A Theory on Media Bias and Elections*

Junze Sun† Arthur Schram‡ Randolph Sloof§

Job Market Paper
October 12, 2018

Abstract

We develop a tractable theory to study the impact of biased media on election outcomes, voter turnout and welfare. News released by media allows voters to infer (i) the relative appeal of candidates, and (ii) the closeness of elections. In large elections, the former determines the election outcome, whereas the latter drives voter turnout. With a single media outlet, an increase in media bias affects the election outcome in a non-monotonic way, and reduces voter welfare by decreasing the probability of electing the efficient candidate and increasing aggregate turnout costs. Media entry can systematically shift the election outcome and voter turnout in either direction, yet it unambiguously improves information transmission and voter welfare. The impact of other ways to strengthen media competition – such as increased polarization and prevention of collusion – critically depends on whether media have commitment power; if not, they can worsen information transmission and voter welfare.

JEL Codes: D72, D82, D83

Keywords: media bias, voting, Poisson games, media competition, commitment

*This is the job market paper of the first author, Junze Sun. We thank Sourav Bhattacharya, Aaron Kamm, Navin Kartik, David K. Levine, Zara Sharif and Joel Sobel for insightful comments at various stages. We also thank audiences in the 2017 NYU-CESS Political Science Conference, the 2017 ESA World meeting in UCSD, the 2018 CES Annual Conference in Hefei, the 2018 EEA-ESEM Conference in Cologne, and also seminar participants from University of Amsterdam, Tinbergen Institute, EUI, NYU-Abu Dhabi, and Hong Kong University of Science and Technology (HKUST), for helpful suggestions and comments. All errors are our own.

†CREED, University of Amsterdam and Tinbergen Institute. Email: J.Sun@uva.nl.

‡CREED, University of Amsterdam and Tinbergen Institute, and Robert Schumann Center for Advanced Studies, EUI, Florence. Email: A.J.H.C.Schram@uva.nl.

§CREED, University of Amsterdam and Tinbergen Institute. Email: R.Sloof@uva.nl.

1 Introduction

Mass media are important information sources that voters rely on to make political decisions. However, media may have their own agendas and be biased in news reporting.[1] Over the past few decades, empirical studies have indeed confirmed the pervasiveness of media bias and documented a significant impact of media bias on voting behavior.[2] For example, DellaVigna and Kaplan (2007) find that exposure to Fox News increased the Republican vote share and voter turnout in the 2000 U.S. presidential election. Enikolopov et al. (2011) find that exposure to NTV, the only national TV channel independent of the Russian government in the late 1990s, increased the vote share for major opposition parties and decreased voter turnout in the 1999 parliamentary election. At the same time, empirical evidence shows that media entry may either increase or decrease voter turnout, but it does not systematically affect party vote shares.[3] Why does media exposure sometimes increase voter turnout while in other cases it does the opposite? Why are media outlets able to affect party vote shares in a particular election, but unable to affect it systematically in the long run? Importantly, even if media cannot systematically affect party vote shares, does this imply that media outlets cannot systematically manipulate election outcomes as well? These are crucial questions for understanding the role of media in modern democracies. Yet, theories providing unified answers are still absent.

We take a step towards filling this gap by developing a tractable theory to explain how biased media affect election outcomes, voter turnout and welfare.[4] Methodologically, we combine the pivotal voter model with Poisson games (Myerson, 2000) and *analytically* derive voting equilibria in large elections. Subsequently, we model the interaction between media and voters via a sender-receiver game, and derive communication equilibria under both cheap

[1]The literature identifies various sources of media bias. From the supply side, media bias could come from the biased motives of journalists and editors (Baron, 2006; Sobbrio, 2014), capture by politicians (Besley and Prat, 2006; McMillan and Zoido, 2004) as well as financial advertising (Ellman and Germano, 2009). From the demand side, media bias can arise from catering readers' preferences for like-minded news (Mullainathan and Shleifer, 2005) or from media outlets' reputation concerns (Gentzkow and Shapiro, 2006). See Prat and Stromberg (2013) and Gentzkow et al. (2015) for comprehensive reviews.

[2]Groseclose and Milyo (2005), Gentzkow and Shapiro (2010), Larcinese et al. (2011), Puglisi and Snyder (2011), and Puglisi and Snyder (2015a) provide empirical evidence and measures for media bias in the US. See Puglisi and Snyder (2015b) for a comprehensive review.

[3]Gentzkow et al. (2011) and Drago et al. (2014) find evidence that newspaper entry increases voter turnout in the U.S. and in Italy, respectively. In contrast, Cagé (2017) find evidence that media entry has a negative impact on voter turnout in France. Gentzkow et al. (2011) find that newspaper entry/exit does not systematically influence party vote share in the U.S. during 1868–1928.

[4]Our analysis primarily concerns traditional, large-scale media like newspapers and TV-channels. Obviously, social media play an important role in modern elections (Smith and Anderson, 2018). Yet, traditional media are still major players (Allcott and Gentzkow, 2017). The ongoing rhetoric of Donald Trump vis-a-vis media like CNN and the New York Times is evident for the importance he attributes to them.

talk (Crawford and Sobel, 1982) and Bayesian persuasion (Kamenica and Gentzkow, 2011).

We consider an election (under simple majority rule) with two candidates, A and B, who differ along the *quality* dimension. The quality of a candidate can be interpreted as her integrity, valence, or executive ability, attributes that all voters equally appreciate. Apart from caring about quality, voters also have their idiosyncratic ideological preferences over candidates. In this way, it is possible for a voter to prefer the (in his view) ideologically less-favored candidate if her quality is sufficiently superior to that of the other candidate. Votes can vote for either candidate or abstain, but casting a vote is costly; these costs are private information. We assume the election is symmetric in the sense that *a priori* (i) the qualities of both candidates are identical, and (ii) voters' ideological preferences do not favor either candidate over the other. Voters know precisely their private ideological preferences, but do not observe candidates' qualities. Media outlets precisely observe this information and communicate it to voters by disseminating public news.

The first insight from our analysis is that public information about candidates' qualities allows voters to (i) assess which candidate is more appealing, and (ii) gauge the closeness of the election. In large elections, the former determines party vote shares and the election outcome, whereas the latter drives voter turnout. With a higher perceived relative quality, a candidate is able to 1) convince marginal voters to support her, and 2) increase the willingness-to-vote of her own supporters, relative to her opponent's supporters. These increase the candidate's expected vote share and winning chances. On the other hand, with a large perceived quality difference, the election is supposed to end up in a landslide victory because the quality-superior candidate is expected to gain a substantially larger share of votes than her opponent. In this situation, voter turnout is low because the chances of casting a pivotal vote are low and voters have little incentive to cast costly votes. Therefore, the perceived quality difference is negatively associated with the closeness of elections. Voter turnout is thus a decreasing function of the perceived quality difference, in line with the "competition effect" (Levine and Palfrey, 2007). Moreover, we show that in large and moderately polarized elections this relationship is also *convex*. As explained below, this convexity plays a key role in deriving unambiguous comparative statics predictions on how media bias and entry affect voter turnout.

Our second insight is that media outlets can systematically manipulate election outcomes and voter turnout by strategically providing public information even if voters are Bayesian. Depending on the information received, voters may either update their posterior beliefs towards candidate A having a superior quality or the opposite. But on average their posterior beliefs should – by Bayes' rule – average back to the prior (i.e., no quality differences). Therefore, manipulating public information cannot systematically shift rational voters' posterior beliefs. However, the election outcome is generically *nonlinear* in voters' posterior beliefs; it is in some regions much more sensitive to a shift in voters' beliefs than in others. For this reason, the ex-post election outcomes need not average back to the outcome under voters' common prior. This nonlinearity gives room for media outlets to systematically manipulate election outcomes in a rational voter framework; they may strategically disclose information to shift voters' posterior beliefs in regions with different sensitivities (Kamenica and Gentzkow, 2011). Importantly, this opportunity arises even when party vote shares can hardly be systematically affected. In the same spirit, we also show that voter turnout is generically nonlinear in voters' posterior expectations of candidates' quality difference, which reveals the closeness of elections. By affecting the distribution of voters' perceived closeness, media exposure may also systematically influence voter turnout.

Building on these insights, we derive precise comparative statics results on how media bias and entry affect election outcomes, voter turnout and welfare. We assume that each media outlet has a commonly known ideological position, which may not be aligned with voters. An *unbiased* media outlet always prefers the candidate with a higher quality. A *biased* media outlet, however, may prefer a certain candidate even if her quality is inferior. When media outlets communicate to voters via cheap talk, we show that in equilibrium their news reporting strategies take a simple cutoff structure: endorsing candidate A (B) only if the relative quality of A is above (below) a certain threshold. This threshold is decreasing in an outlet's bias towards candidate A. In this way, media outlets partition the state (i.e., candidate A's relative quality) space into adjacent and disjoint intervals.

We study the impact of media bias on elections in the context with one media outlet. We show that, from the ex-ante perspective, the relationship between media bias (say, towards candidate A) and the election outcome is non-monotonic. On one hand, an outlet with a stronger bias is *a priori* more likely to endorse candidate A; this effect *per se* increases A's electoral prospects. On the other hand, this outlet's endorsement for candidate A becomes less credible, while its endorsement for B becomes more credible, as media bias increases. These effects *per se* reduce A's winning chances. The net effect, therefore, depends on which effect dominates. For a small degree of media bias, the first effect dominates and a marginal increase of bias increases candidate A's winning probability. For a sufficiently large degree of bias the opposite applies. Under all levels of biases, the outlet is able to systematically increase the winning probability of candidate A, whom it favors ex-ante. The probability of electing the quality-superior candidate, however, decreases in media bias.

The impact of media bias on voter turnout follows from what an endorsement reveals about the closeness of elections. With a strong media bias, an outlet's endorsement for A (e.g., a Republican endorsement from the Fox News) is incredible and signals a small quality difference between candidates (if any). This in turn implies that the election may end up in a close race, leading to a high voter turnout. In contrast, this outlet's endorsement for candidate B (e.g., a Democratic endorsement from the Fox News) credibly signals the superiority of B and a substantial quality difference. This in turn implies that the election may end up in a landslide victory of B and voter turnout is thus low. Therefore, as media bias increases voter turnout is higher (lower) conditional on an endorsement for candidate A (B). Ex-ante, the expected voter turnout increases in media bias unambiguously for two reasons. First, an increase in media bias leads to a spread of voters' posterior expectations of quality difference. This effect would increase voter turnout if on average the expected quality differences remain constant, due to the convexity of voter turnout as discussed above. Second, with a stronger bias the media outlet is *a priori* more likely to endorse candidate A, and thus induce a low expected quality difference. Since voter turnout is negatively associated with the expected quality difference, this effect further boosts voter turnout. As a result, an increase in media bias reduces voter welfare by not only decreasing the probability of electing the quality-superior candidate, but also increasing aggregate turnout costs.

We study the impact of media entry by introducing a second media outlet. Under the cutoff endorsement strategy, media entry essentially refines the information partition induced by the existing outlet. Hence, more information must be transmitted to voters after media entry. We show that, and precisely identify conditions under which, media entry can systematically shift the election outcome as well as voter turnout in either direction. Our results may thus reconcile the mixed empirical evidence on the relationship between media entry and voter turnout (Gentzkow et al., 2011; Drago et al., 2014; Falck et al., 2014; Cagé, 2017). To the best of our knowledge, Piolatto and Schuett (2015) is the only other paper that may explain these mixed effects theoretically. However, the mechanism in our theory is completely different from theirs, which is based on the ethical voter framework (Coate and Conlin, 2004; Feddersen and Sandroni, 2006a,b).[5] From the welfare perspective, we show that media entry weakly increases voter welfare as the electorate size grows without bound. This confirms, in an environment where media outlets communicate to voters via cheap talk, the conventional wisdom that media competition improves information transmission and voter welfare (Besley and Prat, 2006; Gentzkow and Shapiro, 2008; Prat, 2018).

We finally explore to what extent this conventional wisdom can be generalized, by varying both the communication protocol used by media outlets and the notion of media competi-

[5]In Piolatto and Schuett (2015), media entry affects voter turnout through influencing independent and partisan voters' decisions to acquire information. They show that information increases the turnout rate of independent voters but decreases the turnout rate of partisan voters in expectation. The net effect is thus ambiguous and critically dependent on the degree of polarization of the electorate. In our model, however, voters are always equally informed and we consider a richer information environment.

tion. In all previous analyses we assumed that media outlets communicate to voters via cheap talk and cannot commit to any specific news reporting strategy. This assumption seems plausible in weak democratic regimes where the state government (or political interest groups) have strong control (e.g., censorship and bribing) over the media outlets (McMillan and Zoido, 2004; Besley and Prat, 2006; Prat and Stromberg, 2013; Enikolopov and Petrova, 2015). However, under other circumstances media outlets may be able to commit because of legal regulations or reputation concerns (Gentzkow and Shapiro, 2006). We study the latter situation under the Bayesian persuasion framework (Gentzkow and Kamenica, 2016, 2017) and find that commitment plays a key role in shaping information transmission in equilibrium. In particular, "conflicting states", where at least two media outlets' partisan preferences disagree, are generically revealed to voters if media commitment is possible while they may not be revealed without commitment.

Analogous to Gentzkow and Kamenica (2016) we consider three different notions of media competition: 1) increasing the number of media outlets, 2) increasing media polarization, and 3) preventing media collusion. In line with Gentzkow and Kamenica (2016), all three ways of strengthening media competition improve information transmission and voter welfare in large elections when commitment is possible. However, in the absence of commitment the opposite may apply under either increased polarization or prevention of collusion. With regard to media entry, both cheap talk and Bayesian persuasion models agree that increasing the number of media outlets improves information transmission and voter welfare in large elections. However, these models yield drastically different implications regarding what types of media outlet should be introduced to maximize these benefits. Taken together, our results suggest that increasing media competition may have perverse impacts on both information transmission and voter welfare in the absence of media commitment, and thus highlight the role of media commitment in the policy debate of media regulation.

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 introduces the model. Section 4 derives voting equilibria in large elections when candidates' qualities are common knowledge. Based on this it also explores the extent to which manipulating public information environment can systematically influence election outcomes. Section 5 presents comparative statics results concerning the influence of media bias and media entry on elections under a cutoff information environment. Section 6 discusses how media commitment affects information transmission and voter welfare. Section 7 concludes and suggests some avenues for future research.

2 Related literature

Our topic relates to a large body of empirical literature studying the electoral impact of mass media. One branch of this literature exploits reasonably exogenous cross-sectional variations in media exposure to identify media's influence in particular elections (DellaVigna and Kaplan, 2007; Gerber et al., 2009; Chiang and Knight, 2011; Enikolopov et al., 2011; Durante and Knight, 2012; Adena et al., 2015). These studies find that exposure to a biased media outlet can 1) either increase or decrease voter turnout, and 2) shift the party vote share and voters' voting intentions in favor of the endorsed candidate, but the strengths of these effects depend on the magnitude of the bias. A second branch of this literature exploits local-level discrete changes in the number of media outlets over a reasonably long time period to identify the systematic impact of media entry on voters' behavior (Gentzkow, 2006; Gentzkow et al., 2011; Drago et al., 2014; Falck et al., 2014; Gavazza et al., 2015; Cagé, 2017; Ellingsen et al., 2018). These studies find mixed evidence on how media entry systematically affects voter turnout. Our theory produces comparative statics predictions that could potentially reconcile these mixed findings. Our paper also relates to theoretical work studying the political impact of media bias (Bernhardt et al., 2008; Duggan and Martinelli, 2011; Piolatto and Schuett, 2015). Unlike these studies, our theory builds on the pivotal voter framework and identifies qualitatively different channels of media influence. A separate strand of literature focuses on the impact of media on electoral competition and political accountability in a setting where candidates' policy choices are endogenous (Besley and Prat, 2006; Chan and Suen, 2008, 2009; Strömberg, 2004; Snyder and Strömberg, 2010). In contrast, we focus only on the selection of candidates.

Our methodology relates to three strands of literature. First, our theoretical framework combines the pivotal voter model (Palfrey and Rosenthal, 1985; Ledyard, 1984) with Poisson games (Myerson, 1998, 2000). We contribute to this literature in two ways. First, we derive analytical voting equilibria in large Poisson elections. Second, we show that in large elections voter turnout varies convexly with the expected quality difference between candidates, which implies a diminishing marginal impact of the "competition effect".

Second, our analysis of media outlets' information transmission without commitment relates to the literature on multi-sender cheap talk games (Gilligan and Krehbiel, 1989; Krishna and Morgan, 2001a,b; Battaglini, 2002; Chan and Suen, 2009). This literature primarily focuses on the payoff structure in Crawford and Sobel (1982) and aims at identifying conditions under which fully revealing equilibria exist. We differ by studying a different setting in which senders' payoffs are monotonic in the receiver's action, conditional on the realized state. We construct the most informative equilibria with an arbitrary number of senders.[6]

Finally, our equilibrium analysis with media having commitment power relates to a recent literature on Bayesian persuasion (Rayo and Segal, 2010; Kamenica and Gentzkow, 2011; Gentzkow and Kamenica, 2016, 2017). Our analysis with multiple media outlets is a special case of Gentzkow and Kamenica (2016). But since we focus on a specific payoff structure, we derive sharper results on how media outlets' biases shape equilibrium information transmission. Our analysis with a single outlet relates to several papers. Alonso and Câmara (2016) study how an information designer can strategically design a policy experiment to persuade voters under super-majority rule and compulsory voting. We instead focus on elections under simple majority rule and assume voting to be voluntary and costly. Kolotilin et al. (2017) and Ginzburg (2017) consider persuasion problems in a similar linear environment but their assumptions on sender preferences are different from ours.

3 The model

Candidates. We consider an election with two candidates, A and B. A commonly valued state k is drawn from a commonly known prior $F(\cdot)$ on $[-1, 1]$. One can interpret k as the relative quality of candidate A. If $k = 0$, the quality of both candidates are identical. If instead $k > (<)0$, than candidate A has a higher (lower) quality than candidate B. Unless explicitly stated otherwise, k is assumed to be unobservable to voters.

Electorate. The electorate consists of a set of voters, N. Following Myerson (1998, 2000), we assume that the electorate size $|N|$ follows a Poisson distribution with mean $n > 0$ (i.e., $Pr[|N| = x] = e^{-n}\frac{n^x}{x!}$ for all non-negative integers x).[7] Each voter can choose an action from $\{A, B, O\}$, representing voting for candidate A, candidate B, and abstaining, respectively. Voter $i \in N$ is characterized by a pair $(v_i, c_i) \in [-\delta, \delta] \times [0, C]$, with $\delta > 1$ and $C > 0$. v_i represents voter i's private ideological preferences. c_i represents her private voting costs, which are incurred only if she casts a vote. We normalize voter i's payoffs, excluding voting costs, to 0 if candidate B is elected and let voter i's payoffs be $k+v_i$ if A is elected. Therefore, voter i prefers candidate A to B if and only if $k + v_i \geq 0$, and $|k + v_i|$ measures voter i's

[6] Rubanov (2014) and Lu (2017) study robust informative equilibria in the presence of multiple senders under Crawford and Sobel type preferences. Ravid and Doron (2017) study a cheap talk game with one sender who has a "transparent motive"; the sender's preferences over the receiver's actions are state-independent. This nests a special case of our model, where voters only hear from an extremely biased media outlet.

[7] $|N|$ denotes the cardinality of set N. As argued by Myerson (1998), in games with a large population it is more realistic to allow for uncertainty over the number of players. In the context of elections, it seems reasonable to assume that voters do not know the exact electorate size.

stake at the election. Formally, given k, v_i and c_i, voter i's utility function equals:[8]

$$u(k, v_i, c_i, a_i, \Omega) = (k + v_i) \cdot \mathbb{1}_{\Omega=A} - c_i \cdot \mathbb{1}_{a_i \neq O} \tag{1}$$

where $\Omega \in \{A, B\}$ indicates the winning candidate, $a_i \in \{A, B, O\}$ is voter i's action, and $\mathbb{1}_E$ is an indicator function that equals 1 (0) if event E is true (false). We assume that both v_i and c_i are independently and identically distributed across voters. v_i follows distribution $G(\cdot)$ on $[-\delta, \delta]$, and c_i is uniformly distributed on $[0, C]$. Both $G(\cdot)$ and C are common knowledge. Although we can derive the voting equilibria for general distributions $F(\cdot)$ and $G(\cdot)$, we impose two assumptions in the remainder of the paper to derive our main results.

Assumption 1 *(Symmetric election) Both $F(\cdot)$ and $G(\cdot)$ are symmetric distributions and admit positive density functions $f(\cdot)$ and $g(\cdot)$, respectively.*

Assumption 2 *(Moderate polarization). $\int_{-\delta}^{-1} G(v)dv \geq 0.04$.*

Assumption 1 states that the distributions of candidates' qualities and of voters' ideological preferences are symmetric. Therefore, *a priori* each candidate is equally likely to have a higher quality and the candidate with a superior quality is efficient in the sense of maximizing the electorate's ex-ante Utilitarian welfare. Assumption 2 specifies a strictly positive lower bound for the expected fraction of *partisan voters*, whose ideology $|v_i| > 1$ such that they prefer one candidate to other regardless of the realized qualities of candidates. Therefore, Assumption 2 implies a moderately polarized electorate, which seems plausible under the current political landscape in the U.S., where the degree of polarization has increased sharply in recent decades (Pew Research Center, 2017; Martin and Yurukoglu, 2017).

Media outlets and media bias. There is a finite set of media outlet(s), M. Each outlet $m \in M$ is precisely informed of k and communicates to voters by sending a public message $s_m \in S$. Unless explicitly stated otherwise, we assume $S = [-1, 1]$ so that the message space coincides with the state space. Media outlets' utility function has a similar structure as voters', except that media outlets have no voting costs and cannot cast any vote:

$$V(k, \chi_m, \Omega) = (k + \chi_m) \cdot \mathbb{1}_{\Omega=A} \tag{2}$$

This implies that outlet m strictly prefers candidate A (B) if and only if $k > (<) - \chi_m$. If $\chi_m = 0$, m is said to be *unbiased* because it always prefers the candidate with the higher

[8] For our analysis to apply, we only need to assume that voters' utility function is monotonically increasing in both k and v, and is additive separable with respect to both arguments. This setup nests the spatial voting model with commonly valued valence (Enelow and Hinich, 1984), which is widely used in the literature (Chiang and Knight, 2011; Durante and Knight, 2012).

quality. If $\chi_m > (<)0$, m is said to be *A(B)-biased* as it may prefer candidate A(B) to be elected even if it has a lower quality. For this reason, we refer to χ_m as *media bias*. We distinguish between two types of communication protocols, cheap talk (CT) and Bayesian persuasion (BP). In the former case media outlets cannot commit to any reporting strategy before observing the realized state k, while in the latter case they can. We primarily focus on CT in deriving the comparative statics of media influence on elections, and compare equilibrium information transmission behavior under CT and BP in Section 6.

Timing and equilibrium concept. The timing of the game is as follows:

1. (BP only) Each outlet $m \in M$ commits to a reporting strategy $\sigma_m : [-1, 1] \mapsto \Delta(S)$.

2. Nature determines the realizations of k, N and the profile of voter types $\{(v_i, c_i)\}_{i \in N}$.

3. Observing k, outlets $m \in M$ simultaneously send public messages $s_m \in S$.

4. Observing the message profile $\{s_m\}_{m \in M}$, voters simultaneously make their voting decisions (vote for A, for B, or abstain).

5. The winning candidate is determined by simple majority rule, with ties broken by a fair coin toss. All payoffs then realize.

A media outlet's reporting strategy maps the realized state to a probability distribution of the message space, $\sigma_m : [-1, 1] \mapsto \Delta(S)$. A voter's voting strategy maps the observed message profile and the realized private type to a probability distribution of possible actions $q : S^{|M|} \times [-\delta, \delta] \times [0, C] \mapsto \Delta(\{A, B, O\})$.[9] Because this is a dynamic game of incomplete information, we derive the Perfect Bayesian Equilibrium (PBE). In a PBE, voters' strategies are best responses to their posterior beliefs conditional on the observed message profile. Media outlets' strategies are best responses to voters' strategies. Voters' posterior beliefs are formed by Bayes' rule whenever possible. Without loss of generality, we focus on type-symmetric PBE where voters of the same type (v_i, c_i) adopt the same voting strategy in equilibrium.[10] Although our analysis generally applies under Assumption 1 and 2, we use Example 1 below to graphically illustrate our main results.

Example 1 *$n = 200$, $k \in U[-1, 1]$, $v_i \sim U[-1.5, 1.5]$ and $c_i \sim U[0, 0.2]$ for all $i \in N$.*

[9] In BP, the voting strategy is also a function of media outlets' reporting strategy profile $\{\sigma_m(\cdot)\}_{m \in M}$.

[10] Myerson (1998) (page 391) shows that all equilibria are type-symmetric under population uncertainty.

4 Voting equilibria in large elections

In Section 4.1, we derive the equilibrium aggregate voting behavior and election outcomes in large elections (i.e., $n \to \infty$) assuming that k is common knowledge. In section 4.2 we assume k is unobservable to voters and analyze the extent to which these aggregate outcomes can be systematically influenced by manipulating voters' information about k. Detailed derivations and proofs are relegated to Appendix A.

4.1 Voting equilibria when k is common knowledge

If k is common knowledge, it is clear from voters' utility function (1) that any voter i with $v_i > (<) - k$ strictly prefers candidate A (B) to be elected, and will vote for candidate A (B) if she votes. Hence, the only strategic decision is whether or not to cast a vote for the preferred candidate. To make this decision a voter has to compare the expected benefits to the costs of casting a vote, and to vote only if the former outweigh the latter. This implies that the optimal turnout strategy takes a cutoff form; vote for the preferred candidate only if the voting costs c_i are below the expected benefits (Palfrey and Rosenthal, 1985; Myerson, 2000). The expected benefits are the product of (i) the individual stake at the election, $|k + v_i|$, and (ii) the probability of casting a pivotal vote. Hence, voter i votes only if:

$$c_i \leq |k + v_i| \cdot Pr[PivA|k, n], \text{ if } v_i > -k$$
$$c_i \leq |k + v_i| \cdot Pr[PivB|k, n], \text{ if } v_i < -k$$

where $Pr[PivA(B)|k, n]$ denotes the probability that a single vote for candidate A (B) is pivotal, conditional on k and n. These pivotal probabilities are endogenously determined in equilibrium. Consistent with common intuition, in equilibrium both pivotal probabilities approach zero as $n \to \infty$. The expected benefits of voting thus also decrease in n and must lie below C for sufficiently large n. In that case voter's turnout probability equals $\frac{|k+v_i| \cdot Pr[PivA(B)|k,n]}{C}$ for $v_i > (<) - k$, since $c_i \sim U[0, C]$.

Let $q_A(k, n)$ and $q_B(k, n)$ be the probability that a randomly sampled voter votes for A and B, respectively. $q_A(k, n)$ and $q_B(k, n)$ can be obtained by integrating the turnout probabilities of "A-supporters" (voters with $v_i > -k$) and "B-supporters" (voters with $v_i < -k$), respectively. For sufficiently large n we have

$$q_A(k, n) = \int_{-k}^{\delta} \frac{|k + v| \cdot Pr[PivA|k, n]}{C} dG(v) \equiv \frac{\alpha(k)}{C} \cdot Pr[PivA|k, n] \tag{3}$$

$$q_B(k, n) = \int_{-\delta}^{-k} \frac{|k + v| \cdot Pr[PivB|k, n]}{C} dG(v) \equiv \frac{\beta(k)}{C} \cdot Pr[PivB|k, n] \tag{4}$$

where $\alpha(k) \equiv \int_{-k}^{\delta} |k+v| dG(v) = k+\int_{-\delta}^{-k} G(k)dk$ and $\beta(k) \equiv \int_{-\delta}^{-k} |k+v| dG(v) = \int_{-\delta}^{-k} G(k)dk$.[11] Both $\alpha(k)$ and $\beta(k)$ have straightforward economic interpretations. $\alpha(k)$ represents the *expected stake* of "A supporters" and $\beta(k)$ represents the *expected stake* of "B supporters". It is clear from the expressions that $\alpha(k)$ increases in k whereas $\beta(k)$ decreases in k. The influence of a rise in k on $\alpha(k)$ and $\beta(k)$ can be decomposed into three effects. First, it convinces the marginal B-supporters to switch to A. Second, it strengthens the expected stakes of the infra-marginal A-supporters. Third, it reduces the expected stakes of the infra-marginal B-supporters. Under Assumption 1 it holds that $\alpha(k) - \beta(k) = k$. Hence, k directly measures of the expected stake difference between A- and B-supporters.

