The Politics of Attention

Lin Hu* Anqi Li†

This Draft: April 2018

Abstract

We develop an equilibrium theory of attention and politics. In a spatial model of electoral competition where candidates can have varying policy preferences, we examine what kinds of political behaviors capture voter's limited attention and how such concern in turn affects political outcomes. Following the seminal work of Downs (1957), we assume that voters are rationally inattentive, meaning that they can process information about candidates' random policies at a cost proportional to entropy reduction. Two salient patterns emerge in equilibrium as we increase the attention cost or garble the media technology: first, arousing and attracting voter's attention becomes harder; second, doing so leads the varying types of the candidates to adopt extreme and exaggerated policy and issue positions. We supplement our analysis with historical accounts, and discuss its relevance in the new era featured with greater media choices and distractions, as well as the rise of partisan media and fake news.

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

†Department of Economics, Washington University in St. Louis. anqili@wustl.edu. We thank Steve Callander, Matt Gentzkow, Kun Li, Ilya Segal, Chris Shannon, Ken Shotts, as well as the seminar audience at Columbia and UC Davis for insightful comments. All errors are our own.

1 Introduction

The 1790's witnessed the separation of two important figures in the U.S. history: Alexander Hamilton and James Madison. Once closed allies who coauthored the Federalist papers and shared huge successes in justifying and marketing the U.S. Constitution, the two figures started to disagree about Hamilton's economic policies, which emphasized finance, manufacturing and trade over agriculture. To fight their war and in particular, to "arouse and attract public attention," the two of them founded political parties and partisan newspapers that adopted extreme and exaggerated positions and rhetorics. To give a vivid example, while Madison himself believed that a healthy country should strike a balance between manufacturing, trade and agriculture, he urged that people look inwards, to the center of the country, to farmers, and go back to the values that made America great, namely the values of low taxes, agriculture and less trade (Feldman (2017)).

This paper develops an equilibrium theory of attention and politics. Specifically, we ask what kinds of political behaviors capture people's limited attention and how this concern in turn affects political outcomes. The premise of our analysis is that attention is a scarce resource and its utilization is governed by meticulous cost-benefit analysis. As Downs famously postulated in *An Economic Theory of Democracy* (1957): "In our model, as in real life, political decisions are made when uncertainty exists and information is obtainable only at a cost. Thus a basic step towards understanding politics is analysis of the economics of being informed, i.e., the rational utilization of scarce resources to obtain data for decision-making."

We examine a spatial model of electoral competition where each of the two candidates can have varying tastes for policies, privately known to himself. Voters are rationally inattentive, meaning that they can process (news about) candidates' random policies at a cost proportional to entropy reduction as in Sims (1998) and Sims (2003). As we increase the attention cost or garble the media technology, capturing voter's limited attention in equilibrium becomes harder and leads the varying types of the candidates to adopt extreme and exaggerated policy and issue positions. We supplement our analysis with historical accounts, and discuss its implications in the new era featured with greater media choices and distractions, as well as the rise of partisan media and fake news.

We begin by adding the above described ingredients into Downs (1957)'s model of electoral competition. For each voter, the profile of the policies proposed by the varying types of the two candidates is a random variable (hereafter, *policy state*), converted by the voter's limited attention into a decision on which candidate to vote for. The cost associated with being attentive represents the time and effort required by the processing and absorption of political contents, or holding conversations and debates in order to think things through. It is assumed to be proportional to the mutual information between the policy state and the voter's decision, scaled by a parameter representing cost shifters such as greater exposures to distractions and media choices, increasing competition for consumer eyeballs, etc. (Baum and Kernell (1999); Prior (2005); Teixeira (2014); Dunaway (2016); Perez (2017)).

We give a full characterization of equilibrium attention allocations and policies when the environment is symmetric around the median voter. As we increase the marginal attention cost, capturing voter's attention becomes harder, and doing so leads the varying types of the candidates to enlarge their policy differences or even to exaggerate the intrinsic differences between themselves. These results require minimal assumptions about voters and virtually no assumption about candidates. They may even carry over to settings where candidates can only partially honor their policy proposals, now regarded as their promises made during the campaign.

Our prediction of *attention-driven extremism and exaggeration* explains Madison's over-emphasis of agriculture over manufacturing and trade, as well as the high level of visibility earned by George W. H. Bush's unequivocal embrace of women's rights during the 1980 Republican Primary (Popkin (1994); Feldman (2017)). It resonates with the fact that campaign messages are extreme and exaggerated (Vavreck (2009)), while offering a potential explanation to the growing political polarization in the U.S. by market conditions that expose us to greater distractions and media choices.

When attention is costly, voters do not bother to notice the small policy differences between candidates, suggesting that the latter enlarge their policy differences as a means of retaining voter's attention. Surprisingly, this intuition holds true in equilibrium where policies and attention allocations can depend endogenously and subtly on each other. A key step in our proof shows that in every symmetric equilibrium, winner is determined the same way as in the original Downsian model, whereby the closest candidate to the median voter wins the election for sure, whereas equally distant candidates from the median voter split the votes evenly. Knowing this fact, voter characteristics, such as marginal attention cost, are irrelevant in the determination of equilibrium policies, and hence the above described intuition is indeed an equilibrium phenomenon.

Under entropic information cost, the monotonicity properties of voters' preferences pass seamlessly over to optimal attention allocations and voting decisions. This feature of the model, together with symmetry, helps transform the problem of aggregating different voters' decisions in a single policy state to that of comparing the same voter's utilities across symmetric policy states. As a result, we can easily compute candidates' winning probabilities and perform comparative statics analysis while bypassing the need to solve for the equilibrium explicitly.

In the first extension of the baseline model, we assume that news is a noisy signal of the policy state released by the media technology. We examine how deteriorations of news quality affect equilibrium attention allocations and policies, a concern that has become more relevant recently due to the rise of partisan media and fake news (Levendusky (2013); Barthel et al. (2016); Lee and Kent (2017); Issac (2018)). In equilibrium, the needed degree of policy difference for capturing voter's attention increases as the media technology becomes less Blackwell-informative, a phenomenon that is hereafter called *media-driven extremism and exaggeration*. This result suggests a potential channel through which the aforementioned concerns, combined with limited voter attention, can have real impact for policy settings, therefore opening the door to new welfare and empirical analyses.

In the case where policies consists of multiple issues, capturing voter's attention in equilibrium requires the use of extreme and exaggerated issue positions. This result provides a formal justification for one of the most sung folklore in political science, namely salient issue positions go hand in hand with high levels of public attention (Popkin (1994)), and it therefore demonstrates the power of our framework.

