The Politics of News Personalization

Lin Hu[*] Anqi Li[†] Ilya Segal[‡]

November 2019

Abstract

We study how news personalization affects policy polarization. In a two-candidate electoral competition model, an attention-maximizing infomediary aggregates information about candidate valence into news, whereas voters decide whether to consume news, trading off the expected utility gain from improved expressive voting against the attention cost. Broadcast news attracts a broad audience by offering a symmetric signal. Personalized news serves extreme voters with skewed signals featuring own-party bias and occasional big surprise. Rational news aggregation yields policy polarization even if candidates are office-motivated. Personalization makes extreme voters the disciplining entity for equilibrium polarization and increases polarization through occasional big surprise.

[*]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

We examine how personalizated news aggregation for rational inattentive voters affects policy polarization and public opinion. The idea that tech-enabled news personalization can affect political actions and outcomes has recently been put forward in the popular press (Sunstein (2009); Pariser (2011); Obama (2017)). However, news consumption has been usually explained by behavioral biases (Mullainathan and Shleifer (2005)) rather than rational choices. The goal of this paper is to examine what kind of personalized news is aggregated for and consumed by rational inattentive voters, and how this affects policy polarization in a model of electoral competition.

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,[1] which in turn provide personalized services based on the voluminous data gathered about individual consumers (e.g., demographic and psychographic attributes, digital footprints, social network positions). In this paper, we abstract from the issue of information generation (e.g., original reporting), focusing instead on the role of infomediaries in aggregating the available information into news that is easy to process and useful for the target audience.

We model an infomediary who has full flexibility in aggregating the available information into news signals. While flexibility is also assumed in the Rational Inattention model (Sims (1998; 2003)), in that model consumers aggregate information optimally themselves and have no need for external aggregators.[2] 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 selectively[3] or aggregate

[1]Recently, news aggregators (e.g., aggregator sites, social media feeds, mobile news apps) have gained prominence as more people get news online, from social media and through mobile devices (Matsa and Lu (2016); Shearer (2018)). Using automated algorithms, aggregators sift through myriad online sources and direct readers to the stories they find interesting. 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 Jeon (2018) for background readings and literature surveys.

[2]Strömberg (2015) first notes this difference between the Rational Inattention model and media models, stipulating that "in the rational inattention model, voters choose what information to pay attention to given their cognitive constraints; in media models, the media chooses what information is most profitable to make available to voters."

[3]Analyses of page activities (e.g., scrolling, viewport time) have established significant levels of

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

The cornerstone of our paper is a model of news aggregation for rational inattentive consumers (NARI). If choosing to consume news, a consumer incurs an attention cost that is posterior separable (Caplin and Dean (2015)) while deriving a value equal to the expected utility gain from improved decisions. Consuming news is optimal if its value exceeds the attention cost. As for the infomediary, we assume that its goal is to maximize the total amount of attention paid by consumers. This objective could be interpreted as the advertising revenue, which is increasing in the attention spent by consumers on the news platform. This stylized assumption captures the key trade-off faced by the infomediary, who offers consumers useful information to attract their attention, while limiting the amount of information to prevent them from tuning out. We focus on the case of a monopolistic infomediary in order to capture the power wielded by the tech giants, yet also consider an extension to perfect competition which, combined with personalization, becomes equivalent to consumers optimally aggregating information themselves as in the standard Rational Inattention model.

We embed our NARI model into an electoral competition game in which two candidates compete for office by proposing policies on a left-right spectrum. Voters vote expressively for their preferred candidate based on policies, as well as a valence state that determine the candidates' relative fitness for office. While voters observe policies, they decide whether to consume news about the valence state before casting votes.[5] Candidates propose policies after the infomediary commits to news signal structures and voters make consumption decisions.

One consequence of NARI is that the infomediary gives binary recommendations as to which candidate to vote for. Indeed, any information beyond voting recommendations would only raise the attention cost without any corresponding benefit and would thus turn away news consumers whose participation constraints bind at the

user attention and engagement in online (even long-form) news reading (Lagun and Lalmas (2016); Mitchell et al. (2016)). There is also evidence that readers go throug most snippets (i.e., headlines plus excerpts), which contain substantial information even if they do not always materialize into click-throughs (Dellarocas et al. (2016)).

[4]Mitchell et al. (2017) reports that 35 percent of online news consumers use aggregators as the preferred pathway for getting most of their online news. Athey and Mobius (2012) and Chiou and Tucker (2017) provide indirect evidence, showing that aggregators reduce search costs and increase traffic to sources compared to direct browsing and web-based searches.

[5]According to Prat and Strömberg (2013), instrumental voting (alongside entertainment and taking private actions) is an important motive for consuming political news.

optimum. Another consequence is that voters strictly prefer to obey the recommendations given to them—a property we will refer to as *strict obedience*. If, instead, a voter is indifferent about obeying one of the recommendations, then he has a weak preference for a candidate that is independent of the recommendation and would therefore abstain from news consumption to save the attention cost.

An important implication of strict obedience is that local deviations from an equilibrium policy profile would not change voters' minds for either recommendation they may receive. Thus, even when, as we assume, candidates are office-motivated, there exist equilibria exhibiting policy divergence. We define *policy polarization* as the maximal distance between candidate positions among all perfect Bayesian equilibrium in which candidates adopt symmetric policy profiles.

A central question of our paper is the effect of news personalization on policy polarization. Specifically, 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 that of *personalized news*, in which she (e.g., personalized news aggregator) can design different news signals for different voters. While polarization arises in either case, its causes and magnitudes differ. Under broadcast news, 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, i.e., make them vote for the candidate unconditionally. Under usual assumptions, a deviation is profitable if and only if it attracts the median voter, so the deviation to the median voter's bliss point constrains equilibrium polarization.

The case of personalized news differs in two respects. First, as the infomediary can now provide conditionally independent signals to individual voters, a deviation is profitable if it attracts non-majorities of voters who are pivotal under some profile of news realizations. In the baseline model featuring three types of voters, each voter is pivotal when voter population is sufficiently dispersed, suggesting that deviations could be more effective and less polarization could be sustained than in the broadcast case. However, there is a countervailing effect, stemming from the skewness of the signals offered in the two cases.

To understand the effect of personalization on skewness, note first that in the broadcast case, the infomediary uses a symmetric signal to attract a broad audience. By contrast, the personalized signals offered to extreme voters are skewed: to maximize usefulness, the recommendation to vote across party lines must be sufficiently strong and, constrained by the attention cost, must be sufficiently rare; by Bayes' rule, the voter must be recommended to vote along party lines most of the time, implying that the news signal exhibits both the *own-party bias*[6] and *occasional big surprise*[7] as documented in the empirical literature.

The skewness of personalized news is crucial for sustaining greater polarization than in the broadcast case. First, skewness makes it hard for a candidate's deviation to attract his base voter in the rare event where the news signal recommends the opposing candidate. Indeed, the recommendation can be so strong that even the most attractive deviation to the base voter won't change his mind. If so, then equilibrium polarization could only be constrained by deviations aiming to attract either the median voter or the opposition voter. Note, however, that the median voter's personalized news is more informative than the broadcast news, which makes him harder to attract. The reason is that in the broadcast case, signal informativeness is determined by the participation constraints of extreme voters, who find symmetric signals less useful. As for the opposition voter, he is also difficult to attract due to his inherent preference for the opposing candidate. For these reasons, deviations may be less effective and equilibrium may be more polarized than in the broadcast case. We find this to be true when the attention cost is either quadratic or Shannon entropy-based.

We examine several comparative statics results. First, we show that mass polarization (defined as a mean-preserving spread of voters' policy preferences) needs not increase policy polarization; in fact, the opposite may occur. Second, we consider the comparative statics of the attention cost parameter, as well as the effect of introducing competition between infomediaries (which turns out to be equivalent to increasing the attention cost parameter). Third, we extend the baseline model to

[6]Own-party bias, or *party sorting*, refers to the positive correlation between voters' views and party identifications. Our prediction is consistent with the stylized fact as documented in Fiorina and Abrams (2008) and Gentzkow (2016), namely the rising party sorting in the past decade has been accompanied by little changes in voters' policy preferences or party identifications.

[7]Chiang and Knight (2011) provides a famous account for occasional big surprise, showing that newspaper endorsements of presidential candidates are most effective in shaping voter decisions when they go against the newspapers' biases. Flaxman et al. (2016) provides suggestive evidence for both own-party bias and occasional big surprise. By investigating the web-browsing patterns of 50,000 US-located Internet users, the authors find that news personalization is associated with an increase in the mean ideological distance between individuals, together with an increase in the degree of cross-partisan exposure. Interestingly, aggregators and social media feeds induce the highest degree of cross-partisan exposure among the major channels through which people discover news stories online.

arbitrary finite types of voters and allow the infomediary to correlate personalized signals across voters. We develop a methodology for analyzing this general model and demonstrate, in particular, that correlation can only increase policy polarization.

1.1 Related Literature

Media bias The literature on media bias is thoroughly surveyed by Prat and Strömberg (2013), Strömberg (2015) and Anderson et al. (2016). We add to the theoretical literature on demand-driven media bias, whose common explanations include but are not limited to: behavioral bias and limited information capacity.[8] Mullainathan and Shleifer (2005) pioneers the idea that people derive psychological utilities from hearing consistent views to their prior beliefs. This idea serves as the basis of the structural estimations of Gentzkow and Shapiro (2010) and Martin and Yurukoglu (2017), as well as the political model of Bernhardt et al. (2008) on electoral errors, holding policies fixed.

The idea that even rational consumers–when constrained by limited information capacities–can exhibit a preference for biased news dates back to Calvert (1985a) and is expanded on by Suen (2004), Burke (2008) and Che and Mierendorff (2018). These Bayesian models predict (but may not emphasize) occasional big surprise but do not examine its consequence for policy polarization. DellaVigna and Gentzkow (2010) surveys the evidence for Bayesian voters.

Rational inattention The literature on rational inattention (RI) pioneered by Sims (1998; 2003) equips individuals with costly communication channels that aggregate source data optimally into signals. To create a role for news aggregators, we assume that the communication channel is designed by an attention-maximizing infomediary rather than voters themselves. Following Caplin and Dean (2015), we work with posterior-separable attention costs that nest mutual information as a special case. Posterior-separability has recently received attention from economists, mainly because of its axiomatic foundations (Zhong (2017); Denti (2018); Caplin et al. (2019); Tsakas (2019)), connections between sequential sampling (Hébert and Woodford (2017); Morris and Strack (2017)) and vindication by lab experiments (Dean and Nelighz (2019)). For other developments in RI, see Caplin (2016) and

[8] See the above surveys for the literature on supply-driven media bias.

Maćkowiak et al. (2018) for literature surveys.

Matějka and Tabellini (2016) examines an electoral competition model in which rational inattentive voters can reduce the variances of normally distributed signals about candidate policies. In that model, policy divergence arises when the cost of variance reduction differs across candidates, who in turn target voters paying different levels of attention and exerting different influences on electoral outcomes.

Media as a flexible and profit-maximizing information channel Our infomediary can aggregate information flexibly and does so to maximize profit.[9] Recent political models sharing these features include Strömberg (2004), Chan and Suen (2008) and Perego and Yuksel (2018).

In Strömberg (2004), the probability that voters discover a government program increases with the latter's press coverage, and newspapers decide how to allocate limited spaces across multiple programs based on readers' revenue potentials. The main research question concerns how newspaper reporting affects government budget allocation rather than platform convergence or divergence.

In Chan and Suen (2008), voters care about whether the realization of a state variable is above or below their personal thresholds, and media outlets partition state realizations using threshold rules. The restriction to threshold rules implies that signal realizations are monotone in voters' types, so the median voter is always disciplining as in our broadcast case despite a plurality of media.

In Perego and Yuksel (2018), media outlets can allocate a fixed information processing capacity between the valence and ideological aspects of an uncertain state variable in a non-RI manner. That model abstracts away from electoral competition and examines how media entry affects opinion disagreement among news consumers.

Strict obedience Strict obedience is a consequence of rational and flexible news aggregation and generates policy divergence even between office-motivated candidates. In the existing literature on electoral competition, voter signals are often drawn from (exogenous) continuous distributions, so even small changes in policy positions can affect candidates' winning probabilities.[10] Under this alternative assumption (and

---

[9] The media in Duggan and Martinelli (2011) and Prat (2018) aggregates information flexibly in order to persuade voters with limited mental capacities.

[10] Likewise, there is no guarantee that the outcome of stylized information acquisition, e.g., pay a fixed cost to draw a signal from an exogenous distribution, would satisfy strict obedience. We do

others), Calvert (1985b), Duggan (2000) and Patty (2005) among others obtain policy convergence between office-motivated candidates.

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 omitted proofs can be found in Appendices A-C.

2 Baseline Model

2.1 Setup

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

Voting Voting is *expressive*. Under any given policy profile $\mathbf{a} = \langle a_L, a_R \rangle$, type k voters earn the following utility difference from choosing candidate R over L:

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

In the above expression, the term

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

captures type k voters' differential valuation of the policies, whereas ω is an uncertain *valence state* that determines the candidates' relative fitness for office.[11] In the

not attempt to review the existing electoral competition models in which signal structures are either exogenous or emerge from stylized information acquisitions. For a partial list of references, see the survey of Duggan (2017), as well as the literature sections of Matějka and Tabellini (2016) and Hu and Li (2018).

[11]E.g., in the ongoing debate about the most effective measure to fight terrorism, $\omega = -1$ if the state favors the use of "soft power" such as diplomatic tactics, and $\omega = 1$ if it favors the use of "hard

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 $\Pi(\cdot \mid \omega)$, $\omega \in \Omega$, specifies a probability distribution over a finite set $\mathcal{Z}$ of signal realizations conditional on the state being ω.

News signals are provided by a *monopolistic* infomediary who is equipped with a *personalization technology* $\mathcal{S}$. $\mathcal{S}$ is a partition of voters' type space $\mathcal{K}$, 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 segment. Focus will be given to the coarsest and finest personalization technologies termed the *broadcast news* $b = \{\mathcal{K}\}$ and *personalized news* $p = \{\{k\} : k \in \mathcal{K}\}$, respectively. The infomediary cannot distinguish between voters at all in the broadcast case but can do so perfectly in the personalized case.

Under personalization technology $\mathcal{S} \in \{b, p\}$, the infomediary *commits* to $|\mathcal{S}|$ news signals, one for each market segment. Within each market segment, voters decide whether to *consume* the news signal that is offered to them. News consumption requires that voters *absorb* the information contained in the news signal by paying an *attention cost* $\lambda \cdot I(\Pi)$. $\lambda > 0$ is a scaling parameter termed the *marginal attention cost*, and $I(\Pi)$ is the needed *attention level* for absorbing news signal Π. Post news consumption, voters observe the signal realization, update beliefs about the state and vote expressively. Ex ante, they prefer to consume news rather than to abstain if and only if the expected utility gain from improved expressive voting exceeds the attention cost.

By paying attention, voters generate revenues to the infomediary, whose net profit equals the total amount of attention paid by voters, minus a fixed operating cost.

Game The game sequence is as follows:

1. (a) the infomediary commits to news signal structures;

   (b) voters decide whether to consume news;

   (c) candidates propose policies;

power" such as military preemption. Candidate L and R is more experienced with the use of soft and hard power, respectively, and if experience matches the true state, has an edge on his opponent.