As detailed in Appendix A.1, both $Pr[PivA|k, n]$ and $Pr[PivB|k, n]$ are functions of $q_A(k, n)$ and $q_B(k, n)$, conditional on k and n. Therefore, the above system of equations (3) and (4) form a self-mapping and a voting equilibrium is defined by its solution, denoted by $(q_A^*(k, n), q_B^*(k, n))$, given k and n. In Appendix A.1 we derive analytical approximations for $q_A^*(k, n)$ and $q_B^*(k, n)$ in large elections and they equal:[12]

$$q_A^*(k,n) \approx \begin{cases} \left(\frac{\alpha(0)}{2\sqrt{\pi}C}\right)^{\frac{2}{3}} \cdot \frac{1}{\sqrt[3]{n}}, & \text{if } k = 0 \\ \left(\frac{\mu(k)}{1-\mu(k)}\right)^2 \cdot \frac{\ln n}{n}, & \text{if } k \neq 0 \end{cases} \tag{5}$$

$$q_B^*(k,n) \approx \begin{cases} \left(\frac{\alpha(0)}{2\sqrt{\pi}C}\right)^{\frac{2}{3}} \cdot \frac{1}{\sqrt[3]{n}}, & \text{if } k = 0 \\ \left(\frac{1}{1-\mu(k)}\right)^2 \cdot \frac{\ln n}{n}, & \text{if } k \neq 0 \end{cases} \tag{6}$$

where $\mu(k) \equiv \sqrt[3]{\frac{\alpha(k)}{\beta(k)}}$ is the cubic root of the expected stakes ratio. Let $VS(k, n) \equiv \frac{q_A^*(k,n)}{q_A^*(k,n)+q_B^*(k,n)}$, $\pi(k, n) \equiv Pr[\text{A wins}|k, n]$ and $T(k, n) \equiv q_A^*(k, n)+q_B^*(k, n)$ be the equilibrium expected vote share, winning probability of candidate A, and the expected voter turnout, conditional on k and n, respectively. Let $W(k, n)$ be the expected utility of a randomly selected voter in equilibrium, conditional on k and n. Theorem 1 precisely characterizes equilibrium voting behavior, election outcomes and voter welfare in large elections.

Theorem 1 *Suppose Assumption 1 holds, for all $k \in [-1, 1]$ there are:*

1. $VS(k, n) \approx \frac{\mu^2(k)}{1+\mu^2(k)}$, *where $\mu(k)$ increases in k, $\mu(0) = 1$ and $\mu(-k) = \frac{1}{\mu(k)}$.*

2. $\pi(k, n) \approx \begin{cases} 1 - C\psi(k)\frac{\ln n}{n}, & \text{if } k > 0 \\ \frac{1}{2}, & \text{if } k = 0 \\ -C\psi(k)\frac{\ln n}{n}, & \text{if } k < 0 \end{cases}$, *where $\psi(k) \equiv \left(\frac{1}{\sqrt[3]{\alpha(k)} - \sqrt[3]{\beta(k)}}\right)^3$ for $k \neq 0$ and*

[11] The second equalities in $\alpha(k)$ and $\beta(k)$ are derived in Appendix A.1, under Assumption 1.

[12] Throughout the paper, the expression $x_n \approx y_n$ denotes $\lim_{n\to\infty} \frac{x_n}{y_n} = 1$.

(a) $\psi(-k) = -\psi(k)$ *for all* $k \in [-1, 0) \cup (0, 1]$*, and* $\lim_{k \downarrow 0} \psi(k) = +\infty$.

(b) $\psi(k)$ *is decreasing and convex on* $(0, 1]$.

3. $T(k,n) \approx \begin{cases} 2\left(\frac{\alpha(0)}{2\sqrt{\pi}C}\right)^{\frac{2}{3}} \frac{1}{\sqrt[3]{n}}, & \text{if } k = 0 \\ \gamma(k)\frac{\ln n}{n}, & \text{if } k \neq 0 \end{cases}$*, where* $\gamma(k) \equiv \frac{1+\mu^2(k)}{(1-\mu(k))^2}$ *for* $k \neq 0$ *and*

(a) $\gamma(k) = \gamma(-k)$ *for all* $k \in (0, 1]$*, and* $\lim_{k \to 0} \gamma(k) = +\infty$.

(b) $\gamma(k)$ *is decreasing in* k *on* $(0, 1]$.

(c) *if Assumption 2 holds, then* $\gamma(k)$ *is convex on* $(0, 1]$ *for all* $G(\cdot)$.

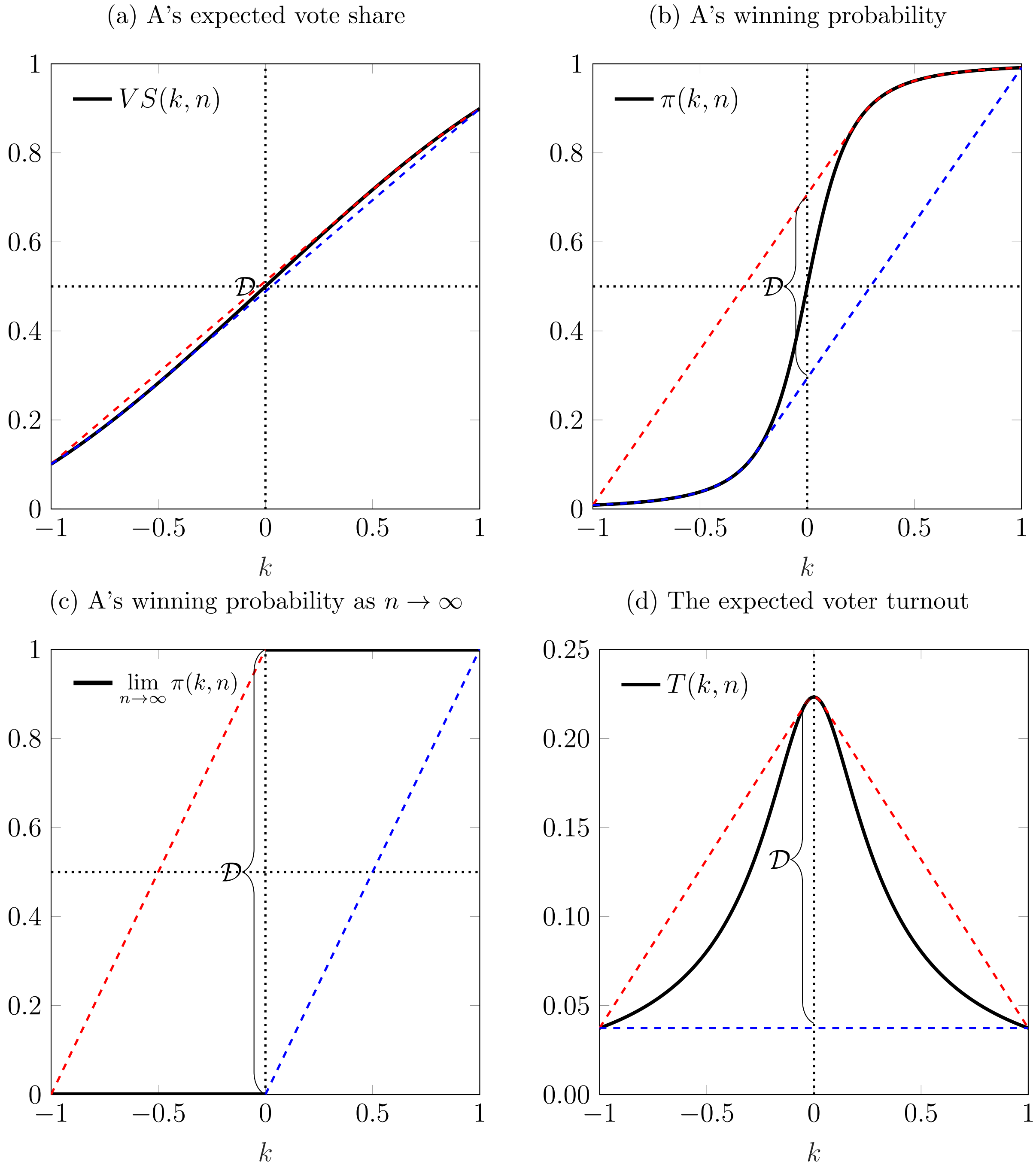
4. $\lim_{n \to \infty} W(k, n) = \max\{k, 0\}$.

Proof. See Appendix A.2. ■

Theorem 1 is graphically illustrated by Figure 1a to 1d, under the model parameters from Example 1. Figure 1a and Figure 1b (black solid lines) depict candidate A's equilibrium vote share and winning probability as functions of k, respectively. As discussed above, with a higher k candidate A is able to (i) convince marginal voters to switch to her, (ii) increase (reduce) the aggregate stakes of her own (opponent's) supporters. For these reasons, both candidate A's expected vote share and winning probability increase in k (Theorem 1.1 and 1.2). If two candidates have identical qualities (i.e., $k = 0$), the aggregate stakes of their supporters are equal, due to the symmetry of ideological preferences. Hence, either candidate's expected vote share and winning probability equals 50%. For positive values of k candidate's A electoral prospects improve, while for negative values of k they deteriorate. As illustrated in Figure 1b, in large elections candidate A's winning probability is an *S-shaped* function of k; it increases in k convexly for $k < 0$ and concavely for $k > 0$ (Theorem 1.2). Under simple majority rule, candidate A only requires a vote share exceeding 50% to assure a victory. Therefore, thanks to the law of large numbers, A's winning probability converges to a step function as $n \to \infty$: $\lim_{n \to \infty} \pi(k, n) = 1(0)$ if and only if $k > (<)0$ (Figure 1c).

Figure 1d (black solid line) depicts the expected voter turnout in equilibrium, $T(k, n)$, as a function of k. $T(k, n)$ is symmetric around $k = 0$ and decreases in quality difference $|k|$ (Theorem 1.3). If $k = 0$, the election is expected to end up in a close race since the two candidates are expected to get equal vote shares. As a result, chances of being pivotal are high and voters have the strongest incentive to cast costly votes. Voter turnout is therefore highest. If instead $|k| \gg 0$, the election is expected to end in a landslide; the quality-superior candidate is expected to get a substantially higher vote share than her opponent. In this case chances of being pivotal are low and, as a result, voter turnout will be low. The negative relationship between the closeness of an election and turnout is known

Figure 1: Equilibrium voting behavior and election outcomes as functions of k



Note: In Panel (a), $VS(k,n)$ denote the equilibrium vote share for candidate A conditional on k, n. In Panel (b) and (c), $\pi(k,n)$ denotes the winning probability of candidate A, conditional on k, n. In Panel (d), $T(k,n)$ denotes the expected voter turnout conditional on k, n. The red (blue) curves in all panels are the upper-concave (lower-convex) envelops of the depicted outcome variable. The distance between the upper-concave and lower-convex envelops, evaluated at the prior mean $k = 0$, is denoted by $\mathcal{D}$. In each panel, $\mathcal{D}$ provides an upper bound on the extent to which the outcome variable can be systematically influence by manipulating public information environments. Model parameters are taken from Example 1.

as the "competition effect" (Levine and Palfrey, 2007).[13] Since the quality difference $|k|$ is negatively associated with the closeness of the election, a higher $|k|$ thus lowers voter turnout through the competition effect.

Theorem 1.3 also claims that, if Assumption 2 holds, the asymptotic voter turnout is generically a *convex* function of $|k|$.[14] This convexity property indicates that the competition effect itself exhibits a diminishing marginal impact; a marginal increase in quality difference has a larger (negative) impact on turnout when the quality difference is smaller. As explained below, this convexity property plays an important role in deriving unambiguous comparative static predictions on how media bias affects voter turnout.

Finally, Theorem 1.4 follows from the fact that with k common knowledge, the election outcome is *asymptotically efficient*; the quality-superior candidate is elected with probability approaching 1 as $n \to \infty$. Moreover, Theorem 1.3 suggests that in large elections the turnout rate vanishes to zero and thus so do the expected voting costs. Recall that k reflects the relative quality of candidate A, i.e. the quality of candidate B is normalized to zero. This explains the expression for the asymptotic welfare in the final part.[15]

4.2 Public information environment and its systematic influence

In this section we assume k is unobservable to voters and explore the extent to which voters' behavior and the election outcome can be systematically influenced by manipulating voters' information about k. Because voters' utility function is linear in k (see (1)), any public information, denoted by I, affects voters' behavior only through its impact on $E[k|I]$, the conditional expectation of k. Consequently, voters behave as if $E[k|I]$ is common knowledge and the voting equilibrium follows immediately from Theorem 1 by replacing k with $E[k|I]$. Therefore, it is convenient to define a *public information environment* as a probability distribution of posterior means that is *Bayes-plausible* (cf. Kamenica and Gentzkow (2011)); the average posterior beliefs must equal the prior $F(\cdot)$. This in turn implies that voters' average posterior expectations of k must equal the prior mean, that is, $k = 0$.

Below we construct a sensitivity measurement to gauge the extent to which the electoral

[13] In fact, our results are more general than the original concept of the competition effect proposed by Palfrey and Rosenthal (1985) and Levine and Palfrey (2007). In their models the individual stakes are fixed and homogeneous; each voter receives a fixed reward if her preferred candidate is elected. So the closeness of an election depends only on the relative number of voters supporting each candidate. In our model the closeness of an election is not only determined by this extensive margin, but also by the average stakes of voters supporting each candidate (an intensive margin), similar to Krishna and Morgan (2011, 2015).

[14] In Figure 1d (with $n = 200$), voter turnout decreases in $|k|$ concavely for $|k|$ close to 0. This is an artifact of the finite electorate size. As $n \to \infty$, the "concave" region vanishes.

[15] Note that, $W(k, n)$, the asymptotic welfare is defined in relative terms; it is relative to a hypothetical scenario where (i) candidate B is elected with probability 1 regardless of k, and (ii) all voters abstain so that the aggregate turnout costs are precisely 0.

outcomes can be systematically influenced by manipulating the public information environment. Let Γ denote the set of all probability distributions of posterior means that average back to 0, the prior mean. Because we do not require elements in Γ to be Bayes-plausible, the set of all public information environments is a proper subset of Γ. Let $\xi(k,n)$ denote the outcome variable of interest, with $\xi \in \{VS, \pi, T\}$. We define:

$$\mathcal{D}(\xi(k,n)) \equiv \max_{\tau\in\Gamma} E_\tau[\xi(k,n)] - \min_{\tau\in\Gamma} E_\tau[\xi(k,n)] \tag{7}$$

In words, $\max_{\tau\in\Gamma} E_\tau[\xi(k,n)]$ is the highest possible expectation of $\xi(k,n)$ that can be induced by a distribution of posterior means that average back to 0, and $\min_{\tau\in\Gamma} E_\tau[\xi(k,n)]$ the lowest. Because these boundaries may be obtained for elements of Γ that are not Bayes-plausible, $\mathcal{D}(\xi(k,n))$ provides an upper bound on how much the ex-ante expectation of $\xi(k,n)$ can be systematically influenced by the public information environment.

Similar to the concavification technology developed in Kamenica and Gentzkow (2011), $\mathcal{D}(\xi(k,n))$ can be straightforwardly identified in Figures 1a to 1d. The red dashed curves in these figures depict the concave closure (or upper concave envelope) of $\xi(k,n)$, while the blue dashed ones reflect the convex closure (lower convex envelope) of $\xi(k,n)$.[16] By construction, $\mathcal{D}(\xi(k,n))$ equals the vertical distance between the blue and the red curve evaluated at the prior mean $k=0$; in these figures this distance is denoted by $\mathcal{D}$.[17]

It is evident from Figure 1a that the expected vote share can hardly be systematically manipulated by (strategically) providing public information, under the model parameters from Example 1. This is because $\mathcal{D}$ is almost negligible in magnitude: $VS_+(0,n) < 0.505$ and $VS_-(0,n) > 0.495$, such that $\mathcal{D} < 0.01$. In other words, no matter how the public information environment varies, *a priori* the expected vote share for any candidate cannot be systematically shifted by more than 1%. This is because, under the model parameters from Example 1, $VS(k,n)$ is almost linear in k; variations in the expected vote share are thus almost proportional to variations in voters' posterior expectations of k. The latter, however, cannot be systematically affected when voters are Bayesian.

The situation is entirely different for the election outcome as depicted in Figure 1b. In Figure 1b it holds that $\pi_+(0,n) = 0.7$, $\pi_-(0,n) = 0.3$ and $\mathcal{D} = 0.4$. Depending on the public information environment, candidate A's ex-ante winning probability can thus vary from around 0.3 to 0.7. This opens up the possibility to systematically influence a candidate's

[16] Formally, let $co(\xi(\cdot,n))$ denote the convex hull of the graph of $\xi(k,n)$: $\{(k,\xi(k,n))\}_{k\in[-1,1]}$. Then $\xi_+(k,n) \equiv \sup\{z|(k,z) \in co(\xi(\cdot,n))\}$ is the *upper-concave envelope* of function $\xi(k,n)$, and $\xi_-(k,n) \equiv \inf\{z|(k,z) \in co(\xi(\cdot,n))\}$ is the *lower-convex envelope* of function $\xi(k,n)$.

[17] In general, let $\Gamma(x)$ denote the set of distributions of posterior means that average back to the prior mean $x \in [-1,1]$. It holds that $\max_{\tau\in\Gamma(x)} E_\tau[\xi(k,n)] = \xi_+(x,n)$ and $\min_{\tau\in\Gamma(x)} E_\tau[\xi(k,n)] = \xi_-(x,n)$. Our result follows from $E_\tau[k] = 0$ for all $\tau \in \Gamma$, by definition of Γ.

winning probability by up to 0.4 through manipulating the public information environment. The higher sensitivity of the election outcome as compared to expected vote shares follows from the fact that $\pi(k,n)$ is generically a highly nonlinear function of k. In particular, given the S-shape of $\pi(k,n)$, candidate A's winning probability is much more sensitive to shifts in $E[k|I]$ when $E[k|I]$ is close to 0 than when $E[k|I]$ is away from 0. This gives the possibility to systematically increase a candidate's winning chances by strategically disclosing information and shifting voters' posterior expectations in regions with different sensitivity.

Opportunities to manipulate the election outcome are largest in large elections. As $n \to \infty$, $\mathcal{D}$ converges to one, as illustrated in Figure 1c. Note that as $n \to \infty$, $\pi(k,n)$ converges to a step function with threshold at the prior mean $k = 0$, under which voters are (in aggregate) indifferent between the two candidates. Therefore, the election outcome is extremely sensitive (insensitive) to shifts in voters' beliefs when $E[k|I] = (\neq)0$, as $n \to \infty$. Exploiting this feature, a media outlet can guarantee an arbitrarily large winning probability of candidate A by endorsing A if $k \in [-1+\varepsilon, 1]$ and endorsing B if $k \in [-1, -1+\varepsilon)$, with $\varepsilon > 0$ arbitrarily close to zero. By doing so, the endorsement for A provides very weak evidence for the superiority of A, yet it is sufficient to break the balance and assure victory of A as $n \to \infty$. Though an endorsement for B provides extraordinarily strong evidence for the superiority of B and assures a certain victory of B as $n \to \infty$, it is sent with vanishing probability as $\varepsilon \to 0$. Therefore, with $\varepsilon \to 0$, an endorsement for A is sent almost surely ex-ante, which in turns guarantees a certain victory of A as $n \to \infty$.

Finally, as with the election outcome, the expected voter turnout is also highly sensitive to the public information environment. In Figure 1d it holds that $T_+(0,n) = 0.22$, $T_-(0,n) = 0.04$ and $\mathcal{D} = 0.18$. Variations in public information environments may thus systematically affect voter turnout by up to 18%. Similar with $\pi(k,n)$, this follows because $T(k,n)$ is also generically nonlinear in k; it is much more sensitive to voters' posterior beliefs in some regions than in others. Taken together, the aforementioned observations provide a possible explanation for why media exposure may systematically affect voter turnout, yet it does not systematically affect party vote shares (Gentzkow et al., 2011; Drago et al., 2014).

5 The influence of media on elections

This section studies the influence of media on voter behavior and election outcomes under a cutoff information environment as explained in Section 5.1. Section 5.2 assumes $|M| = 1$ and explores the impact of media bias on elections. In Section 5.3 we consider an increase in $|M|$ from 1 to 2 to study the electoral impacts of media entry.

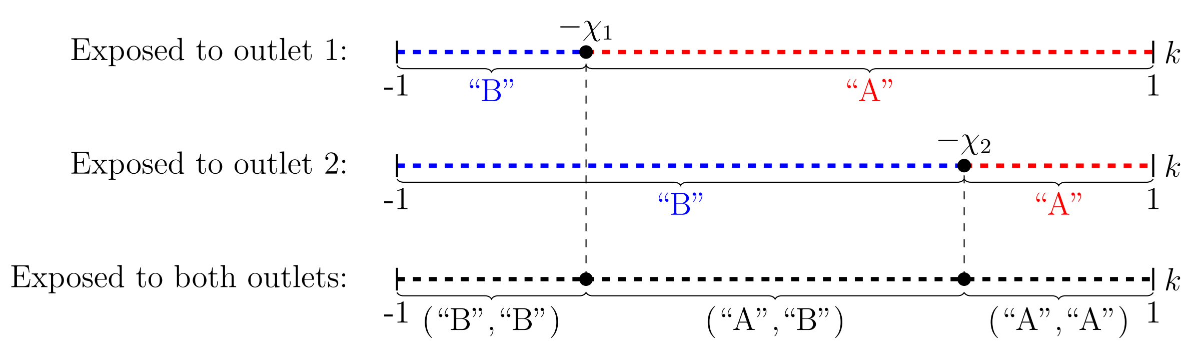
5.1 The cutoff information environment

By construction (cf. equation (2)), the preferences of media outlet $m \in M$ take a simple cutoff form: it prefers candidate A if and only if $k > -\chi_m$. This makes it natural to conjecture that media outlets use a simple cutoff endorsement strategy with a binary message space $S = \{\text{“A”}, \text{“B”}\}$:

$$\sigma_m(k) = \begin{cases} \text{“A”}, & \text{if } k > -\chi_m \\ \text{“B”}, & \text{if } k \leq -\chi_m \end{cases} \tag{8}$$

That is, each media outlet $m \in M$ sends message “A” (“B”) if $k > (\leq) -\chi_m$.[18] Message “A” (“B”) can be literally interpreted as an *endorsement* for candidate A (B). Let $\{\chi^{(m)}\}_{m \in M}$ be a descending permutation of media biases: $\chi^{(1)} \geq \chi^{(2)} \geq \cdots \geq \chi^{(|M|)}$. Under strategy (8), the $|M|$ media outlets jointly induce a “cutoff information environment”, which partitions the state space into adjacent and disjoint intervals with cutoff thresholds $-\chi^{(1)}, -\chi^{(2)}, ..., -\chi^{(|M|)}$. For notational convenience, we denote the information partition induced by the $|M|$ outlets by the set of their cutoff points: $\mathcal{P}_C(\chi_1, \cdots, \chi_{|M|}) = \{-\chi_1, \cdots, -\chi_{|M|}\}$.[19] Figure 2 illustrates the endorsement strategy (8) and the cutoff partition environment in the context where $|M| = 2$, $\chi_1 = 0.5$ and $\chi_2 = -0.5$.

Figure 2: The Cutoff Endorsement Strategy and Information Environment



Note: There are two media outlets with biases $\chi_1 = 0.5$ and $\chi_2 = -0.5$, respectively. The red (blue) segments denote states where media outlet $m \in \{1, 2\}$ will send message “A” (“B”). The pair (s_1, s_2) in the bottom line denotes the message combination the two media sent in each realized state. Since voters read from both outlets, voters' posterior knowledge is a coarse interval partition of the state space.

We focus on this cutoff information environment for two reasons. First, as shown in Section 6 this environment endogenously arises as the most information coordination-proof

[18] Under Assumption 1, $k = -\chi_m$ is a zero probability event and thus ties can be broken arbitrarily.

[19] The subscript C denotes a cutoff partition made up of adjacent and disjoint intervals, and it distinguishes this particular category from other partitions discussed in Section 6 below.

equilibrium outcome if media outlets communicate to the electorate via cheap talk.[20] Second, this environment is widely applied in theoretical as well as empirical papers studying the electoral impact of mass media (Chan and Suen, 2008; Chiang and Knight, 2011; Durante and Knight, 2012; Puglisi and Snyder, 2015a; Prat, 2018).

5.2 The influence of media bias

Let $|M| = 1$ and χ be the bias of this single media outlet. Without loss of generality we assume $\chi \in [0, 1]$, i.e., the media is either unbiased ($\chi = 0$) or A-biased ($\chi > 0$). We study both the *interim* and the *ex-ante* impacts of an increase in media bias on the voting equilibrium; the former impact is conditional on the realized message, while the latter is unconditional. Denote the induced posterior expectations as $k_A(\chi) \equiv E_F[k|k > -\chi]$ after message "A", and $k_B(\chi) \equiv E_F[k|k \leq -\chi]$ after message "B". Since voters' utility function is linear in k, media bias χ affects voting equilibria only through its impact on $k_A(\chi)$ and $k_B(\chi)$ and their relative weights. Lemma 1 summarizes this impact.

Lemma 1 *For all $\chi \in [0, 1]$ it holds that:*

1. $k_B(\chi) < 0 \leq k_A(\chi)$.

2. *both $k_A(\chi)$ and $k_B(\chi)$ decrease in χ.*

3. $E_{\mathcal{P}_C(\chi)}[k] = (1 - F(-\chi)) \cdot k_A(\chi) + F(-\chi) \cdot k_B(\chi) = 0$.

4. $E_{\mathcal{P}_C(\chi)}[|k|] = (1 - F(-\chi)) \cdot |k_A(\chi)| + F(-\chi) \cdot |k_B(\chi)|$ *decreases in χ.*

Proof. See Appendix B.1. ■

In words, Lemma 1.1 states that, regardless of media bias, message "A" always signals that candidate A is superior while message "B" always signals that A is inferior. However, as the media becomes more biased, message "A" becomes weaker in signaling the superiority of candidate A, while message "B" becomes stronger in signaling the inferiority of candidate A (Lemma 1.2). Ex-ante, however, rational voters' average posterior expectations of A's relative quality are not affected by media bias (Lemma 1.3). Nevertheless, an increase in media bias does systematically reduce voters' posterior expectations of candidates' quality difference, $|k|$ (Lemma 1.4). Since candidates' quality differences are negatively associated with the closeness of the election, an increase in media bias systematically raises voters' perceived closeness of elections. Lemma 1 highlights three effects after a marginal increase in media bias, which are crucial for understanding the electoral impacts of media bias:

[20] Alternatively, the cutoff information environment can also arise in a setup where media outlets can commit to a binary reporting strategy (Chan and Suen, 2008; Oliveros and Várdy, 2015).

Effect I A more biased media outlet (larger χ) is *a priori* more likely to send message "A" and induce posterior expectation $k_A(\chi)$ (with probability $1 - F(-\chi)$).

Effect II Both $k_A(\chi)$ and $|k_A(\chi)|$ decrease in χ. Message "A" thus becomes less credible and the election is expected to be more close under message "A", as χ increases.

Effect III $k_B(\chi)$ decreases while $|k_B(\chi)|$ increases in χ. Message "B" thus becomes more credible and the election is expected to be less close under message "B", as χ increases.

Proposition 1 summarizes the interim and ex-ante impacts of increasing media bias on party vote shares and the election outcome.

