L$:

$$\begin{aligned}
& v\left(-t(1), a', -1\right) + \mu_L(-1) \\
& \leq v(-t(1), t(-1), -1) + \mu_L(-1) && \textbf{(inverted V-shape)} \\
& = v(-t(1), -t(1), -1) + \mu_L(-1) && \textbf{(symmetry)} \\
& = 0 + \mu_L(-1) \\
& < 0,
\end{aligned}$$

and the following hold for $z = M, R$:

$$\begin{aligned}
& v\left(-t(1), a', -1\right) + \mu_z(-1) \\
& > v(-t(1), t(1), -1) + \mu_z(-1) && \textbf{(inverted V-shape)} \\
& \geq 0.
\end{aligned}$$

To complete the proof, suppose type -1 voters vote for candidate R when they are indifferent between the two candidates, and we are done. □

References

Allcott, H., and M. Gentzkow. (2017): "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*, 31(2), 211-236.

Ambuehl, S., A. Ockenfels, and C. Stewart. (2019): "Attention and selection effects," *Working Paper.*

Anderson, S. P., D., Strömberg, and J. Waldfogel, eds. (2016): *Handbook of Media Economics,* Elsevier.

Athey, S., and M. Mobius. (2012): "The impact of aggregators on Internet news consumption: The case of localization," *Working Paper.*

Athey, S., M. Mobius, and J. Pal. (2017): "The Impact of aggregators on Internet news consumption," *Working Paper.*

Aumann, R., and M. Maschler. (1995): *Repeated Games with Incomplete Information,* Cambridge, MA: MIT Press.

Barber, M. J., and N. McCarty. (2015): "Causes and Consequences of Polarization," in *Solutions to Polarization in America*, ed. by Persily, N.. Cambridge, U.K.: Cambridge University Press, 15-58.

Barbera, P. (2020): "Social Media, Echo Chambers, and Political Polarization," in *Social Media and Democracy: The State of the Field*, ed. by N. Persily., and J. Tucker.. Cambridge University Press.

Burke, J. (2008): "Primetime spin: Media bias and belief confirming information," *Journal of Economics and Management Strategy*, 17(3), 633-665.

Calvert, R. L. (1985a): "The value of biased information: A rational choice model of political advice," *Journal of Politics*, 47(2), 530-555.

——— (1985b): "Robustness of the multidimensional voting model: Candidate motivations, uncertainty, and convergence," *American Journal of Political Science*, 29(1), 69-95.

Caplin, A., and M. Dean. (2013): "Behavioral implications of rational inattention with Shannon entropy," *NBER working paper*.

——— (2015): "Revealed preference, rational inattention, and costly information acquisition," *American Economic Review*, 105(7), 2183-2203.

Caplin, A., M. Dean, and J. Leahy. (2019): "Rationally inattentive behavior: Characterizing and generalizing Shannon entropy," *Working Paper*.

Chan, J., and W. Suen. (2008): "A spatial theory of news consumption and electoral competition," *Review of Economic Studies*, 75(3), 699-728.

Che, Y-K., and K. Mierendorff. (2019): "Optimal dynamic allocation of attention," *American Economic Review*, 109(8), 2993-3029.

Chiang, C-F., and B. Knight. (2011): "Media bias and influence: evidence from newspaper endorsements," *Review of Economic Studies*, 78(3), 795-820.

Chiou, L., and C. Tucker. (2017): "Content aggregation by platforms: The case of the news media," *Journal of Economics and Management Strategy*, 26, 782-805.

Cover, T. M., and J. A. Thomas. (2006): *Elements of Information Theory,* Hoboken, NJ: John Wiley & Sons, 2nd ed.

Dean, M., and N. Nelighz. (2019): "Experimental tests of rational inattention," *Working Paper*.

Dellarocasm, C., J. Sutanto, M. Calin, and E. Palme. (2016): "Attention allocation in information-rich environments: The case of news aggregators," *Management Science*, 62(9), 2457-2764.

DellaVigna, S., and M. Gentzkow. (2010): "Persuasion: Empirical evidence," *Annual Review of Economics*, 2, 643-669.

Denti, T. (2018): "Posterior-separable cost of information," *Working Paper*.

Downs, A. (1957): *An Economic Theory of Democracy,* New York, NY: Harper and Row, 1st ed.

Duggan, J. (2017): "A survey of equilibrium analysis in spatial model of elections," *Working Paper*.

Duggan, J., and C. Martinelli. (2011): "A spatial theory of media slant and voter choice," *Review of Economic Studies*, 78(2), 640-666.

Dunaway, J. (2016): "Mobile vs. computer: Implications for news audiences and outlets," *Shorenstein Center on Media, Politics and Public Policy*, August 30.

European Parliament and Council of European Union (2016) Regulation (EU) 2016/679, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN, Accessed 08/08/2020.