2. the state is realized;
3. voters observe policies and signal realizations and vote; winner is determined by simple majority rule with even tie-breaking.

The solution concept is *perfect Bayesian equilibrium*. The goal is to characterize all perfect Bayesian equilibria in which candidates propose a *symmetric policy profile* of form $\langle -a, a \rangle$, $a \geq 0$ in stage 1(c) of the game.

2.2 Model Discussions

News signal We take the source data (e.g., original reporting, videos, archived documents) as given and assume that they fully describe the valence state. A signal structure codes the way that source data are assembled and packaged into news. Under signal structure Π,

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

is the probability that the signal realization is $z \in \mathcal{Z}$. In what follows, assume without loss of generality that $\pi_z > 0$ for all $z \in \mathcal{Z}$, and let

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

denote the posterior mean of the state conditional on the signal realization being $z \in \mathcal{Z}$. Bayes' plausibility mandates that the expected posterior mean must equal the prior mean:

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

We allow the infomediary to design any signal structure. After setting up the infrastructure, e.g., the algorithm, the procedure of news aggregation becomes automatic, thus justifying the assumption of commitment power.

Attention cost The marginal attention cost $\lambda > 0$ captures factors that affect the (opportunity) cost of paying attention[12] and is assumed to be constant across voters for convenience (see, however, Section 7 for an exception).

[12] Examples include distractions coming from the internet and mobile devices (Prior (2005); Dunaway (2016)); increasing competition between firms for consumer eyeballs (Teixeira (2014)).

The needed attention level for news consumption satisfies the following properties:

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

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

where the function $h : [-1, 1] \to \mathbb{R}_+$ *satisfies: (i)* $h(0) = 0$ *and strict convexity; (ii) continuity on* $[-1, 1]$ *and twice differentiability on* $(-1, 1)$*; and (iii) symmetry around zero.*

Parts (i) of Assumption 1 is equivalent to weak posterior-separability (Caplin et al. (2019)), implying, in particular, that the needed attention level for absorbing no information equals zero and increases as the posterior belief becomes more distant from the prior. Parts (ii) and (iii) of Assumption 1 imposes regularities on our problem, with Part (iii) saying that only the magnitude of posterior belief matters whereas the direction of it does not.

Assumption 1 relates the needed attention level for consuming news to the reduction in the uncertainty associated with the valence state before and after taking conditional expectations. In the case where $h(\mu) = \mu^2$ (Gentzkow and Kamenica (2014)), $I(\Pi)$ equals the reduction in variance. In the case where $h(\mu) = 1 - H((1 + \mu)/2)$ and H is the binary entropy function, $I(\Pi)$ equals the reduction in entropy, also termed the mutual information of the valence state and news signal (Maćkowiak et al. (2018) surveys recent works making this functional form assumption). Part but not all of the upcoming analysis will make functional form assumptions about h.

Attention-based business model and personalization Modern news aggregators live on consumer eyeballs, which generate advertisement revenues.[13] The main expenditure, apart from the operating cost, is the payment to sources. The existing business models range from no payment, fixed payment to revenue sharing.[14] To encompass these variations, we assume that the infomediary's gross profit equals the total amount of attention paid by news consumers. In fact, we could apply any

[13]An exception is Google News, which displays zero ads and instead funnels readers over to the main Google search engine that displays produce ads.

[14]E.g., Google News displays only snippets and charges/pays nothing to sources. Yahoo! News pays a fixed fee to its newspaper consortium for displaying source contents on its platform. Facebook News Feeds display ads in articles and snippets and share revenues between sources.

strictly increasing transformation to the gross profit function, and none of our results will change qualitatively.

The above attention-based business model is also adopted by some conventional media outlets. An example is commercial TV, whose main revenue source is again advertisement and whose programs are curated to attract viewers with highest revenue potentials (Hamilton (2004)). One of the major differences between modern news aggregators and conventional media is the ability to personalize news. We capture this difference by the fineness of the personalization technology and examine how *news personalization*–modeled as a transition from broadcast news to personalized news–affects equilibrium outcomes. Focusing on the coarsest and finest personalization technologies isn't restrictive: for general partitional technologies, we can construct a representative for each market segment as in the broadcast case and transform the problem to that of designing personalized signals for representative voters.

Game sequence and candidate information We consider equilibria in which candidates propose symmetric policy profiles in stage 1(c) of the game. The current game sequence is specified such that we only have to characterize the best responses of the infomediary and voters to symmetric policy profiles in stages 1(a) and 1(b) of the game (when designing and consuming signals), respectively, and symmetry greatly simplifies the analysis. The assumption that candidates are uninformed about the state is less restrictive as it might seem: so long as the infomediary constitutes the sole news provider, the current equilibria will remain equilibria even if candidates can observe arbitrary signals of the state when proposing policies.

3 Optimal News Signals

In this section, we fix any symmetric policy profile $\mathbf{a} = \langle -a, a \rangle$, $a \geq 0$ and solve for the optimal news signals that maximize the infomediary's profit in stage 1(a) of the game. We formalize the infomediary's problem in Section 3.1 and characterize its solutions in Section 3.2. To facilitate analysis, we say that candidate L and left-wing voters form the *left-wing party*, and that 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 The Infomediary's Problem

Under personalization technology $\mathcal{S}$, any optimal news signal of market segment $s \in \mathcal{S}$ solves the following problem:

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

where $\mathcal{D}\left(\Pi; \mathbf{a}, s\right)$ denotes the demand for news signal Π in market segment s under policy profile $\mathbf{a}$. To figure out $\mathcal{D}(\cdot)$, notice that absent news consumption, extreme voters would always vote along party lines. For these voters, new consumption is useful if it convinces them to vote across party lines when their own-party candidates are less fit for office. Post news consumption, voters strictly prefer candidate R to L (resp. candidate L to R) if and only if $v\left(\mathbf{a}, k\right) + \mu_z > 0$ (resp. $v\left(\mathbf{a}, k\right) + \mu_z < 0$). Ex ante, the expected utility gain from consuming Π is

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

where

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

Therefore,

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

where the term $q\left(k, s\right)$ in the above expression denotes the population of type k voters in market segment s.

If a solution to Problem (s) has zero demand, then it is outcome equivalent to and will be replaced by degenerate signals. This rules out uninteresting situations where the infomediary deters consumption using nondegenerate signals.

3.2 Main Features

3.2.1 Binary Recommendation and Strict Obedience

This section demonstrates that optimal news signals are either degenerate or binary and, in the second case, gives voting recommendations that news consumers strictly prefer to obey.

Formally, we say that a news signal realization z *endorses* candidate R (resp. candidate L) and *disapproves* candidate L (resp. candidate R) if $\mu_z > 0$ (resp. $\mu_z < 0$). For degenerate signals, we write $\mathcal{Z} = \{N\}$. For binary signals, we write $\mathcal{Z} = \{L, R\}$ and assume without loss of generality that $\mu_L < 0 < \mu_R$.[15] In this way, we can interpret each $z \in \mathcal{Z}$ as an endorsement for candidate z and a disapproval of candidate $-z$. In addition, we say that a binary news signal induces *strict obedience* if its consumers strictly prefer the endorsed candidate to the disapproved candidate under all signal realizations:

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

The next lemma states the main result of this section:

Lemma 1. *Fix any symmetric policy profile $\langle -a, a \rangle$, $a \geq 0$ and assume Assumption 1. Then,*

(i) any optimal broadcast signal is either degenerate or binary;

(ii) any optimal personalized signal of any type of voters is either degenerate or binary;

(iii) any optimal news signal, if binary, induces strict obedience.

Proof. See Appendix C.1. □

The proof of Lemma 1 distinguishes between broadcast and personalized news. In the personalized case, any optimal signal has at most two realizations because individual voters make binary voting decisions. Given this, any information beyond binary 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 maximizing the usefulness of news consumption at the maximal attention level.

The proof of the broadcast case is more subtle and involves aggregating voters with binding participation constraints into a representative voter. Since the resulting

[15]Since the state is binary, it is without loss to identify binary signals with posterior means of the state., i.e., $\Pi(z = R \mid \omega = 1) = \frac{-\mu_L(1+\mu_R)}{\mu_R - \mu_L}$ and $\Pi(z = R \mid \omega = -1) = \frac{-\mu_L(1-\mu_R)}{\mu_R - \mu_L}$.

information design problem features binary states and a posterior-separable objective function, applying the concavification method developed by Aumann and Maschler (1995) and Kamenica and Gentzkow (2011) yields at most two signal realizations. In Section 6.2, we relax the assumption of binary states and examine its consequence.

Strict obedience is a consequence of rational and flexible information aggregation. If, instead of (SOB), a consumer of a binary signal has a weakly preferred candidate that is independent of the signal realization, then he would abstain from news consumption to save the attention cost, which leads to a contradiction.

The next lemma gives sufficient conditions that ensure the uniqueness of optimal news signals:

Lemma 2. *Fix any symmetric policy profile* $\langle -a, a \rangle$, $a \geq 0$ *and assume Assumption 1. Then,*

(i) in the broadcast case, if it is optimal to induce consumption from all voters, then the optimal news signal is unique;

(ii) the optimal personalized signal of any type of voters is unique.

Proof. See Appendix C.1. □

The next assumption ensures regularity and will be maintained till Section 6.2.[16] Together with Lemmas 1 and 2, it implies that under all feasible symmetric policy profiles, all optimal news signals are unique and binary, and the posterior means of the state conditional on their realizations take interior values in $(-1, 1)$:

Assumption 2. *Under any symmetric policy profile* $\langle -a, a \rangle$, $a \geq 0$,

(i) any optimal news signal is nondegenerate, and the posterior means of the state conditional on its realizations belong to $(-1, 1)$*;*

(ii) it is optimal to induce consumption from all voters in the broadcast case.

[16] Loosely speaking, Assumption 2 holds when the marginal attention cost is moderate and voters' policy preferences are sufficiently homogeneous. When attention cost is quadratic, i.e., $h(\mu) = \mu^2$, this assumption is equivalent to $\lambda > 1/2$ and $\lambda t(1) < \min\left\{-1 + 3\sqrt{2}/4, \left(\sqrt{3} - 1\right)/12\right\}$. When attention cost is Shannon entropy-based, the numerical solutions presented in Appendix B confirm this intuition.

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 $\mu_z^{\mathcal{S}}(a,k)$ to denote the posterior mean of the state conditional on the signal realization being $z \in \{L, R\}$. The probability that $\Pi^{\mathcal{S}}(a,k)$ endorses candidate R is

$$\pi^{\mathcal{S}}(a,k) = -\frac{\mu_L^{\mathcal{S}}(a,k)}{\mu_R^{\mathcal{S}}(a,k) - \mu_L^{\mathcal{S}}(a,k)}.$$

In the case of $\mathcal{S} = b$, we will suppress the notation of k and write $\Pi^b(a)$, $\mu_z^b(a)$ and $\pi^b(a)$ instead.

3.2.2 Skewness

This section examines the skewness of optimal news signals. The result is threefold. First, the optimal broadcast signal is symmetric and endorses each candidate with .5 percent probability. Second, the optimal personalized signals of extreme voters are skewed and exhibit *own-party bias* and *occasional big surprise*. Third, the optimal broadcast signal attracts less attention per capita than optimal personalized signals—a property we will refer to as the trade-off between *coverage and accuracy*:

Theorem 1. *Fix any symmetric policy profile $\langle -a, a\rangle$, $a \geq 0$ and assume Assumptions 1 and 2. Then,*

(i) $\pi^b(a) = 1/2$ *and* $\mu_L^b(a) + \mu_R^b(a) = 0$*;*

(ii) $\mu_L^p(a,-k) + \mu_R^p(a,k) = 0$ *for all* $k \in \mathcal{K}$*, and*

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

  (b) $\pi^p(a,k) = 1/2$ *and* $\mu_L^p(a,k) + \mu_R^p(a,k) = 0$ *if* $k = 0$*;*

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

(iii) $I(\Pi^p(a,k)) > I\left(\Pi^b(a)\right)$ *for all* $k \in \mathcal{K}$*.*

Proof. See Appendix C.1. □

Theorem 1(i) follows from the assumptions that voters' policy preferences exhibit increasing differences[17] and the marginal attention cost is constant across voters. Under these assumptions, only extreme voters' participation constraints can be binding,

[17] That is, voters more prefer candidate R to L as they become more pro-right, i.e., $v(a_L, a_R, k)$ is increasing in k for all $a_R \geq a_L$.

in which case the representative voter constitutes their Lagrange multiplier-weighted average. To maximize the usefulness of news consumption to the representative voter, the infomediary offers a symmetric signal that endorses each candidate with .5 probability.

To develop intuition for Theorem 1(ii), recall that a signal realization is useful for an extreme voter only if it convinces him to vote across party lines. Since such signal realization moves the posterior mean of the state far away from the prior, it is costly to absorb, can only occur with a small probability and is therefore termed occasional big surprise. By Bayes' plausibility (BP), the news signal must endorse the voter's own-party candidate most of the time, albeit moderately. In this sense, it exhibits an own-party bias, which must go hand in hand with the occasional big surprise.

We finally turn to Theorem 1(iii). From Theorem 1(i), we know that the optimal broadcast signal caters to a representative voter demanding balanced views about candidate fitness. But the decision to consume news is made by extreme voters, who prefer skewed signals as shown in Theorem 1(ii). The conflict of interest underpins the tradeoff between coverage and accuracy.

4 Equilibrium Policies

This section endogenizes candidates' policy positions and gives complete characterizations of equilibrium outcomes.

Under personalization technology $\mathcal{S}$, a policy profile $\langle -a, a \rangle$, $a \geq 0$ and *news profile* $\widetilde{\boldsymbol{\mu}}$ 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 $\widetilde{\boldsymbol{\mu}}$, candidate L's policy $-a$ and voters' behaviors in stages 1(b) and 3 of the game as given.

Our goal is to characterize *all* perfect Bayesian equilibria of the above form. However, the analysis so far pins down only the marginal news distribution within each market segment but leaves the joint news distribution across market segments unspecified. While the joint news distribution is irrelevant to expressive voting, holding marginal news distributions constant, it nevertheless enters the candidates' strategic reasoning. In Appendix A, we consider all joint news distributions that are consistent with the marginal news distributions as solved in Section 3. In the baseline model, we restrict attention to the case where news signals are *conditionally independent* across market segments for any given state realization. The implication of this restriction will soon become clear.

4.1 Key Concepts

This section develops the key concepts for equilibrium characterization.

We first describe how a unilateral deviation of a candidate from a symmetric policy profile in stage 1(c) of the game can affect the expressive voting decisions in stage 3 of the game. By symmetry, consider only candidate R's deviation from $\langle -a, a \rangle$ to a', whose effect on any type k voters is twofold. First, a' can *attract* type k voters, i.e., win their support even if their news signal disapproves candidate R:

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

Alternatively, a' can *repel* type k voters, i.e., lose their support even if their news signal endorses candidate R:[18]

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

If a' attracts (resp. repels) type k voters, then it makes them vote for (resp. against) candidate R unconditionally. Otherwise it has no effect on type k voters' voting decisions.