1.1 Related Literature

Models of electoral competition Downs' (1957) model of electoral competition assumes that candidates are office motivated and that players' preferences and actions are common knowledge. The prediction is a strong one, namely equilibrium policies converge to the median voter's ideal position. A large body of the subsequent literature has been developed to reverse this prediction. The most related to our work are probabilistic voting models.[1]

A common feature of probabilistic voting models is that candidates are uncertain about the median voter's position. As demonstrated by Calvert (1985b), this feature alone does not predict policy divergence. For different candidates to adopt different policies in equilibrium, one need to assume in addition that candidates are either policy motivated as in Wittman (1983) and Calvert (1985b) or are treated differentially by voters for non-policy reasons as in Aragones and Palfrey (2002). Variations of these classical models abound. For references, see the textbook treatment of Persson and Tabellini (2000) and the recent survey of Duggan (2017).

Our model differs from the above described ones in crucial aspects. First, our candidates have random preferences for policies, so our comparison of the policies adopted by the varying types of the same candidate should be distinguished from the policy divergence defined above. Second, our candidates face no aggregate-level uncertainty about voters, despite that individual voters act randomly so long as they choose to be attentive to politics. Thus, factors that affect the randomness of individual votes, such as marginal attention cost, are irrelevant in the determination of equilibrium policies, and they should not be regarded as sources of the aggregate-level uncertainty in probabilistic voting models.

Election with costly information acquisition Several recent papers add costly information acquisition into probabilistic voting models. Matějka and Tabellini (2016) examines the case of office-motivated candidates. Unlike our model where policy uncertainty arises from candidates' policy preferences, there voters face a normal random variable equal to the policy plus an exogenous shock. Contrary to our assumption that voters concern only about the utility difference from choosing one candidate over another, there the cost of information acquisition can differ across candidates due to the latter's differing informational attributes such as transparency and media coverage. Leveraging the flexibility of their framework, these authors obtain rich and detailed insights into how information acquisition interacts with policy settings in equilibrium. Their attribution of policy divergence to the differing informational attributes between candidates follows from the same reason as that given in Aragones and Palfrey (2002), which we briefly discussed in the previous paragraph.

[1]Another commonly mentioned reason for policy divergence is signalling. The idea is that candidates can use divergent policies to signal hidden characteristics that voters care about alongside with policies (see Banks (1990), Kartik and McAfee (2007) and Callander (2008) among others).

Yuksel (2014) examines a probabilistic voting model where candidates are policy motivated and voters can partially learn their aggregate preference shock. In a special case of her model where a policy consists of a single issue, improvements in the learning technology enlarge the information discrepancy between candidates and voters, resulting in greater levels of policy divergence for the same reason as that given in Wittman (1983) and Calvert (1985b).

Prato and Wolton (2016) develops a model of electoral communication where greater attention from the voter increases the probability that he discovers the winning candidate's plan once in office. Surprisingly, too much voter attention can be a curse, as the resulting high probability of successful communication may lead incompetent candidates to make inefficient policy choices.

Strategic voting with costly information acquisition A body of the strategic voting literature asks how the concern of pivotality affects voter's incentive to acquire information ex-ante (see Martinelli (2006), Gerardi and Yariv (2008) and Gershkov and Szentes (2009) among others). Chen and Yang (2011) shows that in the case of entropic cost, this concern disappears and voting becomes informative, because acquiring information that is later ignored incurs redundant costs.

Rational inattention Early development of the literature of rational inattention sought to attribute the stickiness of macroeconomic variables to information processing costs (see Sims (1998), Sims (2003), Maćkowiak and Wiederholt (2009) and Woodford (2009) among others). By contrast, the current analysis explores the flexibility of attention allocation under entropic information cost for equilibrium outcomes as in Yang (2015) and Yang (2016).[2] Under entropic cost, the monotonicity properties of voters' preferences carry seamlessly over to optimal attention allocations and voting decisions. This feature of the model enables us to conduct comparative statics analysis while bypassing the need to solve for the equilibrium explicitly.

The remainder of this paper proceeds as follows: Section 2 describes the model setup; Sections 3 and 4 present main results; Sections 5 and 6 investigate extensions of the baseline model; Section 7 concludes. See Appendix A for omitted proofs and the online appendix for additional results.

[2]Other recent efforts to understand the impact of rational inattention for strategic interactions include, but are not limited to Matějka and McKay (2012), Martin (2017) and Ravid (2017).

2 Baseline Model

2.1 Setup

Primitives There is a unit mass of infinitesimal voters and two candidates named α and β, each endowed with a type that affects his preference for policies in $\Theta = [-1, 1]$. Players are divided into camp α and camp β (e.g., left wing and right wing, pro- and anti-women's rights), depending on whether their types belong to $\Theta_\alpha = [-1, 0]$ or $\Theta_\beta = [0, 1]$. Voters' types are distributed according to a continuous probability function P with support Θ and zero median, whereas each candidate c's type is a finite random variable with support $T_c \subset \Theta_c$ and probability mass function P_c, $c = \alpha, \beta$. Denote the type of a typical voter t and that of candidate c by t_c. Suppose the distribution of types is independent across players.

Each candidate c can implement one of the policies in a finite set $A_c \subset \Theta_c$. In the case where candidate c assumes office and implements policy a, the utility of type t voter, the winning candidate and the losing candidate is $u(a, t)$, $u_+(a, t_c)$ and $u_-(a, t_{-c})$, respectively. As usual, the functions u, u_+ and u_- are continuous, and the function $u(\cdot, t)$ is strictly increasing on $[-1, t]$ and is strictly decreasing on $[t, 1]$ for all t. Contrary to the stylized assumptions made by many existing studies on electoral competition, here candidates can have both office motivations and policy preferences, as their utilities can depend on who is in charge and which policy is being implemented.

Election game The election game evolves as follows:

1. Nature draws types for players;
2. each candidate c privately observes his type and proposes a policy a_c;
3. the press releases news $\boldsymbol{\omega}$ about the *policy state* $\mathbf{a} = (a_\alpha, a_\beta)$;
4. voters attend to politics and vote for one of the candidates;
5. winner is determined by simple majority rule with even tie-breaking and implements his policy proposal in Stage 2.

We add to the original Downsian model three new ingredients: random policy state, limited voter attention and noisy news.

Candidate's strategy Each candidate c's strategy $\sigma_c : T_c \to \Delta(A_c)$ maps each of his types to a lottery on policies. A strategy profile $\sigma = (\sigma_\alpha, \sigma_\beta)$ yields a random policy state, which is non-degenerate if $|\operatorname{supp}(\sigma)| \geq 2$.

Voter's problem We make two assumptions about voters' behaviors. First, voting is expressive as in most election models, meaning that each voter chooses his most preferred candidate based on the information available.[3] Second, attending to politics enables better decisions but also incurs (opportunity) costs, as the time and effort required by the processing and absorption of the voluminous political contents, conversing with colleagues, family members and friends, or even holding debates and deliberations to help think things through, could be spent elsewhere, such as leisure and entertainment. Due to space limitations, we will proceed with model description first and postpone the discussion of supporting evidence till the next section.