Eslami, M., A. Aleyasen, K. G. Karahalios, K. Hamilton, and C. Sandvig. (2015): "FeedVis: A path for exploring news feed curation algorithms," *CSCW'15 Companion: Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, 65-68.

Fiorina, M. P., and S. J. Abrams. (2008): "Political polarization in the American public," *Annual Review of Political Science*, 11, 563-588.

Flaxman, S., S. Goel, and J. M. Rao. (2016): "Filter bubbles, echo chambers and online news consumption," *Public Opinion Quarterly*, 80(S1), 298-320.

Gentzkow, M. (2016): "Polarization in 2016," *Working Paper*. Stanford University.

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

Gerber, A. S., J. G. Gimpel, D. P. Green, and D. R. Shaw. (2011): "How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment," *American Political Science Review*, 105(1), 135-150.

Hamilton, J. (2004): *All the News That Is Fit to Sell: How the Market Transforms Information into News,* Princeton, New Jersey: Princeton University Press.

Hébert, B., and M. D. Woodford. (2017): "Rational inattention in continuous time," *Working Paper.*

Herrera, H., D. K. Levine, and C. Martinelli. (2008): "Policy platforms, campaign spending and voter participation," *Journal of Public Economics*, 92, 501-513.

Hersh, E. D. (2015): *Hacking the Electorate: How Campaigns Perceive Voters,* Cambridge, U.K.: Cambridge University Press.

Jung, J., J. Kim, F. Matějka, and C. A. Sims. (2019): "Discrete actions in information-constrained problems," *Review of Economic Studies*, 86(6), 2643-2667.

Kamenica, E., and M. Gentzkow. (2011): "Bayesian persuasion," *American Economic Review*, 101(6), 2590-2615.

Lagun. D, and M. Lalmas. (2016): "Understanding user attention and engagement in online news reading," *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, 113-122.

Li, A., L. Hu, and I. Segal. (2020): "Electoral accountability and selection with personalized news aggregation," *Working Paper.*

Matějka, F., and A. McKay. (2015): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105(1), 272-298.

MATĚJKA, F., AND G. TABELLINI. (2016): "Electoral competition with rationally inattentive voters," *Working Paper.*

MATSA, K. E., AND K. LU. (2016): "10 Facts about the changing digital news landscape," *Pew Research Center*, September 14.

MITCHELL, A., G. STOCKING, AND K. E. MATSA. (2016): "Long-form of reading shows signs of life in our mobile news world," *Pew Research Center*, May 5.

MORRIS, S., AND P. STRACK. (2017): "The Wald problem and the equivalence of sequential sampling and static information costs," *Working Paper.*

MULLAINATHAN. S, AND A. SHLEIFER. (2005): "The market for news," *American Economic Review*, 95(4), 1031-1053.

OBAMA, B. (2017): "President Obama's farewell address," https://obamawhitehouse.archives.gov/farewell, Accessed 03/28/2019.

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

PEREGO, J., AND S. YUKSEL. (2018): "Media competition and social disagreement," *Working Paper.*

PRAT, A. (2018): "Media power," *Journal of Political Economy*, 126(4), 1747-1783.

PRAT, A., AND D. STRÖMBERG. (2013): "The Political Economy of Mass Media," in *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*, ed. by D. Acemoglu, M. Arellano, and E. Dekel.. Cambridge University Press.

PRATO, C., AND S. WOLTON. (2016): "The voters' curses: Why we need goldilocks voters," *American Journal of Political Science*, 60(3), 726-737.

PRUMMER, A. (2020): "Micro-targeting and polarization," *Journal of Public Economics*, 188, 104210.

Shannon, C. E. (1948): "A mathematical theory of communication," *Bell Labs Technical Journal*, 27(3), 379-423.

Sims, C. A. (1998): "Stickiness," *Carnegie-Rochester Conference Series on Public Policy*, 49(1), 317-356.

——— (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50(3), 665-690.

Strömberg, D. (2004): "Mass media competition, political competition, and public policy," *Review of Economic Studies*, 71(1), 265-284.

——— (2015): "Media and politics," *Annual Review of Economics*, 7, 173-205.

Suen, W. (2004): "The self-perpetuation of biased beliefs," *The Economic Journal*, 114, 377-396.

Sunstein, C. R. (2009): *Republic.com 2.0*, Princeton, NJ: Princeton University Press.

Teixeira, T. S. (2014): "The rising cost of consumer attention: Why you should care, and what you can do about it," *Working Paper*.

The Digital Competition Expert Panel. (2019): *Unlocking Digital Competition,* U.K.

Tsakas, E. (2019): "Robust scoring rules," *Theoretical Economics*, 15, 955-987.

Warren, E. (2019): "Here's how we can break up Big Tech," *Medium*, March 8.

Zhong, W. (2017): "Optimal dynamic information acquisition," *Working Paper*.