Define

$$\phi^{\mathcal{S}}(-a, a', k) = v(-a, a', k) + \mu_L^{\mathcal{S}}(a, k)$$

as type k voters' *susceptibility* to policy deviation a' when news is unfavorable to candidate R. For each $k \in \mathcal{K}$, define the *k-proof set* $\Xi^{\mathcal{S}}(k)$ by the policy a's such that no unilateral deviation of candidate R from the corresponding symmetric policy profile $\langle -a, a \rangle$ attracts type k voters. Since $t(k)$ is type k voters' bliss point, the

[18] Replacing the strict inequalities in the above conditions with weak inequalities won't change the result.

k-proof set constitutes policies that deter candidate R from deviating to $a' = t(k)$:

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

The maximum of the k-proof set

$$\xi^{\mathcal{S}}(k) = \max \Xi^{\mathcal{S}}(k),$$

is called type k voters' *policy latitude* and intuitively decreases with the latter's susceptibility to policy deviations. As we will soon demonstrate, policy latitudes are well-defined and, crucially, strictly positive.

We next describe equilibrium outcomes. Under personalization technology $\mathcal{S}$ and population function q, let $\mathcal{E}^{\mathcal{S},q}$ denote the set of policy a's such that the corresponding symmetric policy profile $\langle -a, a\rangle$ can arise in equilibrium, and define $a^{\mathcal{S},q} = \max \mathcal{E}^{\mathcal{S},q}$ as the *policy polarization*, or *polarization* for short. Type k voters are *disciplining* if their policy latitude dictates policy polarization:

Definition 1. *Under personalization technology $\mathcal{S}$ and population function q, type k voters are* disciplining *if $a^{\mathcal{S},q} = \xi^{\mathcal{S}}(k)$.*

4.2 Main Result

Our main theorem proves existence of disciplining voters and pins down their identities. It shows that policy polarization is strictly positive and the equilibrium policy set satisfies the interval property:

Theorem 2. *Assume Assumptions 1 and 2. Then under all personalization technology $\mathcal{S} \in \{b, p\}$ and population function q, $\mathcal{E}^{\mathcal{S},q} = \left[0, a^{\mathcal{S},q}\right]$ and $a^{\mathcal{S},q} > 0$. In particular,*

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

(ii) in the personalized case,

(a) median voters are disciplining if they constitute a majority of the voter population, i.e., $a^{p,q} = \xi^{p}(0)$ if $q(0) > 1/2$;

(b) otherwise voters with the smallest policy latitude are disciplining, i.e., $a^{p,q} = \min_{k \in \mathcal{K}} \xi^{p}(k)$ if $q(0) \leq 1/2$.

Proof. Theorem 3 of Appendix A proves a more general result. □

4.2.1 Proof Sketch

Broadcast news In this case, all voters receive the same voting recommendation, so a deviation by candidate R is *profitable*, i.e., strictly increases his winning probability, if and only it attracts a majority of voters. Under the assumption that voters' policy preferences exhibit increasing differences, a deviation attracts a majority of voters if and only if it attracts median voters, who are therefore disciplining under all population distributions.

Personalized news In this case, median voters remain disciplining if they constitute a majority of the voter population. Otherwise no type of voters alone constitutes a majority, and a deviation is profitable if it attracts any type k voters, holding other things constant. The reason is pivotality: since the infomediary can now offer conditionally independent signals to different voters, the above deviation strictly increases candidate R's winning probability in the event where types $-k$ voters disagree about candidate fitness.

The above argument leaves open the question of whether attracting some voters would come at the cost of repelling others. The next lemma shows that this concern is unwarranted:

Lemma 3. *Assume Assumptions 1 and 2. In the case where $\mathcal{S} = p$ and $q(0) \leq 1/2$, a symmetric policy profile $\langle -a, a\rangle$, $a \geq 0$ can be attained 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 in $[-a, a]$.*

Proof. Lemma 6 of Appendix A proves a more general result. □

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 deviations, candidate R may successfully attract voters whose bliss points lie outside $[-a, a]$. But this success comes at the cost of repelling opposite types of voters and overall does no good to candidate R's winning probability. Meanwhile, the second kind of deviations repels no voter and attracts no voter whose bliss point lies outside $[-a, a]$. If, in addition, it doesn't attract voters whose bliss points lie inside $[-a, a]$, then $\langle -a, a\rangle$ can be attained in equilibrium.

By Lemma 3, a policy in $[0, t(1))$ can be attained in equilibrium if and only if it deters candidate R from deviating to median voters' bliss point, and so can a policy in $[t(1), \overline{a}]$ if and only if it deters candidate R from deviating to any voter's bliss point. Combining yields $\mathcal{E}^{p,q} = A(0) \cup A(1)$, where $A(0) = [0, t(1)) \cap \Xi^p(0)$ and $A(1) = [t(1), \overline{a}] \cap \bigcap_{k \in \mathcal{K}} \Xi^p(k)$.

Completing the proof (SOB) implies that local deviations from any equilibrium policy profile won't change voters' mind. Therefore, there exist equilibria featuring policy divergence, and no equilibrium of our interest is strict. Further characterizations of policy latitudes establish the interval property and pin down the identities of disciplining voters.

Takeaway News personalization makes attracting non-majorities of voters profitable deviations. Voters with the smallest policy latitude are most susceptible to policy deviations and therefore constitute the disciplining entity for equilibrium polarization. Holding other things constant, deviations could be more effective and less polarization could be sustained in the personalized case than in the broadcast case.

4.2.2 Ranking Policy Latitudes

Which voters have the smallest policy latitude in the personalized case? An initial guess would be right-wing voters, candidate R's *base*, who most prefer candidate R policy-wise. But after taking into account the belief about candidate fitness, the answer to this question becomes less obvious. The next lemma makes this intuition precise:

Lemma 4. *When $\overline{a}$ is large,*

(i) if $\xi^b(0) \geq t(1)$, then $\xi^b(0) = -\mu_L^b := -\mu_L^b(t(1))$;

(ii) for all $k \in \mathcal{K}$, $\xi^p(k) = -[t(k) + \mu_L^p(k)] > |t(k)|$, where $\mu_L^p(k) := \mu_L^p(|t(k)|, k)$.

Proof. See Appendix C.4. □

Lemma 4 decomposes the negative policy latitude into two competing forces: $t(k)$ and $\mu_L^{\mathcal{S}}(k)$. While $t(k)$ is increasing in k, $\mu_L^{\mathcal{S}}(k)$ is decreasing in k. The reason is that right-wing voters seek occasional big surprises and are therefore most pessimistic when news is unfavorable to candidate R. By contrast, left-wing voters, the *opposition* of candidate R, are most optimistic. The combined effect with bliss points can be subtle.

Of particular interest is the question of when base voters have a bigger policy latitude than opposition voters. By Lemma 4, this is the case if and only if the personalized signals of extreme voters are sufficiently skewed, so that the difference in voter beliefs exceeds the difference in bliss points:

$$\mu_L^p(1) + \mu_R^p(1) < -2t(1). \qquad (*)$$

If Condition $(*)$ holds, then candidate R won't target his base when contemplating deviations *off equilibrium path*. Instead, he appeals to either median voters or opposition voters—whoever are the easiest to attract in the event where news is unfavorable. We reduce Condition $(*)$ to model primitives in Section 5.1 and Appendix B.3.

5 Comparative Statics

This section examines how equilibrium polarization depends on the personalization technology $\mathcal{S}$, the marginal attention cost λ and voters' population distribution q. a

5.1 Personalization Technology

The next proposition characterizes the policy polarization effect of news personalization:

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

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

Condition $(**)$ holds if $\xi^b(0) < t(1)$. When $\xi^b(0) \geq t(1)$,[19]

[19]The fact that equilibrium policies can be more polarized than all voters' bliss points is an artifact of the restriction to three types of voters. According to Theorem 3 of Appendix A, policy polarization is generally capped by the bliss point of most extreme voters provided that the latter's population is small enough. In that case, one can think of $t(1)$ as the average position of right-leaning voters, and there is good evidence that candidates can adopt more extreme positions than their own-party voters

(a) if right-wing voters are disciplining under personalized news, then Condition $(**)$ *is equivalent to* $\mu_L^p(1) - \mu_L^b < -t(1)$*;*

(b) if left-wing voters are disciplining under personalized news, then Condition $(**)$ *is equivalent to* $t(-1) < \mu_L^b - \mu_L^p(-1)$*.*

Proof. The proof is immediate from Theorem 2 and is therefore omitted. □

Proposition 1(i) is immediate from the trade-off between coverage and accuracy: since median voters' personalized signal is more Blackwell-informative than the broadcast signal, median voters know more about candidate valence and therefore become more resistant to policy deviations (hereafter the trade-off between *policy and valence*) as news becomes personalized;[20] if, in addition, median voters are disciplining before and after the transition, then personalization increases polarization.

Proposition 1(ii) delivers a more subtle message: in the case where extreme voters are disciplining under personalized news, the skewness of their personalized signals is crucial for sustaining greater polarization than in the broadcast case. Consider two subcases: (a) right-wing voters are disciplining, and (b) left-wing voters are disciplining.

In case (a), compare right-wing voters hearing personalized news and median voters hearing broadcast news. Since right-wing voters most prefer candidate R policy-wise, the only explanation for why they could have a bigger policy latitude than median voters is the occasional big surprise of their personalized signal. That is, in order to satisfy Condition $(*)$, right-wing voters must be significantly more pessimistic than median voters when news is unfavorable to candidate R, i.e., $\mu_L^p(1) - \mu_L^b < -t(1)$.

In case (b), the presumption is that Condition $(*)$ holds, and the explanation offered in Section 4.2.2 hinges on the skewness of personalized signals. Off equilibrium path, candidate R contemplating deviations won't target right-wing voters, his base, because doing so is either needless (when he already captures his base) or futile (when the base is convinced of his unfitness and becomes unpersuadable). Instead, he appeals to left-wing voters, his opposition, which is itself challenging if the opposition has a strong preference against his policies. In the case where $t(-1) < \mu_L^b - \mu_L^p(-1)$, news personalization increases polarization.

on average (see, e.g., the survey article Barber and McCarty (2015) and the references therein).

[20]Barber and McCarty (2015) surveys the empirical evidence on centrist voters becoming more "apathetic and indifferent" to policy changes.

We next reduce Conditions $(*)$ and $(**)$ to model primitives. In general, the comparison between broadcast and personalized news is challenging because the Lagrange multipliers of voters' participation constraints differ across these cases and lack analytical solutions. The next example fully solves the model in the case of quadratic attention cost. Appendix B.3 investigates the case of entropy attention cost, and the numerical results therein confirm and enrich the intuition below.

Example 1. In the case where $h(\mu) = \mu^2$, tedious but straightforward algebra shows that

$$\xi^b(0) = -\mu_L^b = \left(1 + \sqrt{1 - 16\lambda t(1)}\right) / (4\lambda) > t(1),$$

in the broadcast case, and that

$$\mu_L^p(k) = \begin{cases} 4t(1) - 1/(2\lambda) & \text{if } k < 0, \\ -1/(2\lambda) & \text{if } k \geq 0, \end{cases}$$

and

$$\xi^p(k) = \begin{cases} 1/(2\lambda) - 3t(1) & \text{if } k = -1, \\ 1/(2\lambda) & \text{if } k = 0, \\ 1/(2\lambda) - t(1) & \text{if } k = 1, \end{cases}$$

in the personalized case. A casual inspection reveals that left-wing voters have the smallest policy latitude, followed by right-wing voters and then median voters, so Condition $(*)$ always holds. Condition $(**)$ then becomes $t(-1) < \mu_L^b - \mu_L^p(-1)$, and simplifying yields $\lambda t(1) > 1/18$.

Consider first the role of λ in Condition $(**)$. As λ increases, paying attention becomes more costly, so the infomediary makes news signals less Blackwell-informative to prevent voters from tuning out (this is formally shown in the proof of Proposition 2). During the process, the infomediary is more constrained in the broadcast case than in the personalized case, because extreme voters' participation constraints are binding in both cases, and it is harder to attract their attention using a symmetric signal than a skewed signal. As λ increases, the difference gets amplified, so the infomediary feels more urgent to increase μ_L^b than to increase $\mu_L^p(-1)$.

In general, the above intuition is incomplete because the infomediary can adjust two posterior beliefs $\mu_L^p(-1)$ and $\mu_R^p(-1)$ differently in the personalized case. In the current example, this concern is unwarranted, since λ enters $\mu_L^p(-1)$ and $\mu_R^p(-1)$

symmetrically. Combining, we obtain that μ_L^b increases with λ faster than $\mu_L^p(-1)$ does, so increases in λ relax Condition ($**$).

Consider next the role of $t(1)$. As $t(1)$ increases, the left-hand side of Condition ($**$) decreases, suggesting that strong policy preferences alienate left-wing voters from candidate R's deviations. The right-hand side increases, suggesting, at least in the current example, that changes in voter beliefs do not undermine the above intuition.

5.2 Marginal Attention Cost

The next proposition shows that increases in the marginal attention cost make news signals less Blackwell-informative and reduces polarization by the trade-off between policy and valence:

Proposition 2. *Assume Assumption 1, and take any $\lambda' > \lambda > 0$ that satisfy Assumption 2. Then $a^{\mathcal{S},q}(\lambda') < a^{\mathcal{S},q}(\lambda)$ for all $\mathcal{S} \in \{b,p\}$ and q.*

Proof. See Appendix C.4. □

The significance of Proposition 2 lies in its policy implication. In Appendix B.2, we extend the baseline model to perfectly competitive infomediaries that maximize voter expected utility rather than attention. As it turns out, introducing perfect competition between infomediaries is equivalent to increasing the marginal attention cost and, by Proposition 2, reduces polarization compared to the case of monopolistic yet personalized news. The latter overfeeds voters with information about candidate valence and sustains greater polarization by the trade-off between policy and valence.

That said, notice that the effect of competition (as well as personalization) on voter welfare is less clear-cut and much depends on the relative positions of equilibrium policies to voter bliss points. Given this subtlety, we recommend caution when evaluating recent proposals to tame the tech giants centered on (1) banning personalization and (2) introducing competition.[21]

[21] Earlier this year, Senator Elizabeth Warren from Massachusetts called for the big tech companies to be broken up and to meet the standard of nondiscriminatory dealing with users. Meanwhile in Europe, the General Data Protection Regulation (GDPR) was created in 2016 to uphold the protection of personally identifiable information of EU citizens. Finally, the Report of the Digital Competition Expert Panel issued by the British government this March recommended more competition rather than break-ups or stifling regulations of personalization.

5.3 Population Distribution

Recently, a growing body of the literature has been devoted to understanding the potential causes and consequences of voter polarization (also termed *mass polarization*). Among these works include Fiorina and Abrams (2008), which defines mass polarization by the bimodal distribution of voter policy preferences on a liberal-conservative scale; as well as Gentzkow (2016), which promotes the uses of measures such as the average ideological distance between Democrats and Republicans.

Motivated by these works, we ask what would happen to policy polarization (also termed *elite polarization*) if we redistribute voters' population from the center to the margin. Notice that our exercise is purely conceptual, since evidence on increasing mass polarization is mixed at best (as argued forcefully by the above works). Our goal is to call reader's attention to the following message: contrary to popular belief, increase in mass polarization modeled as a mean-preserving spread of voters' policy preferences may decrease rather than increase elite polarization.