Following the literature on rational inattention, we model attention as a channel that converts news about the policy state into decisions on which candidate to vote for, and assume that the cost of being attentive is proportional to the mutual information between news and decisions. The baseline model examines a simple case where news coincides with the policy state itself, i.e., $\boldsymbol{\omega} = \mathbf{a}$. By Lemma 1 of Matějka and McKay (2015), we can summarize the decisions of any arbitrary voter t in this case by an *attention rule* $m_t : \operatorname{supp}(\sigma) \to [0, 1]$, where each $m_t(\mathbf{a})$ specifies the probability that voter t chooses candidate β in policy state $\mathbf{a}$. Let

$$v(\mathbf{a}, t) = u(a_\beta, t) - u(a_\alpha, t)$$

be the voter's utility difference from choosing candidate β over candidate α in state $\mathbf{a}$, and let

$$V_t(m_t, \sigma) = \sum_{\mathbf{a} \in \operatorname{supp}(\sigma)} m_t(\mathbf{a})\, v(\mathbf{a}, t)\, \sigma(\mathbf{a})$$

be the expected utility difference under any profile (m_t, σ) of attention rule and candidates' strategies. The voter's expected payoff is then

$$V_t(m_t, \sigma) - \mu_t \cdot I(m_t, \sigma),$$

[3] As demonstrated by Chen and Yang (2011), this assumption is consistent with strategic voting under entropic information cost.

where the second term of the above expression represents the attention cost and will be further discussed in the next section.

Candidate's payoff In each policy state $\mathbf{a}$, the number of votes earned by candidate β is $\int m_t(\mathbf{a})\, dP(t)$. Define

$$w(\mathbf{a}) = \begin{cases} 0 & \text{if } \int m_t(\mathbf{a})\, dP(t) < \frac{1}{2}, \\ \frac{1}{2} & \text{if } \int m_t(\mathbf{a})\, dP(t) = \frac{1}{2}, \\ 1 & \text{if } \int m_t(\mathbf{a})\, dP(t) > \frac{1}{2}, \end{cases} \tag{2.1}$$

and let the winning probability of candidate α and β in state $\mathbf{a}$ be $w_\alpha(\mathbf{a}) = 1 - w(\mathbf{a})$ and $w_\beta(\mathbf{a}) = w(\mathbf{a})$, respectively. Under any profile of attention rules $m = (m_t)_{t\in\Theta}$ and candidates' strategies σ, the expected payoff of candidate c is equal to

$$V_c(m,\sigma) = \mathbb{E}_{m,\sigma}\left[w_c(\tilde{\mathbf{a}})\, u_+\left(\tilde{a}_c, \tilde{t}_c\right) + (1 - w_c(\tilde{\mathbf{a}}))\, u_-\left(\tilde{a}_{-c}, \tilde{t}_c\right)\right]. \tag{2.2}$$

Equilibrium A strategy profile (m^*, σ^*) constitutes a *Bayesian Nash equilibrium* of the election game if

1. each m_t^* maximizes voter t's expected payoff, taking σ^* as given:

$$m_t^* \in \underset{m_t:\text{supp}(\sigma^*)\to[0,1]}{\arg\max}\; V_t(m_t, \sigma^*) - \mu_t \cdot I(m_t, \sigma^*) \;\; \forall t;$$

2. each σ_c^* maximizes candidate c's expected payoff, taking m^* and σ_{-c}^* as given:

$$\sigma_c^* \in \underset{\sigma_c}{\arg\max}\, V_c\left(m^*, \sigma_c, \sigma_{-c}^*\right) \;\; \forall c = \alpha, \beta.$$

In the case where not all feasible policies occur on the equilibrium path, equilibrium analysis requires us to take a stand on how voters react to deviant policies that they rationally ignored in the first place. For now and only, suppose these policies will lead to (unmodeled) consequences that all players wish to avoid, e.g., a complete voter abstention. This assumption will prove to be irrelevant to our main results, and it will be relaxed when we later introduce noisy news into the analysis.

2.2 Attention Cost

The thesis of Downs (1957) has two main elements: attention is costly and attention allocation is rational. Both elements are deeply grounded in reality.

Evidence In our model, attending to politics refers to activities that improve voters' decision-makings, e.g., the processing and absorption of the voluminous political contents, conversing with colleagues, family members and friends, or even holding debates and deliberations to help think things through (Gentzkow and Shapiro (2011); Mitchell et al. (2014)). As discussed earlier, the time and effort required by these activities could be spent elsewhere such as leisure and entertainment.

Recently, the (opportunity) cost associated with attending to politics seems to have been on the rise, partly because (1) the competition for consumer eyeballs has been greatly intensified in the past two decades, and (2) there is now a plethora of opportunities to entertain and socialize that is available through cable TV, internet and digital media (Teixeira (2014); Perez (2017)).[4] Greater exposures to distractions and media choices have left footprints on our behaviors: results include a shift of attention from politics to entertainment among the majority of voters, as well as the low levels of engagement with political news and contents among mobile device users (Baum and Kernell (1999); Prior (2005); Dunaway (2016)).

Meanwhile, there is ample evidence that voters focus limited attention on the most relevant aspect of reality to their personal well-beings. For example, it is found that farmers and laborers largely disagreed about the effectiveness of Eisenhower's leadership during his first term in office, and that blacks and whites hold differing intensities of opinions on civil rights issues (Campbell et al. (1960); Vavreck (2009)).[5]

Theory To formalize Down's thesis, we model attention as a channel that converts the policy state into a random decision on which candidate to vote for. For an arbitrary voter t, the cost associated with being attentive is proportional to the mutual

[4]Teixeira (2014) estimates that the cost of consumer attention has risen by seven to nine times in the past two decades. Perez (2017) reports that mobile apps for gaming, messaging, music and social alone seize up to three hours of U.S. users' time per day as of 2016.

[5]Eisenhower aimed to discharge the agriculture surplus accumulated by his democratic predecessor and did not subsidize farmers as much. During his first term his office, farmers' votes moved strongly with variations in crop prices, whereas labors reacted more sensitively to changes in labor conditions, rhetorics of unions and major events in big cities (Campbell et al. (1960)).

information between these random variables:

$$\underbrace{\mu_t}_{\text{marginal attention cost}} \cdot \underbrace{I(m_t, \sigma)}_{\text{mutual information}} .$$

In particular, the parameter $\mu_t > 0$ captures the aforementioned cost shifters and is assumed to vary continuously with the voter's type; it is hereafter called the *marginal attention cost*.