Proposition 1 *Suppose $|M| = 1$, the following properties hold:*

1. *$\forall \chi' > \chi \geq 0$, for both $s \in \{A, B\}$ and sufficiently large n it holds that $VS(k_s(\chi'), n) > VS(k_s(\chi), n)$ and $\pi(k_s(\chi'), n) > \pi(k_s(\chi), n)$.*

2. *For a sufficiently large n, $E_{\mathcal{P}_C(\chi)}[\pi(k, n)]$ varies non-monotonically with χ on $[0, 1]$.*[21]

    (a) *$E_{\mathcal{P}_C(\chi)}[\pi(k, n)] = \frac{1}{2}$ for $\chi \in \{0, 1\}$, and $E_{\mathcal{P}_C(\chi)}[\pi(k, n)] > \frac{1}{2}$ for $\chi \in (0, 1)$.*

    (b) *$E_{\mathcal{P}_C(\chi)}[\pi(k, n)]$ increases (decreases) in χ if χ is sufficiently close to 0 (1).*

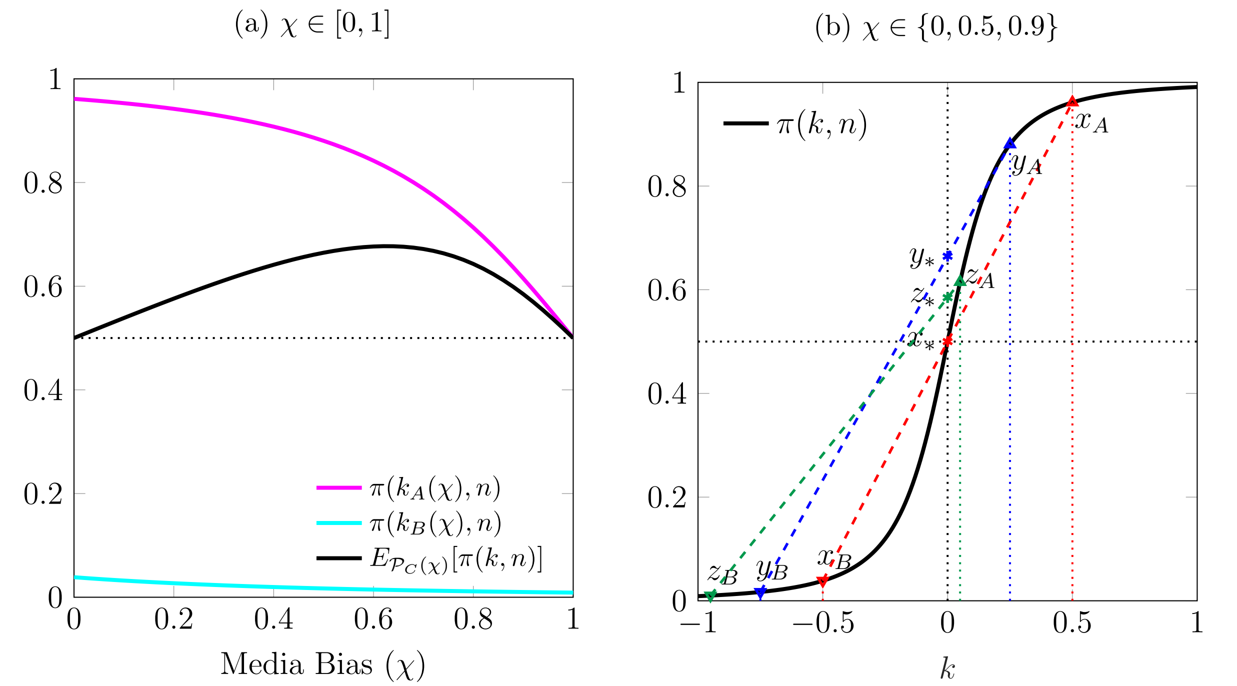
Proof. See Appendix B.1. ■

Figure 3a illustrates both the interim and the ex-ante impacts of increasing media bias on the election outcome, based on model parameters from Example 1. As Effect II and III indicate, both $k_A(\chi)$ and $k_B(\chi)$ decreases in χ. Therefore, both candidate A' expected vote share and winning probability decrease in media bias after either message (Proposition 1.1), since $VS_A(k, n)$ and $\pi(k, n)$ are decreasing functions of k. Effect I, however, provides a force in the opposite direction; a more A-biased outlet is *a priori* more likely to send message "A" to induce posterior expectation $k_A(\chi)$ instead of $k_B(\chi)$. Since $k_A(\chi) > k_B(\chi)$, and hence $\pi(k_A(\chi), n) > \pi(k_B(\chi), n)$, Effect I *per se* improves candidate A's electoral prospects. The overall net effect depends on which of the opposing forces is strongest. For a sufficiently small bias (χ close to 0), Effect I dominates so that a marginal increase in media bias increases candidate A's winning chances. For a sufficiently large degree of media bias (χ close to 1), however, the opposite applies. This explains the non-monotonic relationship between media bias and A's winning probability (Proposition 1.2).

The interim and ex-ante impacts of media bias on election outcomes can also be illustrated through an elegant geometric approach presented in Figure 3b. With k uniformly distributed

[21] Throughout the paper, we use notation $E_{\mathcal{P}}[z(k)]$ to denote the expectation of any real valued function $z(k)$ under the partition information environment characterized by $\mathcal{P}$.

Figure 3: The interim and ex-ante impact of media bias on the election outcome



Note: Panel (a) depicts the interim and the ex-ante impacts of media bias for all $\chi \in [0, 1]$. Panel (b) geometrically demonstrates these impacts for $\chi \in \{0, 0.5, 0.9\}$. The function $\pi(k, n)$ depicted in Panel (b) is identical to $\pi(k, n)$ in Figure 1b. In Panel (b), the selected levels of media bias increase from red to blue to green: $\chi = 0$ (unbiased, nodes x), $\chi = 0.5$ (weakly biased, nodes y), and $\chi = 0.9$ (strongly biased, nodes z). Nodes w_A, w_B and w_*, where $w \in \{x, y, z\}$, represent candidate A's winning probability conditional on message "A", message "B" and unconditionally, respectively. The dashed line segments represent the sets of all convex combinations of w_A and w_B, for $w \in \{x, y, z\}$. Model parameters are taken from Example 1.

on $[-1, 1]$ (cf. Example 1), it holds that $k_A(\chi) = \frac{1-\chi}{2}$ and $k_B(\chi) = \frac{-1-\chi}{2}$. In Figure 3b we consider three levels of media bias $\chi \in \{0, 0.5, 0.9\}$. First suppose $\chi = 0$, that is, the media outlet is unbiased. Then $k_A(0) = 0.5$ and $k_B(0) = -0.5$, and hence the electorate selects candidate with probability $\pi(0.5, n)$ ($\pi(-0.5, n)$) conditional on message "A" ("B"), represented by the red node x_A (x_B) in Figure 3b. Ex-ante, candidate A's winning probability is a weighted average of $\pi(0.5, n)$ and $\pi(-0.5, n)$, and must lie on the red dashed line segment connecting x_A and x_B. By the law of iterated expectations, the posterior means must average back to the prior mean, which is 0 (Lemma 1.3). Therefore, candidate A's ex-ante winning probability can be geometrically represented by the red node x_*, the intersection of line segment $x_A x_B$ with the vertical line $k = 0$.

A similar exercise can be done for a moderately biased outlet with $\chi = 0.5$, yielding interim election probabilities represented by the blue nodes y_A and y_B and an ex-ante winning probability given by y_*. Likewise, for a strongly biased media with $\chi = 0.9$ the election outcomes conditional on message "A" and "B", along with the ex-ante impact, are represented by the green nodes z_A, z_B and z_*, respectively. The interim impacts of an increase in media bias can now be straightforwardly represented by the movement from x_s to y_s to z_s, for $s \in \{A, B\}$. This reveals that candidate A's winning probability decreases as media bias increases conditional on both message "A" and "B", as predicted by Proposition 1.1. The ex-ante impact of increasing media bias is given by the movement from x_* to y_* to z_*. With $x_* < z_* < y_*$, the geometric analysis clearly illustrates the non-monotonic ex-ante impact of media bias on the election outcome, derived in Proposition 1.2.

We next turn to the interim and ex-ante impacts of media bias on voter turnout.

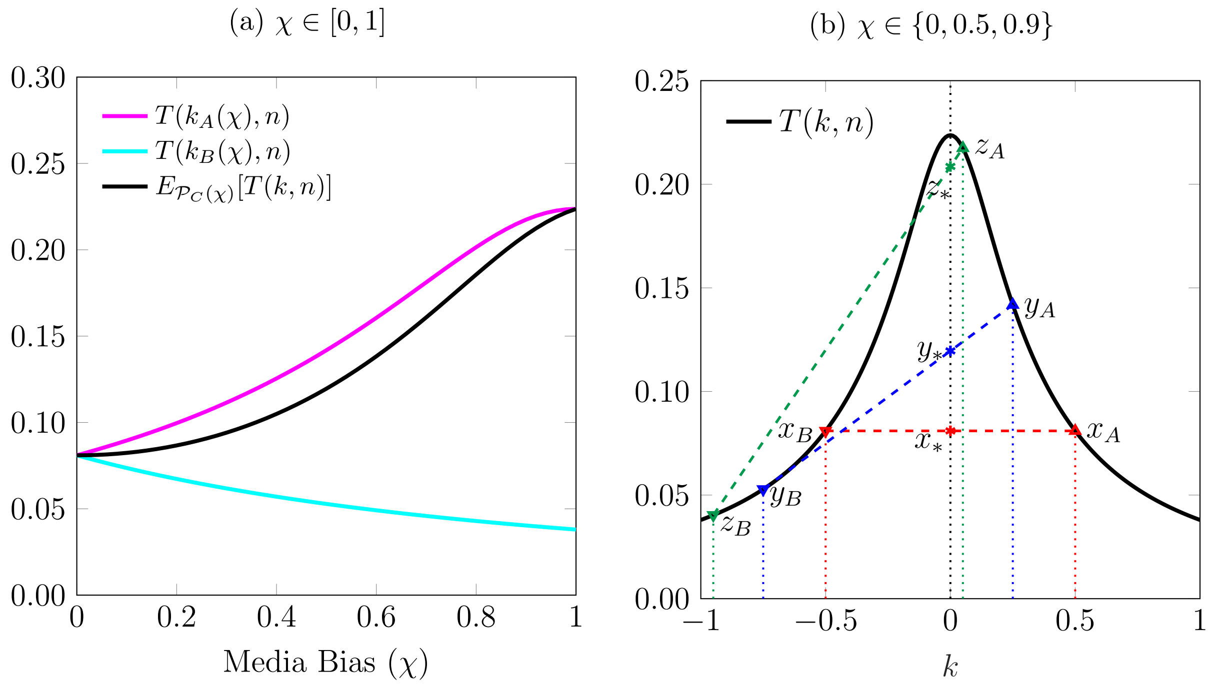
Proposition 2 *Suppose* $|M| = 1$. $\forall \chi' > \chi \geq 0$ *and for sufficiently large* n *there are:*

1. $T(k_s(\chi'), n) > (<) T(k_s(\chi), n)$ *for* $s = A(B)$.

2. $E_{\mathcal{P}_C(\chi')}[T(k, n)] > E_{\mathcal{P}_C(\chi)}[T(k, n)]$.

Proof. See Appendix B.1. ■

Figure 4: The interim and ex-ante impact of media bias on voter turnout



Note: Panel (a) depicts the interim and the ex-ante impacts of media bias for all $\chi \in [0, 1]$. Panel (b) geometrically demonstrates these impacts for $\chi \in \{0, 0.5, 0.9\}$. The function $T(k, n)$ depicted in Panel (b) is identical to $T(k, n)$ in Figure 1d. In Panel (b), the selected levels of media bias increase from red to blue to green: $\chi = 0$ (unbiased, nodes x), $\chi = 0.5$ (weakly biased, nodes y), and $\chi = 0.9$ (strongly biased, nodes z). Nodes w_A, w_B and w_*, where $w \in \{x, y, z\}$, represent the expected voter turnout conditional on message "A", message "B" and unconditionally, respectively. The dashed line segments represent the sets of all convex combinations of w_A and w_B, for $w \in \{x, y, z\}$. Model parameters are taken from Example 1.

Figure 4a illustrates Proposition 2. The interim impact of media bias on voter turnout depends on the message sent by the outlet (Proposition 2.1). Conditional on message "A", the expected voter turnout is higher as media bias increases. This follows because, by Effect II, $|k_A(\chi)|$ decreases in χ and the election is thus expected to end up in closer races, driving up voter turnout. Based on a similar intuition (cf. Effect III), conditional on message "B" the expected voter turnout is lower as media bias increases. Nevertheless, from the ex-ante perspective voter turnout increases with media bias unambiguously (Proposition 2.2) for two reasons. First, with $|k_A(\chi)|$ decreasing and $|k_B(\chi)|$ increasing in χ, the distribution of posterior expectations of $|k|$ becomes more spread out as χ increases. This would lead to a higher expected voter turnout if the average posterior mean of $|k|$ remains constant, due to the convexity of $T(|k|, n)$ in large elections. Second, due to Effect I the average posterior mean of $|k|$ puts more weight on $|k_A(\chi)|$ as media bias increases, since message "A" is more likely to be sent ex-ante. As a result, the average posterior mean of $|k|$ decreases in media bias (Lemma 1.4). This further increases voter turnout because $T(|k|, n)$ is decreasing in $|k|$.

Figure 4b takes the same geometric approach as in Figure 3b. Similar to the previous analysis of election outcomes, the interim impact conditional on message "A" and "B", as well as the ex-ante impact of media bias on voter turnout, are geometrically represented by nodes w_A, w_B and w_*, respectively, for $w \in \{x, y, z\}$ (again corresponding to $\chi \in \{0, 0.5, 0.9\}$). The interim impact conditional on message "A" ("B") again follows from the movement from x_A to y_A to z_A (x_B to y_B to z_B). Ex-ante voter turnout increases in media bias since $x_* < y_* < z_*$.

Finally, we study how media bias affects the probability of electing the first-best (i.e., quality-superior) candidate and voter welfare, as $n \to \infty$. Unlike previous propositions, which focus on $|M| = 1$ and information partition $\mathcal{P}_C(\chi)$, Proposition 3 quantifies these outcomes for generic information partitions $\mathcal{P} \subset [-1, 1]$.

Proposition 3 *For any partition $\mathcal{P} \subset [-1, 1]$, let $x^* \equiv \min\{|x| | x \in \mathcal{P}\}$, it holds that*

1. *the probability of electing the first-best candidate converges to $\frac{1}{2} + F(-x^*)$ as $n \to \infty$.*
2. $\lim\limits_{n \to \infty} E_{\mathcal{P}}[W(k, n)] = \int_{x^*}^{1} k dF(k)$.

Proof. See Appendix B.1. ■

With $|M| = 1$ and $\chi \geq 0$, $\mathcal{P}_C(\chi) = \{-\chi\}$. It follows directly from Proposition 3 that, as $n \to \infty$, the probability of electing the first-best candidate and the expected voter welfare $E_{\mathcal{P}_C}[W(k, n)]$ converge to $\frac{1}{2} + F(-\chi)$ and $\int_{\chi}^{1} k dF(k)$, respectively. Both are decreasing in media bias χ. Moreover, by Proposition 2, increasing media bias also systematically increases voter turnout and thus the aggregate voting costs. This effect further reduces voters' welfare. Proposition 3 also implies that, when the media outlet is unbiased ($\chi = 0$), the election outcome is *asymptotically efficient* in the sense that the asymptotic voter welfare, $\lim\limits_{n \to \infty} E_{\mathcal{P}}[W(k, n)]$, is maximized. This holds because, with an unbiased media outlet, voters know precisely whether the realized state is above or below 0 for sure ex-post, and hence elect the quality-superior candidate with a probability approaching 1 as $n \to \infty$. Conversely, if the media outlet is extremely biased with $\chi = 1$, voters remain uninformed ex-post because the media outlet always endorses candidate A regardless of the realized state. In this case, the election outcome is completely independent of the realized state. Therefore, both the probability of electing the first-best candidate and the asymptotic voter welfare are minimized with $\chi = 1$.

5.3 The influence of media entry

In this section we vary $|M|$ from 1 to 2 to study the ex-ante impact of media entry on voter behavior and election outcomes. Let the biases of the incumbent (outlet 1) and the entrant (outlet 2) be χ_1 and χ_2, respectively (with $\chi_2 \neq \chi_1$). Without loss of generality, we assume $\chi_1 \in [0, 1)$.[22] As Figure 2 illustrates, the incumbent partitions the posterior state space into two intervals $[-1, -\chi_1]$ and $(-\chi_1, 1]$ by sending message "B" and "A", respectively. Media entry affects voters' information environment by introducing a finer partition on $[-1, -\chi_1]$ ($(-\chi_1, 1]$), if $\chi_2 > (<)\chi_1$. We say the incumbent and entrant are *like-minded* (*opposite-minded*) if χ_1 and χ_2 have the same (opposite) sign. Moreover, we say the entrant has a *stronger (weaker)* bias if $|\chi_2| > (<)|\chi_1|$. Proposition 4 precisely characterizes how media entry affects the election outcome from an ex-ante perspective.

Proposition 4 *For all $\chi_1 \in [0, 1)$ and $\chi_2 \neq \chi_1$, it holds for sufficiently large n that $E_{\mathcal{P}_C(\chi_1,\chi_2)}[\pi(k, n)] > (<)E_{\mathcal{P}_C(\chi_1)}[\pi(k, n)]$ if and only if $\chi_2 > (<)\chi_1$.*

Proof. See Appendix B.2. ■

In words, Proposition 4 states that in sufficiently large elections, if the entrant is like-minded and has a stronger bias, then media entry always increases the winning probability of the candidate ex-ante favored by the incumbent. Conversely, if the entrant is opposite-minded, or like-minded with a weaker bias, then media entry always decreases the winning chances of that candidate. Loosely put, in large elections media entry systematically shifts the election outcome in the direction of the entrant's bias (relative to the incumbent's).

Proposition 4 is primarily driven by the S-shape property of $\pi(k, n)$. To see this, suppose $\chi_1 > 0$ so that the incumbent favors candidate A ex-ante. Conditional on the incumbent sending message "B", the realized state must lie in interval $[-1, -\chi_1]$, where $\pi(k, n)$ is convex in k. If the entrant is like-minded with a stronger bias (i.e., $\chi_2 > \chi_1$), it refines the information partition on $[-1, -\chi_1]$ and hence induces a mean-preserving spread of voters'

[22] We ignore $\chi_1 = 1$ because in this case voters are effectively uninformed and the equilibrium behavior is equivalent to the scenario with outlet 2 only. This single media case is analyzed in Section 5.2.

posterior expectations on the same interval, which necessarily increases the expectation of any convex function. If instead $\chi_2 < \chi_1$, then the entrant refines the information partition on $(-\chi_1, 1]$ (i.e., conditional on the incumbent sending message "A"), where $\pi(k, n)$ is neither convex nor concave in k. In this case, the argument for why media entry unambiguously decreases A's winning probability is more subtle.

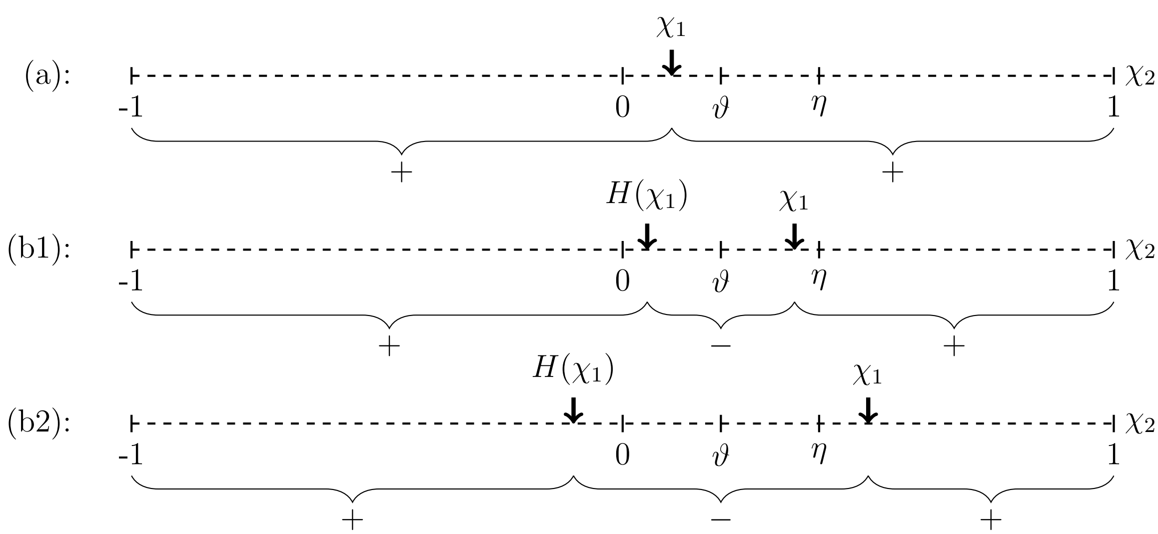
Our next proposition concerns how media entry systematically affects voter turnout. Essentially it shows that entry may potentially only decrease turnout if the incumbent is sufficiently extreme and the entrant has a weaker bias. Otherwise entry necessarily increases ex ante voter turnout.

Proposition 5 *For a sufficiently large n, there exists $\vartheta \in (0, 1)$ and a decreasing function $\hbar(\chi)$ satisfying (i) $-\chi < \hbar(\chi) < \chi$ for all $\chi > \vartheta$ and (ii) $\hbar(\eta) = 0$ for some $\eta > \vartheta$ (with $\eta < 1$), such that:*

1. *if $0 \leq \chi_1 \leq \vartheta$ and $\chi_2 \neq \chi_1$, then $E_{\mathcal{P}_C(\chi_1,\chi_2)}[T(k, n)] > E_{\mathcal{P}_C(\chi_1)}[T(k, n)]$.*
2. *if $\chi_1 \in (\vartheta, 1)$, then $E_{\mathcal{P}_C(\chi_1,\chi_2)}[T(k, n)] < E_{\mathcal{P}_C(\chi_1)}[T(k, n)]$ if $\chi_2 \in (\hbar(\chi_1), \chi_1)$ and the opposite holds otherwise.*

Proof. See Appendix B.2. ■

Figure 5: A graphical illustration of Proposition 5



Note: The intervals in curly braces with a "+" ("−") sign denote the sets of χ_2 under which media entry increases (decreases) ex-ante voter turnout, for a given $\chi_1 \geq 0$.

Figure 5 graphically illustrates Proposition 5. Panel (a) corresponds to part 1 stating that if the bias of the incumbent is sufficiently low (i.e., $\chi_1 < \vartheta$), then media entry systematically increases voter turnout regardless of χ_1. In contrast, if the incumbent is at least moderately biased (i.e., $\chi_1 > \vartheta$), then media entry systematically decreases voter turnout if the entrant has a weaker bias sufficiently close to the bias of the incumbent (see panel (b1)). For a rather strongly biased incumbent (i.e., $\chi_1 > \eta$), this less extreme entrant could even be opposite minded (see panel (b2)). Note that in the latter case also entry of an unbiased entrant can systematically decrease voter turnout.

Proposition 5 is also primarily driven by the curvature of $T(k, n)$. If the entrant is like-minded and has a stronger bias (i.e., $\chi_2 > \chi_1$), then media entry induces a finer information partition on $[-1, -\chi_1]$ and necessarily increases the expectation of $T(k, n)$, which is convex $[-1, -\chi_1]$. If instead the entrant is opposite-minded (i.e., $\chi_2 < 0$), or like-minded with a weaker bias (i.e., $0 < \chi_2 < \chi_1$), then the entrant induces a finer partition on interval $[-\chi_1, 1]$, where voter turnout $T(k, n)$ is *not* convex in k. In this case, how media entry influences ex-ante voter turnout depends on the biases of both outlets.[23]

The literature suggests some alternative mechanisms through which media entry may affect voter turnout. First, media entry may increase voter turnout by simply notifying voters about an upcoming election (Gentzkow et al., 2011). Second, media entry may lead to a substitution effect; the entry of commercial TV channels and the Internet may shift voters' interests from politics to entertainment, decreasing their exposure to political information and thereby voter turnout (Gentzkow, 2006; Ellingsen et al., 2018; Falck et al., 2014). Third, competition in the marketplace may affect media outlets' provision of political information and, in turn, affect voter turnout (Cagé, 2017).[24] All these mechanisms rely on the assumption that informed voters are more likely to turnout than their uninformed counterparts. This assumption finds support from the literature on individual decision making theory (Matsusaka, 1995), the Swing Voter's Curse (Feddersen and Pesendorfer, 1996, 1999; Battaglini et al., 2008, 2010) and government's responsiveness to the demand of informed voters (Strömberg, 2004; Gavazza et al., 2015). Yet, theories explaining, in a unified framework, why media entry might sometimes increase and in other cases decrease voter turnout, are rare. To the best of our knowledge, Piolatto and Schuett (2015) is the only other paper doing so, in an ethical voter framework. We provide a different explanation based on how media entry affects (voters' perception of) the closeness of elections. The comparative statics results we obtain are not only consistent with the mixed empirical evidence on the relation-

[23] As in Section 5.2, both interim and ex-ante impacts of media entry on election outcomes and voter turnout can be illustrated by the geometric approach. Details are available from the authors.

[24] Cagé (2017) studies the political impact of media entry from an industrial organization perspective. Depending on the level of heterogeneity of consumers' willingness to pay for different categories of news (e.g., politics and entertainment), media entry (i.e., the transition from monopoly to duopoly) may result in market segregation and under-provision of news in certain categories. Because the author assumes that information is positively correlated with voter turnout, media competition affects voter turnout merely via its impact on information provision.

ship between media entry and voter turnout, but also with the observation that media entry does not systematically affect party vote shares (Gentzkow et al., 2011; Drago et al., 2014).

Given that media entry can systematically drive both the election outcome and voter turnout in either direction, how does it affect voter welfare? *A priori* one would expect that it makes voters better off, because it improves information transmission and thus allows rational voters to make better decisions. Proposition 3 confirms this conventional wisdom in large elections under the cutoff information environment; with $|M| = 2$ and $\mathcal{P}_C(\chi_1, \chi_2) = \{-\chi_1, -\chi_2\}$, the probability of electing the welfare-superior candidate and voter welfare converges to $\frac{1}{2} + F(-x^*)$ and $\int_{x^*}^{1} kdF(k)$, respectively, where $x^* = \min\{|-\chi_1|, |-\chi_2|\}$. Because $x^* \leq |\chi_m|$ for both $m = 1, 2$, media entry unambiguously improves these outcomes compared to the scenario with only one outlet. Moreover, it also suggests that the asymptotic voter welfare is determined solely by the *least biased* outlet, whose χ_m is closest to 0.

6 Media commitment, information transmission and voter welfare

In Section 5 we analyzed media outlets' influence on voting equilibria under the assumption that all media outlets use the cutoff endorsement strategy (8). In this section we endogenize media outlets' reporting strategies and characterize equilibrium information transmission under two different communication protocols: cheap talk (CT) and Bayesian persuasion (BP). In the former setup, media outlets cannot commit to any reporting strategy before observing the realized state whereas in the latter setup they can. The comparison between the two communication protocols informs us how media commitment affects equilibrium information transmission and voter welfare. Indeed, we find that media commitment plays a key role in determining the amount of information that can be transmitted in equilibrium (cf. Section 6.1). This in turn begs the question to what extent the conventional wisdom – media competition improves information revelation and voter welfare – can be generalized under different communication protocols. This is explored in Section 6.2.

6.1 Equilibrium information transmission under CT and BP

We use a sender-receiver game framework to study the information transmission between media outlets and the electorate. Since media outlets send public information, all voters form a common posterior conditional on the message profile sent by all media outlet(s). So from the media outlets' perspective, the entire electorate is equivalent to a single receiver who rationally updates belief to obtain posterior $E[k|I]$ and subsequently selects candidate A with probability $\pi(E[k|I], n)$. Recall that we depict media m's reporting strategy by $\sigma_m(k)$, which is a mapping from the state space ($[-1, 1]$) to the message space S. In this section, we assume $S = [-1, 1]$ so that a message can be literally interpreted as a *report of the state*. As in all cheap talk games, there are multiple equilibria. For tractability we impose the *monotonicity constraint* on the strategies $\sigma_m(\cdot)$ for all $m \in M$:

$$\forall k, k' \in [-1, 1], k > k' \implies \sigma_m(k) \geq \sigma_m(k') \tag{9}$$

In words, (9) requires that each outlet $m \in M$ must announce weakly higher messages in higher states. Therefore, the monotonicity constraint implies that (i) a strictly higher message must reveals strictly higher states, and (ii) if a message s is sent in two different states, then the same s must also be sent in all states in between. Taken together, any reporting strategy $\sigma_m(\cdot)$ must induce a *partition*, denoted by $\mathcal{P}^m$, of the state space. As in Section 5.1, we characterize $\mathcal{P}^m$ by the set of cutoff points; $x \in \mathcal{P}^m$ if and only if $\sigma_m(k) > (\leq)\sigma_m(x)$ implies $k > (\leq)x$ for all $k, x \in [-1, 1]$. In other words, $x \in \mathcal{P}^m$ if and only if the electorate is precisely informed of whether the realized k is strictly above, or weakly below x, by reading from outlet m.[25] The information environment created by all media outlets $m \in M$ is the joint of all partitions $\mathcal{P} = \cup_{m \in M} \mathcal{P}^m$. Any $\mathcal{P}$ partitions the state space $[-1, 1]$ into adjacent and disjoint *separating* or *pooling* intervals. The realized state k is precisely revealed in a separating interval, while unrevealed in a pooling interval.

We focus on the most informative equilibria, i.e., equilibria that induce the finest partition $\mathcal{P}$, and do not distinguish equilibria inducing an identical $\mathcal{P}$. When $|M| \geq 2$, we also derive the most informative *coordination-proof* equilibria, in which no subset of media outlets can strictly benefit by jointly deviating from their equilibrium reporting strategy profile. Theorem 2 precisely characterize the most informative partition in equilibrium under CT.

Theorem 2 *Suppose media outlets communicate via cheap talk and* $\chi_1 \geq \chi_2 \geq \cdots \geq \chi_{|M|}$.