Consider the personalized case, since policy polarization is independent of voters' population distribution in the broadcast case:

Proposition 3. *Under Assumptions 1 and 2, $a^{p,q} \geq a^{p,q'}$ for all population functions q and q' such that $q(0) > q'(0)$, and the inequality is strict if and only if $q(0) > 1/2 \geq q'(0)$ and $0 \notin \arg\min_{k \in \mathcal{K}} \xi^p(k)$.*

Proof. The proof is immediate from Theorem 2 and is thus omitted. □

The idea behind Proposition 3 is simple: as we keep redistributing voters' population from the center to the margin, attracting extreme voters eventually becomes profitable deviations besides attracting median voters. If, in addition, extreme voters have smaller policy latitudes than median voters (as in Example 1), then a decrease in policy polarization ensues. While caution should be exercised when extrapolating this result to general environments, its warning message nevertheless warrants attention. Appendix A.4.2 proves a similar result assuming arbitrary finite types of voters and quadratic attention cost.

6 Extensions

This section previews the main extensions of the baseline model. The formal analyses are relegated to Appendices A and B.

6.1 General Model

In Appendix A, we extend the baseline model to arbitrary finite types of voters holding general policy preferences. Rather than assuming that news is conditionally independent across market segments, we consider all joint news distributions that are consistent with the marginal news distributions as solved in Section 3.

The analysis leverages a new concept called *influential coalition*. Loosely speaking, a set of voters is influential if attracting all its members, holding other things constant, strictly increases the deviating candidate's winning probability. Under broadcast news, influential coalitions coincide with majorities of voters. Under personalized news, non-majorities of voters can be influential, due to the imperfect correlation between the news signals consumed by different voters. The next table compiles the influential coalitions in the baseline model:

	$\mathcal{S} = b$	$\mathcal{S} = p$
$q(0) > 1/2$	majorities	majorities
$q(0) < 1/2$	majorities	$2^{\mathcal{K}} - \emptyset$

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

As it turns out, influential coalitions depend on the joint news distribution only through the *news configuration*—a matrix that compiles all news profiles that are realized with positive probabilities. Under certain regularity conditions, the set of policies that can be attained in equilibrium under personalization technology $\mathcal{S}$, news configuration $\boldsymbol{\chi}$ and population function q is

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

In the above expression, $\mathcal{C}$'s denote the influential coalitions formed under the pair $\langle \boldsymbol{\chi}, q \rangle$, whereas $\xi^{\mathcal{S}}\left(\mathcal{C}\right)$ is the policy latitude of $\mathcal{C}$ and depends on the joint news distribution only through marginal news distributions. Thus in general, the influential coalition with the smallest policy latitude constitutes the disciplining entity, and factors that enrich influential coalitions through the news configuration or the population function reduces polarization, holding marginal news distributions constant.

Two implications are immediate. First, news personalization changes marginal news distributions and enriches influential coalitions. Since the second channel reduces polarization, the first channel is essential to the increase in polarization as shown in Proposition 1. Second, in the personalized case, influential coalitions become richer as news become more independent across voters and as voters' population distribution becomes more uniform across types. Therefore, relaxing the assumption of conditional independence can only increase polarization, and the term constitutes $\min_{k\in\mathcal{K}} \xi^p(k)$ (as in Proposition 1 ii (b)) constitutes the exact lower bound for the polarization that can be attained across all scenarios.[22]

6.2 Beyond the Binary Case

This section relaxes two assumptions: (1) the state is binary, (2) optimal news signals are nondegenerate (thus binary and induce strict obedience) under all feasible policy profiles.

General state distribution We first extend the analysis to general state distributions. The material in Appendix B.1 assumes arbitrary finite types of voters.

In the personalized case, any optimal news signal remains either degenerate or binary. As in Matějka and McKay (2015), our voters face binary decision problems and can acquire any signal of the state at a posterior-separable cost. The only difference is that in our setting, information acquisition is subsidized by the infomediary to maximize attention rather than voter expected utility. In the case where voter's participation constraint is binding, maximizing attention is equivalent to minimizing the attention cost, given the usefulness of news consumption. As shown in Matějka and McKay (2015), any solution to this problem provides no extraneous information beyond a binary voting recommendation.

The broadcast case is more complicated. As in Section 3.2.2, we aggregate individual voters facing binary decision problems into a representative voter, counting

[22]So long as this term doesn't vanish, changes in the environment–e.g., enriching voters' type space, artificially dividing the same type of voters into multiple subgroups–will not render polarization trivial.

only those with binding participation constraints. In the case where it is optimal to include all voters in the market (as in the baseline model), only the participation constraints of most extreme voters can be binding, and aggregating their decision problems yields three decision variables: LL, LR and RR.[23] By Matějka and McKay (2015), any optimal broadcast signal has at most three realizations.

In the case of two signal realizations, results so far remain valid, at least qualitatively.[24] In the case of three signal realizations, news personalization unambiguously increases polarization. From symmetry, we know that when the signal realization is LR, the posterior mean of the state equals zero and makes median voters indifferent between the candidates. So by deviating to $a' = 0$, candidate R can attract median voters without repelling any voter, implying that the only symmetric policy profile that can be attained in equilibrium is $\langle 0, 0 \rangle$.

Relax Assumption 2 Assumption 2 requires that all optimal signals be nondegenerate (thus binary and induce strict obedience) under all feasible policy profiles. The examples below disprove its necessity for sustaining positive polarization in equilibrium.

Example 2. Let everything be as in the baseline model except that extreme voters abstain form news consumption. In Appendix C.4, we demonstrate that all policy profiles $\langle -a, a \rangle$, $a \in \left[0, \underline{\xi}\right]$, can be attained in equilibrium, where $\underline{\xi} = \min\left\{t\left(1\right), \xi^{\mathcal{S}}\left(0\right)\right\}$.[25]

Example 3. In the personalized case, suppose, instead, that extreme voters observe three signal realizations L, M and R and are indifferent between the candidates given M.[26] In Appendix C.4, we demonstrate that the policy profile $\langle -t\left(1\right), t\left(1\right) \rangle$ can be attained in equilibrium if median voters' policy latitude exceeds $t\left(1\right)$.

6.3 Skewness Effect vs. Level Effect

Two things happen to marginal news distributions as news becomes personalized. First, the signal becomes skewed, holding the attention level constant. Second, the

[23] LL and RR mean that both extreme voters choose the same candidate, whereas LR means that the left-leaning voter chooses candidate L and the right-leaning one chooses candidate R.

[24] The only caveat is that the comparative statics with respect to λ becomes less regular.

[25] Thus mistaken exclusions of some voters from news consumption due to, e.g., model misspecification, do not necessarily render polarization trivial.

[26] This could happen if the state has more than two realizations and voters have other goals besides expressive voting, e.g., form coarse opinions about the state.

attention level increases, causing the signal to become more Blackwell-informative. The corresponding effects on polarization are termed the *skewness effect* and *level effect*, respectively. The level effect increases polarization because of the trade-off between policy and valence. Below we demonstrate that the skewness effect per se can increase polarization:

Example 1 (Continued)**.** Fix the attention level to $I\left(\Pi^b\left(\xi^b(0)\right)\right)$ and solve for the optimal personalized signals.[27] Denote the policy latitude of left-wing voters–which is the smallest among all voters–by $-\left[t(-1)+\widehat{\mu}_L(-1)\right]$, and decompose Condition $(**)$ as follows:

$$t(-1) < \underbrace{\mu_L^b - \widehat{\mu}_L(-1)}_{\text{skewness}} + \underbrace{\widehat{\mu}_L(-1) - \mu_L^p(-1)}_{\text{level}}.$$

Absent the level effect–which is positive–the skewness effect per se increases polarization if and only if

$$t(-1) < \mu_L^b - \widehat{\mu}_L(-1).$$

As depicted in Figure 1, the above condition holds when λ and $t(1)$ are sufficiently large.

7 Concluding Remarks

An unintended consequence of assuming a constant marginal attention cost across voters is that median voters pay more attention than extreme voters in the personalized case. Relaxing this assumption better reconciles our findings with the stylized fact that extreme voters could be more attentive to politics than centrist voters (see, e.g., Barber and McCarty (2015) and the references therein). The only qualitative difference this change might (but not necessarily) cause is to make median voters' participation constraint binding in the broadcast case. If so, then the optimal broadcast signal would coincide with median voters' optimal personalized signal, and new personalization would weakly reduce polarization.

Tech-enabled personalization is ubiquitous and seems to maximize social surplus by best catering to consumers' needs. The caveat, in our opinion, is that the tech

[27] Alternatively, one can first increase the attention level and then vary the skewness. The insight gained through this exercise is similar to the one above.

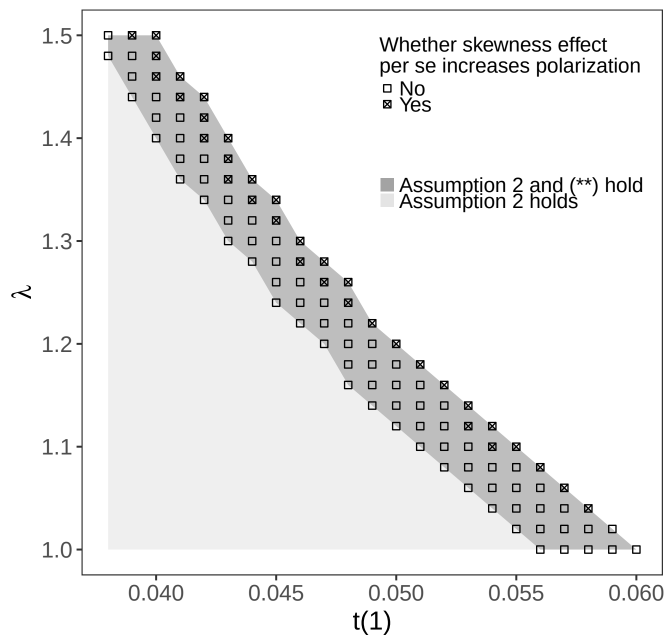

Figure 1: Skewness effect: $u(a,k) = -|t(k) - a|$ and $h(\mu) = \mu^2$.

giants now constitute major providers of political news and could therefore affect political decisions and outcomes. The current theory formalizs a particular channel through which news personalization for rational inattentive voters could affect policy polarization. The comparative statics results shed light on plausible ways of quantifying the highlighted channel, as well as the effectiveness of recent policy proposals to tame the tech giants. The significance of the highlighted channel, as well as its relevance in other settings–e.g., how personalized product recommendation affects product differentiation and consumer welfare–await future investigations.

A General Model

In this appendix, we extend the baseline model to arbitrary finite types of voters holding general policy preferences, and relax the assumption that news signals are conditionally independent across market segments. To that end, we assume that $\mathcal{K} = \{-K, \cdots, 0, \cdots, K\}$ where K can be any positive integer, that voters' population function $q : \mathcal{K} \to \mathbb{R}_{++}$ is symmetric around zero, and that their utility function satisfies the following properties:

Assumption 3. *The function $u : \mathcal{A} \times \mathcal{K} \to \mathbb{R}$ satisfies the following properties:*

concavity $u(\cdot, k)$ *is continuous and concave for all* $k \in \mathcal{K}$*;*

symmetry $u(a, k) = u(-a, -k)$ *for all* $a \in \mathcal{A}$ *and* $k \in \mathcal{K}$*;*

inverted V-shape *there exists* $t : \mathcal{K} \to int(\mathcal{A})$ *that is strictly increasing and satisfies* $t(k) = -t(-k)$ *such that* $u(\cdot, k)$ *is strictly increasing on* $[-\overline{a}, t(k)]$ *and is strictly decreasing on* $[t(k), \overline{a}]$ *for all* $k \in \mathcal{K}$*;*

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

Assumption 3 says that voters' utility function is concave in policies, symmetric across types, inverted V-shaped in policies and exhibits increasing differences in its arguments. It is satisfied by standard utility functions in the election literature, e.g., $-|t(k) - a|$, $-(t(k) - a)^2$, etc..

In the remainder of this appendix, we develop new concepts in Appendix A.1 and conduct equilibrium analyses in Appendices A.2-A.4.

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 $\boldsymbol{\chi}$ compiles the profiles of voting recommendations to types $-K, \cdots, K$ voters that occur with positive probabilities. Each column of $\boldsymbol{\chi}$ constitutes a voting recommendation profile and is therefore a $|\mathcal{K}|$-vector. Each entry of $\boldsymbol{\chi}$ is either 0 or 1, where 0 means that candidate R is disapproved and 1 means 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 $\mathbf{b}^+$ and $\mathbf{b}^-$ are strictly positive and add up to one.

We restrict attention to *symmetric* joint news distributions. Formally, let $\mathbf{x}$ be any 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 by

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

so that $\mathbf{x}$ and $\Sigma \circ \mathbf{x}$ expose opposite types of voters to symmetric situations, i.e., $\mathbf{x}$ 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. Symmetry requires that $\Sigma \circ \mathbf{x}$ be a recommendation profile, too, and that the probability $\mathbf{x}$ occurs in state $\omega = 1$ equals that of $\Sigma \circ \mathbf{x}$ in state $\omega = -1$.

With a slight abuse of notation, let $[\cdot]_m$ denote both the m^{th} entry of a column vector and the m^{th} column of a matrix. Then,

Definition 2. *A news configuration $\boldsymbol{\chi}$ is* symmetric *if for all m, there exists n 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 all m, n such that $\Sigma \circ [\boldsymbol{\chi}]_m = [\boldsymbol{\chi}]_n$.*

Consistency We consider all symmetric joint news distributions that are consistent with the marginal news distributions as solved in Section 3. In Footnote 15, we solved for the probabilities that any binary news signal endorses candidate R in states $\omega = 1$ and $\omega = -1$, respectively. For any personalization technology $\mathcal{S} \in \{b, p\}$ and policy profile $\langle -a, a \rangle$, $a \geq 0$, we compile these probabilities across types $-K, \cdots, K$ voters into two $|\mathcal{K}|$-vectors $\boldsymbol{\pi}^{\mathcal{S},+}(a)$ and $\boldsymbol{\pi}^{\mathcal{S},-}(a)$, respectively. Then,

Definition 3. *Under personalization technology $\mathcal{S} \in \{b, p\}$, a joint news distribution $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^- \rangle$ is $\langle \mathcal{S}, a \rangle$-*consistent *for some $a \geq 0$ if*

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

*A news configuration $\boldsymbol{\chi}$ is $\langle \mathcal{S}, a \rangle$-*consistent *for some $a \geq 0$ if there exist probability vectors $\mathbf{b}^+$ and $\mathbf{b}^-$ such that $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^- \rangle$ is $\langle \mathcal{S}, a \rangle$-consistent. It is $\mathcal{S}$-*consistent *if it is $\langle \mathcal{S}, a \rangle$-consistent for all $a \geq 0$.*

Notice two things. First, $\boldsymbol{\chi}^*$ is the only $\langle b, a \rangle$-consistent news configuration for any $a \geq 0$, and it is b-consistent. $\boldsymbol{\chi}^{**}$ is p-consistent but is in general not uniquely p-consistent. Second, $\mathcal{S}$-consistency is a stronger notion than $\langle \mathcal{S}, a \rangle$-consistency, and both notions will be covered in the upcoming analysis.

Influential coalition The upcoming analysis builds on the concept of influential coalition:

Definition 4. *Fix any personalization technology $\mathcal{S} \in \{b, p\}$, population function q and policy $a \geq 0$, 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. A set of voters constitutes an R-*influential coalition*, or* influential coalition *for short, if attracting all its members, holding other things constant, strictly increases candidate R's winning probability compared to the default.*

By definition, majorities of voters are influential, and supersets of influential coalitions are influential. Under broadcast news, signals are perfectly correlated among voters, so influential coalitions and majorities of voters coincide. Under personalized news, non-majorities of voters can be influential—see Table 1 for an illustration.