Intuitively, if a voter pays no attention to politics, then the uncertainty he faces can be captured by the entropy of the policy state:

$$H(\sigma) = -\sum_{\mathbf{a} \in \text{supp}(\sigma)} \sigma(\mathbf{a}) \log \sigma(\mathbf{a}).$$

If he pays attention and acts according to m_t, then the residual uncertainty conditional on his decision being of $s = \alpha, \beta$ is

$$H(p(\cdot \mid s)) = -\sum_{\mathbf{a} \in \text{supp}(\sigma)} p(\mathbf{a} \mid s) \log p(\mathbf{a} \mid s),$$

where

$$p(\mathbf{a} \mid s) = \begin{cases} \dfrac{m_t(\mathbf{a})\sigma(\mathbf{a})}{p_s} & \text{if } s = \beta, \\ \dfrac{(1 - m_t(\mathbf{a}))\sigma(\mathbf{a})}{p_s} & \text{if } s = \alpha, \end{cases}$$

is the conditional probability that the policy state is $\mathbf{a}$, and

$$p_s = \begin{cases} \displaystyle\sum_{\mathbf{a} \in \text{supp}(\sigma)} m_t(\mathbf{a})\sigma(\mathbf{a}) & \text{if } s = \beta, \\ \displaystyle\sum_{\mathbf{a} \in \text{supp}(\sigma)} (1 - m_t(\mathbf{a}))\sigma(\mathbf{a}) & \text{if } s = \alpha, \end{cases}$$

is the marginal probability that the decision is s. The reduction in entropy, or mutual information:

$$I(m_t, \sigma) = H(\sigma) - \sum_s p_s H(p(\cdot \mid s)),$$

is increasing in the time and effort that the voter spends on information processing and digestion and thus constitutes a legitimate measure of attention cost.

To build more intuition, consider two extreme situations: (1) the voter votes unconditionally for one of the candidates, i.e., $m_t(\mathbf{a}) \equiv 0$ or 1, and (2) the voter always votes for his preferred candidate, i.e.,

$$m_t(\mathbf{a}) = \begin{cases} 1 & \text{if } v(\mathbf{a}, t) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

In the first situation, the decision requires no attention on the voter's part, and the mutual information between the random variables of our interest is equal to zero. In the second situation, the decision calls for perfect attention and the mutual information is equal to plus infinity.

The use of entropic cost has information theoretic foundations. As Shannon (1948) famously argues, "The fundamental problem of communication is for the receiver to be able to identify what data was generated by the source, based on the signal it receives through the channel," where source and channel here refer to the policy state and attention, respectively. Shannon's entropy determines the minimum channel capacity needed to losslessly transmit the source data as encoded in binary digits (Cover and Thomas (2006)), meaning that if a voter asks a series of yes-or-no questions about the policy state at a fixed unit cost, then the minimum expected cost is entropic cost.

3 Optimal Attention Allocation

As discussed earlier, paying attention enables voters to differentiate between the varying policy states and to act differently accordingly. The next lemma characterizes the optimal attention rule of an arbitrary voter while holding candidates' strategies fixed:

Lemma 1. *Fix any σ and t. For any $\mathbf{a} \in \operatorname{supp}(\sigma)$,*

(i) $m_t^*(\mathbf{a}) = 0$ *if and only if* $\mathbb{E}_\sigma\left[\exp\left(\mu_t^{-1} v(\tilde{\mathbf{a}}, t)\right)\right] < 1$;

(ii) $m_t^*(\mathbf{a}) = 1$ *if and only if* $\mathbb{E}_\sigma\left[\exp\left(-\mu_t^{-1} v(\tilde{\mathbf{a}}, t)\right)\right] < 1$;

(iii) otherwise $m_t^*(\mathbf{a}) \in (0,1)$ *and satisfies the following first-order condition:*

$$v(\mathbf{a}, t) = \mu_t \cdot \log\left(\frac{m_t^*(\mathbf{a})}{1 - m_t^*(\mathbf{a})} \cdot \frac{1 - \mathbb{E}_\sigma\left[m_t^*(\tilde{\mathbf{a}})\right]}{\mathbb{E}_\sigma\left[m_t^*(\tilde{\mathbf{a}})\right]}\right).$$

Proof. See Proposition 2 of Yang (2016). □

According to Lemma 1, the solution to voter's optimization problem can take two forms, depending on whether the gain from paying attention justifies the cost or not. In Cases (i) and (ii), the voter clearly prefers one candidate over another and makes a deterministic decision without being attentive. In Case (iii), the voter has no clear preference ex-ante, and the result of him paying attention is a random decision that equalizes the marginal gain from distinguishing the two candidates and the marginal cost, state by state. For notational ease, we shall hereafter write $m_t^* = 0$, $m_t^* = 1$ and $m_t^* \in (0, 1)$ in Case (i), (ii) and (iii), respectively.

A closer look at the first-order condition reveals interesting patterns. First, $m_t^*(\mathbf{a})$ depends on $\mathbf{a}$ only through $v(\mathbf{a}, t)$, meaning that one should waste no time and effort on differentiating states with the same value of v, because doing so does no good to decision-making but incurs redundant costs.

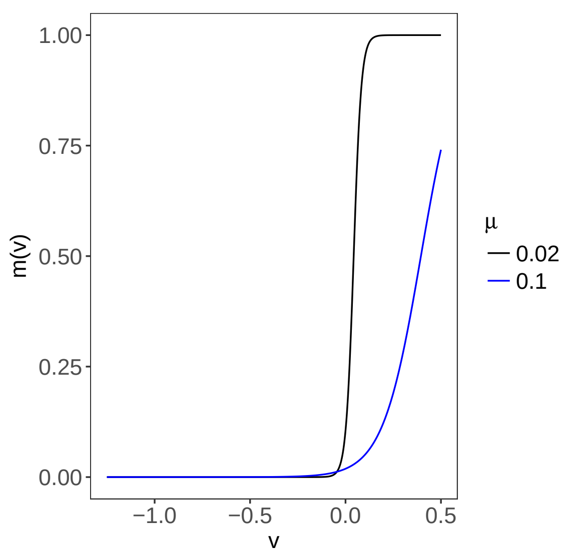

Figure 1: Plot $m_t^*(\mathbf{a})$ against $v(\mathbf{a}, t)$ for $t = -.25$: a_c is uniformly distributed on Θ_c, $c = \alpha, \beta$, and $u(a, t) = -|t - a|$.

Second, the optimal attention rule, when formulated as a function of v, is increasing in its argument, reflecting the simple fact that one should vote for a candidate more often as the gain from doing this increases. When attention is cheap, the optimal attention rule depicted in Figure 1 resembles a step function that jumps roughly from zero to one at around $v = 0$, meaning that in this case, the voter should clearly identify which candidate is more preferable than the other but nothing else. As attention cost increases, the voter no longer differentiates the varying policy states as much, and the optimal attention rule flattens out accordingly.