1. *(Discipline equilibrium) The most informative partition in equilibrium equals*

$$\mathcal{P}_D(\chi_1, \cdots, \chi_{|M|}) = \begin{cases} \{-\chi_1\}, & \textit{if } |M| = 1 \\ [-1, -\chi_1] \cup [-\chi_2, 1], & \textit{if } |M| = 2 \\ [-1, 1], & \textit{if } |M| \geq 3 \end{cases}$$

2. *(Cutoff equilibrium) Focusing on coordination-proof equilibria only, the most informative partition in equilibrium equals* $\mathcal{P}_C(\chi_1, \cdots, \chi_{|M|}) = \{-\chi_1, -\chi_2, \cdots, -\chi_{|M|}\}$.

[25] To simply the presentation and proofs, we assume any cutoff point x is downward pooled. Under Assumption 1, $k = x$ is a zero-probability event and the specification of x in the partition is inconsequential.

Proof. See Appendix C.1. ■

To illustrate Theorem 2, we start with $|M| = 1$. In this case, the most informative equilibrium yields a unique information partition $\{-\chi_1\}$, which consists of two pooling intervals $[-1, -\chi_1]$ and $(-\chi_1, 1]$. This is equivalent to the information environment created by the cutoff endorsement strategy introduced in Section 5.1, illustrated by Figure 2 therein. Since the outlet prefers candidate A(B) to be elected if $k > (\leq) - \chi_1$ and A's winning probability increases in voters' posterior expectations of k, it is optimal for this outlet to send the highest (lowest) message when $k > (\leq) - \chi_1$. Therefore, whether k lies above or below $-\chi_1$ must be precisely communicated, yielding the aforementioned pooling intervals.

In the presence of two media outlets ($|M| = 2$) with any biases $\chi_1 \geq \chi_2$, the most informative CT equilibrium, referred to as the Discipline equilibrium (Theorem 2.1), yields an information partition $\mathcal{P}_D(\chi_1, \chi_2) = [-1, -\chi_1] \cup [-\chi_2, 1]$. This partition consists of two separating intervals $[-1, -\chi_1]$ and $(-\chi_2, 1]$, and a pooling interval $(-\chi_1, -\chi_2]$ in between. In line with the literature on multiple-sender cheap talk games, in a Disciplining equilibrium the receiver (electorate) can discipline senders (media outlets) to reveal more information by cross-checking their messages and punish senders by a strictly undesirable action if their messages do not agree (Gilligan and Krehbiel, 1989; Krishna and Morgan, 2001a,b). The electorate can discipline media outlets to reveal all "aligned states" (i.e., $k \in [-1, -\chi_1) \cup (-\chi_2, 1]$), conditional on which both outlets prefer the same candidate to be elected, by the following cross-checking strategy: if $s_1 = s_2 = s$ the electorate believes $k = s$, otherwise they form posterior expectation $E_F[k|k \in (\chi_1, -\chi_2)]$. With such a strategy, both outlets are worse off in aligned states if their messages do not match, as sending inconsistent signals would increase the winning chances of their unfavored opponent. In "conflicting states", where the two outlets support different candidates (i.e, $k \in (-\chi_1, -\chi_2)$), such a common punishment does not exist and they cannot be disciplined into revealing any information. With more than two media outlets, full information revelation is generically possible under the Discipline equilibrium. This is because for all realized state k, there always exist (at least) two media outlets whose interests are aligned. These interest alignments can always be exploited to construct a fully revealing equilibrium under appropriate cross-checking strategies.

The constructions of Discipline equilibria depend critically on the premise that senders cannot coordinate their messages. To see this, note that when $k > -\chi_2$ and coordination is possible, both outlets would be better off by coordinating on $s_1 = s_2 = 1$ (rather than on $s_1 = s_2 = k$) instead to maximize A's winning chances, given the aforementioned cross-checking strategy of the electorate. Therefore, if we require media outlets' reporting strategy profile to be coordination-proof, then aligned states can no longer be credibly communicated in equilibrium (Theorem 2.2). In fact, the most informative equilibrium under coordination-proofness generates the cutoff information environment described in Section 5.1 (illustrated by Figure 2 therein) and is thus referred to as the Cutoff equilibrium.

Theorem 3 characterizes the most informative equilibria in the BP setup, under a single-crossing condition defined in Appendix C.2 (Assumption 3 therein).

Theorem 3 *Suppose media outlets can commit to any reporting strategy profile and $\chi_1 \geq \chi_2 \geq \cdots \geq \chi_{|M|}$. Under the single-crossing condition defined in Appendix C.2 it holds that:*

1. *The most informative partition in equilibrium equals*

$$\mathcal{P}(\chi_1, \cdots, \chi_{|M|}) = \begin{cases} [a(\chi_1; n), b(\chi_1; n)], & \textit{if } |M| = 1 \\ [-1, 1], & \textit{if } |M| > 1 \end{cases}$$

2. *(BP equilibrium) Focusing on coordination-proof equilibria only, the equilibrium information partition is unique and equals $\mathcal{P}_{BP}(\chi_1, \cdots, \chi_{|M|}) = [a(\chi_1; n), b(\chi_{|M|})]$.*

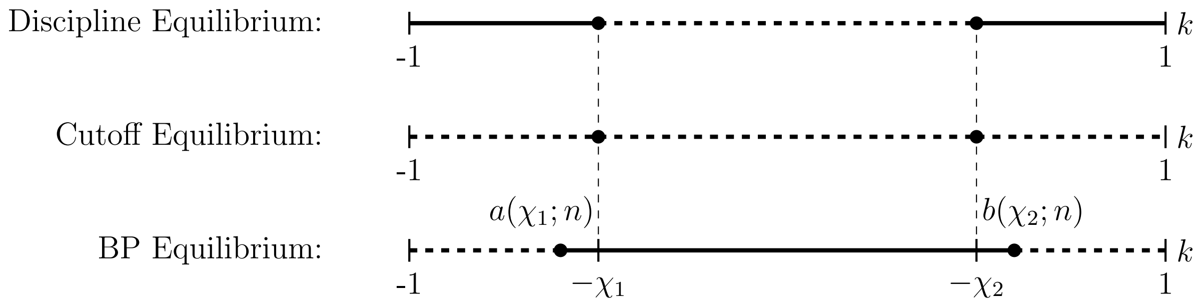
where functions $a(x; n)$ and $b(x; n)$ are implicitly defined in Appendix C.2 and satisfy for all $x \in [-1, 1]$: (i) $-1 \leq a(x; n) < -x < b(x; n) \leq 1$; (ii) both $a(x; n)$ and $b(x; n)$ are weakly decreasing in x.

Proof. See Appendix C.2. ■

Under BP, an information partition can be supported in equilibrium if and only if no senders can profit from unilaterally committing to any finer partition (Gentzkow and Kamenica, 2016). If $|M| = 1$, the equilibrium partition is unique and equals $[a(\chi_1; n), b(\chi_1, n)]$; it consists of two pooling intervals $[-1, a(\chi_1; n)]$ and $(b(\chi; n), 1]$, and a non-trivial separating interval $(a(\chi_1; n), b(\chi_1, n)]$ which contains $-\chi_1$. In words, the optimal reporting strategy precisely reveals states with low payoff relevance (i.e., k sufficiently close to $-\chi_1$), and pool all remaining high states ($k > b(\chi_1; n)$) into a single message "A" and low states ($k \leq a(\chi_1; n)$) in to a message "B".[26] In deciding whether to separate the boundary state (say, $b(\chi_1; n)$) from the pooling message (say, "A"), the outlet balances two opposite forces. On one hand, doing so lowers candidate A's winning probability when the marginal $b(\chi_1; n)$ realizes. On the other hand, separating $b(\chi_1; n))$ from "A" strengthens the credibility of message "A" and thus increases A's electoral prospects in all states $k > b(\chi_1; n)$. For k sufficiently close to $-\chi_1$ the latter effect always dominates so $b(\chi_1; n) > -\chi_1$ generically holds. This is because, by (2), the outlet's stake at election equals to $|k + \chi_1|$ and it vanishes as $k \to -\chi_1$. Hence, the outlet is almost indifferent between candidates for k close to $-\chi_1$ and the marginal losses

[26]It is possible for an optimal reporting strategy to have $a(\chi_1; n) = -1$ or $b(\chi_1; n) = 1$. In these cases, the outlet precisely reveal all states $k \in [-1, -\chi_1]$ or $k \in (-\chi_1, 1]$, respectively. However, $a(\chi_1; n) = -1$ and $b(\chi_1; n) = 1$ never simultaneously hold, that is, full information revelation is impossible under $|M| = 1$.

Figure 6: Equilibrium information partitions under CT and BP when $|M| = 2$



Note: The solid line segments and black nodes include all elements in each equilibrium information partition. As an alternative explanation, the solid (dashed) line segments also represent separating (pooling) intervals.

from the former effect is negligible compared to the marginal gains from the latter. The choice of the other boundary state $a(\chi_1; n)$ follows the analogous principle.

If $|M| \geq 2$ and without coordination-proofness, fully revealing equilibria generically exist under BP (Theorem 3.1). Nevertheless, these equilibria are implemented in weakly dominated strategies.[27] For this reason, the literature on Bayesian persuasion with multiple senders typically focuses on the least informative equilibria (Gentzkow and Kamenica, 2016, 2017), which coincide with the unique equilibrium under BP and coordination-proofness, referred to as the BP equilibrium (Theorem 3.2). The critical impact of adding media outlets on the BP equilibrium is that all "conflicting states" must be fully separated. Because $\pi(k, n)$ is generically nonlinear, in any non-trivial pooling interval containing conflicting states, inducing a finer information partition will systematically push the election outcome in favor of one candidate. In those conflicting states, at least two media outlets have opposite partisan preferences, and hence there must exist (at least one) media outlets finding unilaterally committing to a finer information partition profitable.[28]

For ease of comparison, we graphically illustrate the most informative equilibrium information partition with $|M| = 2$ under the Discipline, Cutoff and BP equilibria in Figure 6. Figure 6 highlights two sharp differences between the equilibrium information partitions under CT and BP setups, when $|M| = 2$. First, in aligned states where all media outlets' partisan preferences agree (e.g., intervals $[-1, -\chi_1]$ and $(-\chi_2, 1]$) the information partition exhibits "semi-separating" in the BP equilibrium. In the two classes of cheap talk equilibria we investigate, however, these states are either fully separated (the Discipline equilibrium) or pooled (the Cutoff equilibrium). Second, the "conflicting states" (i.e., interval $(-\chi_1, -\chi_2]$) are fully revealed in all BP equilibria. In sharp contrast, however, these conflicting states

[27] Suppose outlet 1 unilaterally commit to full revelation, it is weakly optimal for all other outlets to commit to the same strategy because they cannot unilaterally influence voters' information.

[28] See also Gentzkow and Kamenica (2016, 2017) for similar insights.

can never be revealed in any cheap talk equilibrium when $|M| = 2$.

In what follows, we use these insights to analyze how information transmission and the asymptotic voter welfare (i.e., as $n \to \infty$) are affected as we increase media competition.

6.2 Competition, information transmission and voter welfare

Following Gentzkow and Kamenica (2016), we consider the following three notions of increasing media competition, under $|M| = 2$. First, increasing the number of media outlets. Second, increasing media polarization. Third, preventing media collusion.

To compare the informativeness of two information partitions $\mathcal{P}_1$ and $\mathcal{P}_2$, we say $\mathcal{P}_1$ is *more informative* than $\mathcal{P}_2$ if $\mathcal{P}_1$ is a finer partition of $\mathcal{P}_2$ (i.e., $\mathcal{P}_2 \subset \mathcal{P}_1$). To compare the asymptotic voter welfare, we say $\mathcal{P}_1$ is *welfare-superior (inferior)* to $\mathcal{P}_2$ if $\lim_{n\to\infty} E_{\mathcal{P}_1}[W(k,n)] >$ $(<) \lim_{n\to\infty} E_{\mathcal{P}_2}[W(k,n)]$. Recall from Proposition 3 that $\lim_{n\to\infty} E_{\mathcal{P}}[W(k,n)] = \int_{x^*}^{1} kdF(k)$ with $x^* = \min\{|x||x \in \mathcal{P}\}$. The asymptotic voter welfare is maximized if and only if $0 \in \mathcal{P}$; voters learn precisely whether the realized state is above or below 0 in equilibrium. For this reason, we say a partition $\mathcal{P}$ is *asymptotically efficient* if $0 \in \mathcal{P}$. Clearly, $\mathcal{P}$ is welfare-superior to all other information partitions if it is asymptotically efficient.

6.2.1 Increasing the number of media outlets

In this section we show that both CT and BP models agree that increasing the number of media outlets weakly improves information transmission and voter welfare. However, these models disagree sharply on what type of media outlet should be introduced to maximize these benefits in large elections. We illustrate this insight by increasing $|M|$ from 1 to 2. Let the biases of the incumbent and entrant outlets be χ_1 and χ_2, respectively. As in Section 5.3, we assume $\chi_1 \geq 0$ and say the two outlets are like-minded (opposite-minded) if χ_1 and χ_2 have the same (opposite) sign. Define $\chi_+ \equiv \max\{\chi_1, \chi_2\}$ and $\chi_- \equiv \min\{\chi_1, \chi_2\}$.

In a Discipline equilibrium it holds that $\mathcal{P}_D(\chi_1, \chi_2) = [-1, -\chi_+] \cup [-\chi_-, 1]$ (cf. the top panel of Figure 6). As is evident graphically, $\mathcal{P}_D(\chi_1, \chi_2)$ is asymptotic efficient (i.e., $0 \in \mathcal{P}_D(\chi_1, \chi_2)$) if and only if the two outlets are like-minded. Moreover, with $\chi_2 = \chi_1$, $\mathcal{P}_D(\chi_1, \chi_2) = [-1, 1]$ so that the Discipline equilibrium is fully revealing.

In a Cutoff equilibrium there are $\mathcal{P}_C(\chi_1, \chi_2) = \{-\chi_1, -\chi_2\}$ (cf. the central panel of Figure 6) and $\lim_{n\to\infty} E_{\mathcal{P}_C(\chi_1,\chi_2)}[W(k,n)] = \int_{\min\{|\chi_1|,|\chi_2|\}}^{1} kdF(k)$. Hence, $\mathcal{P}_C(\chi_1, \chi_2)$ is asymptotically efficient if and only if $\chi_j = 0$ for some $j = 1, 2$, i.e., at least one outlet is unbiased.

Finally, in a BP equilibrium there are $\mathcal{P}_{BP}(\chi_1, \chi_2) = [a(\chi_+; n), b(\chi_-; n)] \supset [-\chi_+, -\chi_-]$ (cf. the bottom panel of Figure 6). It is evident graphically that, if the two outlets are opposite-minded, $\mathcal{P}_{BP}(\chi_1, \chi_2)$ must be asymptotically efficient since $0 \in [-\chi_+, -\chi_-]$. Moreover, $\mathcal{P}_{BP}(\chi_1, \chi_2)$ must be fully revealing if $\chi_1 = 1$ and $\chi_2 = -1$, i.e., the two outlets are opposite-minded with extreme biases. We summarize these observations in Proposition 6.

Proposition 6 *Let $|M| = 2$ and $\chi_1 > 0$, it holds that:*

1. *$\mathcal{P}_D(\chi_1, \chi_2)$ is asymptotically efficient if and only if $\chi_2 \geq 0$, and it is fully revealing if $\chi_2 = \chi_1$.*
2. *$\mathcal{P}_C(\chi_1, \chi_2)$ is asymptotically efficient if and only if $\chi_2 = 0$.*
3. *$\mathcal{P}_{BP}(\chi_1, \chi_2)$ is asymptotically efficient if $\chi_2 \leq 0$, and it is fully revealing if $\chi_1 = 1$ and $\chi_2 = -1$.*

Proposition 6 highlights the critical role of media commitment in identifying the types of welfare and information maximizing entrant outlets. Under CT (i.e., without media commitment), asymptotic efficiency is achieved only if the entrant outlet is unbiased or like-minded (under the Discipline equilibrium only). Under BP (i.e., with media commitment), however, asymptotic efficiency is generically achieved with an opposite-minded entrant. The implication of media commitment on information revelation is in sharp contrast; in a Discipline (BP) equilibrium, it is optimal to have an entrant outlet that is *like-minded with an identical bias* (*opposite-minded with an extreme bias*) to maximize information revelation.

6.2.2 Increasing media polarization

In this section, we assume $|M| = 2$ and ask how information transmission and the asymptotic voter welfare is affected by increasing the misalignment of interests between the two media outlets. This informs us how voters are affected from exposure to a polarized media environment. We study the impact of increasing media polarization in a context with two opposite-minded and equally biased outlets; $|M| = 2$, $\chi_1 = \chi$ and $\chi_2 = -\chi$ with $\chi \geq 0$. We say *media polarization increases* if χ increases.

The Discipline and Cutoff equilibria yield partitions $\mathcal{P}_D(\chi, -\chi) = [-1, -\chi] \cup [\chi, 1]$ and $\mathcal{P}_C(\chi, -\chi) = \{-\chi, \chi\}$, respectively. The BP equilibrium produces partition $\mathcal{P}_{BP}(\chi, -\chi) = [a(\chi; n), b(-\chi; n)] \supset [-\chi, \chi]$. It is evident from Figure 6 that, as media polarization increases, $\mathcal{P}_D(\chi, -\chi)$ becomes less informative whereas $\mathcal{P}_{BP}(\chi, -\chi)$ becomes more informative.[29] In both Discipline and Cutoff equilibria, the asymptotic voter welfare equals $\int_\chi^1 k dF(k)$, which is strictly decreasing in media polarization (measured by χ). Nevertheless, the BP equilibrium is always asymptotically efficient, as $0 \in \mathcal{P}_C(\chi, -\chi)$ generically holds. We summarize these observations in Proposition 7.

[29]Note that our notion of informativeness is a partial order, and it is not defined for $\mathcal{P}_C(\chi, -\chi)$ as χ varies. For this reason, we do not compare the informativeness of $\mathcal{P}_C(\chi, -\chi)$.

Proposition 7 *Let $|M| = 2$ and $\chi \geq 0$, $\chi_1 = \chi$ and $\chi_2 = -\chi$. As χ increases, it holds that:*

1. *$\mathcal{P}_D(\chi, -\chi)$ becomes less informative and welfare-inferior.*
2. *$\mathcal{P}_C(\chi, -\chi)$ becomes welfare-inferior.*
3. *$\mathcal{P}_{BP}(\chi, -\chi)$ becomes more informative, and it is always asymptotically efficient.*

6.2.3 Preventing media collusion

In this section, we characterize the equilibrium information transmission and voter welfare depending on whether media collusion is possible. This comparison allows us to understand how media collusion (i.e., media merge) may affect the amount of information that can be transmitted and voter welfare. Suppose $|M| = 2$ and $\chi_1 > \chi_2$. We assume that the payoff function of colluded media outlets equals the average payoff of both outlets:

$$\bar{V}(k, \chi_1, \chi_2, \Omega) = \frac{V(k, \chi_1, \Omega) + V(k, \chi_2, \Omega)}{2} = \left(k + \frac{\chi_1 + \chi_2}{2}\right) \cdot \mathbb{1}_{\Omega = A} \tag{10}$$

In this way, media collusion essentially yields a single "colluded outlet" with the average bias $\chi \equiv \frac{\chi_1+\chi_2}{2}$. Under CT, by Theorem 2, the information partition produced by this colluded outlet equals $\mathcal{P}_D(\frac{\chi_1+\chi_2}{2}) = \mathcal{P}_C(\frac{\chi_1+\chi_2}{2}) = \{\frac{\chi_1+\chi_2}{2}\}$. Hence, in both Discipline and Cutoff equilibria, media collusion results in an asymptotic voter welfare equals $\int_{|\frac{\chi_1+\chi_2}{2}|}^1 k dF(k)$.

Suppose collusion is not possible and the two outlets are like-minded. It follows from Section 6.2.1 that the Discipline equilibrium is asymptotically efficient and hence welfare-superior to all other equilibria. In a Cutoff equilibrium the asymptotic voter welfare equals $\int_{\min\{|\chi_1|,|\chi_2|\}}^1 k dF(k)$. Because χ_1 and χ_2 have the same sign, it holds that $\min\{|\chi_1|, |\chi_2|\} < |\frac{\chi_1+\chi_2}{2}|$ and hence collusion of two like-minded outlets unambiguously reduces asymptotic voter welfare in either Discipline or Cutoff equilibria. If instead the two outlets are opposite-minded, then (without collusion) the asymptotic voter welfare in both Discipline and Cutoff equilibria equals $\int_{\min\{|\chi_1|,|\chi_2|\}}^1 k dF(k)$. Therefore, media collusion can improve voter welfare in large election if $|\frac{\chi_1+\chi_2}{2}| < \min\{|\chi_1|, |\chi_2|\}$. This is possible with two opposite-minded outlets with sufficiently close biases; for example, $\chi_1 = 0.5$ and $\chi_2 = -0.5$. Therefore, media collusion may either increase or decrease voter welfare under cheap talk.

In a BP equilibrium, Theorem 3 implies that the colluded outlet generates information partition $\mathcal{P}_{BP}(\frac{\chi_1+\chi_2}{2}) = \left[a(\frac{\chi_1+\chi_2}{2}; n), b(\frac{\chi_1+\chi_2}{2}; n)\right]$. Without media collusion there are $\mathcal{P}_{BP}(\chi_1, \chi_2) = [a(\chi_1; n), b(\chi_2; n)]$. Since $\chi_1 \geq \frac{\chi_1+\chi_2}{2} \geq \chi_2$ and both $a(\chi; n)$ and $b(\chi; n)$ are weakly decreasing in χ, it holds that $\mathcal{P}_{BP}(\frac{\chi_1+\chi_2}{2}) \subset \mathcal{P}_{BP}(\chi_1, \chi_2)$. Media collusion thus yields a strictly less informative and weakly welfare-inferior information partition, compared to that without collusion. We summarize these observations in Proposition 8.

Proposition 8 *Let $|M| = 2$ and $\chi_1 > \chi_2$, it holds that:*

1. *in both Discipline and Cutoff equilibria, media collusion increases asymptotic voter welfare if and only if $|\frac{\chi_1+\chi_2}{2}| < \min\{|\chi_1|, |\chi_2|\}$.*
2. *the BP equilibrium is less informative, and the asymptotic voter welfare weakly decreases if media outlets collude.*

Consistent with Gentzkow and Kamenica (2016, 2017), we show that media collusion can never improve information transmission and voter welfare under BP. Under CT, however, media collusion can be welfare improving if the two media outlets are opposite-minded and have sufficiently symmetric biases. The collusion of two like-minded media outlets, however, unambiguously reduces voter welfare.

7 Conclusion

We develop a tractable theory to study the influence of media bias and entry on election outcomes, voter turnout and welfare. We derive three key insights. First, news released by media outlets allows voters to infer (i) the relative appeal of candidates, and (ii) the closeness of elections. In large elections, the former information determines party vote shares and the election outcome, while the latter drives voter turnout. Second, media outlets can systematically manipulate election outcomes and voter turnout by strategically providing public information, even if voters are Bayesian. Based on these insights, we derive precise comparative statics predictions on the influence of media bias and entry on election outcomes and voter turnout. These results provide a possible rationale to organize the mixed empirical evidence on the relationship between media exposure and voter turnout. Third, the conventional wisdom – media competition improves information transmission and voter welfare – may fail to hold if media outlets cannot commit to their news reporting strategies. This insight highlights the role of media commitment in the policy debate of media regulation.

We suggest three avenues for future research. First, instead of reading from all media outlets, in real life voters are highly selective in media exposure (Gentzkow and Shapiro, 2010, 2011; Durante and Knight, 2012). Exploring how media consumption correlates with voter types and whether large elections can aggregate dispersed voter information is undoubtedly a salient issue. Second, voters may also have private information beyond the public news provided by public media (Liu, 2018). Our approach cannot directly extend to these contexts since voters no longer have common posterior beliefs, due to idiosyncratic private information. Third, our Poisson game setup precludes communication among voters and turnout mobilization through social norms (e.g., peer pressure), which are empirically relevant in determining political participations (Grosser and Schram, 2006; DellaVigna et al., 2016). The interactions between media outlets' electoral impact with various communication networks and social norms remain insufficiently understood. These questions might be explored by embedding media outlets in a pivotal voter framework allowing for correlated equilibria (Pogorelskiy, 2014) or in a group-based framework allowing for explicitly modeling of social norms (Levine and Mattozzi, 2016).

A Equilibrium derivations and proofs in Section 4

A.1 Deriving the asymptotic voting equilibria

Since voters are ex-ante identical, we use q_A and q_B to denote the probabilities that any randomly sampled voter votes for candidate A and B, respectively.[30] Let N_A and N_B be the realized number of voters voting for A and B, respectively. In addition, from the perspective of any voter i, let $\tilde{N}$ denote the number of other eligible voters in the electorate (not counting voter i). Since the electoral size N is a random variable, so are N_A, N_B and $\tilde{N}$. Lemma 2, proved in Myerson (1998), summarizes fundamental properties of the Poisson games.

Lemma 2 *(Myerson, 1998) In a Poisson game with mean population size n, there are*

1. Independent actions. *N_A and N_B are independently distributed.*
2. Decomposition property. *N_A and N_B are Poisson distributed with mean $n_A \equiv nq_A$ and $n_B \equiv nq_B$, respectively.*
3. Environmental equivalence. *$\tilde{N}$ is Poisson distributed with mean n.*

Let $T_l \equiv \{(N_A, N_B) \in Z_+^2 | N_A - N_B = l, l \in Z\}$ denote the set of events where candidate A gets exactly l more votes than B (if l is negative, then A gets $|l|$ fewer votes than B). Under the Poisson distribution, $Pr[T_l|n_A, n_B] = e^{-(n_A+n_B)} \sum_{i=\min\{0,l\}}^{\infty} \frac{{n_A}^{i+l}{n_B}^i}{(i+l)!i!}$. Conditional on the expected numbers of votes for A and B (n_A and n_B, respectively), the pivotal probabilities of voting for A and for B are[31]

$$\begin{aligned} Pr[PivA|n_A, n_B] &\equiv \frac{Pr[T_0|n_A, n_B] + Pr[T_{-1}|n_A, n_B]}{2} \\ &= \frac{e^{-(n_A+n_B)}}{2} \sum_{i=0}^{\infty} \frac{(n_A)^i (n_B)^i}{i!i!} \left(1 + \frac{n_B}{i+1}\right) \end{aligned} \tag{11}$$

$$\begin{aligned} Pr[PivB|n_A, n_B] &\equiv \frac{Pr[T_0|n_A, n_B] + Pr[T_1|n_A, n_B]}{2} \\ &= \frac{e^{-(n_A+n_B)}}{2} \sum_{i=0}^{\infty} \frac{(n_A)^i (n_B)^i}{i!i!} \left(1 + \frac{n_A}{i+1}\right) \end{aligned} \tag{12}$$

If k is common knowledge, voters with $v_i > (<) - k$ always vote for candidate A(B), conditional on turnout. As for their turnout decision, voters vote only if their private voting

[30] The probability of a randomly sampled voter abstaining is then $1 - q_A - q_B$.

[31] Note that a vote for A is pivotal only if it breaks a tie (in event T_0) or creates a tie (in event T_{-1}). In both cases, it changes the election outcome in favor of A with probability 0.5. The calculation of the pivotal probability of a vote for B is similar.

costs are sufficiently low, that is, below some threshold. The cost threshold must be equal to the expected benefit of voting to make the marginal voters indifferent between voting and abstaining. Let $\rho(x) \equiv \min\{\frac{x}{C}, 1\}$, $x \geq 0$, be the cumulative distribution function (CDF) of voting costs. Since voting costs are independently drawn from distribution $\rho(\cdot)$, *a priori* a voter with ideology $v_i > (<) - k$ votes for candidate A(B) with probability $\rho(|k+v| \cdot Pr[PivA|n_A, n_B])$ $(\rho(|k+v| \cdot Pr[PivB|n_A, n_B]))$. Hence, ex-ante, the probabilities of a randomly sampled voter voting for A and B, respectively, are given by

$$q_A = \int_{-k}^{\delta} \rho(|k+v| \cdot Pr[PivA|n_A, n_B]) dG(v) \tag{13}$$

$$q_B = \int_{-\delta}^{-k} \rho(|k+v| \cdot Pr[PivB|n_A, n_B]) dG(v) \tag{14}$$

A *voting equilibrium* is characterized by a solution of the above system of equations. Since both $Pr[PivA|n_A, n_B]$ and $Pr[PivB|n_A, n_B]$ are continuous functions of q_A and q_B, and the system of equations is a self-mapping on the compact and convex set $[0, 1]^2$, the existence of equilibrium is guaranteed by Brouwer's fixed-point theorem.[32] We focus on voting equilibria in large elections (i.e., $n \to \infty$). Formally, let $\{q_A(k, n), q_B(k, n)\}$ be a voting equilibrium for a given pair of k and n, and $\{q_A(k, n), q_B(k, n)\}_{n\to\infty}$ be a sequence of voting equilibria. Let $n_\Omega(k, n) \equiv nq_\Omega(k, n)$ be the expected number of votes for candidate $\Omega \in \{A, B\}$ in equilibrium. Myerson (2000) and Krishna and Morgan (2015) show that along any sequence of voting equilibria, the expected total number of votes goes to infinity while the pivotal probabilities vanish to zero as $n \to \infty$, as summarized in Lemma 3.[33]

Lemma 3 *(Myerson, 2000) For all $k \in [-1, 1]$ and any $\{q_A(k, n), q_B(k, n)\}_{n\to\infty}$,*

1. $\lim_{n\to\infty} n_A(k, n) + n_B(k, n) = +\infty$;
2. $\lim_{n\to\infty} Pr[Piv\Omega|n_A(k, n), n_B(k, n)] = 0$ *for both* $\Omega = A, B$.