As it turns out, 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 (as noted in Table 1):

Lemma 5. *Let everything be as in Definition 4. Then influential coalitions depend on* $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^- \rangle$ *and* q *only through the pair* $\langle \boldsymbol{\chi}, q \rangle$*, and they are independent of* a *if* $\boldsymbol{\chi}$ *is* $\mathcal{S}$*-consistent.*

Proof. See Appendix C.2. □

A.2 Main Lemma

The next lemma gives a full characterization of equilibrium policies:

Lemma 6. *Fix any 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 $\langle -a, a \rangle$*,* $a \geq 0$ *can arise in an equilibrium where the joint news distribution* $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^- \rangle$ *is* $\langle \mathcal{S}, a \rangle$*-consistent;*

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

Proof. See Appendix C.2. □

Let $\mathcal{E}^{\mathcal{S}, \boldsymbol{\chi}, q}$ denote the set of the policy a's that such that the corresponding 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 6 prescribes the exact algorithm for computing this set:

Step 1. Compute the influential coalitions formed under the pair $\langle \boldsymbol{\chi}, q \rangle$. For every $a \geq 0$, 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 ideological 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 every element a of $\widetilde{\mathcal{E}}^{\mathcal{S}, \boldsymbol{\chi}, q}$, check if $\boldsymbol{\chi}$ is $\langle \mathcal{S}, a \rangle$-consistent. If the answer is negative, then remove a from $\widetilde{\mathcal{E}}^{\mathcal{S}, \boldsymbol{\chi}, q}$.

When the above procedure terminates, the output is $\mathcal{E}^{\mathcal{S}, \boldsymbol{\chi}, q}$. If $\boldsymbol{\chi}$ is $\mathcal{S}$-consistent, then no element of $\widetilde{\mathcal{E}}^{\mathcal{S}, \boldsymbol{\chi}, q}$ is removed in Step 2, so $\widetilde{\mathcal{E}}^{\mathcal{S}, \boldsymbol{\chi}, q} = \mathcal{E}^{\mathcal{S}, \boldsymbol{\chi}, q}$.

A.3 Main Theorem

This appendix gives a full characterization of the set $\mathcal{E}^{\mathcal{S},\boldsymbol{\chi},q}$ for $\mathcal{S}$-consistent news configuration $\boldsymbol{\chi}$'s.

We first modify the existing concepts. Under personalization technology $\mathcal{S}$, define

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

as the *susceptibility* of a set $\mathcal{D} \subseteq \mathcal{K}$ of voters to policy deviation a' when news is unfavorable to candidate R, and define the *$\mathcal{D}$-proof set* $\Xi^{\mathcal{S}}(\mathcal{D})$ by the policy a's such that no unilateral deviation from the symmetric policy profile $\langle -a, a \rangle$ by candidate R attracts all members of $\mathcal{D}$, or attracts $\mathcal{D}$ for short. By Assumption 3 **inverted V-shape**,

$$\Xi^{\mathcal{S}}(\mathcal{D}) = \left\{ a \geq 0 : \max_{a' \in [\min t(\mathcal{D}), \max t(\mathcal{D})]} \phi^{\mathcal{S}}(-a, a', \mathcal{D}) \leq 0 \right\},$$

where $t(\mathcal{D})$ denotes the image of the set $\mathcal{D}$ under the mapping t. The maximum of the $\mathcal{D}$-proof set

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

is called $\mathcal{D}$'s *policy latitude*, and $\mathcal{D}$ is *disciplining* if its policy latitude dictates the equilibrium policy polarization:

Definition 5. *Under personalization technology $\mathcal{S}$, $\mathcal{S}$-consistent news configuration $\boldsymbol{\chi}$ and population function q, a set $\mathcal{D}$ of voters is disciplining if* $\max \mathcal{E}^{\mathcal{S},\boldsymbol{\chi},q} = \xi^{\mathcal{S}}(\mathcal{D})$.

We next state the main assumptions. In addition to Assumptions 1-3, we require that the susceptibility function $\phi^{\mathcal{S}}(-a, a', k)$ be increasing in a in a local region:

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

$\phi^{\mathcal{S}}(-a, a', k)$ is the sum of $v(-a, a', k)$ and $\mu_L^{\mathcal{S}}(a, k)$. Since $v(-a, a', k)$ is increasing in a on $[|t(k)|, \overline{a}]$ by Assumption 3 **inverted V-shape**, $\phi^{\mathcal{S}}(-a, a', k)$ is, too, if $\mu_L^{\mathcal{S}}(a, k)$ doesn't vary much with a. Lemma 9 of Appendix C.2.1 shows that this is the case if $\mathcal{S} = b$ or if $\mathcal{S} = p$ and either $u(a, k) = -|t(k) - a|$ or $h(\mu) = \mu^2$.

We finally state the main theorem, whose message is twofold. First, in general, the influential coalition with the smallest policy latitude constitutes the disciplining entity. Second, marginal news distributions affect policy polarization through policy latitudes, whereas the joint news distribution does so through the news configuration and influential coalitions:

Theorem 3. *Assume Assumptions 1-4. Then for all personalization technology $\mathcal{S} \in \{b, p\}$, $\mathcal{S}$-consistent news configuration $\boldsymbol{\chi}$ and population function q,*

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

where $\mathcal{C}$'s denote the influential coalitions formed under the pair $\langle \boldsymbol{\chi}, q \rangle$.

Proof. See Appendix C.2. □

An immediate consequence of Theorem 3 is that median voters are disciplining under broadcast news:

Corollary 1. *Under Assumption 1-3, $\mathcal{E}^{b,\boldsymbol{\chi}^*,q} = \left[0, \xi^b\left(0\right)\right]$ for all q.*

A.4 Comparative Statics

This appendix examines how equilibrium polarization depends on model primitives.

A.4.1 Richness of Influential Coalitions

Our starting observation is that factors that enrich influential coalitions through $\boldsymbol{\chi}$ or q reduces polarization, holding marginal news distributions constant. Below we examine three implications of this observation.

First, under personalized news, enriching the news configuration reduces polarization, and the minimal polarization is attained when news is conditionally independent across voters. Formally, we say that $\boldsymbol{\chi}$ is *richer than* $\boldsymbol{\chi}'$ and write $\boldsymbol{\chi} \succeq \boldsymbol{\chi}'$ if $\boldsymbol{\chi}$ contains all the recommendation profiles as compiled in $\boldsymbol{\chi}'$:

Definition 6. *$\boldsymbol{\chi} \succeq \boldsymbol{\chi}'$ if every column of $\boldsymbol{\chi}'$ is a column of $\boldsymbol{\chi}$.*

Proposition 4. *Fix any population function q and let everything be as in Theorem 3. Then for all p-consistent news configurations $\boldsymbol{\chi}$ and $\boldsymbol{\chi}'$ such that $\boldsymbol{\chi} \succeq \boldsymbol{\chi}'$,*

$$\min_{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}^{**}, q \rangle} \xi^p\left(\mathcal{C}\right) \leq \min_{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}, q \rangle} \xi^p\left(\mathcal{C}\right) \leq \min_{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}', q \rangle} \xi^p\left(\mathcal{C}\right).$$

Proof. The proof combines two facts. First, enriching the news configuration creates new influential coalitions while preserving the old ones, i.e., if $\boldsymbol{\chi} \succeq \boldsymbol{\chi}'$, then $\{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}, q\rangle\} \supseteq \{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}', q\rangle\}$. Second, the news configuration $\boldsymbol{\chi}^{**}$ is the richest among all, i.e., $\boldsymbol{\chi}^{**} \succeq \boldsymbol{\chi}$ for all $\boldsymbol{\chi}$'s. □

Second, under personalized news, polarization is minimized when the news configuration is $\boldsymbol{\chi}^{**}$ and voters' population distribution is uniform across types:

Proposition 5. *Let everything be as in Theorem 3. Then for all population function q,*

$$\min_{k \in \mathcal{K}} \xi^p(k) = \min_{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}^{**}, uniform\rangle} \xi^p(\mathcal{C}) \leq \min_{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}^{**}, q\rangle} \xi^p(\mathcal{C}).$$

Proof. Under $\boldsymbol{\chi}^{**}$ and uniform population distribution, attracting any type of voters, holding other things constant, strictly increases the deviating candidate's winning probability when the remaining voters disagree about candidate fitness. Therefore, every type of voters is influential, and the collection of influential coalitions $\{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}^{**}, \text{uniform}\rangle\} = 2^{\mathcal{K}} - \emptyset$ is the richest across all scenarios. □

Third, recall the exercise in Section 6.1 that decomposes the policy polarization effect of news personalization into (1) changing marginal news distributions and (2) enriching influential coalitions. The second channel can now be formalized as follows:

Proposition 6. *$\{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}^{*}, q\rangle\} \subseteq \{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}, q\rangle\}$ for all p-consistent news configuration $\boldsymbol{\chi}$ and population function q.*

Proof. $\{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}^{*}, q\rangle\} = \{\text{majorities of voters}\} \subseteq \{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}, q\rangle\}$ for all $\boldsymbol{\chi}$ and q as above. □

A.4.2 Population Distribution

In this appendix, we continue to investigate the relationship between mass polarization and elite polarization. Inspired by Fiorina and Abrams (2008) and Gentzkow (2016), we say that the mass becomes more polarized and write $q \overset{SOSD}{\succeq} q'$ if q has second-order stochastic dominance over q'. Compared to q, q' has a thicker tail and pushes the average distance between the left-leaning and right-leaning voters further apart:

Definition 7. $q \overset{SOSD}{\succeq} q'$ *if* $\sum_{k=m}^{K} q(k) \leq \sum_{k=m}^{K} q'(k)$, $\forall m = 1, \cdots, K$.

The analysis assumes quadratic attention cost:

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

The next proposition extends Proposition 3 to arbitrary finite types of voters and p-consistent news configurations:

Proposition 7. *Under Assumptions 1-3 and 5, the following holds for all p-consistent news configuration $\boldsymbol{\chi}$ and population functions q and q' such that $q \overset{SOSD}{\succeq} q'$,*

$$\min_{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}, q \rangle} \xi^p(\mathcal{C}) \geq \min_{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}, q' \rangle} \xi^p(\mathcal{C}).$$

Proof. See Appendix C.2. □

B Other Extensions

B.1 General State Distribution

Let everything be as in Appendix A except that the state ω is distributed symmetrically around zero according to a cumulative density function $G \in \Delta(\mathbb{R})$. A news signal is a mapping $\Pi : \Omega \to \Delta(\mathbb{R})$, where each $\Pi(\cdot \mid \omega)$ specifies a probability distribution over a finite set $\mathcal{Z}$ of signal realizations. Under signal structure Π,

$$\pi_z = \int_{\omega} \Pi(z \mid \omega) G(d\omega)$$

is the probability that the signal realization is $z \in \mathcal{Z}$. Assume without loss of generality that $\pi_z > 0$ for all $z \in \mathcal{Z}$, and let μ_z denote 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 attention level for consuming $\Pi : \Omega \to \Delta(\mathbb{R})$ is*

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

where $H(G)$ is the entropy of the state and $H(G(\cdot \mid z))$ is conditional entropy of the state given signal realization z.

Let $\mathcal{Z}^{\mathcal{S}}(a,k)$ denote the support of any optimal news signal $\Pi^{\mathcal{S}}(a,k)$ that type k voters consume under personalization technology $\mathcal{S}$ and symmetric policy profile $\langle -a,a\rangle$, $a\geq 0$. Drop the notation of k in the broadcast case. Then,

Proposition 8. *Fix any $\langle -a,a\rangle$, $a\geq 0$ and assume Assumptions 3 and 6. Then,*

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

(ii) any optimal broadcast signal $\Pi^{b}(a)$ that is nondegenerate, induces consumption from all voters and makes some voters' participation constraints 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 properties as in Theorem 1(i);*

  (b) if $|\mathcal{Z}^{b}(a)|=3$, then we can write $\mathcal{Z}^{b}(a)=\{LL,LR,RR\}$, whereby $\mu^{b}_{LL}(a)<\mu^{b}_{LR}(a)=0<\mu^{b}_{RR}(a)$, $\mu^{b}_{LL}(a)+\mu^{b}_{RR}(a)=0$, and $v(\mathbf{a},k)+\mu^{b}_{LL}(a)<0<v(\mathbf{a},k)+\mu^{b}_{RR}(a)$ for all $k\in\mathcal{K}$.

Proof. See Appendix C.3. □

B.2 Perfectly Competitive Infomediaries

In this appendix, we divide every type k voters into $m(k)\geq 2$ subpopulations and serve them all by distinct infomediaries. The population mass of type k voters in subpopulation $i=1,\cdots,m(k)$ is $\rho(k,i)$, where $\rho(k,i)>0$ and $\sum_{i=1}^{m(k)}\rho(k,i)=q(k)$. Symmetry requires that $m(k)=m(-k)$ for all $k\in\mathcal{K}$ and that $\rho(k,i)=\rho(-k,i)$ for all $k\in\mathcal{K}$ and $i=1,\cdots,m(k)$.

The perfect competition between infomediaries leads to the maximization of voter expected utility, holding policies fixed. Under any symmetric policy profile $\mathbf{a}=\langle -a,a\rangle$, $a\geq 0$, the competitive news signal $\Pi^{c}(a,k)$ of type k voters solves

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

When the above problem admits multiple solutions, select the most Blackwell informative one in order to make the upcoming result most difficult to establish. Repeating the argument for Lemma 1 shows that the selection is unique, has at most two realizations and, if binary, induces strict obedience among type k voters.

In the current setting, a joint news distribution $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^- \rangle$ is a tuple, where the news configuration $\boldsymbol{\chi}$ compiles the voting recommendation profiles to all subpopulations of voters that occur with strictly positive probabilities, and $\mathbf{b}^+$ and $\mathbf{b}^-$ compile the probabilities that columns of $\boldsymbol{\chi}$ occur in states $\omega = \pm 1$, respectively. As in Appendix A.1, we can define the c-consistency between the joint news distribution and the marginal news distributions as solved above. We can also redefine the p-consistency between the joint news distribution and the marginal news distributions as solved in Section 3, subject to the restriction that news signals are perfectly correlated among the same type of voters. The exercise is omitted for brevity's sake.

The next proposition shows that policy polarization decreases as we transition from optimal personalized news to competitive news:

Proposition 9. *Fix any function ρ and assume Assumptions 1-4 for $\mathcal{S} \in \{c, p\}$. Then for all c-consistent news configuration $\boldsymbol{\chi}$ and p-consistent news configuration $\boldsymbol{\chi}'$ such that $\boldsymbol{\chi} \succeq \boldsymbol{\chi}'$,*

$$\mathcal{E}^{c,\boldsymbol{\chi},\rho} = \left[0, \min_{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}, \rho \rangle} \xi^c(\mathcal{C})\right] \subsetneq \mathcal{E}^{p,\boldsymbol{\chi}',\rho} = \left[0, \min_{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}', \rho \rangle} \xi^p(\mathcal{C})\right].$$

Proof. See Appendix C.3. □

Two forces are acting in the same direction. First, competitive news maximizes voter expected utility rather than attention and is less Blackwell-informative than optimal personalized news. The latter overfeeds voters with information about candidate valence and sustains greater polarization by the trade-off between policy and valence. Interestingly, the effect of introducing perfect competition between infomediaries is equivalent to increasing the marginal attention cost in the case of monopolistic yet personalized news.