Perhaps most interestingly, optimal attention allocation exhibits endogenous confirmatory biases.[6] This can be best seen from Figure 2, which depicts the optimal attention rules of voters $t = 0$ and $-.25$. Compared to the median voter, the camp α voter more prefers candidate α than candidate β a priori. When attention is costly, such bias manifests itself through attention allocation, whereby the voter, constrained by limited attention, ignores candidate β unless the latter is significantly closer to the median voter than candidate α is.[7]

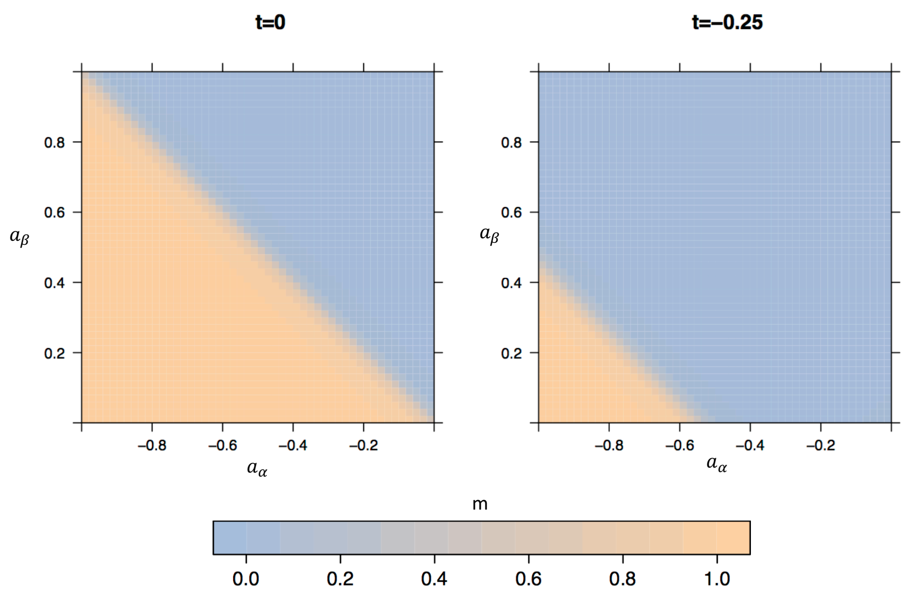

Figure 2: Plot $m_t^*(\mathbf{a})$ against $\mathbf{a}$ for $t = 0$ and $t = -.25$: a_c is uniformly distributed on Θ_c, $c = \alpha, \beta$, $u(a,t) = -|t-a|$ and $\mu = .02$. m_0^* jumps roughly from zero to one around $a_\alpha + a_\beta = 0$, whereas $m_{-.25}^*$ does so around $a_\alpha + a_\beta = -.5$.

[6] Confirmatory bias refers to the tendency to search for, interpret, favor and recall information in a way that confirms one's preexisting beliefs or hypotheses. The question of whether the consumption of political news and contents exhibits confirmatory bias has attracted much attention of economists and political scientists. For example, Gentzkow and Shapiro (2011) estimates indicies of ideological segregation using large-scale data on online and off-line news consumption. Peterson et al. (2018) updates these estimates using the data collected during the 2016 U.S. presidential campaign. We refer the reader to these papers and the references therein for a thorough review of the literature.

[7] An antecedent of this result appears in Calvert (1985a), which finds that biased voters may gain most from consulting biased experts when facing limited sources of information.

4 Equilibrium Analysis

This section conducts equilibrium analysis under the assumption that the environment is symmetric around the media voter:

Assumption 1. *For all a and t, (i) $\tilde{u}(a,t) = \tilde{u}(-a,-t)$ for all $\tilde{u} \in \{u, u_+, u_-\}$, (ii) $\mu_t = \mu_{-t}$, (iii) $dP(t) = dP(-t)$ and (iv) $P_c(t) = P_{-c}(-t)$.*

We focus on equilibrium where players' strategies are symmetric around the median voter, i.e., $\sigma_c(a \mid t) = \sigma_{-c}(-a \mid -t)$ for all $a \in A_c$ and $t \in T_c$, and $m_t(-a, a') = 1 - m_{-t}(-a', a)$ for all $t \in \Theta$ and $(-a, a') \in \text{supp}(\sigma)$.

We begin with a preview of the main results. The exposition below is kept deliberately brief; details are relegated to Appendix B.1.

Example 1. Each of the two candidates can be either the centrist type ($t = \pm 1/4$) or the extreme type ($t = \pm 3/4$) with probability $1/2$. Utility functions are $u(a,t) = -|a-t|$, $u_+(a,t) = R - \delta_+|t-a|$ and $u_-(a,t) = -\delta_-|t-a|$, where $R \geq 0$ represents the office rent, and $\delta_+, \delta_- \geq 0$ candidates' preference weights on policies.

Suppose candidates use pure symmetric strategies and the policy state is a non-degenerate random variable. Let $-a_2 < -a_1 \leq 0 \leq a_1 < a_2$ denote the policies adopted by the varying types of candidates, each realized with probability $1/2$. Figure 3 depicts equilibrium outcomes in the case of $R = 6$, $\delta_+ = 9$ and $\delta_- = 1$. In particular, the union of yellow and blue areas compiles policy profiles (a_1, a_2) that can be attained in equilibrium and, according to Theorem 1, is independent of voter characteristics such as marginal attention cost. Meanwhile, the black line depicts the boundary at which an arbitrary camp α voter, say $t = -.001$, is exactly indifferent between (1) the yellow area, where the voter pays no attention and votes unconditionally for candidate α based on ideology, and (2) the blue area, where he attends to politics and acts informatively. Theorem 2 shows that the intercept of this straight line is strictly increasing in the voter's marginal attention cost.

As marginal attention cost increases, the regime boundary moves northwest, which causes an expansion of the yellow area and a contraction of the blue area. Thus in equilibrium, arousing and attracting voter's attention becomes harder, and doing so leads the varying types of candidates to enlarge their policy differences or even to exaggerate the intrinsic differences between themselves (as in the case of $a_2 - a_1 > 3/4 - 1/4 = 1/2$). We shall hereafter refer to this phenomenon as *attention-driven extremism* and *exaggeration*.