To get the intuition for part 1), note that if the expected total number of votes is bounded in large elections, then the expected numbers of votes for each candidate are also bounded. As a result, the pivotal probabilities are strictly positive (though very small) in the limit. Hence, voters with sufficiently low but strictly positive voting costs will vote, and this yields a strictly positive fraction of the electorate since $\rho(\cdot)$ has a lower support at 0. The expected total number of votes then goes to infinity as $n \to \infty$, leading to a contradiction. Therefore, at least one of candidate A or B must get an infinite expected number of votes as $n \to \infty$.

[32]The continuity of $Pr[PivA|n_A, n_B]$ and $Pr[PivB|n_A, n_B]$ follows from Theorem 7.13 of Rudin (1964).

[33]See Theorem 4 in Myerson (2000) or Lemma 6 to 9 in Krishna and Morgan (2015) for the formal proof. The latter work shows that Lemma 3 holds true under a broader category of population uncertainty.

This implies that both pivotal probabilities vanishes 0 as $n \to \infty$, validating part 2). The following Lemma 4 characterizes the properties of pivotal probabilities as $n \to \infty$.

Lemma 4 *(Myerson, 2000) If $\lim_{n\to\infty} n_A(k,n) + n_B(k,n) = +\infty$, then*[34]

1. $\lim_{n\to\infty} \frac{Pr[PivA|n_A,n_B]}{Pr[PivB|n_A,n_B]} = \lim_{n\to\infty} \sqrt{\frac{n_B}{n_A}}$;
2. *If in addition* $0 < \lim_{n\to\infty} \frac{n_B(k,n)}{n_A(k,n)} < +\infty$, *then*

$$Pr[PivA|n_A,n_B] \approx \frac{e^{-(\sqrt{n_A}-\sqrt{n_B})^2}}{4\sqrt{\pi\sqrt{n_A n_B}}}\left(1+\sqrt{\frac{n_B}{n_A}}\right)$$

$$Pr[PivB|n_A,n_B] \approx \frac{e^{-(\sqrt{n_A}-\sqrt{n_B})^2}}{4\sqrt{\pi\sqrt{n_A n_B}}}\left(1+\sqrt{\frac{n_A}{n_B}}\right)$$

Since pivotal probabilities vanish as $n \to \infty$, $|k+v| \cdot Pr[Piv\Omega|nq_A(k,n), nq_B(k,n)] < C$ for sufficiently large n. Recall that $\rho(x) = \min\{\frac{x}{C}, 1\}$, equations (13) and (14) imply (15) and (16), respectively, for sufficiently large n.

$$q_A(k,n) = \int_{-k}^{\delta} \frac{|k+v| \cdot Pr[PivA|n_A,n_B]}{C} dG(v) = \frac{\alpha(k)}{C} Pr[PivA|n_A,n_B] \tag{15}$$

$$q_B(k,n) = \int_{-\delta}^{-k} \frac{|k+v| \cdot Pr[PivB|n_A,n_B]}{C} dG(v) = \frac{\beta(k)}{C} Pr[PivB|n_A,n_B] \tag{16}$$

where $\alpha(k) \equiv \int_{-\delta}^{\delta} |k+v| dG(v)$ and $\beta(k) \equiv \int_{-\delta}^{-k} |k+v| dG(v)$. Under the assumption that $G(\cdot)$ is a symmetric distribution on $[-\delta, \delta]$, there are[35]

$$\alpha(k) = \int_{-k}^{\delta} |k+v| dG(v) = k + \delta - \int_{-k}^{\delta} G(v)dv = k + \int_{-\delta}^{-k} G(v)dv \tag{17}$$

$$\beta(k) = \int_{-\delta}^{-k} |k+v| dG(v) = -\int_{-\delta}^{-k} (k+v) dG(v) = \int_{-\delta}^{-k} G(v)dv \tag{18}$$

Equations (15) and (16), together with Lemma 4, imply for all $\alpha(k)$ and $\beta(k)$ that

$$\frac{q_A(k,n)}{q_B(k,n)} \approx \frac{\alpha(k)}{\beta(k)} \sqrt{\frac{q_B(k,n)}{q_A(k,n)}} \tag{19}$$

[34]In these expressions we suppress arguments k and n to avoid cluttering.

[35]In the derivation of $\alpha(k)$, the second step follows from integration by part, and the third step follows from the fact that $E_G[v] = \int_{-\delta}^{\delta} v dG(v) = \delta - \int_{-\delta}^{\delta} G(v)dv = 0$ (as $G(v)$ is symmetric). The derivation of $\beta(k)$ is analogous.

Therefore, $\lim_{n\to\infty} \frac{q_A(k,n)}{q_B(k,n)} = \left(\frac{\alpha(k)}{\beta(k)}\right)^{\frac{2}{3}} = \mu^2(k)$, where $\mu(k) \equiv \sqrt[3]{\frac{\alpha(k)}{\beta(k)}}$. Namely, along any sequence of equilibria the expected vote ratio $\frac{q_A(k,n)}{q_B(k,n)}$ must converge to $\mu^2(k)$. Since we assume $\delta > 1$, both $\alpha(k)$ and $\beta(k)$ are strictly positive for all $k \in [-1,1]$. Hence, $0 < \mu(k) < +\infty$ holds for all $k \in [-1,1]$. This allows us to apply part 2) of Lemma 4 and approximate equations (15) and (16) by

$$q_A(k,n) = \frac{n_A(k,n)}{n} \approx \frac{\alpha(k)}{C} \cdot \frac{e^{-\left(\sqrt{n_A(k,n)}-\sqrt{n_B(k,n)}\right)^2}}{4\sqrt{\pi\sqrt{n_A(k,n)n_B(k,n)}}}\left(1+\sqrt{\frac{n_B(k,n)}{n_A(k,n)}}\right) \tag{20}$$

$$q_B(k,n) = \frac{n_B(k,n)}{n} \approx \frac{\beta(k)}{C} \cdot \frac{e^{-\left(\sqrt{n_A(k,n)}-\sqrt{n_B(k,n)}\right)^2}}{4\sqrt{\pi\sqrt{n_A(k,n)n_B(k,n)}}}\left(1+\sqrt{\frac{n_A(k,n)}{n_B(k,n)}}\right) \tag{21}$$

Any sequence of equilibria $\{q_A(k,n), q_B(k,n)\}_{n\to\infty}$ must converge to a sequence of solutions of the equation system (20) and (21) for any $k \in [-1,1]$, $\{\widehat{q_A}(k,n), \widehat{q_B}(k,n)\}_{n\to\infty}$. If $0 < \mu(k) < +\infty$ and $\mu(k) \neq 1$, the solution is unique and given by:[36]

$$\widehat{q_A}(k,n) = \frac{\widehat{n_A}(k,n)}{n} = \frac{3}{2}\left(\frac{\mu(k)}{1-\mu(k)}\right)^2 \frac{\omega(z(k,n))}{n} \approx \left(\frac{\mu(k)}{1-\mu(k)}\right)^2 \frac{\ln n}{n} \tag{22}$$

$$\widehat{q_B}(k,n) = \frac{\widehat{n_B}(k,n)}{n} = \frac{3}{2}\left(\frac{1}{1-\mu(k)}\right)^2 \frac{\omega(z(k,n))}{n} \approx \left(\frac{1}{1-\mu(k)}\right)^2 \frac{\ln n}{n} \tag{23}$$

where $\omega(x)$ is the Wright Omega function that satisfies $\omega(x) + \ln \omega(x) = x$ for all real $x > 0$ and $z(k,n) \equiv \frac{2}{3}\ln n + \frac{2}{3}\ln\frac{\mu(k)+1}{\sqrt{\mu(k)}} + 2\ln|\sqrt[3]{\alpha(k)} - \sqrt[3]{\beta(k)}| + \frac{2}{3}\ln\frac{1}{4\sqrt{\pi}C} - \ln\frac{3}{2}$. The approximations in the second steps follow from $\lim_{x\to+\infty}\frac{\omega(x)}{x} = 1$, which implies $\omega(z(k,n)) \approx \frac{2}{3}\ln n$ since $z(k,n) \approx \frac{2}{3}\ln n$. Hence, both $\widehat{q_A}(k,n)$ and $\widehat{q_B}(k,n)$ converge to zero at a rate of $\frac{\ln n}{n}$ as $n \to \infty$. If $k = 0$ and thus $\mu(0) = 1$, the (unique) solution takes the form:

$$\widehat{q_A}(0,n) = \widehat{q_B}(0,n) = \left(\frac{\alpha(0)}{2\sqrt{\pi}C}\right)^{\frac{2}{3}} \frac{1}{\sqrt[3]{n}} \tag{24}$$

Hence, both $\widehat{q_A}(0,n)$ and $\widehat{q_B}(0,n)$ converge to zero at a rate of $\frac{1}{\sqrt[3]{n}}$. With equations (22) to (24), the asymptotic equilibrium vote share $VS(k,n)$ and $T(k,n)$ can be directly approx-

[36] We compute this in two steps. First, dividing (20) by (21) yields $\frac{\widehat{q_A}(k,n)}{\widehat{q_B}(k,n)} = \mu^2(k)$ (see also equation (6.3) in Myerson (2000)). Second, let $x \equiv \widehat{q_B}(k,n)$ and define $\xi \equiv \ln\left(\frac{\beta(k)(1+\mu(k))}{4C\sqrt{\pi n\mu(k)}}\right)$ and $\eta \equiv n(1-\mu(k))^2$. Then $\widehat{q_A}(k,n) = \mu^2(k)x$ and (20), (21) jointly imply $\frac{3}{2}\ln(x) = \xi - \eta x$. This equation has a unique solution $x = \frac{3}{2\eta}\omega\left(\frac{2}{3}\xi - \ln\left(\frac{3}{2\eta}\right)\right)$, where $\omega(\cdot)$ is the White Omega function. This is equivalent to expression (23). $\widehat{q_A}(k,n) = \mu^2(k)\widehat{q_B}(k,n)$, which yields (22).

imated according to their definitions:

$$VS(k,n) \approx \frac{\widehat{q_A}(k,n)}{\widehat{q_A}(k,n)+\widehat{q_B}(k,n)} = \frac{\mu^2(k)}{1+\mu^2(k)} \tag{25}$$

$$T(k,n) \approx \widehat{q_A}(k,n)+\widehat{q_B}(k,n) = \begin{cases} 2\big(\frac{\alpha(0)}{2\sqrt{\pi}C}\big)^{\frac{2}{3}}\frac{1}{\sqrt[3]{n}}, & \text{if } k=0 \\ \gamma(k)\frac{\ln n}{n}, & \text{if } k\neq 0 \end{cases} \tag{26}$$

where $\gamma(k) \equiv \frac{1+\mu^2(k)}{(1-\mu(k))^2}$. Candidate A's equilibrium winning probability, $\pi(k,n)$, can be calculated by expression (27).

$$\pi(k,n) = \frac{1}{2}Pr[T_0|n_A(k,n),n_B(k,n)] + \sum_{l=1}^{\infty} Pr[T_l|n_A(k,n),n_B(k,n)] \tag{27}$$

To approximate $\pi(k,n)$ in large elections, we apply Theorem 2 of Myerson (2000): for all $l \in Z$, it holds that

$$\frac{Pr[T_l|n_A(k,n),n_B(k,n)]}{Pr[T_0|n_A(k,n),n_B(k,n)]} \approx \left(\sqrt{\frac{n_A(k,n)}{n_B(k,n)}}\right)^l \approx \mu^l(k)$$

Note that $\mu(k)\in(0,1)$ for $k<0$, equation (27) implies that[37]

$$\begin{aligned}\pi(k,n) &= \frac{1}{2}Pr[T_0|n_A(k,n),n_B(k,n)] + \sum_{l=1}^{\infty} Pr[T_l|n_A(k,n),n_B(k,n)] \\ &\approx \frac{e^{-\left(\sqrt{n_A}-\sqrt{n_B}\right)^2}}{2\sqrt{\pi\sqrt{n_A n_B}}}\left(\frac{1}{2}+\sum_{l=1}^{\infty}\mu^l(k)\right) = \frac{e^{-\left(\sqrt{n_A}-\sqrt{n_B}\right)^2}}{4\sqrt{\pi\sqrt{n_A n_B}}}\frac{1+\mu(k)}{1-\mu(k)}\end{aligned}$$

For $\mu(k)<1$, equations (22) and (23) imply that $\sqrt{n_A(k,n)} \approx \frac{\mu(k)}{1-\mu(k)}\sqrt{\frac{3}{2}\omega(z(k,n))}$ and $\sqrt{n_B(k,n)} \approx \frac{1}{1-\mu(k)}\sqrt{\frac{3}{2}\omega(z(k,n))}$. Plugging these into the above equation yields

$$\pi(k,n) \approx \frac{1}{2\sqrt{6\pi}}\frac{e^{-\frac{3}{2}\omega(z(k,n))}}{\sqrt{\omega(z(k,n))}}\frac{1+\mu(k)}{\sqrt{\mu(k)}} = \frac{1}{2\sqrt{6\pi}}\omega(z(k,n))e^{-\frac{3}{2}z(k,n)}\frac{1+\mu(k)}{\sqrt{\mu(k)}}$$

The second step holds because $\omega(x)+\ln\omega(x)=x$ holds for all real x, which implies $\frac{e^{-\frac{3}{2}\omega(z)}}{\sqrt{\omega(z)}} = \frac{e^{-\frac{3}{2}(z-\ln\omega(z))}}{\sqrt{\omega(z)}} = \omega(z)e^{-\frac{3}{2}z}$. Note that $z(k,n) = \frac{2}{3}\ln n + \frac{2}{3}\ln\frac{\mu(k)+1}{\sqrt{\mu(k)}} + 2\ln\left(\sqrt[3]{\beta(k)}-\sqrt[3]{\alpha(k)}\right) + \frac{2}{3}\ln\frac{1}{4\sqrt{\pi}C} - \ln\frac{3}{2}$ for $k<0$. It follows immediately that $e^{-\frac{3}{2}z(k,n)} = \frac{\sqrt{\mu(k)}}{1+\mu(k)}\frac{3\sqrt{6\pi}C}{(\sqrt[3]{\beta(k)}-\sqrt[3]{\alpha(k)})^3}\frac{1}{n}$ for

[37] Again, in the last two expression k and n are suppressed for clarity.

$k < 0$. Therefore, if $k < 0$, $\pi(k,n)$ can be approximated by

$$\pi(k,n) \approx -\frac{3C}{2} \cdot \left(\frac{1}{\sqrt[3]{\alpha(k)} - \sqrt[3]{\beta(k)}} \right)^3 \cdot \frac{\omega(z(k,n))}{n} \approx -C\psi(k)\frac{\ln n}{n}$$

where $\psi(k) \equiv \left(\frac{1}{\sqrt[3]{\alpha(k)} - \sqrt[3]{\beta(k)}}\right)^3$. The last step holds since $\lim\limits_{z\to+\infty} \frac{\omega(z)}{z} = 1$ and $z(k,n) \approx \frac{2}{3}\ln n$. $\pi(k,n)$ thus vanishes to zero at a rate of $\frac{\ln n}{n}$. This proves part 3) for $k < 0$. If $k > 0$ so that $\mu(k) > 1$, using analogous reasoning we can show that $1 - \pi(k,n) \approx C\psi(k)\frac{\ln n}{n}$ and hence $\pi(k,n) \approx 1 - C\psi(k,n)\frac{\ln n}{n}$. If $k = 0$, then $\mu(k) = 1$ and $q_A(0,n) = q_B(0,n)$. $\pi(0,n) = \frac{1}{2}$ immediately follows from symmetry. Combined together, we obtain

$$\pi(k,n) \approx \begin{cases} 1 - C\psi(k)\frac{\ln n}{n}, & \text{if } k > 0 \\ \frac{1}{2}, & \text{if } k = 0 \\ -C\psi(k)\frac{\ln n}{n}, & \text{if } k < 0 \end{cases} \tag{28}$$

A.2 Proof of Theorem 1

The asymptotic approximations of $VS(k,n)$, $T(k,n)$ and $\pi(k,n)$ directly follow from expressions (25), (26) and (28), respectively. All mentioned properties of $\mu(k)$, $\psi(k)$ and $\gamma(k)$ are proved in Lemma 5.

Lemma 5 *Under Assumption 1, for all $k \in [-1,1]$ there are*

1. *$\mu(k)$ is increasing in k, $\mu(-k) = \frac{1}{\mu(k)}$ and $\mu(0) = 1$;*
2. *$\psi(-k) = -\psi(k)$, $\lim\limits_{k\downarrow 0} \psi(k) = +\infty$ and $\psi(k)$ is decreasing and convex on $(0,1]$;*
3. *$\gamma(k) = \gamma(-k)$, $\lim\limits_{k\to 0} \gamma(k) = +\infty$ and $\gamma(k)$ is decreasing in $|k|$;*
4. *if in addition Assumption 2 holds, then $\gamma(k)$ is convex on $(0,1]$.*

Proof. We first show part 1). By the definitions of $\alpha(k)$ and $\beta(k)$ (see (17) and (18)), there are $\alpha'(k) = 1 - G(-k) > 0$ and $\beta'(k) = -G(-k) < 0$. So $\alpha(k)$ increases in k whereas $\beta(k)$ decreases in k. $\mu(k) = \sqrt[3]{\frac{\alpha(k)}{\beta(k)}}$ thus increases in k. Moreover, note that[38]

$$\alpha(-k) = \int_k^\delta (-k+v)dG(v) = \int_{-\delta}^{-k} (-k-v)dG(v) = \beta(k)$$

[38]The second step follows from symmetry of $G(\cdot)$.

So $\mu(-k) = \sqrt[3]{\frac{\alpha(-k)}{\beta(-k)}} = \sqrt[3]{\frac{\beta(k)}{\alpha(k)}} = \frac{1}{\mu(k)}$. This also implies $\mu(0) = 1$. In part 2), $\psi(-k) = -\psi(k)$, $\lim_{k \downarrow 0} \psi(k) = +\infty$ and the decreasing property follow immediately from the aforementioned properties of $\alpha(k)$ and $\beta(k)$. We show $\psi(k)$ is convex on $(0, 1]$ by directly verifying its second order derivative. Taking first order derivative yields (we suppress k for clarity)

$$\psi'(k) = \underbrace{\left(\frac{1}{\sqrt[3]{\alpha} - \sqrt[3]{\beta}}\right)^4}_{A(k)} \times \underbrace{\left[\beta^{-\frac{2}{3}}\beta' - \alpha^{-\frac{2}{3}}\alpha'\right]}_{B(k)}$$

Because $\alpha'(k) > 0$ and $\beta'(k) < 0$, $A(k) > 0$ and $B(k) < 0$ for all $k > 0$. So $\psi'(k) < 0$ on $(0, 1]$. Taking first order derivatives of $A(k)$ and $B(k)$ yields

$$\begin{aligned}
A'(k) &= \frac{4}{3}\left(\frac{1}{\sqrt[3]{\alpha} - \sqrt[3]{\beta}}\right)^5 B(k) \\
B'(k) &= \frac{2}{3}\alpha^{-\frac{5}{3}}\alpha'^2 - \frac{2}{3}\beta^{-\frac{5}{3}}\beta'^2 - \alpha^{-\frac{2}{3}}\alpha'' + \beta^{-\frac{2}{3}}\beta'' \\
&= \frac{2}{3}\alpha^{-\frac{5}{3}}\beta'^2\left[\left(\frac{\alpha'}{\beta'}\right)^2 - \left(\frac{\alpha}{\beta}\right)^{\frac{5}{3}}\right] - \left[\alpha^{-\frac{2}{3}} - \beta^{-\frac{2}{3}}\right]g(-k)
\end{aligned}$$

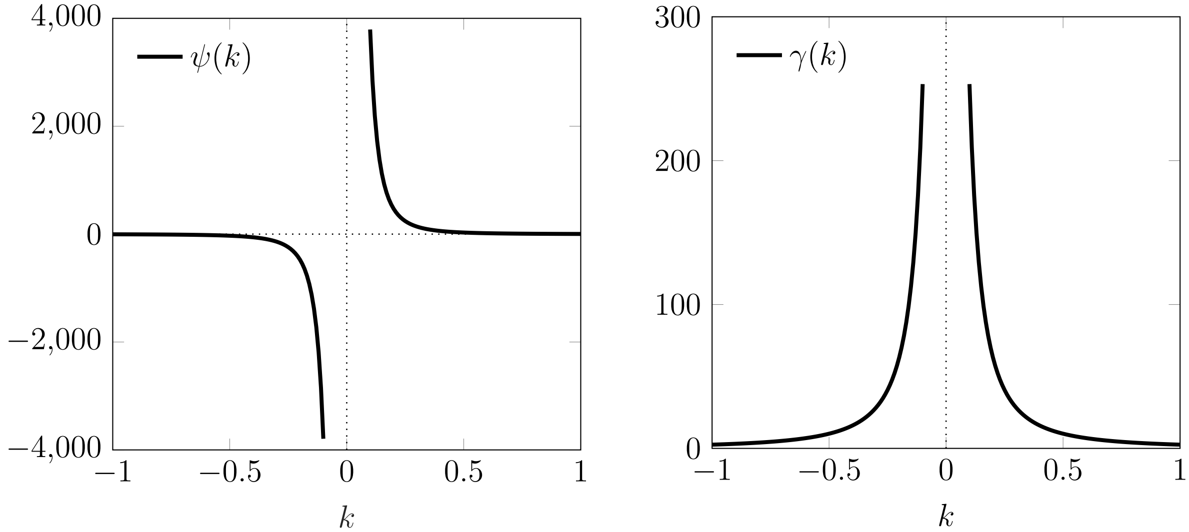
The second step in deriving $B'(k)$ exploits the fact that $\alpha''(k) = \beta''(k) = g(-k)$. Since $\alpha(k) > \beta(k)$ for $k > 0$, $A'(k) < 0$ and $B'(k) > 0$ for $k > 0$. Hence, $\psi''(k) = A'(k)B(k) + A(k)B'(k) > 0$ for $k > 0$. Because $\psi(-k) = -\psi(k)$, $\psi(k)$ is decreasing and concave on $[-1, 0)$ (cf. left panel of Figure 7). To show part 3), note that $\gamma(k) = h(\mu(k))$, where $h(x) \equiv \frac{1+x^2}{(1-x)^2}$ and satisfies $h(x) = h(\frac{1}{x})$ for all $x > 0$ and $x \neq 1$. $h(x)$ is convex and decreasing in x for $x > 1$ and satisfy $\lim_{x \to 1} h(x) = +\infty$. Part 3) then follows from combining part 1) and the aforementioned properties of $h(x)$ (cf. right panel of Figure 7). Finally, to show part 4), we rewrite $\mu(k)$ as

$$\mu(k) = \sqrt[3]{\frac{\alpha(k)}{\beta(k)}} = \sqrt[3]{\frac{k + \int_{-\delta}^{-k} G(v)dv}{\int_{-\delta}^{-k} G(v)dv}} = \sqrt[3]{\frac{k + \int_{-1}^{-k} G(v)dv + \nu}{\int_{-1}^{-k} G(v)dv + \nu}}$$

where $\nu \equiv \beta(1) = \int_{-\delta}^{-1} G(v)dv$. Below we show that $\gamma(k)$ is convex in k for all $k > 0$ if $\nu \geq 0.04$ (i.e., Assumption 2 holds). By symmetry of $G(\cdot)$, $G(v) \leq \frac{1}{2}$ for all $v < 0$ and thus $\int_{-1}^{-k} G(v)dv \leq \int_{-1}^{-k} \frac{1}{2}dv = \frac{1-k}{2}$. This implies the following inequality

$$\mu(k) = \sqrt[3]{\frac{k + \int_{-1}^{-k} G(v)dv + \nu}{\int_{-1}^{-k} G(v)dv + \nu}} \geq \sqrt[3]{\frac{k + \frac{1-k}{2} + \nu}{\frac{1-k}{2} + \nu}} = \sqrt[3]{\frac{1 + k + 2\nu}{1 - k + 2\nu}} \equiv \tilde{\mu}(k)$$

Figure 7: Graphical illustrations of $\psi(k)$ and $\gamma(k)$



Note: $\psi(k)$ and $\gamma(k)$ are calculated based on model parameters from Example 1.

In fact, $\tilde{\mu}(k)$ represents the cubic root of the expected stakes ratio for an *extremely polarized electorate*; the distribution $G(\cdot)$ puts all probability weights on extreme ideologies $[-\delta,-1)\cup(1,\delta]$. Let $\bar{\gamma}(k)\equiv h(\tilde{\mu}(k))$. Since $h(x)$ decreases in x and $\tilde{\mu}(k)$ increases in k, $\bar{\gamma}(k)$ decreases in k. Calculating the second order derivative of $\bar{\gamma}(k)$ reveals that $\bar{\gamma}(k)$ is convex on $(0,1]$ if and only if $\nu\geq 0.04$. Below we show that the convexity of $\bar{\gamma}(k)$ implies the convexity of $\gamma(k)$, and thus complete the proof. Consider any $k_1,k_2,k_3\in(0,1]$ with $k_1<k_2<k_3$. Define $y(k)\equiv\tilde{\mu}^{-1}(\mu(k))$, then $y(k)$ increases in k and

$$\frac{\gamma(k_2)-\gamma(k_1)}{k_2-k_1}=\frac{\bar{\gamma}(y(k_2))-\bar{\gamma}(y(k_1))}{y(k_2)-y(k_1)}\frac{y(k_2)-y(k_1)}{k_2-k_1}$$
$$\frac{\gamma(k_3)-\gamma(k_2)}{k_3-k_2}=\frac{\bar{\gamma}(y(k_3))-\bar{\gamma}(y(k_2))}{y(k_3)-y(k_2)}\frac{y(k_3)-y(k_2)}{k_3-k_2}$$

Convexity of $\bar{\gamma}(\cdot)$ implies $\frac{\bar{\gamma}(y(k_2))-\bar{\gamma}(y(k_1))}{y(k_2)-y(k_1)}<\frac{\bar{\gamma}(y(k_3))-\bar{\gamma}(y(k_2))}{y(k_3)-y(k_2)}<0$. Hence, $\frac{\gamma(k_2)-\gamma(k_1)}{k_2-k_1}<\frac{\gamma(k_3)-\gamma(k_2)}{k_3-k_2}$ if $y(k)$ is concave on $(0,1]$, i.e., $\frac{y(k_2)-y(k_1)}{k_2-k_1}>\frac{y(k_3)-y(k_2)}{k_3-k_2}>0$. Note that $y(k)=\tilde{\mu}^{-1}(\mu(k))=(2\nu+1)\frac{\mu^3(k)-1}{\mu^3(k)+1}=(2\nu+1)\frac{\alpha(k)-\beta(k)}{\alpha(k)+\beta(k)}=(2\nu+1)\frac{k}{k+2\int_{-\delta}^{-k}G(v)dv}$, where the last two steps follow from the definitions of $\alpha(k)$, $\beta(k)$ and $\mu(k)$. Concavity of $y(k)$ can be straightforwardly confirmed by computing its second order derivative. ■

Finally, let $E_A\equiv\{\{(v_i,c_i)\}_{i\in N}|k,n,\text{A wins}\}$ denote the set of voter type profiles conditional on k, n and the event that candidate A wins, were all voters behave according to the equilibrium strategies. Let $E[v|E_A]$ denote the expected ideology v conditional on E_A. Recall from equation (1) that any voter i's utility equals $k+v_i$ (0) if candidate A (B) is elected. So the expected utility of a randomly sampled voter, given k and n, equals

$$W(k,n) = \pi(k,n)\big(k + E[v|E_A]\big) - \frac{C \cdot T(k,n)}{2}$$

The first term represents a randomly sampled voter's expected payoff from the election outcome. The second term $\frac{C \cdot T(k,n)}{2}$ represents the expected voting costs paid by a randomly selected voter, and it vanishes in large elections as $T(k,n) \to 0$. Since $v_i \in [-\delta, \delta]$ for all voter $i \in N$ and $\delta < +\infty$, $E[v|E_A]$ must be bounded. By (28), $\pi(k,n) \approx 1(0)$ for $k > (<)0$. For $k < 0$, $\pi(k,n) \approx 0$ and the first term vanishes so $W(k,n) \approx 0$. For $k > 0$, there are $\pi(k,n) \to 1$ and $\pi(k,n)E[v|E_A] = -\big(1 - \pi(k,n)\big)E[v|\widetilde{E_A}] \to 0$.[39] The second step holds because $E_G(v) = 0$ (Assumption 1). The third step holds because $E[v|\widetilde{E_A}]$ is bounded. As a result $W(k,n) \approx k$ for $k > 0$. This completes the proof.