Second, in the case where news signals are imperfectly correlated among the same type of voters, competition enriches the news configuration and, by Proposition 4, reduce polarization. An example is where competitive signals are conditionally independent across all subpopulations of voters, whereas optimal personalized signals are only conditionally independent across different types of voters.

B.3 Entropy Attention Cost

This appendix solves the baseline model in the case of entropy attention cost. Detailed algebra are available upon request.

Condition ($*$) The left-hand side of Condition ($*$):

$$\mu_L^p(1) + \mu_R^p(1) < -2t(1)$$

captures the skewness of right-wing voters' personalized signal and can be shown to be decreasing in λ and $t(1)$. As λ increases, the infomediary makes news signals less Blackwell-informative to prevent voters from tuning out. During the process, she is reluctant to raise $\mu_L^p(1)$, which makes news consumption useful for right-wing voters, and instead reduces $\mu_R^p(1)$ significantly as this doesn't affect the usefulness of news consumption as much. Meanwhile as $t(1)$ increases, right-wing voters become more biased policy-wise and seek stronger occasional big surprises than before, i.e., $\mu_L^p(1)$ decreases. At the same time, they derive less expected utilities from news consumption, so $\mu_R^p(1)$ must decrease, too, to prevent them from tuning out. The combined effect on $\mu_L^p(1) + \mu_R^p(1)$ is negative and seems to relax Condition ($*$) when λ and $t(1)$ are both large (see Figure 2 for numerical solutions).

Condition ($**$) The cases in which left-wing voters and median voters are disciplining under personalized news have already been covered in the main text. Here we focus on the case in which right-wing voters are disciplining and Condition ($**$) becomes:

$$\mu_L^p(1) - \mu_L^b < -t(1).$$

As discussed above, the infomediary is reluctant to raise $\mu_L^p(1)$ and instead reduces $\mu_R^p(1)$ significantly as λ increases. Such flexibility is absent in the broadcast case, where the adjustments of μ_L^b and μ_R^b must be symmetric. Combining suggests that increases in λ relax Condition ($**$), and the numerical solutions presented in Figure 2 confirm this intuition.

As for $t(1)$, notice that while $\mu_L^p(1)$ is decreasing in $t(1)$, $\mu_L^b(1)$ is increasing in it: as extreme voters become more biased policy-wise, they derive less expected utilities from news consumption, so the broadcast signal must become less Blackwell-informative to prevent them from tuning out. Thus $\mu_L^p(1) - \mu_L^b(1)$ is decreasing in

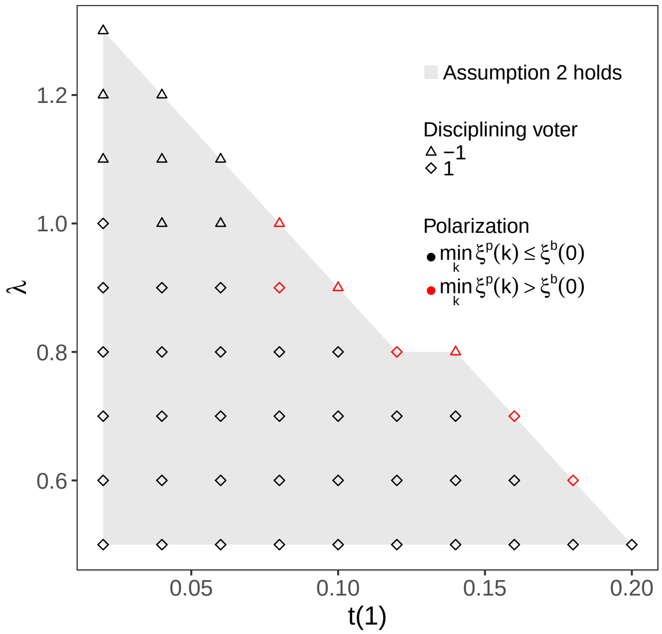

Figure 2: Numerical solutions: entropy attention cost.

$t\,(1)$, and the effect seems to relax Condition $(**)$ when λ and $t\,(1)$ are both large (see again Figure 2).

C Mathematical Proofs

C.1 Proofs of Section 3

This appendix proves the results of Section 3 in the general context laid out in Appendix A, taking any symmetric policy profile $\mathbf{a} = \langle -a, a\rangle$, $a \geq 0$ as given. Since the state is binary, it is without loss to identify signal realizations with posterior means of the state. A signal structure is then a profile $\langle \mu_z, \pi_z \rangle_{z \in \mathcal{Z}}$ of posterior mean μ_z's and probability π_z's, subject to the restriction of Bayes' plausibility (BP): $\sum_{z\in\mathcal{Z}} \pi_z \cdot \mu_z = 0$. As noted in Footnote 15, it is without loss to identify binary signals with the pair $\langle \mu_L, \mu_R \rangle$, where μ_R is realized with probability $-\mu_L / \left(\mu_R - \mu_L\right)$ by (BP).

Proof of Lemmas 1 and 2:

Proof. We proceed in four steps.

Step 1. Show that the optimal personalized signal of any type k voters is unique and has at most two realizations.

If an optimal personalized signal makes type k voters' participation constraint slack, then it coincides with the true (binary) state and therefore constitutes the unique solution to the attention-maximization problem. If it violates type k voters' participation constraint, then it induces zero demand and is therefore replaced by degenerate signals. Otherwise let $\gamma(k) > 0$ denote the Lagrange multiplier associated with type k voters' participation constraint, and consider the following relaxed problem:

$$\max_{\mathcal{Z},\langle \pi_z,\mu_z\rangle_{z\in\mathcal{Z}}} \sum_{z\in\mathcal{Z}} \pi_z \left[\nu(\mu_z;\mathbf{a},k) - \left(\lambda - \frac{1}{\gamma(k)}\right) h(\mu_z)\right] \text{ s.t. (BP)}.$$

Since $\lambda - 1/\gamma(k) > 0$ is needed for the solution(s) to the above problem to differ from the true state, it follows that $\nu(\mu;\mathbf{a},k) - (\lambda - 1/\gamma(k))\, h(\mu)$ is the maximum of two strictly concave functions of μ: (1) $-(\lambda - 1/\gamma(k))\, h(\mu)$, (2) $v(\mathbf{a},k) + \mu - (\lambda - 1/\gamma(k))\, h(\mu)$ if $k \leq 0$ and $-v(\mathbf{a},k) - \mu - (\lambda - 1/\gamma(k))\, h(\mu)$ if $k > 0$. Applying the concavification method developed by Aumann and Maschler (1995) and Kamenica and Gentzkow (2011) yields a unique solution with at most two signal realizations.

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

If an optimal broadcast signal makes all voters' participation constraints slack, then it coincides with the true (binary) state and therefore constitutes the unique solution to the attention-maximization problem. If it violates all voters' participation constraints, then it induces zero demand and is therefore replaced by degenerate signals. Otherwise let $\mathcal{B}$ denote the set of voters with binding participation constraints and $\gamma(k) > 0$ denote the Lagrange multiplier associated with the participation constraint of type $k \in \mathcal{B}$ voters. Consider the following relaxed problem:

$$\max_{\mathcal{Z},\langle \pi_z,\mu_z\rangle_{z\in\mathcal{Z}}} \sum_{z\in\mathcal{Z}} \pi_z \cdot f(\mu_z) \text{ s.t. (BP)},$$

where

$$f(\mu_z) = \sum_{k\in\mathcal{B}} \frac{\gamma(k)}{\sum_{k\in\mathcal{B}}\gamma(k)} \cdot \nu(\mu_z;\mathbf{a},k) - \left(\lambda - \frac{1}{\sum_{k\in\mathcal{B}}\gamma(k)}\right) h(\mu_z).$$

Let $f^+(\mu)$ denote the concave closure of $f(\mu)$. If $f(0) = f^+(0)$, then the solution to the above problem is degenerate. Otherwise define $\mu_L = \sup\{\mu < 0 : f^+(\mu) > f(\mu)\}$ and $\mu_R = \inf\{\mu > 0 : f^+(\mu) > f(\mu)\}$. Among the solution(s) to the above problem, the binary signal $\langle \mu_L, \mu_R \rangle$ is most Blackwell-informative and therefore constitutes the unique solution to the original attention-maximization problem.[28]

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

Under Assumption 3 **increasing difference**, the value $V(\Pi; \mathbf{a}, k)$ of consuming any news signal Π is increasing in k when $k \leq 0$ and is decreasing in k when $k \geq 0$. Depending on whether the participation constraints of types $-K$ and K voters are binding or slack, there are three cases to consider:

Case 1. *Both participation constraints are slack.*

In this case, all voters' participation constraints are slack by Assumption 3 **increasing difference**, so the optimal broadcast signal coincides with the true state, which is unique and symmetric.

Case 2. *One participation constraint is binding and the other is slack.*

Without loss of generality, suppose that type $-K$ voters' participation constraint is binding whereas that of type K voters is slack. If so, then the current signal cannot be type $-K$ voters' optimal personalized signal, as the latter would violate type K voters' participation constraint rather than making it slack. Therefore there exists a perturbation to the current signal such that after the perturbation: (1) type $-K$ voters' participation constraint remains binding, (2) type K voters' participation constraint remains slack, and (3) consuming the perturbed signal requires a higher attention level than consuming the current signal. Under Assumption 3 **increasing difference**, (1) and (2) imply that the perturbed signal satisfies all voters' participation constraints, which combined with (3) leads to a contradiction.

Case 3. *Both participation constraints are binding.*

[28] While the solution to the original problem is unique for any given $\mathcal{B}$, the multiplicity of $\mathcal{B}$'s can still cause a multiplicity of solutions.

In this case, if the optimal broadcast signal is degenerate, then we are done. Otherwise take any optimal binary signal $\langle \mu_L, \mu_R \rangle$, which by assumption yields same consumption value to types $-K$ and K voters:

$$\frac{-\mu_L}{\mu_R - \mu_L} \left(v(\mathbf{a}, -K) + \mu_R \right) = \frac{-\mu_R}{\mu_R - \mu_L} \left(v(\mathbf{a}, K) + \mu_L \right) = \lambda \cdot I \left(\langle \mu_L, \mu_R \rangle \right).$$

Simplifying the above expression using $v(\mathbf{a}, K) = -v(\mathbf{a}, -K)$ (Assumption 3 **symmetry**) yields symmetry in the distribution of posterior means, i.e., $\mu_L + \mu_R = 0$. Taken together, $\langle \mu_L, \mu_R \rangle$ is the optimal personalized signal of type $-K$ voters (equivalently, type K voters), subject to the symmetry restriction. In particular, μ_L solves the following problem:

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

and the uniqueness of solution (existence is implicit; otherwise the optimal signal is degenerate rather than binary) is guaranteed by Assumption 1.

Step 4. Show that any optimal news signal, if binary, must induce strict obedience among its consumers.

Let $\Pi^{\mathcal{S}}(a,k)$ be as above and notice that $I\left(\Pi^{\mathcal{S}}(a,k)\right) > 0$ by Assumption 1. If, instead of (SOB), we have either $v(\mathbf{a}, k) + \mu_R^{\mathcal{S}}(a,k) > v(\mathbf{a}, k) + \mu_L^{\mathcal{S}}(a,k) \geq 0$ or $v(\mathbf{a}, k) + \mu_L^{\mathcal{S}}(a,k) < v(\mathbf{a}, k) + \mu_R^{\mathcal{S}}(a,k) \leq 0$, then type k voters have a weakly preferred candidate that is independent of the signal realization and would therefore abstain from news consumption to save the attention cost, a contradiction. □

Proof of Theorem 1

Proof. Part (i): See Step 3 in the proof of Lemma 1.

Part (ii): We only prove the result for an arbitrary type $k < 0$ voters, for whom we write $v(\mathbf{a}, k) = v$ and note that $v < 0$ by Assumption 3 **increasing difference**. Consider the following relaxed problem:

$$\max_{\langle \mu_L, \mu_R \rangle} \pi (v + \mu_R) - (\lambda - 1/\gamma) \left[(1-\pi) h(\mu_L) + \pi h(\mu_R) \right] \text{ s.t. } \pi = -\mu_L / (\mu_R - \mu_L),$$

where $\gamma > 0$ denotes the Lagrange multiplier associated with type k voters' participation constraint. Assuming interior solutions (Assumption 2), we obtain $\lambda - 1/\gamma > 0$ and the following first-order conditions:

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

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

$$h'(\mu_R) - h'(\mu_L) = \frac{\mu_R - \mu_L}{(\lambda - 1/\gamma)\,\Delta\mu}$$

and thus $\mu_R > -\mu_L$ by Assumption 1.

Part (iii): See the verbal argument after the statement of Theorem 1. $\square$

C.2 Proofs of Appendix A

C.2.1 Useful Lemmas

Proof of Lemma 5

Proof. Fix any personalization technology $\mathcal{S}$, population function q, policy $a \geq 0$ and $\langle \mathcal{S}, a\rangle$-consistent news distribution $\langle \boldsymbol{\chi}, \mathbf{b}^+, \mathbf{b}^- \rangle$. Let $\mathbf{q}$ denote the $|\mathcal{K}|$-vector that compiles the populations of types $-K, \cdots, K$ of voters. 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 above 1/2 with 1 and those below 1/2 with zero.

By definition, the row vector $\mathbf{q}^\top \boldsymbol{\chi}$ (resp. $\widehat{\mathbf{q}^\top \boldsymbol{\chi}}$) lists the number of votes (resp. winning probability) earned by candidate R under the voting recommendation profiles as compiled in $\boldsymbol{\chi}$. Meanwhile, $\boldsymbol{\chi}_{\mathcal{C}}$ is the vote configuration matrix after candidate R commits a deviation as in Definition 4, which makes all voters in $\mathcal{C}$ vote unconditionally for him while holding other things constant. From $\widehat{\mathbf{q}^\top \boldsymbol{\chi}_{\mathcal{C}}} \geq \widehat{\mathbf{q}^\top \boldsymbol{\chi}}$, it follows that the above deviation strictly increases candidate R's winning probability in expectation if and only if it does so under some voting recommendation profile, i.e., $\widehat{\mathbf{q}^\top \boldsymbol{\chi}_{\mathcal{C}}} \neq \widehat{\mathbf{q}^\top \boldsymbol{\chi}}$. The last condition is equivalent to $\mathcal{C}$ being influential and depends only on $\boldsymbol{\chi}$ and $\mathbf{q}$. In particular, if $\boldsymbol{\chi}$ is independent of a, then so are the influential coalitions formed under the pair $\langle \boldsymbol{\chi}, q \rangle$. $\square$

Proof of Lemma 6

Proof. Fix any personalization technology $\mathcal{S}$, population function q, policy $a \geq 0$ and $\langle \mathcal{S}, a \rangle$-consistent news distribution $\langle \boldsymbol{\chi}, \mathbf{b}^{+}, \mathbf{b}^{-} \rangle$. 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 in four steps that a' cannot increase candidate R's winning probability if and only if (1) $a' \notin [-a, a]$ or (2) $a' \in [-a, a)$ and $\phi^{\mathcal{S}}(-a, a', \mathcal{C}) \leq 0$ for all $\mathcal{C}$'s such that $t(\mathcal{C}) \subseteq [-a, a]$.