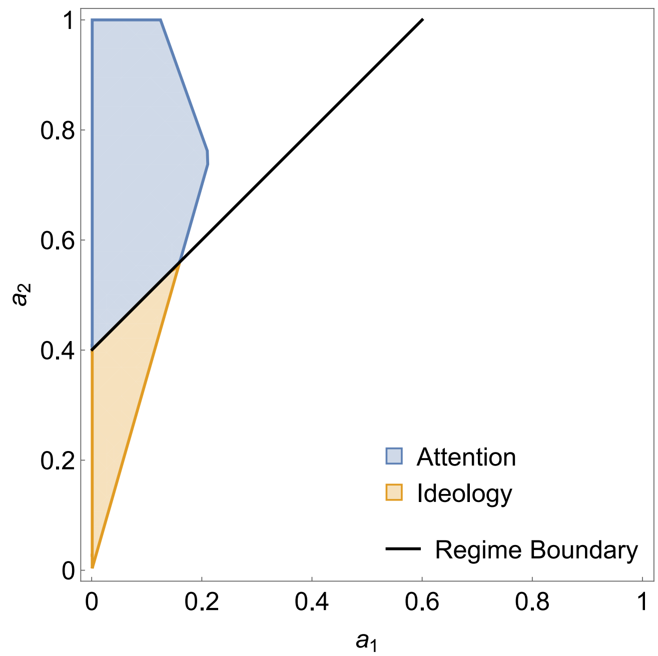

Figure 3: Equilibrium outcomes: $R = 6$, $\delta_+ = 9$, $\delta_- = 1$. The regime boundary is drawn for voter $-.001$.

When attention is costly, voters do not bother to distinguish the small policy differences between the varying types of candidates, suggesting that the latter enlarge their policy differences as a means of retaining voter's attention. Interestingly, this intuition holds true in equilibrium where policies settings can depend endogenously and subtly on attention allocations and vice versa.

The remainder of this section is organized as follows: Section 4.1 develops a matrix representation of equilibrium outcomes; Section 4.2 conducts equilibrium analysis; Section 4.3 delineates proof sketches; Section 4.4 discusses related issues.

4.1 Matrix Representation

Let $-a_N < \cdots < -a_1 \leq 0 \leq a_1 < \cdots < a_N$ denote the policies proposed by the varying types of the candidates, where $N \geq 2$ if and only if the policy state is a non-degenerate random variable. Let $\mathbf{A}$, $\mathbf{\Sigma}$ and $\mathbf{W}$ be square matrices of order N, and denote their ij^{th} entries by $\mathbf{a}_{ij}$, σ_{ij} and w_{ij}, respectively. Assume throughout that $\mathbf{a}_{ij} = (-a_i, a_j)$, $\sigma_{ij} \geq 0$, $\sigma_{ij} = \sigma_{ji}$ and $w_{ij} \in \{0, 1/2, 1\}$ for all i and j, and that $\sum_{i,j} \sigma_{ij} = 1$.

Intuitively, $\mathbf{A}$ and $\mathbf{\Sigma}$ compile all policy states and their probabilities, and are hereafter referred to as the *policy matrix* and the *probability matrix*, respectively. Likewise, $\mathbf{W}$ collects candidate β's winning probability in each policy state and is henceforth called the *winning probability matrix*.

In Example 1, the probability matrix of our interest is $\frac{1}{4}\mathbf{J}_2$, where $\mathbf{J}_2$ denotes the 2×2 matrix of all ones. In the case where policies can be observed by voters without errors, the winning probability matrix is a square matrix $\widehat{\mathbf{W}}_N$ where for all $i, j = 1, \cdots, N$,

$$\widehat{w}_{ij} = \begin{cases} 0 & \text{if } i > j, \\ \frac{1}{2} & \text{if } i = j, \\ 1 & \text{if } i < j, \end{cases} \tag{4.1}$$

meaning that in each policy state, the closest candidate to the median voter wins the election for sure, whereas equally distant candidates from the median voter split the votes evenly.

A tuple $[\mathbf{A}, \mathbf{\Sigma}, \mathbf{W}]$ of matrices can be attained in a symmetric equilibrium of the election game if $[\mathbf{A}, \mathbf{\Sigma}]$ is *incentive compatible for candidates under* $\mathbf{W}$ (hereafter, $\mathbf{W}$-IC), and if $\mathbf{W}$ can be *rationalized by optimal attention allocations under* $[\mathbf{A}, \mathbf{\Sigma}]$ (hereafter, $[\mathbf{A}, \mathbf{\Sigma}]$-rationalizable). Formally,

Definition 1. $[\mathbf{A}, \mathbf{\Sigma}]$ *is* $\mathbf{W}$-IC *if there exists a profile* σ *of candidates' strategies such that*

(i) the probabilities of policy states under σ *are given by* $\mathbf{\Sigma}$*, i.e.,*

$$\sigma(\mathbf{a}_{ij}) = \sigma_{ij} \ \forall i, j;$$

(ii) each σ_c *maximizes candidate* c*'s expected payoff, taking the winning probability matrix* $\mathbf{W}$ *and the other candidate's strategy* σ_{-c} *as given, i.e.,*

$$\sigma_c \in \arg\max_{\sigma'_c} V_c\left(\mathbf{W}, \sigma'_c, \sigma_{-c}\right),$$

where $V_c\left(\mathbf{W}, \sigma\right)$ *can be obtained from plugging* $\mathbf{W}$ *and* σ *into Equation (2.2).*

Definition 2. $\mathbf{W}$ is $[\mathbf{A}, \mathbf{\Sigma}]$-rationalizable *if*

$$w_{ij} = w(\mathbf{a}_{ij}) \ \forall i, j,$$

where $w(\mathbf{a}_{ij})$ can be obtained from plugging $m_t^*(\mathbf{a}_{ij})$, $t \in \Theta$ under $[\mathbf{A}, \mathbf{\Sigma}]$ into Equation (2.1).

An implicit assumption here, as in many existing studies of electoral competition, is that candidates care about voters' behaviors only through the winning probabilities. This observation, albeit a subtle one, has profound implications for the upcoming analysis.

4.2 Analysis

4.2.1 Equilibrium Policies

For any integer N and any probability matrix $\mathbf{\Sigma}$ of order N, define $\mathcal{E}(\mathbf{\Sigma})$ as the set of policy matrices which, when paired with $\mathbf{\Sigma}$, can arise in the symmetric equilibrium of the election game, i.e.,

$$\mathcal{E}(\mathbf{\Sigma}) = \left\{ \mathbf{A} : \exists \mathbf{W} \text{ s.t. } \begin{array}{c} [\mathbf{A}, \mathbf{\Sigma}] \text{ is } \mathbf{W} - \text{IC} \\ \mathbf{W} \text{ is } [\mathbf{A}, \mathbf{\Sigma}] - \text{rationalizable} \end{array} \right\}.$$

Our first main result gives a full characterization of the set $\mathcal{E}(\mathbf{\Sigma})$, under the assumption that voters have concave preferences for policies:

Assumption 2. *$u(a, t)$ is concave in a for all t.*

Theorem 1. *Assume Assumptions 1 and 2. Then for any integer N and any probability matrix $\mathbf{\Sigma}$ of order N,*

$$\mathcal{E}(\mathbf{\Sigma}) = \left\{ \mathbf{A} : [\mathbf{A}, \mathbf{\Sigma}] \text{ is } \widehat{\mathbf{W}}_N - \text{IC} \right\}.$$

Theorem 1 shows that in every symmetric equilibrium, winner is determined the same way as in the original Downsian model, meaning that in each policy state, the closest candidate to the median voter wins for sure, whereas equally distant candidates from the median voter split the votes evenly.