B All proofs in Section 5

B.1 Proofs in Section 5.2

Proof of Lemma 1. Part 1) and 2) follow from Assumption 1 and the definitions of $k_A(\chi)$ and $k_B(\chi)$. Part 3) follows from the law of iterated expectations. Below we show part 4). Since $k_B(\chi) < 0 \leq k_A(\chi)$, we have $E_{\mathcal{P}_C(\chi)}[|k|] = [1 - F(-\chi)]k_A(\chi) - F(-\chi)k_B(\chi)$. Note that

$$\begin{aligned}
[1 - F(-\chi)]k_A(\chi) &= \int_{-\chi}^{1} kdF(k) \\
F(-\chi)k_B(\chi) &= \int_{-1}^{-\chi} kdF(k) = -\int_{\chi}^{1} kdF(k)
\end{aligned}$$

The second equality follows from symmetry of $F(k)$. Hence,

$$E_{\mathcal{P}_C(\chi)}[|k|] = \int_{-\chi}^{1} kdF(k) + \int_{\chi}^{1} kdF(k) = \int_{-\chi}^{\chi} kdF(k) + 2\int_{\chi}^{1} kdF(k) = 2\int_{\chi}^{1} kdF(k)$$

where the second equality follows from the symmetry of $F(\cdot)$ (Assumption 1). $E_{\mathcal{P}_C(\chi)}[|k|]$ decreases in χ since $\int_{\chi}^{1} kdF(k)$ decreases in χ for $\chi \in [0,1]$. ■

Proof of Proposition 1. Part 1) follows from the fact that both $VS(k,n)$ and $\pi(k,n)$ are increasing in k (Theorem 1) and that both $k_A(\chi)$ and $k_B(\chi)$ are decreasing in χ (Lemma 1).

To show part 2a), we first show that $E_{\mathcal{P}_C(\chi)}[\pi(k,n)] = \frac{1}{2}$ for $\chi \in \{0,1\}$. If $\chi = 0$,

[39] Here $\widetilde{E}$ denotes the complement of set E.

by Assumption 1 there are $F(0) = \frac{1}{2}$, $k_A(0) = -k_B(0)$ and $\pi(k_A(0), n) = 1 - \pi(k_B(0), n)$. Hence, $E_{\mathcal{P}_C(0)}[\pi(k,n)] = \frac{\pi(k_A(0),n)+\pi(k_B(0),n)}{2} = \frac{1}{2}$. If $\chi = 1$, then $F(-1) = 0$ and $k_A(1) = 0$. So $E_{\mathcal{P}_C(1)}[\pi(k,n)] = \pi(0,n) = \frac{1}{2}$. Next, we show that $E_{\mathcal{P}_C(\chi)}[\pi(k,n)] > \frac{1}{2}$ for all $\chi \in (0,1)$ if n is sufficiently large. For any $\chi \in (0,1)$, there are $F(-\chi) < \frac{1}{2}$ and $k_A(\chi) \in (0,1)$. Note that

$$E_{\mathcal{P}_C(\chi)}[\pi(k,n)] = [1 - F(-\chi)]\pi(k_A(\chi), n) + F(-\chi)\pi(k_B(\chi), n)$$

The first term $[1 - F(-\chi)]\pi(k_A(\chi), n) > \frac{1}{2}$ must hold for sufficiently large n since $1 - F(-\chi) > \frac{1}{2}$ and $\lim_{n\to\infty} \pi(k_A(\chi), n) = 1$. As the second term $F(-\chi)\pi(k_B(\chi), n)$ is positive, $E_{\mathcal{P}_C(\chi)}[\pi(k,n)] > \frac{1}{2}$ must hold for sufficiently large n and this completes the proof of part 2a). To show part 2b), we take the first order derivative of $E_{\mathcal{P}_C(\chi)}[\pi(k,n)]$ and obtain[40]

$$\begin{aligned}\frac{\partial E_{\mathcal{P}_C(\chi)}[\pi(k,n)]}{\partial \chi} =& f(-\chi)\Bigg\{ \underbrace{\big(\pi(k_A(\chi), n) - \pi(k_B(\chi), n)\big)}_{\text{Effect I, } > 0} \\ & - \underbrace{\big(\chi + k_A(\chi)\big)\frac{\partial \pi}{\partial k}\Big|_{k=k_A(\chi)}}_{\text{Effect II}} - \underbrace{\big|\chi + k_B(\chi)\big|\frac{\partial \pi}{\partial k}\Big|_{k=k_B(\chi)}}_{\text{Effect III}} \Bigg\}\end{aligned} \tag{29}$$

Equation (29) explicitly reflects the three effects of a marginal increase in media bias on candidate A's ex-ante winning probability, as explained in the text. The aggregate impact of a marginal increase in media bias on A's winning probability depends on which is stronger, Effect I or the sum of Effect II and III. Let $\lambda(\chi, n) \equiv \big(\pi(k_A(\chi), n) - \pi(k_B(\chi), n)\big) - (\chi + k_A(\chi))\frac{\partial \pi}{\partial k}\big|_{k=k_A(\chi)} - |\chi + k_B(\chi)|\frac{\partial \pi}{\partial k}\big|_{k=k_B(\chi)}$ denote the term in the curly bracket of equation (29). We show below that $\lambda(0,n) > 0$ and $\lambda(1,n) < 0$ and hence $\frac{\partial E_{\mathcal{P}_C(\chi)}[\pi(k,n)]}{\partial \chi} > (<) 0$ for χ close to 0 (1), and hence complete the proof of part 2b).

- If $\chi = 0$, then $k_A(0) = -k_B(0)$ and hence $|k_A(0)| = |k_B(0)|$. By Theorem 1, $\frac{\partial \pi}{\partial k}\big|_{k=k_A(0)} \approx -C\psi'(k_A(0))\frac{\ln n}{n}$ and $\frac{\partial \pi}{\partial k}\big|_{k=k_B(0)} \approx -C\psi'(k_B(\chi))\frac{\ln n}{n} = -C\psi'(k_A(0))\frac{\ln n}{n}$ because $|k_A(0)| = |k_B(0)|$ and $\psi'(x) = \psi'(-x)$ for all $x \in (0,1]$. It then holds that

  $$\begin{aligned}\lambda(0,n) &= \big(\pi(k_A(0,n) - \pi(k_B(0), n)\big) - 2k_A(0)\frac{\partial \pi}{\partial k}\Big|_{k=k_A(0)} \\ &\approx \big(\pi(k_A(0), n) - \pi(k_B(0), n)\big) + 2k_A(0)C\psi'(k_A(0))\frac{\ln n}{n}\end{aligned}$$

  The second term captures the sum of Effect II and III (which is negative as $\psi'(k_A(0)) < 0$); it vanishes as $n \to \infty$ if $\chi = 0$. Effect I is always positive because $\pi(k_A(0), n) -$

[40] We use the fact that $k_A'(\chi) = -\frac{f(-\chi)}{1-F(-\chi)}\big(\chi + k_A(\chi)\big)$ and $k_B'(\chi) = \frac{f(-\chi)}{F(-\chi)}\big(\chi + k_B(\chi)\big)$ in the derivation.

$\pi(k_B(0), n)$ approaches 1 as $n \to \infty$. Hence, $\lambda(0, n) > 0$ and $\frac{\partial E_{\mathcal{P}_C(\chi)}[\pi(k,n)]}{\partial \chi} > 0$ at $\chi = 0$ for sufficiently large n. By continuity, $\frac{\partial E_{\mathcal{P}_C(\chi)}[\pi(k,n)]}{\partial \chi} > 0$ for χ sufficiently close to 0.

- If $\chi \uparrow 1$, then $k_A(\chi) \downarrow 0$. Because $\lim_{n\to\infty} \frac{\partial \pi}{\partial k}\big|_{k=0} = +\infty$, Effect II is unbounded as $n \to \infty$. Since Effect I and III are always bounded, $\lambda(\chi, n) < 0$ and $\frac{\partial E_{\mathcal{P}_C(\chi)}[\pi(k,n)]}{\partial \chi} < 0$ for sufficiently large n and for χ close to 1.

This completes the proof. ■

Proof of Proposition 2. By Theorem 1, $T(k, n)$ decreases in $|k|$ for large n and part 1) holds because $|k_A(\chi)|$ decreases in χ whereas $|k_B(\chi)|$ increases in χ. To show part 2), note for all $\chi \in [0, 1)$ that

$$\begin{aligned} E_{\mathcal{P}_C(\chi)}[T(k,n)] &= [1 - F(-\chi)]T(k_A(\chi), n) + F(-\chi)T(k_B(\chi), n) \\ &\approx \left\{[1 - F(-\chi)]\gamma(k_A(\chi)) + F(-\chi)\gamma(|k_B(\chi)|)\right\} \cdot \frac{\ln n}{n} = E_{\mathcal{P}_C(\chi)}[\gamma(|k|)] \cdot \frac{\ln n}{n} \end{aligned}$$

It then suffices to show $E_{\mathcal{P}_C(\chi)}[\gamma(|k|)]$ increases in χ on $[0, 1)$. Take any $\chi, \chi' \in [0, 1)$ such that $\chi' > \chi$. Lemma 1 implies $0 < k_A(\chi') < k_A(\chi) \leq |k_B(\chi)| < |k_B(\chi')|$ and $E_{\mathcal{P}_C(\chi)}[|k|] > E_{\mathcal{P}_C(\chi')}[|k|]$. Hence, there exists a unique $\lambda > F(-\chi')$ such that $(1-\lambda)\cdot k_A(\chi')+\lambda\cdot|k_B(\chi')| = E_{\mathcal{P}_C(\chi)}[|k|]$. Let τ denote a lottery that induces posterior expectation $k_A(\chi)$ with probability $1 - \lambda$ and $k_B(\chi)$ with probability λ. Below we show $E_{\mathcal{P}_C(\chi)}[\gamma(\cdot)] < E_\tau[\gamma(\cdot)] < E_{\mathcal{P}_C(\chi')}[\gamma(\cdot)]$.

$E_{\mathcal{P}_C(|k|)}[\gamma(|k|)] < E_\tau[\gamma(\cdot)]$ holds because the latter yields a mean-preserving spread of $\gamma(|k|)$ compared to the form and $\gamma(|k|)$ is convex. To show the second inequality, note that both $E_\tau[\gamma(\cdot)]$ and $E_{\mathcal{P}_C(\chi')}[\gamma(\cdot)]$ are convex combinations of $\gamma(k_A(\chi'))$ and $\gamma(|k_B(\chi')|)$. $E_\tau[\gamma(|k|)] < E_{\mathcal{P}_C(\chi')}[\gamma(|k|)]$ because the latter puts more weight on $\gamma(k_A(\chi'))$ and $\gamma(k_A(\chi')) > \gamma(|k_B(\chi')|)$. Finally, if $\chi = 1$, then $F(-1) = 0$ and $k_A(1) = 0$. As a result, $E_{\mathcal{P}_C(1)}[T(k, n)] = T(0, n)$, which converges to 0 at a rate of $\frac{1}{\sqrt[3]{n}}$, much slower than $\frac{\ln n}{n}$. Therefore, $E_{\mathcal{P}_C(1)}[T(k, n)] > E_{\mathcal{P}_C(\chi)}[T(k, n)]$ for any given $\chi \in [0, 1)$ if n is sufficiently large. ■

Proof of Proposition 3. Candidate A (B) is quality-superior if $k > (<)0$. By $\lim_{n\to\infty} W(k, n) = \max\{k, 0\}$, we obtain for any partition $\mathcal{P}$ that $\lim_{n\to\infty} E_\mathcal{P}[W(k, n)] = E_\mathcal{P}[\max\{k, 0\}]$. Let $x^* = \min\{|x||x \in \mathcal{P}\}$. If $x^* = 0$ then voters know ex-post whether k is above or below 0. Hence, their posterior expectation must be positive (negative) for $k > (<)0$. It then follows from Theorem 1.2 that candidate A(B) is elected almost surely if $k > (<)0$ and the superior candidate is elected with probability approaching 1 as $n \to \infty$. Because $\max\{k, 0\}$ is a convex and piece-wise linear on $[-1, 0]$ and $(0, 1]$, $E_\mathcal{P}[\max\{k, 0\}]$ is maximized and equal to $\int_0^1 kdF(k)$ if $x^* = 0 \in \mathcal{P}$. Below we consider $x^* > 0$. For any $x^* > 0$ there must be $(-x^*, x^*) \cap \mathcal{P} = \emptyset$, that is, partition $\mathcal{P}$ must reveal no information on interval $(-x^*, x^*)$. We show part 1) in two steps.

1. Candidate A is elected almost surely when $k \geq x^*$ as $n \to \infty$. This is because in all $k \geq -x^*$ voters' posterior expectation, denoted by $\theta(k)$, must be nonnegative. Suppose instead for some $k \geq x^*$ it holds that $\theta(k) \leq 0$, then k must be pooled with some state $k' < -k$ and thus $[k', k] \cap \mathcal{P} = \emptyset$. This contradicts with the fact that $x^* \leq k$ and $x^* \in \mathcal{P}$. Analogously, candidate B must be elected almost surely when $k \leq -x^*$ as $n \to \infty$. Taken together, the quality-superior candidate is elected almost surely in large elections for $k \in [-1, -x^*] \cup [x^*, 1]$.

2. Since voters cannot distinguish states in $(-x^*, x^*)$, their posterior expectation condition on $k \in (-x^*, x^*)$, denoted by θ^* must be constant. By Theorem 1, $\lim\limits_{n\to\infty} \pi(\theta, n) = 1(0)$ if $\theta > (<)0$, and $\lim\limits_{n\to\infty} \pi(\theta, n) = \frac{1}{2}$ if $\theta = 0$. In these cases, the probability of electing the quality-inferior candidate converges to $F(0) - F(-x^*)$ (for $\theta > 0$), $F(x^*) - F(0)$ (for $\theta < 0$) and $\frac{F(x^*)-F(-x^*)}{2}$ (for $\theta = 0$), respectively, as $n \to \infty$. Under Assumption 1 they all equal $\frac{1}{2} - F(-x^*)$. This in turn implies that the probability of electing the quality-superior candidate converges to $1 - (\frac{1}{2} - F(-x^*)) = \frac{1}{2} + F(-x^*)$ as $n \to \infty$.

To show part 2), note that $(-x^*, x^*) \cap \mathcal{P} = \emptyset$ for all $x^* > 0$, so $E_{\mathcal{P}}[\max\{k, 0\}|k \in (-x^*, x^*)] = \max\{E_F[k|k \in (-x^*, x^*)], 0\} = 0$. The last step holds because $E_F[k|k \in (-x^*, x^*)] = 0$, by Assumption 1. Since $\max\{k, 0\}$ is piece-wise linear on $[-1, 0]$ and $(0, 1]$, it holds that $E_{\mathcal{P}}[\max\{k, 0\}|k \leq -x^*] = 0$ and $E_{\mathcal{P}}[\max\{k, 0\}|k \geq x^*] = E_F[k|k \geq x^*]$. In aggregate, $E_{\mathcal{P}}[\max\{k, 0\}] = (1 - F(x^*)) \cdot E_F[k|k \geq x^*] = \int_{x^*}^1 kdF(k)$. ■

B.2 Proofs in Section 5.3

In this section we prove Proposition 4 and 5. Let the biases of the incumbent and the entrant be χ_1 and χ_2, respectively. Without loss of generality, let $\chi_1 \in [0, 1)$. Define $\chi_+ \equiv \max\{\chi_1, \chi_2\}$ and $\chi_- \equiv \min\{\chi_1, \chi_2\}$. For any given pair (χ_1, χ_2), let $k_{AA} \equiv E_F[k|k > \chi_-]$, $k_{AB} \equiv E_F[k|k \in (-\chi_+, -\chi_-]]$ and $k_{BB} = E_F[k|k \leq -\chi_+]$ denote voters' posterior expectations of k conditional on message profiles ("A","A"), ("A","B") (or ("B","A")) and ("B","B"), respectively. Let k_A (k_B) denote $k_A(\chi_1)$ ($k_B(\chi_1)$), it holds for the concerned outcome $\xi \in \{\pi, T\}$ that

$$E_{\mathcal{P}_C(\chi_1)}[\xi(k,n)] = [1 - F(-\chi_1)]\xi(k_A,n) + F(-\chi_1)\xi(k_B,n) \tag{30}$$

$$\begin{aligned} E_{\mathcal{P}_C(\chi_1,\chi_2)}[\xi(k,n)] =& [1 - F(-\chi_-)]\xi(k_{AA},n) + [F(-\chi_-) - F(-\chi_+)]\xi(k_{AB},n) \\ &+ F(-\chi_+)\xi(k_{BB},n) \end{aligned} \tag{31}$$

Define $d_\xi(x,y) \equiv \frac{\xi(x,n)-\xi(y,n)}{x-y}$ for all $1 \geq x > y \geq -1$. Geometrically, $d_\xi(x,y)$ measures the slope of the line segment connecting two points x and y on the graph of function $\xi(k,n)$. Lemma 6 provides geometrically straightforward conditions to compare $E_{\mathcal{P}_C(\chi_1,\chi_2)}[\xi(k,n)]$ and $E_{\mathcal{P}_C(\chi_1)}[\xi(k,n)]$, and hence determine whether media entry systematically increase or decrease the expectation of $\xi(k,n)$ from the ex-ante perspective.

Lemma 6 *For any given pair (χ_1, χ_2), it holds that*

1. *If $\chi_2 > \chi_1$, then $E_{\mathcal{P}_C(\chi_1,\chi_2)}[\xi(k,n)] > (<) E_{\mathcal{P}_C(\chi_1)}[\xi(k,n)]$ if and only if $d_\xi(k_{AB}, k_B; n) > (<) d_\xi(k_B, k_{BB}; n)$.*

2. *If $\chi_2 < \chi_1$, then $E_{\mathcal{P}_C(\chi_1,\chi_2)}[\xi(k,n)] > (<) E_{\mathcal{P}_C(\chi_1)}[\xi(k,n)]$ if and only if $d_\xi(k_{AA}, k_A; n) > (<) d_\xi(k_A, k_{AB}; n)$.*

Proof. For $\chi_2 > \chi_1$, $\chi_+ = \chi_2$ and $\chi_- = \chi_1$. It follows from (30) and (31) that

$$\begin{aligned}
&\frac{E_{\mathcal{P}_C(\chi_1,\chi_2)}[\xi(k,n)] - E_{\mathcal{P}_C(\chi_1)}[\xi(k,n)]}{F(-\chi_1)} \\
=&\frac{F(-\chi_1) - F(-\chi_2)}{F(-\chi_1)}\xi(k_{AB},n) + \frac{F(-\chi_2)}{F(-\chi_1)}\xi(k_{BB},n) - \xi(k_B,n) \\
=&\frac{F(-\chi_1) - F(-\chi_2)}{F(-\chi_1)}\Big(\xi(k_{AB},n) - \xi(k_B,n)\Big) - \frac{F(-\chi_2)}{F(-\chi_1)}\Big(\xi(k_B,n) - \xi(k_{BB},n)\Big) \\
=&\frac{F(-\chi_1) - F(-\chi_2)}{F(-\chi_1)}\Big(k_{AB} - k_B\Big)d_\xi(k_{AB},k_B;n) - \frac{F(-\chi_2)}{F(-\chi_1)}\Big(k_B - k_{BB}\Big)d_\xi(k_B,k_{BB};n) \\
=&\frac{F(-\chi_1) - F(-\chi_2)}{F(-\chi_1)}\Big(k_{AB} - k_B\Big)\Big(d_\xi(k_{AB},k_B;n) - d_\xi(k_B,k_{BB};n)\Big)
\end{aligned}$$

By law of iterated expectations there are $\frac{F(-\chi_1)-F(-\chi_2)}{F(-\chi_1)}k_{AB} + \frac{F(-\chi_2)}{F(-\chi_1)}k_{BB} = k_B$. This equality implies $\frac{F(-\chi_1)-F(-\chi_2)}{F(-\chi_1)}\big(k_{AB}-k_B\big) = \frac{F(-\chi_2)}{F(-\chi_1)}\big(k_B-k_{BB}\big)$, validating the last step of the derivation. With $\chi_2 > \chi_1$, there are $F(-\chi_1) > F(-\chi_2)$ and $k_{AB} > k_B$. Therefore, $E_{\mathcal{P}_C(\chi_1,\chi_2)}[\xi(k,n)] - E_{\mathcal{P}_C(\chi_1)}[\xi(k,n)]$ must be sign equivalent to $d_\xi(k_{AB},k_B;n) - d_\xi(k_B,k_{BB};n)$, which establishes part 1). The proof for part 2) is analogous. ■

Lemma 6 implies that for $\chi_2 > \chi_1$ and sufficiently large n, the comparison between $E_{\mathcal{P}_C(\chi_1,\chi_2)}[\xi(k,n)]$ and $E_{\mathcal{P}_C(\chi_1)}[\xi(k,n)]$ is equivalent to the comparison between $d_\xi(k_{AB},k_B;n)$

and $d_\xi(k_B, k_{BB}; n)$. For $\chi_2 < \chi_1$, the comparison between $E_{\mathcal{P}_C(\chi_1,\chi_2)}[\xi(k,n)]$ and $E_{\mathcal{P}_C(\chi_1)}[\xi(k,n)]$ is equivalent to the comparison between $d_\xi(k_{AA}, k_A; n)$ and $d_\xi(k_A, k_{AB}; n)$.

Proof of Proposition 4. We discuss three different cases.

Case 1: $\chi_2 > \chi_1$. In this case, media entry induces a finer partition on interval $[-1, -\chi_1] \subset [-1, 0]$ and there are $-1 \le k_{BB} < k_B < k_{AB} < 0$. By Lemma 6, we only need to show $d_\pi(k_{AB}, k_B; n) > d_\pi(k_B, k_{BB}; n)$ to establish $E_{\mathcal{P}_C(\chi_1,\chi_2)}[\pi(k,n)] > E_{\mathcal{P}_C(\chi_1)}[\pi(k,n)]$. By Theorem 1.2, $\pi(k,n) \approx -C\psi(k)\frac{\ln n}{n}$ for $k < 0$. It then holds that

$$d_\pi(k_{AB}, k_B; n) \approx -\frac{\psi(k_{AB}) - \psi(k_B)}{k_{AB} - k_B} C \frac{\ln n}{n}$$
$$d_\pi(k_B, k_{BB}; n) \approx -\frac{\psi(k_B) - \psi(k_{BB})}{k_B - k_{BB}} C \frac{\ln n}{n}$$

Therefore, for sufficiently large n, $d_\pi(k_{AB}, k_B; n) > d_\pi(k_B, k_{BB}; n)$ if and only if

$$\frac{\psi(k_{AB}) - \psi(k_B)}{k_{AB} - k_B} < \frac{\psi(k_B) - \psi(k_{BB})}{k_B - k_{BB}}$$

This inequality holds because $-1 \le k_{BB} < k_B < k_{AB} < 0$ and $\psi(k)$ is concave on $[-1, 0)$ (see Lemma 5). Hence, $E_{\mathcal{P}_C(\chi_1,\chi_2)}[\pi(k,n)] > E_{\mathcal{P}_C(\chi_1)}[\pi(k,n)]$ holds for $\chi_2 > \chi_1$ and sufficiently large n.

Case 2: $\chi_2 < -\chi_1$. In this case it holds that $0 < k_{AB} < k_A < k_{AA} \le 1$. By Lemma 6, we only need to show $d_\pi(k_{AA}, k_A; n) < d_\pi(k_A, k_{AB}; n)$ to establish $E_{\mathcal{P}_C(\chi_1,\chi_2)}[\pi(k,n)] < E_{\mathcal{P}_C(\chi_1)}[\pi(k,n)]$. By Theorem 1.2, $\pi(k,n) \approx 1 - C\psi(k)\frac{\ln n}{n}$ for $k > 0$. Following the analogous calculus as in Case 1, we obtain that $d_\xi(k_{AA}, k_A; n) < d_\xi(k_A, k_{AB}; n)$ for sufficiently large n if and only if

$$\frac{\psi(k_{AA}) - \psi(k_A)}{k_{AA} - k_A} > \frac{\psi(k_A) - \psi(k_{AB})}{k_A - k_{AB}}$$

This inequality holds by convexity of $\psi(k)$ on $(0, 1]$ (see Lemma 5).

Case 3: $\chi_2 \in [-\chi_1, \chi_1)$. Suppose first $\chi_2 \in (-\chi_1, \chi_1)$, there are $k_{AB} < 0 < k_A < k_{AA}$. Therefore, $\pi(k_{AB}, n) \approx -C\psi(k_{AB})\frac{\ln n}{n}$ whereas $\pi(k_A, n) \approx 1 - C\psi(k_A)\frac{\ln n}{n}$ and $\pi(k_{AA}, n) \approx 1 - C\psi(k_{AA})\frac{\ln n}{n}$. These imply

$$d_\pi(k_{AA}, k_A; n) \approx -\frac{\psi(k_{AA}) - \psi(k_A)}{k_{AA} - k_A} C \frac{\ln n}{n} \to 0$$
$$d_\pi(k_A, k_{AB}; n) \approx \frac{1 - C\big(\psi(k_A) - \psi(k_{AB})\big)\frac{\ln n}{n}}{k_A - k_{AB}} \to \frac{1}{k_A - k_{AB}}$$

Since $k_A - k_{AB} > 0$, the limit of $d_\pi(k_A, k_{AB}; n)$ is positive. Therefore, $d_\xi(k_{AA}, k_A; n) < d_\xi(k_A, k_{AB}; n)$ must hold for sufficiently large n, which also implies $E_{\mathcal{P}_C(\chi_1,\chi_2)}[\pi(k,n)] < E_{\mathcal{P}_C(\chi_1)}[\pi(k,n)]$. Finally, if $\chi_2 = -\chi_1$ then $k_{AB} = 0$ and $\pi(k_{AB}, n) = \frac{1}{2}$ for all n. It then holds that $d_\pi(k_A, k_{AB}; n) \to \frac{1}{2k_A} > 0$ and the above argument applies. ■

Proof of Proposition 5. As with the previous proof, we consider the same three cases.

Case 1: $\chi_2 > \chi_1$. As argued in the proof of Proposition 4, it holds that $-1 \le k_{BB} < k_B < k_{AB} < 0$ and that $E_{\mathcal{P}_C(\chi_1,\chi_2)}[T(k,n)] - E_{\mathcal{P}_C(\chi_1)}[T(k,n)]$ is sign equivalent to $d_T(k_{AB}, k_B; n) - d_T(k_B, k_{BB}; n)$. By Theorem 1.3, $T(k,n) \approx \gamma(k)\frac{\ln n}{n}$ for all $k \neq 0$ and $\gamma(k)$ is convex on $[-1, 0)$. These imply

$$d_T(k_{AB}, k_B; n) \approx \frac{\gamma(k_{AB}) - \gamma(k_B)}{k_{AB} - k_B}\frac{\ln n}{n}$$
$$d_T(k_B, k_{BB}; n) \approx \frac{\gamma(k_B) - \gamma(k_{BB})}{k_B - k_{BB}}\frac{\ln n}{n}$$

Hence, for sufficiently large n the sign of $d_T(k_{AB}, k_B; n) - d_T(k_B, k_{BB}; n)$ is determined by the sign of $\frac{\gamma(k_{AB})-\gamma(k_B)}{k_{AB}-k_B} - \frac{\gamma(k_B)-\gamma(k_{BB})}{k_B-k_{BB}}$, which is positive by convexity of $\gamma(k)$.