Step 1. Show that no $a' > a$ increases candidate R's winning probability.

Let a' be as above. Notice first that no type $k \leq 0$ voters find a' attractive, because under Assumption 3,

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

Below we demonstrate that if a' attracts any type $k > 0$ voters, i.e., $v(-a, a', k) + \mu_L^{\mathcal{S}}(a, k) > 0$, then it repels type $-k$ voters, i.e., $v(-a, a', -k) + \mu_R^{\mathcal{S}}(a, -k) < 0$. If so, then combining with the symmetries of the population function and joint news distribution gives the desired result.

The argument exploits the symmetry of marginal news distributions (as shown in Theorem 1(ii)), i.e., $\mu_R^{\mathcal{S}}(a, -k) = -\mu_L^{\mathcal{S}}(a, k)$, which combined with Assumption 3 **symmetry** yields

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

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

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

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

Step 2. Show that no $a' < -a$ increases candidate R's winning probability, either.

The proof resembles that of Step 1. First, a' cannot attract any type $k \geq 0$ voters by Assumption 3 **inverted V-shape**:

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

Second, if a' attracts any type $k < 0$ voters, then it repels type $-k$ voters as shown in Step 1. Combining gives the desired result.

Step 3. Show that no $a' \in [-a, a)$ repels any voter.

(SOB), together with Assumption 3 **inverted V-shape**, implies that if $t(k) \leq a'$, then

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

Meanwhile, Assumption 3 **inverted V-shape** alone implies that if $t(k) > a'$, then

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

Combining yields $v(-a, a', k) + \mu_R^{\mathcal{S}}(a, k) > 0$ for all k, thus completing the proof.

Step 4. Show that no $a' \in [-a, a)$ attracts voters whose ideological bliss points lie outside $[-a, a]$.

Let a' be as above. Assumption 3 **inverted V-shape** implies that if $t(k) < -a$, then

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

Meanwhile, Assumption 3 **inverted V-shape** and (SOB) together imply that if $t(k) > a$, then

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

Combining yields $v(-a, a', k) + \mu_L^{\mathcal{S}}(a, k) < 0$ for all k, thus completing the proof. □

Lemma 7. *Let everything be as in Theorem 3. Then for all* $k = 0, \cdots, K$ *and* $\mathcal{D} \subseteq \{-k, \cdots, k\}$ *such that* $\mathcal{D} \cap \{-k, k\} \neq \emptyset$, $\xi^{\mathcal{S}}(\mathcal{D}) > t(k)$ *and* $\left[t(k), \xi^{\mathcal{S}}(\mathcal{D})\right] = [t(k), \overline{a}] \cap \Xi^{\mathcal{S}}(\mathcal{D})$.

Proof. Let k and $\mathcal{D}$ be as above, and take any $a' \in [\min t(\mathcal{D}), \max t(\mathcal{D})] \subseteq [t(-k), t(k)]$. From Assumption 3 and (SOB), we know that

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

and that

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

Combining yields

$$\phi^{\mathcal{S}}(-t(k), a', \mathcal{D}) = \min_{k' \in \mathcal{D}} \phi^{\mathcal{S}}(-t(k), a', k') < 0.$$

Meanwhile, Assumption 4 implies that the function $\phi^{\mathcal{S}}(-a, a', \mathcal{D})$ is increasing in a on $[t(k), \overline{a}]$. Taken together, we obtain that

$$\xi^{\mathcal{S}}(\mathcal{D}) = \max\left\{a \geq 0 : \max_{a' \in [\min t(\mathcal{D}), \max t(\mathcal{D})]} \phi^{\mathcal{S}}(-a, a', \mathcal{D}) \leq 0\right\} > t(k),$$

and that

$$[t(k),\bar{a}]\cap\Xi^{\mathcal{S}}(\mathcal{D})=\left\{a\geq t(k):\max_{a'\in[\min t(\mathcal{D}),\max t(\mathcal{D})]}\phi^{\mathcal{S}}(-a,a',\mathcal{D})\leq 0\right\}=\left[t(k),\xi^{\mathcal{S}}(\mathcal{D})\right].$$

□

Lemma 8. *Assume Assumptions 1-3 and 5. Then for all $a\geq 0$ and $a'\in[-a,a]$,*

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

(ii) $\phi^{p}(-a,a',k)\leq\phi^{p}(-a,a',-k)$ for all $k>0$.

Proof. Let a and a' be as above. Tedious but straightforward algebra shows that in the personalized case, Assumption 2 is equivalent to

$$2\lambda>1 \text{ and } 4\lambda v(-a,a,K)<1,$$

and

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

Part (i): For all $k\leq 0$,

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

where the last line is decreasing in k by Assumption 3 **increasing difference**. For all $k>0$,

$$\phi^{p}(-a,a',k)=v(-a,a',k)-\frac{1}{2\lambda}$$

and is thus increasing in k by Assumption 3 **increasing difference**.

Part (ii): For all $k > 0$, Assumption 3 implies that

$$\begin{aligned}
&\phi^p(-a, a', k) - \phi^p(-a, a', -k) \\
&= v(-a, a', k) - 2v(-a, a, k) - v(-a, a', -k) \\
&= v(-a, a', k) - 2v(-a, a, k) - v(a, -a', k) && (\textbf{symmetry}) \\
&= [u(-a, k) - u(-a', k)] - [u(a, k) - u(a', k)] \\
&= v(a', a, -k) - v(a', a, k) && (\textbf{symmetry}) \\
&\quad \leq 0. && (\textbf{increasing difference})
\end{aligned}$$

□

Lemma 9. *Under Assumptions 1-3, Assumption 4 holds if* $\mathcal{S} = b$ *or* $\mathcal{S} = p$ *and either* $u(a, k) = -|t(k) - a|$ *or* $h(\mu) = \mu^2$.

Proof. Let a, a' and k be as in Assumption 4. From Assumption 3 **inverted V-shape**, we know that $v(-a, a', k)$ is strictly increasing in a on $[|t(k)|, \overline{a}]$. Thus, $\phi^{\mathcal{S}}(-a, a', k) = v(-a, a', k) + \mu_L^{\mathcal{S}}(a, k)$ is increasing in a on $[|t(k)|, \overline{a}]$ if $\mu_L^{\mathcal{S}}(a, k)$ is nondecreasing in a on $[|t(k)|, \overline{a}]$. Consider three cases:

$\mathcal{S} = b$ When proving Lemma 1, we demonstrated that $\mu_L^b(a)$ is the unique solution to the following problem:

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

where h is strictly convex and strictly decreasing on $[-1, 0]$. Meanwhile, Assumption 3 implies that for all $a' > a \geq 0$:

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

i.e., $v(-a, a, -K)$ is decreasing in a. Combining gives the desired result.

$\mathcal{S} = p$ and $u(a,k) = -|t(k) - a|$ In this case, plugging $v(-a,a,k) \equiv 2t(k)$ for all $a \geq |t(k)|$ into the proof of Theorem 1(ii) gives the desired result.

$\mathcal{S} = p$ and $h(\mu) = \mu^2$ In this case, exploiting the algebra in the proof of Lemma 8 gives the desired result. $\square$

C.2.2 Proofs of Theorems and Propositions

Proof of Theorem 3

Proof. Let $\mathcal{C}$'s denote the influential coalitions formed under $\langle \boldsymbol{\chi}, q \rangle$. Define

$$A(k) = \begin{cases} [t(k), t(k+1)) \cap \bigcap_{\mathcal{C} \subseteq \{-k,\cdots,k\}} \Xi^{\mathcal{S}}(\mathcal{C}) & \text{if } k = 0, \cdots, K-1, \\ [t(K), \overline{a}] \cap \bigcap_{\mathcal{C}} \Xi^{\mathcal{S}}(\mathcal{C}) & \text{if } k = K. \end{cases}$$

Lemma 6 implies that

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

It remains to show that $\cup_{k=0}^{K} A(k) = \left[0, \min_{\mathcal{C}} \xi^{\mathcal{S}}(\mathcal{C})\right]$. By Lemma 7, we can prove this result by induction:

Step 0. The set $A(0)$ equals $\left[0, \xi^{\mathcal{S}}(0)\right]$ if $\{0\}$ is influential and $\xi^{\mathcal{S}}(0) < t(1)$, and it equals $[0, t(1))$ otherwise. In the first case, $\xi^{\mathcal{S}}(\mathcal{C}) > t(1)$ for all $\mathcal{C} \neq \{0\}$ by Lemma 7, so $\min_{\mathcal{C}} \xi^{\mathcal{S}}(\mathcal{C}) = \xi^{\mathcal{S}}(0)$, $A(k) = \emptyset$ for all $k \geq 1$, and taking union yields $\cup_{k=0}^{K} A(k) = \left[0, \min_{\mathcal{C}} \xi^{\mathcal{S}}(\mathcal{C})\right]$. In the second case, we proceed to the next step.

Step m. Since $\cup_{k=0}^{m-1} A(k) = [0, t(m))$, and $[t(m), \overline{a}] \cap \Xi^{\mathcal{S}}(\mathcal{C}) = \left[t(m), \xi^{\mathcal{S}}(\mathcal{C})\right]$ for all $\mathcal{C} \subseteq \{-m, \cdots, m\}$ such that $\mathcal{C} \cap \{-m, m\} \neq \emptyset$ (Lemma 7), the set $\cup_{k=0}^{m} A(k)$ equals $\left[0, \min_{\mathcal{C} \subseteq \{-m,\cdots,m\}} \xi^{\mathcal{S}}(\mathcal{C})\right]$ if $\min_{\mathcal{C} \subseteq \{-m,\cdots,m\}} \xi^{\mathcal{S}}(\mathcal{C}) < t(m+1)$ and $[0, t(m+1))$ otherwise. In the first case, $t(m+1) < \xi^{\mathcal{S}}(\mathcal{C}')$ for all $\mathcal{C}' \nsubseteq \{-m, \cdots, m\}$ by Lemma 7, so $\min_{\mathcal{C} \subseteq \{-m,\cdots,m\}} \xi^{\mathcal{S}}(\mathcal{C}) = \min_{\mathcal{C}} \xi^{\mathcal{S}}(\mathcal{C})$, $A(k) = \emptyset$ for all $k \geq m+1$, and taking union yields $\cup_{k=0}^{K} A(k) = \left[0, \min_{\mathcal{C}} \xi^{\mathcal{S}}(\mathcal{C})\right]$. 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 7

Proof. We proceed in two steps.

Step 1. Show that $\xi^{p}(\mathcal{D}) > t(K)$ for all $\mathcal{D} \subseteq \mathcal{K}$.

Fix any $a' \in [t(-K), t(K)]$ and $\mathcal{D} \subseteq \mathcal{K}$, and notice two things. First,

$$
\begin{aligned}
&\phi^{p}(-t(K), a', \mathcal{D}) \\
&= \min_{k \in \mathcal{D}} \phi^{p}(-t(K), a', k) \\
&\quad \leq \max_{k \in \mathcal{D}} \phi^{p}(-t(K), a', k) \\
&\qquad \leq \phi^{p}(-t(K), a', -K) && \text{(Lemma 8)} \\
&\qquad\quad \leq \phi^{p}(-t(K), t(-K), -K) && \text{(Assumption 3 \textbf{inverted V-shape})} \\
&\qquad\quad = v(-t(K), t(-K), K) + \mu_{L}^{p}(t(K), -K) \\
&\qquad\quad = 0 + \mu_{L}^{p}(t(K), -K) && \text{(Assumption 3 \textbf{symmetry})} \\
&\qquad\qquad < 0.
\end{aligned}
$$

Second, Lemma 9 shows that $\phi^{p}(-a, a', \mathcal{D})$ is increasing in a on $[t(K), \overline{a}]$. Taken together, we obtain that

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

or, equivalently,

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

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 all $a \geq t(K)$ and $a' \in [-a, a]$: (1) $\phi^{p}(-a, a', \mathcal{C}) \leq 0$,

(2) $\phi^p(-a, a', \max \mathcal{C}) \le 0$, and (3) $\phi^p(-a, a', \{k \le \max \mathcal{C}\}) \le 0$. Second, since $\mathcal{C}$ is influential and $\mathcal{C} \subseteq \{k \le \max \mathcal{C}\}$, $\{k \le \max \mathcal{C}\}$ is influential, too. Combining yields

$$\min_{\substack{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}, q \rangle: \\ \max \mathcal{C} \le 0}} \xi^p(\mathcal{C}) = \min_{\substack{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}, q \rangle: \\ \mathcal{C} = \{k \le \alpha\}, \alpha \le 0}} \xi^p(\mathcal{C}). \tag{C.1}$$

Consider the right-hand side of Equation (C.1). By Lemma 8, the term

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

is increasing in α on $\alpha \le 0$. Meanwhile, every set $\{k \le \alpha\}, \alpha < 0$ is more likely to be influential under q' than under q because $q \overset{SOSD}{\succeq} q'$. Combining yields

$$\min_{\substack{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}, q \rangle: \\ \mathcal{C} = \{k \le \alpha\}, \alpha \le 0}} \xi^p(\mathcal{C}) \ge \min_{\substack{\mathcal{C}'s \text{ formed under } \langle \boldsymbol{\chi}, q' \rangle: \\ \mathcal{C} = \{k \le \alpha\}, \alpha \le 0}} \xi^p(\mathcal{C}),$$

thus completing the proof of case (a). The proofs of cases (b) and (c) are similar and are therefore omitted. □

C.3 Proofs of Appendix B

Proof of Proposition 8

Proof. The upcoming analysis assumes that u has strictly increasing differences or, equivalently, $v(\mathbf{a}, k)$ is strictly increasing in k. This is without loss of generality, since we already assume that $v(\mathbf{a}, k)$ is increasing in k and can thus identify adjacent types k and $k+1$ of voters as the same person if $v(\mathbf{a}, k) = v(\mathbf{a}, k+1)$.

Part (i): Consider the following relaxed problem:

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

where $\gamma(k) > 0$ denotes the Lagrange multiplier associated with type k voters' participation constraint. Notice that $\lambda - 1/\gamma(k) > 0$, because otherwise the solution to the above problem coincides with the true state and hence $\gamma(k) = 0$, a contradiction. By Matějka and McKay (2015), any solution has at most two signal realizations and must therefore be binary if nondegenerate. (SOB) is then immediate.

Write $\lambda' = \lambda - 1/\gamma(k)$ and $\mathcal{L}(k)$ as the likelihood that type k voters choose candidate R over L. By Matějka and McKay (2015), the probability type k voters choose candidate R in state ω equals

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

Thus, the above problem is equivalent to

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

where $I(\mathcal{L})$ denotes the mutual information of the state variable and voting decision given likelihood ratio $\mathcal{L}$. Since the objective function has strict increasing differences in $\mathcal{L}$ and k, $\mathcal{L}(k)$ is increasing in k. Symmetry then implies that $\mathcal{L}(k) > 1/2$ if $k > 0$, $\mathcal{L}(k) = 1/2$ if $k = 0$ and $\mathcal{L}(k) < 1/2$ if $k < 0$, which combined with Bayes' plausibility (BP) gives the desired result.