The proof sketch of Theorem 1 can be found in Section 4.3. Here we discuss its implication, which is twofold. First, the problem of finding equilibrium policies boils down to solving candidates' best responses to a known winning probability matrix, and equilibrium existence follows from standard arguments. Second, knowing that the winning probability matrix is $\widehat{\mathbf{W}}_N$, voter characteristics such as marginal attention cost are irrelevant in the determination of equilibrium policies. While these characteristics certainly affect the randomness of individual voters' decisions, it imposes *no* aggregate-level uncertainty on candidates, for whom it is as if policies could be observed by voters without errors.

4.2.2 Comparative Statics and Applications

We say that voter t *pays active attention to politics* if $m_t^* \in (0,1)$, and that he *acts based on ideology* if

$$m_t^* = \begin{cases} 1 & \text{if } t > 0, \\ 0 & \text{if } t < 0. \end{cases}$$

For any probability matrix $\mathbf{\Sigma}$ and any level $\mu > 0$ of marginal attention cost, define the *attention set* of an arbitrary voter $t < 0$ by the set of policy matrices which, when paired with $\mathbf{\Sigma}$, leads the voter to pay active attention to politics:

$$\mathcal{A}_t\left(\mathbf{\Sigma}, \mu\right) = \left\{\mathbf{A} : \mathbb{E}_{[\mathbf{A},\mathbf{\Sigma}]}\left[\exp\left(\mu^{-1} v\left(\tilde{\mathbf{a}}, t\right)\right)\right] \geq 1\right\}.$$

Intersecting this set with $\mathcal{E}\left(\mathbf{\Sigma}\right)$, we obtain the set of policy matrices that draws the voter's attention to politics in equilibrium:

$$\mathcal{E}\mathcal{A}_t\left(\mathbf{\Sigma}, \mu\right) = \left\{\mathbf{A} \in \mathcal{E}\left(\mathbf{\Sigma}\right) : \mathbb{E}_{[\mathbf{A},\mathbf{\Sigma}]}\left[\exp\left(\mu^{-1} v\left(\tilde{\mathbf{a}}, t\right)\right)\right] \geq 1\right\},$$

as well as the set of policy matrices that leads the voter to act based on ideology in equilibrium:

$$\mathcal{E}\mathcal{I}_t\left(\mathbf{\Sigma}, \mu\right) = \left\{\mathbf{A} \in \mathcal{E}\left(\mathbf{\Sigma}\right) : \mathbb{E}_{[\mathbf{A},\mathbf{\Sigma}]}\left[\exp\left(\mu^{-1} v\left(\tilde{\mathbf{a}}, t\right)\right)\right] < 1\right\}.$$

Below we examine how these sets vary with μ.

Assumption 3. *$u(a',t) - u(a,t)$ is strictly increasing in t for all a, a'.*

Assumption 4. *There exist $t < 0$ and $\kappa > 0$ such that $\left|v\left(\mathbf{a}, t\right) - v\left(\mathbf{a}, 0\right)\right| > \kappa |t|$ for all $\mathbf{a}$.*

Assumption 3 says that voters' preferences exhibit strict increasing differences between policies and types (Milgrom and Shannon (1994)), meaning that pro-β voters more prefer pro-β policies to pro-α policies than pro-α voters do. Assumption 4 goes one step further, stipulating that for some voters, the utility difference from choosing one candidate over another differs significantly than that of the median voter. These assumptions hold true for many standard utility functions in the literature. Notable examples include $u(a, t) = -|t - a|$, for which $\kappa = 2$ when $t \approx 0$, as well as $u(a, t) = -(a - t)^2$, for which $\kappa = 2 \min A_\beta$.

Theorem 2. *Assume Assumption 1-3, and let $\mathbf{\Sigma}$ be any probability matrix of order $N \geq 2$ for which the set $\mathcal{E}(\mathbf{\Sigma})$ is non-empty. For any $t < 0$ that satisfies Assumption 4 and where $v(\mathbf{a}, t) > 0$ for some $\mathbf{a} \in \mathbf{A} \in \mathcal{E}(\mathbf{\Sigma})$, the following happen as as we increase μ from zero to infinity:*

(i) the set $\mathcal{E}\mathcal{A}_t(\mathbf{\Sigma}, \mu)$ shrinks and the set $\mathcal{E}\mathcal{I}_t(\mathbf{\Sigma}, \mu)$ expands;

(ii) $\min\{u(a_1, 0) - u(a_N, 0) : \mathbf{A} \in \mathcal{E}\mathcal{A}_t(\mathbf{\Sigma}, \mu)\}$ *increases;*

(iii) the above described objects do not always stay constant.

Theorem 2 shows that as voter's marginal attention cost increases, arousing and attracting his attention in equilibrium becomes harder, and doing so leads the varying types of the candidates to enlarge their policy differences or even to exaggerate the intrinsic differences between themselves. Noticeably, these results hold true for any probability matrix that can emerge in equilibrium, as well as any non-extreme voter who at least prefers the opposing candidate in some policy state.

The above described results resonate with Madison's deliberate over-emphasis of agriculture over finance, manufacturing and trade, contrary to his belief that a healthy country should excell in all these areas (Feldman (2017)). They attribute the high levels of visibility earned by George H. W. Bush during the 1980 Republican primary to the candidate's unequivocal embrace of women's rights, as well as the sensation stirred by Gary Hart during the 1984 Democratic primary to the extreme centrist position that the candidate adopted (Popkin (1994)). Last but not least, they explain why the conformity in the 50's, hallmarked by Eisenhower's embrace of New Deal, led to the historically low levels of political attention and knowledge among voters – a conclusion drawn by Campbell (1960) based on the finding that most subjects of their survey struggled to tell the exact stances of candidates on critical issues.

4.2.3 Voter Behaviors

Two patterns emerge as we increase the marginal attention cost of an arbitrary voter while holding candidates' equilibrium strategies fixed. First, lower levels of attention and political knowledge ensue, both measured in terms of the mutual information between the policy state and the voter's decision. Voter welfare first decreases and then stays constant after voting becomes driven solely by ideology.