Case 2: $\chi_2 < -\chi_1$. Following the same argument from the proof of Proposition 4 and the approximation of $T(k,n)$, we obtain that (i) $0 < k_{AB} < k_A < k_{AA} \le 1$, and (ii) $d_T(k_{AA}, k_A; n) - d_T(k_A, k_{AB}; n)$ is sign equivalent to $\frac{\gamma(k_{AA})-\gamma(k_A)}{k_{AA}-k_A} - \frac{\gamma(k_A)-\gamma(k_{AB})}{k_A-k_{AB}}$ for sufficiently large n. Since $\gamma(k)$ is convex on $(0, 1]$, $\frac{\gamma(k_{AA})-\gamma(k_A)}{k_{AA}-k_A} - \frac{\gamma(k_A)-\gamma(k_{AB})}{k_A-k_{AB}} > 0$ and so does $d_T(k_{AA}, k_A; n) - d_T(k_A, k_{AB}; n)$. By Lemma 6, $E_{\mathcal{P}_C(\chi_1,\chi_2)}[T(k,n)] > E_{\mathcal{P}_C(\chi_1)}[T(k,n)]$ for sufficiently large n.

Case 3: $\chi_2 \in [-\chi_1, \chi_1)$. If $\chi_2 = -\chi_1$, then $k_{AB} = 0$ and $T(k,n) \approx 2\left(\frac{\alpha(0)}{2\sqrt{\pi}C}\right)^{\frac{2}{3}}\frac{1}{\sqrt[3]{n}}$. By (31), $E_{\mathcal{P}_C(\chi_1,\chi_2)}[T(k,n)]$ vanishes to zero at a rate of $\frac{1}{\sqrt[3]{n}}$, whereas $E_{\mathcal{P}_C(\chi_1)}[T(k,n)]$ vanishes at a rate of $\frac{\ln n}{n}$. So $E_{\mathcal{P}_C(\chi_1,\chi_2)}[T(k,n)] > E_{\mathcal{P}_C(\chi_1)}[T(k,n)]$ must hold for sufficiently large n. Below we focus on $\chi_2 \in (-\chi_1, \chi_1)$, where $k_{AB} < 0 < k_A < k_{AA}$. It holds that

$$\frac{d_T(k_{AA}, k_A; n)}{\ln n/n} \approx \frac{\gamma(k_{AA}) - \gamma(k_A)}{k_{AA} - k_A} \equiv \phi(\chi_1, \chi_2) \tag{32}$$
$$\frac{d_T(k_A, k_{AB}; n)}{\ln n/n} \approx \frac{\gamma(k_A) - \gamma(k_{AB})}{k_A - k_{AB}} \equiv \varphi(\chi_1, \chi_2) \tag{33}$$

Hence, $d_T(k_{AA}, k_A; n) - d_T(k_A, k_{AB}; n)$ is sign equivalent to $\phi(\chi_1, \chi_2) - \varphi(\chi_1, \chi_2)$, for sufficiently large n. Because $k_A = k_A(\chi_1)$ depends on χ_1, and both k_{AA}, k_{AB} depend on χ_1 and χ_2, both (32) and (33) are functions of χ_1 and χ_2. Let $H(\chi_1, \chi_2) \equiv \phi(\chi_1, \chi_2) - \varphi(\chi_1, \chi_2)$, then $E_{\mathcal{P}_C(\chi_1,\chi_2)}[T(k,n)] - E_{\mathcal{P}_C(\chi_1)}[T(k,n)]$ is sign equivalent to $H(\chi_1, \chi_2)$ as $n \to \infty$. Lemma 7, 8, and 9 below summarizes properties for $\phi(\chi_1, \chi_2)$, $\varphi(\chi_1, \chi_2)$ and $H(\chi_1, \chi_2)$, respectively.

Lemma 7 *[$\phi(\chi_1, \chi_2)$] It holds for all $\chi_1 \in [0, 1)$ and $\chi_2 \in (-\chi_1, \chi_1)$ that*

1. *$\phi(\chi_1, \chi_2) < 0$ and it is decreasing in both arguments.*

2. *$\lim\limits_{\chi_2 \uparrow \chi_1} \phi(\chi_1, \chi_2) = \gamma'(k_A)$.*

Proof. Recall that $k_A = k_A(\chi_1) > 0$ and it is decreasing in χ_1. With $\chi_2 < \chi_1$, it holds that $k_{AA} = k_A(\chi_2) > k_A(\chi_1)$ and k_{AA} is decreasing in χ_2. Part 1) then follows from the decreasing and convexity property of $\gamma(k)$ on $(0, 1]$. Part 2) follows from $\lim\limits_{\chi_2 \uparrow \chi_1} k_{AA} = k_A$. ■

Lemma 8 *[$\varphi(\chi_1, \chi_2)$] It holds for all $\chi_1 \in [0, 1)$ and $\chi_2 \in (-\chi_1, \chi_1)$ that*

1. *If $\varphi(\chi_1, \chi_2) \leq 0$, then $\varphi(\chi_1, \chi_2') > (<) \varphi(\chi_1, \chi_2)$ if $\chi_2' > (<) \chi_2$.*

2. *$\lim\limits_{\chi_2 \uparrow \chi_1} \varphi(\chi_1, \chi_2) = \frac{\gamma(k_A) - \gamma(-\chi_1)}{k_A + \chi_1}$ and $\lim\limits_{\chi_2 \downarrow -\chi_1} \varphi(\chi_1, \chi_2) = -\infty$.*

Proof. By (33), $\varphi(\chi_1, \chi_2) \leq 0$ implies $\gamma(k_A) \leq \gamma(k_{AB})$. Note that (i) $k_{AB} < 0$ increases in χ_2, and (ii) $\gamma(k)$ is increasing in k on $[-1, 0)$, $\gamma(k_A) - \gamma(k_{AB})$ decreases in χ_2. Consider any $\chi_2' > \chi_2$. If it holds under χ_2' that $\gamma(k_A) > \gamma(k_{AB})$, then $\varphi(\chi_1, \chi_2') > 0 \geq \varphi(\chi_1, \chi_2)$. Otherwise, $\gamma(k_A) - \gamma(k_{AB})$ is negative and its absolute value decreases with χ_2. The fact that k_{AB} implies $k_A - k_{AB}$ increases in χ_2. These results jointly imply $\varphi(\chi_1, \chi_2') > \varphi(\chi_1, \chi_2)$ for $\chi_2' > \chi_1$. $\varphi(\chi_1, \chi_2') < \varphi(\chi_1, \chi_2)$ for $\chi_2' < \chi_1$ can be proved analogously. To show part 2), note that $k_{AB} = E_F[k | k \in (-\chi_1, \chi_2]]$, which is decreasing in χ_2 and converges to $-\chi_1$ (0) as $\chi_2 \uparrow \chi_1$ ($\chi_2 \downarrow -\chi_1$). Since $\lim\limits_{k \to 0} \gamma(k) = +\infty$, the limit properties in part 2) hold. ■

Lemma 9 *[$H(\chi_1, \chi_2)$] It holds for all $\chi_1 \in [0, 1)$ and $\chi_2 \in (-\chi_1, \chi_1)$ that*

1. *If $H(\chi_1, \chi_2) \geq 0$, then $H(\chi_1, \chi_2') > 0$ if $\chi_2' < \chi_2 < \chi_1$.*

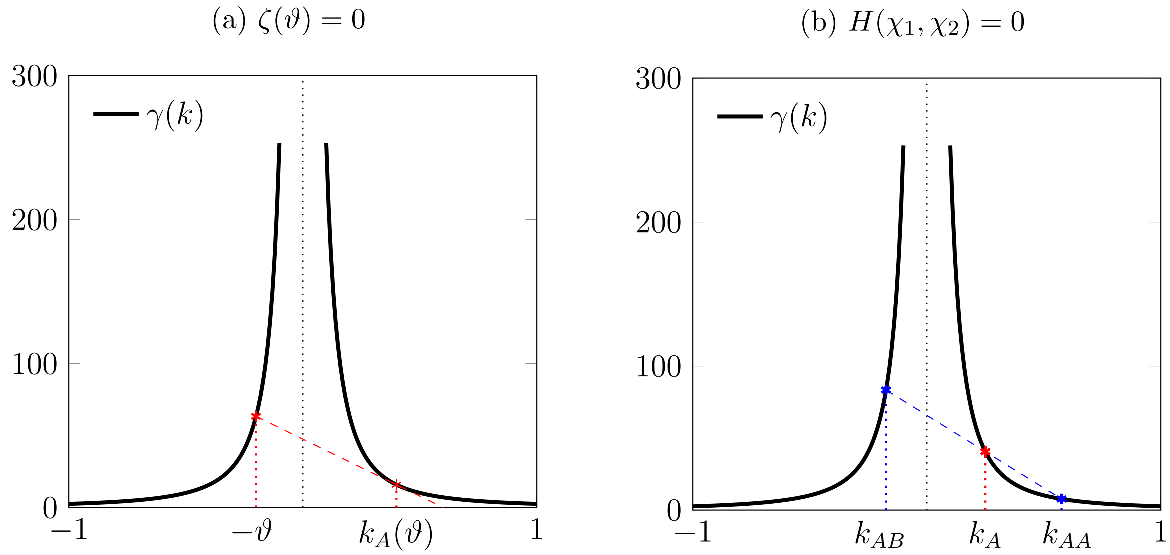
2. *If $H(\chi_1, \chi_2) \leq 0$, then $H(\chi_1, \chi_2') < 0$ if $\chi_2 < \chi_2' < \chi_1$.*

3. *If $H(\chi_1, \chi_2) \geq 0$, then $H(\chi_1', \chi_2) > 0$ if $\chi_2 < \chi_1' < \chi_1$.*

4. *If $H(\chi_1, \chi_2) \leq 0$, then $H(\chi_1', \chi_2) < 0$ if $\chi_2 < \chi_1 < \chi_1'$.*

Proof. To show part 1), note that $H(\chi_1, \chi_2) \geq 0$ implies $\varphi(\chi_1, \chi_2) \leq \phi(\chi_1, \chi_2) < 0$. Consider any $\chi_2' < \chi_2$. Lemma 7 and 8 imply $\phi(\chi_1, \chi_2') > \phi(\chi_1, \chi_2)$ and $\varphi(\chi_1, \chi_2') < \varphi(\chi_1, \chi_2)$, respectively. Therefore, $H(\chi_1, \chi_2') = \phi(\chi_1, \chi_2') - \varphi(\chi_1, \chi_2') > \phi(\chi_1, \chi_2) - \varphi(\chi_1, \chi_2) \geq 0$. Part 2) follows from analogous reasoning.

Figure 8: Geometrical illustrations for the proof of Proposition 5



To show part 3), define $\kappa(\chi_1,\chi_2) \equiv \frac{\gamma(k_{AA})-\gamma(k_{AB})}{k_{AA}-k_{AB}}$. Note that $\kappa(\chi_1,\chi_2)$ can be rewritten as $\frac{(\gamma(k_{AA})-\gamma(k_A))-(\gamma(k_A)\gamma(k_{AB}))}{(k_{AA}-k_A)-(k_A-k_{AB}})$, it follows from arithmetic necessity that $\kappa(\chi_1,\chi_2)$ must lie between $\phi(\chi_1,\chi_2)$ and $\varphi(\chi_1,\chi_2)$. Hence, $H(\chi_1,\chi_2) \geq 0$ implies $0 > \phi(\chi_1,\chi_2) \geq \kappa(\chi_1,\chi_2)$ and thus $\gamma(k_{AB}) > \gamma(k_{AA})$. Using similar arguments in Lemma 8, we can show that with $\kappa(\chi_1,\chi_2) < 0$ it holds that $\kappa(\chi_1',\chi_2) > (<)\kappa(\chi_1,\chi_2)$ if $\chi_1' > (<)\chi_1$. On the other hand, Lemma 7 implies that $\phi(\chi_1',\chi_2) > \phi(\chi_1,\chi_2)$ if $\chi_1' < (>)\chi_1$. Hence, for $\chi_1' < \chi_1$ it holds that $\phi(\chi_1',\chi_2) - \kappa(\chi_1',\chi_2) > \phi(\chi_1',\chi_2) - \kappa(\chi_1',\chi_2) \geq 0$, or equivalently, this implies $H(\chi_1',\chi_2) > 0$. Part 4) can be analogous proved. ■

It follows from the limit properties in Lemma 7 and 8 that for all $\chi_1 \in (0,1)$ it holds that $\lim\limits_{\chi_2 \downarrow -\chi_1} H(\chi_1,\chi_2) = +\infty$ and $\lim\limits_{\chi_2 \uparrow \chi_1} H(\chi_1,\chi_2) = \zeta(\chi_1)$ where

$$\zeta(\chi_1) \equiv \gamma'(k_A) - \frac{\gamma(k_A)-\gamma(-\chi_1)}{k_A+\chi_1} \tag{34}$$

It can be verified that $\lim\limits_{\chi_1 \downarrow 0} \zeta(\chi_1) = -\infty$ and $\lim\limits_{\chi_1 \uparrow 0} \zeta(\chi_1) = +\infty$. Hence, there exists a $\vartheta \in (0,1)$ such that $\zeta(\vartheta) = 0$. Moreover, $\zeta(\chi_1)$ is single-crossing at ϑ: $\zeta(\chi_1) > (<)0$ if and only if $\chi_1 > (<)\vartheta$. Geometrically, at the critical threshold ϑ the line segment connecting $-\vartheta$ and $k_A(\vartheta)$ on the graph of $\gamma(k)$ is tangent to $\gamma(k)$ at $k = k_A(\vartheta)$, as illustrated in Figure 8a.

By continuity of $H(\chi_1,\chi_2)$, $H(\chi_1,\chi_2) > 0$ for χ_2 sufficiently close to $-\chi_1$, and hence $E_{\mathcal{P}_C(\chi_1,\chi_2)}[T(k,n)] > E_{\mathcal{P}_C(\chi_1)}[T(k,n)]$ for large n. For χ_2 close to χ_2, the sign of $H(\chi_1,\chi_2)$ depends on the sign of $\zeta(\chi_1)$. If $\zeta(\chi_1) \geq 0$, or equivalently $\chi_1 \in [0,\vartheta]$, then $\lim\limits_{\chi_2 \uparrow \chi_1} H(\chi_1,\chi_2) \geq 0$. By Lemma 9 it holds that $H(\chi_1,\chi_2) > 0$, and hence $E_{\mathcal{P}_C(\chi_1,\chi_2)}[T(k,n)] > E_{\mathcal{P}_C(\chi_1)}[T(k,n)]$

for sufficiently large n, for all $\chi_2 \in (-\chi_1, \chi_1)$. Combined with all previous analyses, we establish that $E_{\mathcal{P}_C(\chi_1,\chi_2)}[T(k,n)] > E_{\mathcal{P}_C(\chi_1)}[T(k,n)]$ holds for any $\chi_2 \neq \chi_1$ if n is sufficiently large. This proves Proposition 5.1.

Below we focus on $\chi_1 > \vartheta$ so that $\zeta(\chi_1) < 0$. This implies that $H(\chi_1, \chi_2) < 0$ for χ_2 sufficiently close to χ_1. By continuity of $H(\chi_1, \chi_2)$, for any given $\chi_1 \in (0,1)$ there must exist some $\upsilon \in (-\chi_1, \chi_2)$ such that $H(\chi_1, \upsilon) = 0$; with $\chi_2 = \upsilon$ it holds that $E_{\mathcal{P}_C(\chi_1,\chi_2)}[T(k,n)] = E_{\mathcal{P}_C(\chi_1)}[T(k,n)]$ and thus media entry does not systematically affect voter turnout. Geometrically, this condition is illustrated by Figure 8b, where the three points on the graph of $\gamma(k)$, k_{AB}, k_A and k_{AA}, lie on a same straight line. Parts 1) and 2) of Lemma 9 imply that υ is unique for each eligible χ_1, and $H(\chi_1, \chi_2) > (<) 0$ if and only if $\chi_2 > (<) \upsilon$. Therefore, υ is a function of χ_1 and we denote it by $\hbar(\chi_1)$, which is implicitly solved from $H(\chi_1, \hbar(\chi_1)) = 0$. It then holds that $E_{\mathcal{P}_C(\chi_1,\chi_2)}[T(k,n)] < E_{\mathcal{P}_C(\chi_1)}[T(k,n)]$ for sufficiently large n if $\chi_2 \in (\hbar(\chi_1), \chi_1)$, and the opposite applies otherwise. To show that $\hbar(\chi_1)$ is decreasing in χ_1, note that for any $\chi_1' > \chi_1$ it holds that $H(\chi_1', \hbar(\chi_1)) < 0$ (cf. Lemma 9.4). $\hbar(\chi_1') < \hbar(\chi_1)$ must hold because $H(\chi_1', \chi_2) < 0$ for all $\chi_2 \geq \hbar(\chi_1)$ (cf. Lemma 9.2).

Finally, we show that for sufficiently large χ_1 it holds that $\hbar(\chi_1) < 0$. Fix $\chi_2 = 0$ and consider $H(\chi_1, 0)$. As $\chi_1 \to 0 = \chi_2$ it holds that (i) $k_{AB} \to 0$ and $\gamma(k_{AB}) \to +\infty$, and (ii) $k_A \to k_{AA} = k_A(0)$. These imply that (i) $\lim_{\chi_1 \to 0} \varphi(\chi_1, 0) = -\infty$, and (ii) $\lim_{\chi_1 \to 0} \phi(\chi_1, 0) = \gamma'(k_A(0))$. Therefore $\lim_{\chi_1 \to 0} H(\chi_1, 0) = +\infty$ and, by continuity, $H(\chi_1, 0) > 0$ for χ_1 close to 0. On the other hand, as $\chi_1 \to 1$ it holds that (i) $k_A \to 0$ and $\gamma(k_A) \to \infty$, (ii) $k_{AB} \to k_B(0)$. Hence, $\lim_{\chi_1 \to 1} \phi(\chi_1, 0) = -\infty$ and $\lim_{\chi_1 \to 1} \varphi(\chi_1, 0) = +\infty$. These imply $\lim_{\chi_1 \to 1} H(\chi_1, 0) = -\infty$ and thus $H(\chi_1, 0) < 0$ for χ_1 sufficiently close to 1. By continuity, there exists $\eta \in (\vartheta, 1)$ such that $H(\eta, 0) = 0$, or equivalently, $\hbar(\eta) = 0$. Since $\hbar(\chi_1)$ is decreasing, $\hbar(\chi_1) < 0$ if $\chi_1 > 0$. This completes the proof for Proposition 5.2. ■

C All proofs in Section 6

C.1 Proof of Theorem 2

Let $X \equiv \prod_{m \in M} S$ be the set of possible message profiles sent by all media outlets $m \in M$. Let $\theta(x) : X \mapsto [-1, 1]$ denote the electorate's posterior expectation of k after observing any message profile $x \in X$. Under the monotonicity constraint (9), for all $x, x' \in X$, $x \neq x'$ implies $\theta(x) \neq \theta(x')$. Let $\underline{x} \equiv \arg\min_{x \in X} \theta(x)$ and $\bar{x} \equiv \arg\min_{x \in X} \theta(x)$ denote the message profiles that induce the lowest and highest posterior expectations, respectively. In any informative equilibrium, there must be $\theta(\bar{x}) > \theta(\underline{x})$.

We start with $|M| = 1$ and let χ be the bias of this single media outlet. Note that the media prefers candidate A(B) to be elected if $k > (<) -\chi$ (see (2)). Since candidate A's winning probability is increasing in the electorate's posterior expectation, the outlet must send message $\bar{x}$ to induce the highest posterior if $k > -\chi$ and send message $\underline{x}$ to induce the lowest posterior if $k < -\chi$. Voters' posterior expectations are formed by Bayes' rule, given the media outlet's strategy; $\theta(\bar{x}) = E_F[k|k > -\chi]$ and $\theta(\underline{x}) = E_F[k|k < -\chi]$. In the most informative equilibrium, the media outlet strictly prefers to separate state $k = -\chi$, where she is indifferent between both candidates, from message $\bar{x}$ and $\underline{x}$ to make these messages more convincing.[41] In what follows we complete the proofs for $|M| \geq 2$.

C.1.1 Proof of Theorem 2.1

We complete the proof by construction. Suppose $|M| = 2$ and $\chi_1 \geq \chi_2$. We show that the reporting strategy profile and the posterior formation function below yield a PBE.

$$\sigma_1(k) = \sigma_2(k) = \begin{cases} k, & \text{if } k \leq -\chi_1 \text{ or } k \geq -\chi_2 \\ E[k|k \in (-\chi_1, -\chi_2)], & \text{if } k \in (-\chi_1, -\chi_2) \end{cases} \tag{35}$$

$$\theta(s_1, s_2) = \begin{cases} s_1, & \text{if } s_1 = s_2 \\ E[k|k \in (-\chi_1, -\chi_2)], & \text{if } s_1 \neq s_2 \end{cases} \tag{36}$$

Under the reporting strategy profile (35) and posterior formation function (36), neither media outlets can profit from any unilateral deviation and the electorate's posterior beliefs are consistent. If $k \geq -\chi_2$, then both media outlets prefer candidate A to be elected and wish to induce a high posterior expectation. If both outlets follow strategy (35) and report messages $s_1 = s_2 = k$, then the electorate forms the posterior expectation $\theta(s_1, s_2) = k \geq -\chi_2$. If any media outlet unilaterally deviates such that $s_1 \neq s_2$, then by (36) the induced posterior is $\theta(s_1, s_2) = E[k|k \in (-\chi_1, -\chi_2)] < -\chi_2$, which is lower than k. Hence, any unilateral deviation necessarily induces a lower posterior expectation, making both media outlets worse off. Analogously, if $k \leq -\chi_1$ any unilateral deviation by either media outlet is also not profitable. Finally, when $-\chi_1 < k < -\chi_2$, the interests of the two media outlets conflict; media outlet 1 (2) wishes candidate A (B) to be elected and thus wishes to induce a high (low) posterior expectation. Under (36), media outlet 1 can guarantee a posterior expectation no lower than $E[k|k \in (-\chi_1, -\chi_2)]$ by sending any message $s_1 \geq E[k|k \in (-\chi_1, -\chi_2)]$. Similarly, media outlet 2 can guarantee a posterior expectation no higher than $E[k|k \in (-\chi_1, -\chi_2)]$ by sending any message $s_2 \leq E[k|k \in (-\chi_1, -\chi_2)]$. As a result, in all equilibria

[41] Under Assumption 1 the specification of reporting strategy at the marginal state $k = -\chi$ is inconsequential since it occurs with zero probability.

the electorate's posterior expectation must equal $E[k|k \in (-\chi_1, -\chi_2)]$ were $k \in (-\chi_1, -\chi_2)$. With the reporting strategies constructed by (35), this is achieved without disagreement in equilibrium. The argument presented in Section C.1.2 below shows that any finer information partition in interval $(-\chi_1, -\chi_2)$ is impossible in any equilibrium. This completes the proof for the case $|M| = 2$.

If $|M| = 3$ and $\chi_1 \geq \chi_2 \geq \chi_3$, the following reporting strategy profile and posterior formation function yield a fully revealing PBE.[42]

$$\sigma_1(k) = \begin{cases} -1, & \text{if } k < -\chi_2 \\ k, & \text{if } k \geq -\chi_2 \end{cases} \tag{37}$$

$$\sigma_2(k) = k, \forall k \in [-1, 1] \tag{38}$$

$$\sigma_3(k) = \begin{cases} k, & \text{if } k < -\chi_2 \\ 1, & \text{if } k > -\chi_3 \end{cases} \tag{39}$$

$$\theta(s_1, s_2, s_3) = \begin{cases} s_2, & \text{if } s_2 = s_1 \geq -\chi_2 \text{ or } s_2 = s_3 < -\chi_2 \\ -\chi_2, & \text{otherwise} \end{cases} \tag{40}$$

This construction requires outlet 2 to tell the true state, and voters infer the true state by cross-checking the state reported by outlet 2 with the state reported by either outlet 1 or 3. More specifically,

- If media outlet 2 reports a high state (i.e., $s_2 \geq -\chi_2$), then voters compare s_2 with s_1, and believe $k = s_2$ only if s_1 and s_2 agree.

- If media outlet 2 reports a low state (i.e., $s_2 < -\chi_2$), then voters compare s_2 with s_3, and believe $k = s_2$ only if s_2 and s_3 agree.

Otherwise, voters form posterior expectation $\theta(s_1, s_2, s_3) = -\chi_2$. If $k \leq -\chi_2$, both outlet 2 and 3 prefer candidate B to be elected and wish to induce a low posterior. It is optimal for outlet 2 to send a low message $s_2 < -\chi_2$.[43] When $s_2 < -\chi_2$, the message from outlet 1 is irrelevant since it does not affect voters' posterior. If both outlet 2 and 3 obey the reporting strategy profile constructed above, then the induced posterior expectation is k, lower than $-\chi_2$. If any of them deviate such that $s_2 \neq s_3$, then voters' posterior expectation increases to $-\chi_2$, making both outlets worse off. Hence when $k \leq -\chi_2$ no media outlet can profit from any unilateral deviations. The argument for $k > -\chi_2$ is similar. This fully

[42]Clearly, there are many fully reveal PBE and we only construct a simple example here.

[43]By doing so she guarantees $\theta(s_1, s_2, s_3) \leq -\chi_2$. If instead she sends $s_2 \geq -\chi_2$ then $\theta(s_1, s_2, s_3) \geq -\chi_2$ for sure.

revealing PBE is constructed by exploring the alignment of interests among certain subsets of media outlets in different realized states. When $|M| > 3$, a fully revealing equilibrium can always be constructed by having three media outlets use the above strategy profile and ignore messages from all other media outlets. This completes the proof.

C.1.2 Proof of Theorem 2.2

If $|M| \geq 2$, we proceed the proof in two steps. First, we show that partition $\mathcal{P}_C = \{-\chi_1, \cdots, -\chi_{|M|}\}$ can be supported in equilibrium. We do this by explicitly constructing a profile of media outlets' reporting strategies and the electorate's posterior formation function, presented below, that induce partition $\mathcal{P}_C$ in a PBE.

$$\sigma_m(k) = \begin{cases} 1, & \text{if } k > -\chi_m \\ -1, & \text{if } k \leq -\chi_m \end{cases}, \text{for all } m \in M \tag{41}$$

$$\theta(x) = \begin{cases} E_F[k|k \in [-1, -\chi_1]], & \text{if } j = 0 \\ E_F[k|k \in (-\chi_j, -\chi_{j+1}]], & \text{if } j \in \{1, \cdots, |M|-1\} \\ E_F[k|k \in (-\chi_{|M|}, 1]], & \text{if } j = |M| \end{cases}, \; j \equiv \sum_{m \in M} \mathbb{1}_{x_m = 1} \tag{42}$$

The reporting strategy (41) is equivalent to the cutoff endorsement strategy (8) in Section 5.1, with message "A" ("B") replaced by 1 (-1). This reporting strategy profile partitions the state space into $|M| + 1$ intervals, and the electorate infers which interval contains the realized state by counting the number of message 1, or "A", sent by all media outlets. It is straightforward to verify that the posterior expectation formed by (42) is consistent with strategy profile (41). Conversely, given (42) media outlet m's best response is to send message 1 (or "A") only if $k > -\chi_m$, where she indeed prefers candidate A to be elected. This is consistent with (41) and thus the above reporting strategy profile and posterior formation function yields a PBE, which induces information partition $\mathcal{P}_C = \{-\chi_1, \cdots, -\chi_{|M|}\}$.

Second, we show that $k \notin \mathcal{P}_C$ if $k \notin \{-\chi_1, \cdots, -\chi_{|M|}\}$. In other words, no partition strictly finer than $\mathcal{P}_C = \{-\chi_1, \cdots, -\chi_{|M|}\}$ can be supported in equilibrium when coordination is possible. Essentially, media outlets with aligned partisan preferences can coordinate their messages to collectively induce the most desirable outcomes.

1. If $k \in [-1, -\chi_1)$, then all media outlets prefer candidate B to be elected and coordinate on the message profile $\underline{x}$ to induce the lowest posterior expectation. Any other message profile $x \neq \underline{x}$ will never be used since $\theta(x) > \theta(\underline{x})$. Similarly, if $k \in (-\chi_{|M|}, 1]$ then all media outlets prefer candidate A to be elected and coordinate on message profile $\bar{x}$ to induce the highest posterior expectation.