Part (ii): Strict increasing differences, plus the reason given in the proof of Lemma 1, imply that voters in $\mathcal{B} = \{-K, K\}$ have binding participation constraints and those in $\mathcal{K}-\mathcal{B}$ have slack participation constraints. Consider the following relaxed problem:

$$\max_{\mathcal{Z},\Pi:\mathbb{R}\to\Delta(\mathcal{Z})} \sum_{z\in\mathcal{Z}}\pi_z\sum_{k\in\mathcal{B}}\frac{\gamma(k)}{\sum_{k\in\mathcal{B}}\gamma(k)}\cdot\nu(\mu_z;\mathbf{a},k)-\left(\lambda-\frac{1}{\sum_{k\in\mathcal{B}}\gamma(k)}\right)I(\Pi),$$

where $\gamma(k) > 0$ denotes the Lagrange multiplier associated with type $k \in \mathcal{B}$ voters' participation constraint. Notice the equivalence of this problem to that of optimal information acquisition as in Matějka and McKay (2015), whereby a representative voter acts on behalf of voters in $\mathcal{B}$ and makes three decisions under complete information: LL, LR and RR (see Footnote 23 for the definitions of these decisions). The information cost parameter $\lambda - \frac{1}{\sum_{k\in\mathcal{B}}\gamma(k)}$ must be positive, because otherwise the solution to the above problem would coincide with the true state and hence $\gamma(k) = 0$ for all $k \in \mathcal{B}$, a contradiction.

By Matějka and McKay (2015), any solution Π to the above problem has at most three signal realizations. In the case of two signal realizations, the remainder of the proof is the same as that of Lemma 1. In the case of three signal realizations, obedience as shown in Matějka and McKay (2015) implies that $v(\mathbf{a},k)+\mu_{LL}<0<v(\mathbf{a},k)+\mu_{RR}$ for $k\in\mathcal{B}$ and that $v(\mathbf{a},-K)+\mu_{LR}\leq 0\leq v(\mathbf{a},K)+\mu_{LR}$. If, in addition, the distribution of posterior means is symmetric around zero, then $\mu_{LR}=0$ and $\mu_{LL}+\mu_{RR}=0$, and we are done.

To demonstrate symmetry, define a new signal structure Π' by $\Pi'(LL\mid\omega)=\Pi(RR\mid-\omega)$, $\Pi'(RR\mid\omega)=\Pi(LL\mid-\omega)$ and $\Pi'(LR\mid\omega)=\Pi(LR\mid-\omega)$. Under Π', the posterior means of the state are $\mu'_{LL}=-\mu_{RR}$, $\mu'_{LR}=-\mu_{LR}$ and $\mu'_{RR}=-\mu_{LL}$, and the corresponding probabilities are $\pi'_{LL}=\pi_{RR}$, $\pi'_{LR}=\pi_{LR}$ and $\pi'_{RR}=\pi_{LL}$. By Assumption 3 **symmetry**, $V(\Pi';\mathbf{a},-k)=V(\Pi;\mathbf{a},k)$ for $k\in\mathcal{B}$. In addition, $I(\Pi)=I(\Pi')$ by the symmetry of mutual information, and combining yields $V(\Pi';\mathbf{a},k)=V(\Pi;\mathbf{a},k)=\lambda\cdot I(\Pi)=\lambda\cdot I(\Pi')$ for $k\in\mathcal{B}$:

$$\begin{aligned}
&V(\Pi';\mathbf{a},k)-\lambda\cdot I(\Pi')\\
&=V(\Pi;\mathbf{a},-k)-\lambda\cdot I(\Pi)\\
&=0 && (-k\in\mathcal{B})\\
&=V(\Pi;\mathbf{a},k)-\lambda\cdot I(\Pi) && (k\in\mathcal{B})\\
&=V(\Pi;\mathbf{a},k)-\lambda\cdot I(\Pi').
\end{aligned}$$

To complete the proof, suppose, to the contrary, that $\mu_{LR}\neq 0$ and hence $\Pi\neq\Pi'$. If so, then the signal structure $\frac{1}{2}\circ\Pi+\frac{1}{2}\circ\Pi'$ yields the same consumption value to voters in $\mathcal{B}$ (because $V(\Pi;\cdot,\cdot)$ is linear in Π) but incurs a lower attention cost than Π (because $I(\Pi)$ is strictly convex in Π). Thus Π is not a solution to the relaxed problem, which leads to a contradiction. □

Proof of Proposition 9

Proof. We proceed in two steps.

Step 1. Show that $\Pi^p(a,k)$ is more Blackwell-informative than $\Pi^c(a,k)$ for all $a\geq 0$ and $k\in\mathcal{K}$.

Write $\Pi^p(a,k,\lambda)$ and $\Pi^c(a,k,\lambda)$ to make their dependence on λ explicit. When proving Lemma 1, we demonstrated that $\Pi^p(a,k,\lambda)=\Pi^c(a,k,\lambda-1/\gamma(k))$, where $\gamma(k)>0$ is the Lagrange multiplier associated with type k voters' participation constraints under $\mathcal{S} = p$. It remains to show that $\Pi^c(a, k, \lambda)$ becomes less Blackwell-informative as λ increases.

We only prove the result for an arbitrary type $k \leq 0$ voters, for whom we write $v(-a, a, k) = v$. The first-order conditions of welfare maximization are

$$v + \mu_R = \lambda \left[\Delta h - h'(\mu_L) \Delta\mu\right], \tag{C.2}$$

$$\text{and} \quad -(v + \mu_L) = \lambda \left[h'(\mu_R) \Delta\mu - \Delta h\right], \tag{C.3}$$

where $\Delta h = h(\mu_R) - h(\mu_L)$ and $\Delta\mu = \mu_R - \mu_L$. Summing up Equations (C.2) and (C.3) yields

$$h'(\mu_R) - h'(\mu_L) = 1/\lambda, \tag{C.4}$$

and using this result when differentiating Equation (C.3) with respect to λ yields

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

Therefore,

$$\frac{d\mu_R}{d\lambda} = \frac{\Delta h - h'(\mu_R) \Delta\mu}{\lambda h''(\mu_R) \Delta\mu} < 0, \tag{C.5}$$

where the inequality follows from $h'' > 0$ and $\Delta\mu > 0$. Meanwhile, differentiating (C.4) with respect to λ yields

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

and simplifying using Equation (C.5), $h'' > 0$ and $\Delta\mu > 0$ yields

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

Step 2. From Step 1, we know that $\mu_L^c(a,k) > \mu_L^p(a,k)$ and hence that

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

for all $a \geq 0$, a' and $\mathcal{D} \subseteq \mathcal{K}$. Plugging this result into the proof of Lemma 7 yields $\xi^c(\mathcal{D}) < \xi^p(\mathcal{D})$. Then,

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

thus completing the proof. □

C.4 Proofs of Sections 4-6

This appendix proves the remaining results of Sections 4-6 in the context laid out in the baseline model. The analysis exploits the functional form assumption $u(a,k) = -|t(k) - a|$, under which

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

is invariant with a on $[|t(k)|, \overline{a}]$. Using this property in the proof of Lemma 1 yields $\mu_L^b(a) \equiv \mu_L^b := \mu_L^b(t(1))$ on $[t(1), \overline{a}]$ and $\mu_L^p(a,k) \equiv \mu_L^p(k) := \mu_L^p(|t(k)|, k)$ on $[|t(k)|, \overline{a}]$.

Proof of Lemma 4

Proof. $\mathcal{S} = b$: When proving Lemma 9, we demonstrated that $\mu_L^b(a)$ is increasing in a, so the function $\phi^b(-a,0,0) = a + \mu_L^b(a)$ is strictly increasing in a. Then from

$\phi^b(-a,0,0)\Big|_{a=0} = \mu_L^b(0) < 0$, it follows that $\xi^b(0) = \max\left\{a \geq 0 : \phi^b(-a,0,0) \leq 0\right\} > 0$, and that $\xi^b(0)$ is the unique root of $\phi^b(-a,0,0)$ when $\overline{a}$ is large. If, in addition, $\xi^b(0) \geq t(1)$, then $\xi^b(0) = -\mu_L^b\left(\xi^b(0)\right) = -\mu_L^b$.

$\mathcal{S} = p$: In the case of $k = 1$, notice that $\phi^p(-a,t(1),1) = a + t(1) + \mu_L^p(1)$ when $a \geq t(1)$. Combining with (SOB): $\phi^p(-t(1),t(1),1) = v(-t(1),t(1),1) + \mu_L^p(t(1),1) < 0$, yields $\xi^p(1) = \max\{a \geq 0 : \phi^p(-a,t(1),1) \leq 0\} > t(1)$, as well as $\xi^p(1) = -(t(1) + \mu_L^p(1))$ when $\overline{a}$ is large.

In the case of $k = -1$, notice that (1) $\phi^p(-a,t(-1),-1) = a + t(-1) + \mu_L^p(-1)$ when $a \geq t(1)$, and that (2) $\phi^p(-t(1),t(-1),-1) = \mu_L^p(-1) < 0$. Combining yields $\xi^p(-1) = \max\{a \geq 0 : \phi^p(-a,t(-1),-1) \leq 0\} > t(1)$, as well as $\xi^p(-1) = -(t(-1) + \mu_L^p(-1))$ when $\overline{a}$ is large.

The remaining case is $k = 0$. The proof resembles that of $k = \pm 1$ and is therefore omitted. □

Proof of Proposition 2

Proof. Fix any $\mathbf{a} = \langle -a, a\rangle$, $a \geq 0$ and k. When proving Proposition 9, we demonstrated that (1) $\Pi^c(a,k,\lambda)$ becomes less Blackwell-informative as λ increases and that (2) $\Pi^p(a,k,\lambda) = \Pi^c(a,k,\lambda - 1/\gamma(\lambda))$, where $\gamma(\lambda)$ denotes the Lagrange multiplier of type k voters' participation under $\mathcal{S} = p$ (as a function of λ). Write $\beta(\lambda) = \lambda - 1/\gamma(\lambda)$. If $\beta(\lambda)$ is increasing in λ, then $\Pi^p(a,k,\lambda)$ becomes less Blackwell-informative as λ increases, and combining with the proof of Proposition 9 gives the desired result.

Suppose, to the contrary, that $\beta(\lambda'') < \beta(\lambda')$ for some $\lambda'' > \lambda' > 0$. Write $\Pi^p(a,k,\lambda') = \Pi'$ and $\Pi^p(a,k,\lambda'') = \Pi''$. From (1) and (2), we know that Π'' is more Blackwell-informative than Π', so $V(\Pi'';\mathbf{a},k) > V(\Pi';\mathbf{a},k)$ and $I(\Pi'') > I(\Pi')$. Then from $V(\Pi';\mathbf{a},k) = \lambda' \cdot I(\Pi')$ and $V(\Pi'';\mathbf{a},k) = \lambda'' \cdot I(\Pi'')$, it follows that $V(\Pi'';\mathbf{a},k) - \lambda' \cdot I(\Pi'') > 0 = V(\Pi';\mathbf{a},k) - \lambda' \cdot I(\Pi')$, which combined with $I(\Pi'') > I(\Pi')$ implies that Π' is not optimal when $\lambda = \lambda'$, a contradiction. □

Completing Example 2

Proof. Let everything be as in the baseline model, except that extreme voters abstain from news consumption and vote along party lines when being indifferent between the candidates. Take any symmetric policy profile $\langle -a, a\rangle$, $a \in \left[0, \underline{\xi}\right]$, where $\underline{\xi} = \min\left\{t(1), \xi^{\mathcal{S}}(0)\right\}$. From Lemma 4, we know that $\left[0, \underline{\xi}\right] \subseteq \left[0, \xi^{\mathcal{S}}(0)\right] = \Xi^{\mathcal{S}}(0)$, so no unilateral deviation from $\langle -a, a\rangle$ attracts median voters. In addition, no such deviation affects the total number of votes that extreme voters cast to the deviating candidate, and combining gives the desired result. □

Completing Example 3

Proof. Fix the policy profile to be $\langle -t(1), t(1)\rangle$. For any type $k \in \{-1, 1\}$ voters, write the posterior mean of the state conditional on signal realization $z \in \{L, M, R\}$ as $\mu_z(k)$, where $v(-t(1), t(1), k) + \mu_M(k) = 0$ by assumption. To satisfy type k voters' participation constraint, a necessary condition is $v(-t(1), t(1), k) + \mu_{z'}(k) < 0 < v(-t(1), t(1), k) + \mu_{z''}(k)$ for $z', z'' \neq M$. Hereafter assume without loss of generality that $z' = L$ and $z'' = R$.

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 even $a' = 0$ cannot attract 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.

By symmetry, consider type -1 voters only, for whom $v(-t(1), t(1), -1) < 0$ and thus $\mu_L(-1) < 0 < \mu_M(-1) = -v(-t(1), t(1), -1) < \mu_R(-1)$ by (BP). From Assumption 3, it follows that

$$\begin{aligned} v(-t(1), a', -1) + \mu_L(-1) &\leq v(-t(1), t(-1), -1) + \mu_L(-1) \\ &= v(-t(1), -t(1), -1) + \mu_L(-1) = 0 + \mu_L(-1) < 0, \end{aligned}$$

and that

$$v(-t(1), a', -1) + \mu_z(-1) > v(-t(1), t(1), -1) + \mu_z(-1) \geq 0$$

when $z = M, R$. To complete the proof, let extreme voters vote across party lines when being indifferent between the 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.

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.*

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

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

Bernhardt, D., S. Krasa and M. Polborn (2008): "Political Polarization and the Electoral Effects of Media Bias," *Journal of Public Economics*, 92, 1092-1104.

Blackwell, D. (1953): "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 24, 265-272.

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. (2016): "Measuring and Modeling Attention," *Annual Review of Economics*, 8, 379-403.

Caplin, A., and M. Dean (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 (2018): "Optimal Dynamic Allocation of Attention," *American Economic Review*, forthcoming.

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.*

Duggan, J. (2000): "Equilibrium Equivalence Under Expected Plurality and Probability of Winning Maximization," *Working Paper.*

——— (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.

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*.

Gentzkow, M., and E. Kamenica (2014): "Costly Persuasion," *American Economic Review: Papers & Proceedings*, 104(5), 457-462.

——— (2010): "What Drives Media Slant? Evidence from U.S. Daily Newspapers," *Econometrica*, 78(1), 35-71.

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. Woodford (2017): "Rational Inattention in Continuous Time," *Working Paper*.

Hu, L., and A. Li (2018): "The Politics of Attention," *Working Paper*.

Jeon, D. S. (2018): "Economics of News Aggregators," *Working Paper*.

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.

Maćkowiak, B., F. Matějka and M. Wiederholt (2018): "Rational Inattention: A Disciplined Behavioral Model," *Working Paper*.

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

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., J. Gottfried, E. Shearer and K. Lu (2017): "How Americans Encounter, Recall and Act Upon Digital News," *Pew Research Center*, February 9.

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

Patty, J. (2005): "Local Equilibrium Equivalence in Probabilistic Voting Models," *Games and Economic Behavior*, 51, 523-536.

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*, eds. by D. Acemoglu, M. Arellano and E. Dekel. Cambridge University Press.

Prior, M. (2005): "News vs. Entertainment: How Increasing Media Choice Widens Gaps in Political Knowledge and Turnout," *American Journal of Political Science*, 49(3), 577-592.

Shearer, E. (2018): "Social Media Outpaces Print Newspapers in the U.S. as a News Source," *Pew Research Center*, December 10.

Sims, C. (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 (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.*

Tsakas, E. (2019): "Robust Scoring Rules," *Working Paper.*

Zhong, W. (2017): "Optimal Dynamic Information Acquisition," *Working Paper.*