2. If $k \in (-\chi_m, -\chi_{m+1})$ for any $m \in \{1, 2, ..., |M| - 1\}$, then media outlets $j = 1, 2, ..., m$ prefer candidate A to be elected while the remaining media outlets prefer candidate B to be elected. Denote the former coalition by $M_A = \{1, ..., m\}$ and the latter coalition by $M_B = \{m+1, ..., |M|\}$. Media outlets in coalition M_A (M_B) wish to induce a posterior expectation as high (low) as possible. Define $X_{M_A} \equiv \prod_{m \in M_A} S$ and $X_{M_B} \equiv \prod_{m \in M_B} S$ the sets of message profiles possibly sent by coalition M_A and M_B, respectively. Let x_A and x_B be typical elements of M_A and M_B, respectively. Suppose an informative equilibrium exists such that for some $k_1, k_2 \in (-\chi_m, -\chi_{m+1})$, $k_1 < k_2$, message profile $x^1 \equiv (x_A^1, x_B^1) \in X_{M_A} \times X_{M_B}$ is sent in state k_1 and message profile $x^2 \equiv (x_A^2, x_B^2) \in X_{M_A} \times X_{M_B}$ is sent in state k_2. Under the monotonicity constraint (9) there must be $\theta(x^1) < \theta(x^2)$. Let $\widetilde{x} \equiv (x_A^2, x_B^1)$. The incentive compatibility constraints for coalitions M_A and M_B require that[44]

$$\theta(x^1) \geq \theta(\widetilde{x}), \text{in state } k_1$$
$$\theta(x^2) \leq \theta(\widetilde{x}), \text{in state } k_2$$

The former inequality ensures that media outlets in coalition M_A cannot profitably deviate by jointly sending message profile x_A^2 in state k_1, conditional on coalition M_B sending x_B^1. The latter inequality ensures that media outlets in coalition M_B cannot profitably deviate by jointly sending message profile x_B^1 in state k_2, conditional on coalition M_A sending x_A^2. These inequalities together implies $\theta(x^1) \geq \theta(x^2)$, contradicting with the premise $\theta(x^1) < \theta(x^2)$. Therefore, any finer information partition in interval $(-\chi_m, -\chi_{m+1})$ necessarily violates incentive compatibility constraints for coalitions, and hence cannot be supported in equilibrium. This completes the proof.

C.2 Proof of Theorem 3

Our proof builds on Proposition 3 of Gentzkow and Kamenica (2016), which states that a partition can be supported in equilibrium if and only if no senders have incentive to unilaterally induce finer partitions. Under BP, media outlet m can commit to any reporting strategy $\sigma_m(\cdot)$ prior to observing the realized k. Without loss of generality we focus on the "direct reporting strategy" $\sigma_m(\cdot)$ such that for each message s sent on equilibrium path, $s = E[k|\sigma_m(\cdot), s]$. Namely, message s precisely indicates the posterior expectation conditional on it being sent, given reporting strategy $\sigma_m(\cdot)$. We focus on a specific class of reporting

[44] We only list incentive compatibility constraints that are relevant for our proof.

strategy characterized by two parameters $(a,b) \in [-1,1]^2$ with $a<b$ as below

$$\sigma_m(k) = \begin{cases} \bar{k}(b), & \text{if } k > b \\ k, & \text{if } k \in [a,b] \\ \underline{k}(a), & \text{if } k < a \end{cases} \tag{43}$$

where $\underline{k}(a) \equiv E_F[k|k<a] = \frac{1}{F(a)}\int_{-1}^{a} kdF(k)$ and $\bar{k}(b) \equiv E_F[k|k>b] = \frac{1}{1-F(b)}\int_{b}^{1}(kdF(k))$. Strategy (43) has the feature of "pooling at tails"; the intermediate states $k \in [a,b]$ are precisely revealed whereas states at two extreme ends, $k \in [-1,a)$ and $k \in (b,1]$, are pooled. For ease of exposition, we use $\pi(k)$ to denote $\pi(k,n)$ for any fixed n. We start with $|M|=1$ and let χ be the bias of this outlet. For any given pair (a,b), we denote the media outlet's expected utility from using strategy (43) by $\widetilde{V}(a,b,\chi)$, which is given by

$$\begin{aligned} \widetilde{V}(a,b,\chi) =& \pi(\underline{k}(a))\int_{-1}^{a}(k+\chi)dF(k) + \int_{a}^{b}(k+\chi)\pi(k)dF(k) \\ &+ \pi(\bar{k}(b))\int_{b}^{1}(k+\chi)dF(k) \end{aligned} \tag{44}$$

Taking first order derivatives with respect to a and b yields

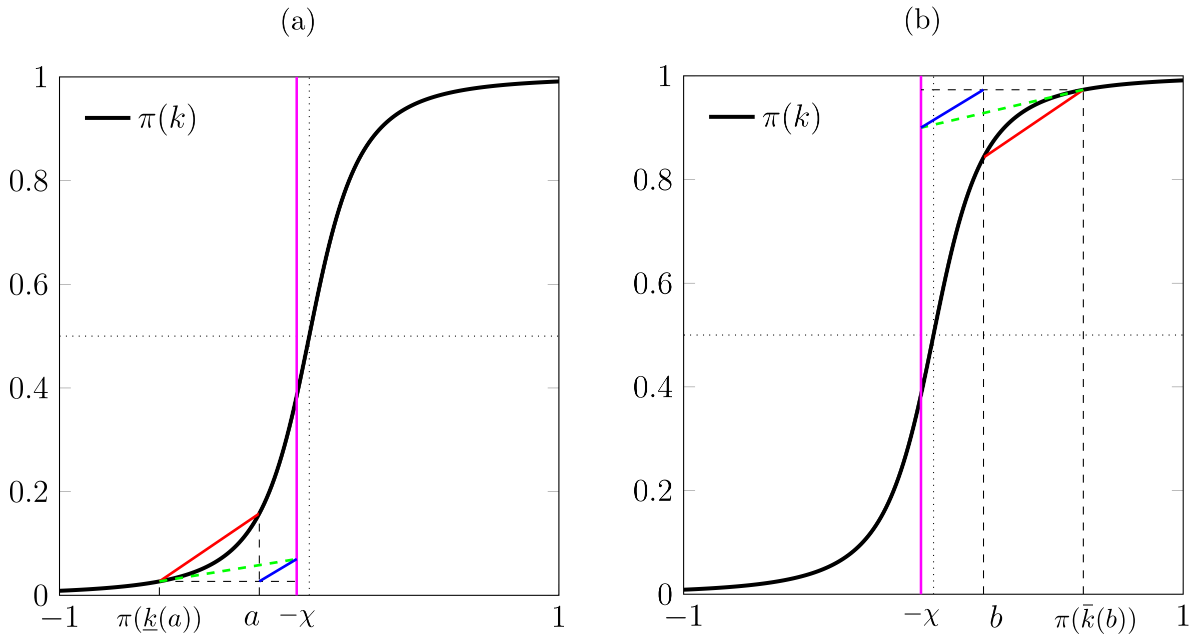
$$\begin{aligned} \frac{\partial \widetilde{V}}{\partial a} &= \pi'(\underline{k}(a))\underline{k}'(a)\int_{-1}^{a}(k+\chi)dF(k) + \pi(\underline{k}(a))(a+\chi)f(a) - (a+\chi)\pi(a)f(a) \\ &= \pi'(\underline{k}(a))(a-\underline{k}(a))(\underline{k}(a)+\chi)f(a) - (\pi(a)-\pi(\underline{k}(a)))(a+\chi)f(a) \\ &= f(a)\cdot(a+\chi)\cdot\underbrace{(a-\underline{k}(a))}_{>0}\cdot\left\{\pi'(\underline{k}(a))\frac{\underline{k}(a)+\chi}{a+\chi} - \frac{\pi(a)-\pi(\underline{k}(a))}{a-\underline{k}(a)}\right\} \end{aligned} \tag{45}$$

and

$$\begin{aligned} \frac{\partial \widetilde{V}}{\partial b} &= \pi'(\bar{k}(b))\bar{k}'(b)\int_{b}^{1}(k+\chi)dF(k) - \pi(\bar{k}(b))(b+\chi)f(b) + (b+\chi)\pi(b)f(b) \\ &= \pi'(\bar{k}(b))(\bar{k}(b)-b)(\bar{k}+\chi)f(b) - (\pi(\bar{b})-\pi(b))(b+\chi)f(b) \\ &= f(b)\cdot(b+\chi)\cdot\underbrace{(\bar{k}(b)-b)}_{>0}\cdot\left\{\pi'(\bar{k}(b))\frac{\bar{k}(b)+\chi}{b+\chi} - \frac{\pi(\bar{k}(b))-\pi(b)}{\bar{k}(b)-b}\right\} \end{aligned} \tag{46}$$

In the above derivations of (45) and (46) we implicitly use the fact that $\underline{k}'(x) = \frac{f(x)}{F(x)}(a-\underline{k}(x))$ and $\bar{k}'(x) = \frac{f(x)}{1-F(x)}(\bar{k}(b)-b)$, respectively. So the media outlet has no incentive to reveal

Figure 9: The geometric illustrations of conditions (47) and (48)



Note: Panel (a) illustrates the geometric condition when the first order condition (47) is binding. The RHS of (47) is equal to the slope of the red line segment, while the LHS of (47) is equal to the slope of the blue line segment. Condition (47) is binding if and only if the slopes of the red and blue line segments are equal. Similarly, Panel (b) illustrates the geometric feature when condition (48) is binding; the slopes of red and blue line segments are equal.

extra information at the margins (both downwards and upwards) if and only if[45]

$$\pi'(\underline{k}(a))\frac{\underline{k}(a)+\chi}{a+\chi} \geq \frac{\pi(a)-\pi(\underline{k}(a))}{a-\underline{k}(a)} \tag{47}$$

and

$$\pi'(\bar{k}(b))\frac{\bar{k}(b)+\chi}{b+\chi} \geq \frac{\pi(\bar{k}(b))-\pi(b)}{\bar{k}(b)-b} \tag{48}$$

The optimal choices $a(\chi)$ and $b(\chi)$ are obtained when both first order conditions (47) and (48) are binding. Recall that $\pi(k)$ is a brief notation for $\pi(k,n)$, $a(\chi)$ and $b(\chi)$ are also functions of n and are corresponding to $a(\chi;n)$ and $b(\chi;n)$ stated in Theorem 3, respectively. Figure 9 provides geometric illustrations for the optimality conditions (47) and (48).

To guarantee that these first order conditions do yield the optimal reporting strategy, we assume the following single-crossing conditions.

Assumption 3 $\forall\chi \in [-1,1]$, *the follow two single-crossing conditions hold:*

1. $\pi'(\underline{k}(a))\frac{\underline{k}(a)+\chi}{a+\chi} - \frac{\pi(a)-\pi(\underline{k}(a))}{a-\underline{k}(a)}$ *crosses* 0 *at most once and from below for* $a \in [-1,-\chi)$.

[45]It is straightforward to verify that $\frac{\partial\widetilde{V}}{\partial a} > 0$ and $\frac{\partial\widetilde{V}}{\partial b} < 0$ for $a > -\chi$ and $b < -\chi$. Hence, it is sufficient to focus on $a < -\chi$ and $b > -\chi$.

2. $\pi'(\bar{k}(b))\frac{\bar{k}(b)+\chi}{b+\chi} - \frac{\pi(\bar{k}(b))-\pi(b)}{\bar{k}(b)-b}$ *cross* 0 *at most once and from above for* $b \in (-\chi, 1]$.

Assumption 3 also has straightforward geometric implications. The first part implies that, in Figure 9a, as a decreases from $-\chi$ to -1, the slope of the red line segment exceeds the slope of the blue line segment only once and from below. The second statement implies that in Figure 9b, as b increases from $-\chi$ to 1, the slope of the red line segment exceeds the slope of the blue line segment only once and from below. Under Assumption 3, we show that in the optimal reporting strategy $a(\chi)$ and $b(\chi)$ satisfy Lemma 10.

Lemma 10 *Suppose Assumption 3 holds and n is sufficiently large, there are*

1. $1 \leq a(\chi) < -\chi < b(\chi) \leq 1$ *for all* $\chi \in [-1, 1]$.

2. *Both* $a(\chi)$ *and* $b(\chi)$ *decrease in* χ.

Proof. To show part 1), let $a \uparrow -\chi$ and $b \downarrow -\chi$, the limits of the LHS of (47) and (48) are both $+\infty$. Because the RHS of (47) and (48) are finite and unrelated to χ, these conditions cannot be binding for a and b sufficiently close to $-\chi$. To show part 2), note that conditions (47) and (48) are binding at $a = a(\chi)$ and $b = b(\chi)$ by constructions of $a(\chi)$ and $b(\chi)$. Since $\underline{k}(a) < a < -\chi$ and $-\chi < b < \bar{k}(b)$, $\frac{\underline{k}(a)+\chi}{a+\chi}$ is increasing in χ while $\frac{\bar{k}(b)+\chi}{b+\chi}$ is decreasing in χ. Hence, a marginal increase in χ (holding $a = a(\chi)$ fixed) implies that condition (47) holds with strict inequality so that the media outlet has incentive to reveal more information downwards by lowering a. On the contrary, a marginal increase in χ (holding $b = b(\chi)$ fixed) violates condition (48) so that the media outlet can profitably deviate by lowering b to pool information at the margin. As a result, both $a(\chi)$ and $b(\chi)$ must decrease in χ. ■

Therefore, when $|M| = 1$ the equilibrium information transmission can be characterized as below. States in the interval $[a(\chi), b(\chi)]$ is fully revealed, whereas states in $[-1, a(\chi))$ and $(b(\chi), 1]$ are pooled respectively. Lemma 10 guarantees that the "revealing interval" $[a(\chi), b(\chi)]$ never degenerates to a singleton. When $|M| \geq 2$ and $\chi_1 \geq \chi_2 \geq \cdots \geq \chi_{|M|}$, Lemma 10 implies $a(\chi_1) \leq a(\chi_m)$ and $b(\chi_{|M|}) \geq b(\chi_m)$ for all $m \in M$. Therefore, for all $m \in M$ conditions (47) and (48) are satisfied for $k \in [-1, a(\chi_1))$ and $k \in (b(\chi_{|M|}), 1]$, respectively, and no media outlet has any incentive to induce finer information partition in these intervals. It remains to show that states $k \in [a(-\chi_1), b(\chi_{|M|})]$ must be fully revealed in any equilibrium.

It is already clear from condition (47) and (48) that media outlet 1 and $|M|$ strictly prefer to fully reveal states in intervals $[a(\chi_1), -\chi_1)$ and $(-\chi_{|M|}, a(\chi_{|M|})]$, respectively. It remains to show that states $k \in [-\chi_1, -\chi_{|m|}]$ must be fully revealed in any equilibrium. Suppose this is not true and there exists a non-degenerate interval $[x, y] \subset [-\chi_1, -\chi_{|M|}]$ with $y > x$

such that states $k \in [x, y]$ is pooled by all media outlets. Let $\tilde{k}(x, y) \equiv E_F[k|k \in [x, y]]$. The expected utility for media outlet 1 to reveal states $k \in [x, z]$, $\forall z \in [x, y]$, is

$$\widetilde{V}_{\chi_1}(z, x, y) \equiv \int_x^z (k + \chi_1)\pi(k)dF(k) + \pi(\tilde{k}(z, y)) \int_z^y (k + \chi_1)dF(k) + Cons$$

where $Cons$ summarizes the expected utility in event $k \notin [x, y]$. Taking first order derivative over z at $z = x$ yields

$$\left.\frac{\partial \widetilde{V}_{\chi_1}}{\partial z}\right|_{z=x} = f(x) \cdot \underbrace{(\tilde{k}(x, y) - x)}_{>0} \cdot \underbrace{(x + \chi_1)}_{>0} \cdot \left\{ \pi'(\tilde{k}(x, y)) \frac{\tilde{k}(x, y) + \chi_1}{x + \chi_1} - \frac{\pi(\tilde{k}(x, y)) - \pi(x)}{\tilde{k}(x, y) - x} \right\}$$

So media outlet 1 has incentive to reveal more information upwards, namely $\left.\frac{\partial \widetilde{V}_{\chi_1}}{\partial z}\right|_{z=x} > 0$, if and only if

$$\pi'(\tilde{k}(x, y)) \frac{\tilde{k}(x, y) + \chi_1}{x + \chi_1} > \frac{\pi(\tilde{k}(x, y)) - \pi(x)}{\tilde{k}(x, y) - x} \tag{49}$$

Similarly, for media outlet $|M|$ the expected utility from revealing information on interval $[w, y]$, $\forall w \in [x, y]$, equals

$$\widetilde{V}_{\chi_{|M|}}(w, x, y) \equiv \pi(\tilde{k}(x, w)) \int_x^w (k + \chi_{|M|})dF(k) + \int_y^w (k + \chi_{|M|})\pi(k)dF(k) + Cons$$

Taking derivative over w at $w = y$ yields

$$\left.\frac{\partial \widetilde{V}_{\chi_{|M|}}}{\partial w}\right|_{w=y} = f(y) \cdot \underbrace{(y - \tilde{k}(x, y))}_{>0} \cdot \underbrace{(y + \chi_{|M|})}_{<0} \cdot \left\{ \pi'(\tilde{k}(x, y)) \frac{\tilde{k}(x, y) + \chi_{|M|}}{y + \chi_{|M|}} - \frac{\pi(y) - \pi(\tilde{k}(x, y))}{y - \tilde{k}(x, y)} \right\}$$

So media outlet $|M|$ has incentive to reveal more information downwards, namely $\left.\frac{\partial \widetilde{V}_{\chi_{|M|}}}{\partial w}\right|_{w=y} < 0$, if and only if

$$\pi'(\tilde{k}(x, y)) \frac{\tilde{k}(x, y) + \chi_{|M|}}{y + \chi_{|M|}} > \frac{\pi(y) - \pi(\tilde{k}(x, y))}{y - \tilde{k}(x, y)} \tag{50}$$

We show that for all possible values of $\tilde{k}(x, y)$, at least one of conditions (49) and (50) must hold so that there either outlet 1 or $|M|$ is strictly better off by unilaterally revealing more information on $[x, y]$.

- If $\tilde{k}(x, y) < 0$, by convexity of $\pi(k)$ on $[-1, 0)$ and $\tilde{k}(x, y) > x$, there are $\pi'(\tilde{k}(x, y)) > \frac{\pi(\tilde{k}(x,y)) - \pi(x)}{\tilde{k}(x,y) - x}$. Since $\frac{\tilde{k}(x,y) + \chi_1}{x + \chi_1} > 1$, condition (49) holds and media outlet 1 can profit by unilaterally inducing a finer partition.

- If $\tilde{k}(x,y) \geq 0$, by concavity of $\pi(k)$ on $[0,1]$ and $\tilde{k}(x,y) < y$, there are $\pi'(\tilde{k}(x,y)) > \frac{\pi(y)-\pi(\tilde{k}(x,y))}{y-\tilde{k}(x,y)}$. Since $\frac{\tilde{k}(x,y)+\chi_{|M|}}{y+\chi_{|M|}} > 1$, condition (50) holds and media outlet $|M|$ can profit by unilaterally inducing a finer partition.

This completes the proof.

References

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., and Zhuravskaya, E. (2015). Radio and the Rise of the Nazis in Prewar Germany. *Quarterly Journal of Economics*, 130(4):1885–1939.

Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36.

Alonso, R. and Câmara, O. (2016). Persuading voters. *American Economic Review*, 106(11):3590–3605.

Baron, D. P. (2006). Persistent media bias. *Journal of Public Economics*, 90(1):1–36.

Battaglini, M. (2002). Multiple referrals and multidimensional cheap talk. *Econometrica*, 70(4):1379–1401.

Battaglini, M., Morton, R., and Palfrey, T. (2008). Information and Pivotal Voter Models in Large Laboratory Elections. *American Economic Review*, 98(2):194–200.

Battaglini, M., Morton, R. B., and Palfrey, T. R. (2010). The swing voter's curse in the laboratory. *Review of Economic Studies*, 77(1):61–89.

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

Besley, T. and Prat, A. (2006). Handcuffs for the grabbing hand? The role of the media in political accountability. *American Economic Review*, 96(3):720–736.

Cagé, J. (2017). Media Competition, Information Provision and Political Participation: Evidence from French Local Newspapers and Elections, 1944-2014. (August 2017). CEPR Discussion Paper No. DP12198. Available at SSRN: https://ssrn.com/abstract=3016784.

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

Chan, J. and Suen, W. (2009). Media as watchdogs: The role of news media in electoral competition. *European Economic Review*, 53(7):799–814.

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

Coate, S. and Conlin, M. (2004). A Group Rule–Utilitarian Approach to Voter Turnout: Theory and Evidence. *American Economic Review*, 94(5):1476–1504.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica*, pages 1431–1451.

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

DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2016). Voting to tell others. *Review of Economic Studies*, 84(1):143–181.

Drago, F., Nannicini, T., and Sobbrio, F. (2014). Meet the press: How voters and politicians respond to newspaper entry and exit. *American Economic Journal: Applied Economics*, 6(3):159–188.

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

Durante, R. and Knight, B. (2012). Partisan control, media bias, and viewer responses: Evidence from Berlusconi's italy. *Journal of the European Economic Association*, 10(3):451–481.

Ellingsen, S., Hernæs, Ø., et al. (2018). The impact of commercial television on turnout and public policy: Evidence from Norwegian local politics. *Journal of Public Economics*, 159(C):1–15.

Ellman, M. and Germano, F. (2009). What do the papers sell? A model of advertising and media bias. *Economic Journal*, 119(537):680–704.

Enelow, J. M. and Hinich, M. J. (1984). *The spatial theory of voting: An introduction.* Boston: Cambridge Univ. Press.

Enikolopov, R. and Petrova, M. (2015). Media capture: Empirical evidence. In *Handbook of media Economics*, volume 1, pages 687–700. Elsevier.

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

Falck, O., Gold, R., and Heblich, S. (2014). E-lections: Voting Behavior and the Internet. *American Economic Review*, 104(7):2238–2265.

Feddersen, T. and Sandroni, A. (2006a). Ethical voters and costly information acquisition. *Quarterly Journal of Political Science*, 1(3):287–311.

Feddersen, T. and Sandroni, A. (2006b). A Theory of Participation in Elections. *American Economic Review*, 96(4):1271–1282.

Feddersen, T. J. and Pesendorfer, W. (1996). The Swing Voter's Curse. *American Economic Review*, pages 408–424.

Feddersen, T. J. and Pesendorfer, W. (1999). Abstention in elections with asymmetric information and diverse preferences. *American Political Science Review*, 93(02):381–398.

Gavazza, A., Nardotto, M., and Valletti, T. M. (2015). Internet and politics: Evidence from UK local elections and local government policies. (December 2015). CEPR Discussion Paper No.DP10991. Available at SSRN: https://ssrn.com/abstract=3039013.

Gentzkow, M. (2006). Television and voter turnout. *Quarterly Journal of Economics*, 121(3):931–972.

Gentzkow, M. and Kamenica, E. (2016). Competition in persuasion. *Review of Economic Studies*, 84(1):300–322.

Gentzkow, M. and Kamenica, E. (2017). Bayesian persuasion with multiple senders and rich signal spaces. *Games and Economic Behavior*, 104:411–429.

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

Gentzkow, M. and Shapiro, J. M. (2008). Competition and Truth in the Market for News. *Journal of Economic Perspectives*, 22(2):133–154.

Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? Evidence from us daily newspapers. *Econometrica*, 78(1):35–71.

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

Gentzkow, M., Shapiro, J. M., and Sinkinson, M. (2011). The effect of newspaper entry and exit on electoral politics. *American Economic Review*, 101(7):2980–3018.

Gentzkow, M., Shapiro, J. M., and Stone, D. F. (2015). Media bias in the marketplace: Theory. In *Handbook of media economics*, volume 1, pages 623–645. Elsevier.

Gerber, A. S., Karlan, D., and Bergan, D. (2009). Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, 1(2):35–52.

Gilligan, T. W. and Krehbiel, K. (1989). Asymmetric information and legislative rules with a heterogeneous committee. *American Journal of Political Science*, pages 459–490.

Ginzburg, B. (2017). Optimal limits on hate speech. (December 12, 2017). Available at SSRN: https://ssrn.com/abstract=2127877 or http://dx.doi.org/10.2139/ssrn.2127877.

Groseclose, T. and Milyo, J. (2005). A measure of media bias. *Quarterly Journal of Economics*, 120(4):1191–1237.

Grosser, J. and Schram, A. (2006). Neighborhood information exchange and voter participation: An experimental study. *American Political Science Review*, 100(2):235–248.

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

Kolotilin, A., Mylovanov, T., Zapechelnyuk, A., and Li, M. (2017). Persuasion of a privately informed receiver. *Econometrica*, 85(6):1949–1964.

Krishna, V. and Morgan, J. (2001a). Asymmetric information and legislative rules: Some amendments. *American Political Science Review*, 95(2):435–452.

Krishna, V. and Morgan, J. (2001b). A model of expertise. *Quarterly Journal of Economics*, 116(2):747–775.

Krishna, V. and Morgan, J. (2011). Overcoming ideological bias in elections. *Journal of Political Economy*, 119(2):183–211.

Krishna, V. and Morgan, J. (2015). Majority Rule and Utilitarian Welfare. *American Economic Journal: Microeconomics*, 7(4):339–375.

Larcinese, V., Puglisi, R., and Snyder, J. M. (2011). Partisan bias in economic news: Evidence on the agenda-setting behavior of us newspapers. *Journal of Public Economics*, 95(9):1178–1189.

Ledyard, J. O. (1984). The pure theory of large two-candidate elections. *Public choice*, 44(1):7–41.

Levine, D. K. and Mattozzi, A. (2016). Voter Turnout with Peer Punishment. Mimeo, European University Institute.

Levine, D. K. and Palfrey, T. R. (2007). The paradox of voter participation? A laboratory study. *American Political Science Review*, 101(01):143–158.

Liu, S. (2018). Voting with public information. (June 22, 2018). University of Zurich, Department of Economics, Working Paper No.191, Revised version. Available at SSRN: https://ssrn.com/abstract=2600507 or http://dx.doi.org/10.2139/ssrn.2600507.

Lu, S. E. (2017). Coordination-free equilibria in cheap talk games. *Journal of Economic Theory*, 168:177–208.

Martin, G. J. and Yurukoglu, A. (2017). Bias in Cable news: Persuasion and polarization. *American Economic Review*, 107(9):2565–99.

Matsusaka, J. G. (1995). Explaining voter turnout patterns: An information theory. *Public Choice*, 84(1-2):91–117.

McMillan, J. and Zoido, P. (2004). How to subvert democracy: Montesinos in peru. *Journal of Economic Perspectives*, 18(4):69–92.

Mullainathan, S. and Shleifer, A. (2005). The Market for News. *American Economic Review*.

Myerson, R. B. (1998). Population uncertainty and Poisson games. *International Journal of Game Theory*, 27(3):375–392.

Myerson, R. B. (2000). Large Poisson games. *Journal of Economic Theory*, 94(1):7–45.

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

Palfrey, T. R. and Rosenthal, H. (1985). Voter participation and strategic uncertainty. *American Political Science Review*, 79(1):62–78.

Pew Research Center (2017). Political Polarization, 1994-2017. Washington, DC: Pew Research Center. http://www.people-press.org/interactives/political-polarization-1994-2017/.

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

Pogorelskiy, K. (2014). Correlated equilibria in voter turnout games. (October 23, 2014). APSA 2014 Annual Meeting Paper. Available at SSRN: https://ssrn.com/abstract=2451740.

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

Prat, A. and Stromberg, D. (2013). The Political Economy of Mass Media. In *Advances in Economics and Econometrics: Volume 2, Applied Economics: Tenth World Congress*, volume 50, page 135. Cambridge University Press.

Puglisi, R. and Snyder, J. M. (2011). Newspaper coverage of political scandals. *Journal of Politics*, 73(3):931–950.

Puglisi, R. and Snyder, J. M. (2015a). The balanced US press. *Journal of the European Economic Association*, 13(2):240–264.

Puglisi, R. and Snyder, J. M. (2015b). Empirical studies of media bias. In *Handbook of media economics*, volume 1, pages 647–667. Elsevier.

Ravid, E. and Doron, L. (2017). Cheap Talk with Transparent Motives. (September 11, 2017). Available at SSRN: https://ssrn.com/abstract=2941601 or http://dx.doi.org/10.2139/ssrn.2941601.

Rayo, L. and Segal, I. (2010). Optimal information disclosure. *Journal of Political Economy*, 118(5):949–987.

Rubanov, O. (2014). Asymptotic Full Revelation in Cheap Talk with Large Number of Senders. (February 7, 2014). Available at SSRN: https://ssrn.com/abstract=2479293 or http://dx.doi.org/10.2139/ssrn.2479293.

Rudin, W. (1964). *Principles of mathematical analysis*, volume 3. McGraw-hill New York.

Smith, A. and Anderson, M. (2018). Social Media Use in 2018. Pew Internet and Technology, March 1, 2018. Accessed at http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ on September 11, 2018.

Snyder, J. M. and Strömberg, D. (2010). Press coverage and political accountability. *Journal of Political Economy*, 118(2):355–408.

Sobbrio, F. (2014). Citizen-editors' endogenous information acquisition and news accuracy. *Journal of Public Economics*, 113:43–53.

Strömberg, D. (2004). Radio's impact on public spending. *Quarterly Journal of Economics*, 119(1):189–221.