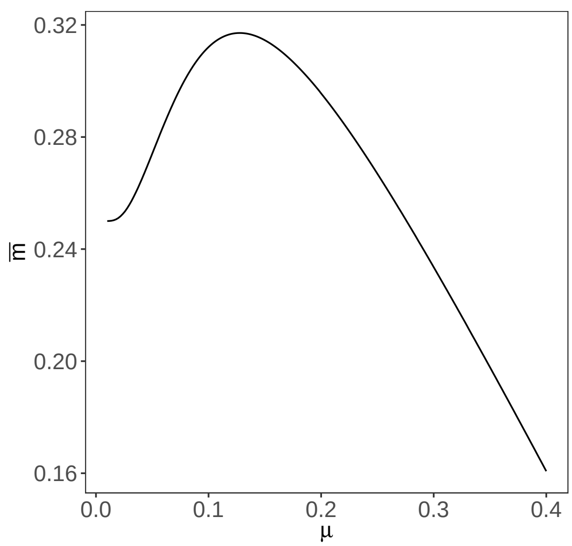

Figure 4: Plot $\mathbb{E}_{\sigma^*}[m_t^*(\tilde{\mathbf{a}})]$ against μ for $t = -.05$: policies equal $\pm 1/4$ and $\pm 3/4$ with equal probability and $u(a,t) = -|t-a|$.

Second, the degree of opinion polarization, measured in terms of the average propensity that the voter supports the opposing camp, eventually grows to one, a phenomenon that is hereafter called *attention-driven opinion polarization*. Nevertheless, the trajectory can be non-monotone, showcasing the subtlety in which optimal attention allocation adjusts to changes in the marginal attention cost (see Figure 4 for a graphical illustration).

4.3 Proof Sketches

Proof of Theorem 1 The proof of Theorem 1 exploits only the assumption of symmetry, as well as the fact that voters are rationally inattentive and have concave preferences for policies. To build intuition, let us take an arbitrary policy state

$\mathbf{a}_{ij} = (-a_i, a_j)$ where $i > j$ and argue why the number of votes $\int m_t^*(\mathbf{a}_{ij})dP(t)$ earned by candidate β exceeds $1/2$.

At first glance, a successful characterization of equilibrium seems to require that we compute m_t^* explicitly, and yet this task has proven to be astoundingly difficult by the existing studies on rational inattention. Our proof strategy circumvents this challenge:

Step 1 By symmetry, we can transform the problem of aggregating different voters' decisions in a single state into that of comparing a single voter's decisions across symmetric states:

$$\begin{aligned}\int m_t^*\left(\mathbf{a}_{ij}\right)dP(t) &= \int_{t<0} 1 - m_{-t}^*\left(\mathbf{a}_{ji}\right)dP(t) + \int_{t>0} m_t^*\left(\mathbf{a}_{ij}\right)dP(t) \\ &= \int_{t>0} m_t^*\left(\mathbf{a}_{ij}\right) - m_t^*\left(\mathbf{a}_{ji}\right)dP(t) + \frac{1}{2}.\end{aligned}$$

Step 2 Since attention allocation is rational, we can further reduce the above problem to that of comparing a single voter's utilities from choosing one candidate over another across symmetric states:

$$\operatorname{sgn} m_t^*\left(\mathbf{a}_{ij}\right) - m_t^*\left(\mathbf{a}_{ji}\right) = \operatorname{sgn} v\left(\mathbf{a}_{ij}, t\right) - v\left(\mathbf{a}_{ji}, t\right).$$

Step 3 Since voters have concave preferences for policies, it follows that a candidate is more preferable to his opponent when his policy is more centrist than the latter's rather than the other way around, i.e., for all $i > j$ and t,

$$v\left(\mathbf{a}_{ij}, t\right) - v\left(\mathbf{a}_{ji}, t\right) = u\left(a_j, t\right) + u\left(-a_j, t\right) - \left[u\left(a_i, t\right) + u\left(-a_i, t\right)\right] \geq 0.$$

Combining Steps one to three yields $\int m_t^*\left(\mathbf{a}_{ij}\right)dP(t) \geq 1/2$.

Step 4 It remains to show that the above inequality is strict for a positive measure of voters. A key step in the proof shows that so long as the policy state is non-degenerate, the median voter — and by continuity, a neighborhood around him — chooses to be attentive to politics:

Lemma 2. *Assume Assumption 1. Then $m_t^* \in (0,1)$ in a neighborhood of $t = 0$ in any symmetric equilibrium that yields a non-degenerate policy state.*

Proof of Theorem 2 We first demonstrate that the attention set of a typical voter shrinks as μ increases from zero to infinity. We next argue that all elements of the attention set must satisfy

$$u(a_1, 0) - u(a_N, 0) \geq \text{a constant strictly increasing in } \mu.$$

Combining these results with Theorem 1 yields the pattern delineated in Example 1.

4.4 Discussions

Preference for attention and equilibrium selection To be sure, our theorems provide no clue as to which element of the set $\mathcal{E}(\mathbf{\Sigma})$ are more likely to emerge in equilibrium than others. While this feature of the model follows from the assumption that candidates care about voters only through the winning probabilities, there is reason to believe that in reality, candidates care about voters' attention levels, too, e.g., because the dissemination of political information may involve entities that care directly about voters' eyeballs, e.g., journalists constrained by norms of neutrality. If only those equilibrium that capture the attention of a critical mass of voters meet our selection criterion, e.g., the blue area in Figure 3, then we obtain a stronger result: as marginal attention cost increases, the set of equilibrium policy profiles shrinks and exhibits a greater degree of attention-driven extremism or exaggeration.

Equilibrium refinement So far our assumption is that any policy outside the support of equilibrium strategies can be perfectly discerned and severely penalized by voters. While this assumption plays no role in the proofs of Theorems 1 and 2, it certainly affects the determination of the set $\mathcal{E}(\mathbf{\Sigma})$.

The equilibrium refinement considered below requires that candidates take each of their feasible policies with at least probability $\epsilon > 0$. By Theorem 1, the winning probability matrix in this case is $\widehat{\mathbf{W}}_{|A_c|}$, where $|A_c|$ denotes the cardinality of the set A_c. Based on this result, we can then solve for equilibrium policies while bypassing the need to compute voters' best responses. Our results are reported below:

Example 1 (Continued)**.** In Example 1, now suppose, instead, that candidates must take each of their 20 policies with at least probability $\epsilon = .0005$. Figure 5 depicts the profiles of policies (a_1, a_2) that emerge in equilibrium where each of a_1 and a_2 is taken by a distinct type of candidate β with at least probability .99. Comparing this figure with Figure 3 reveals two insights: first, only a few policy profiles survive our refinement; second, as in the baseline model, whether or not candidates will exaggerate the intrinsic differences between themselves in equilibrium depends much on candidates' utility functions.

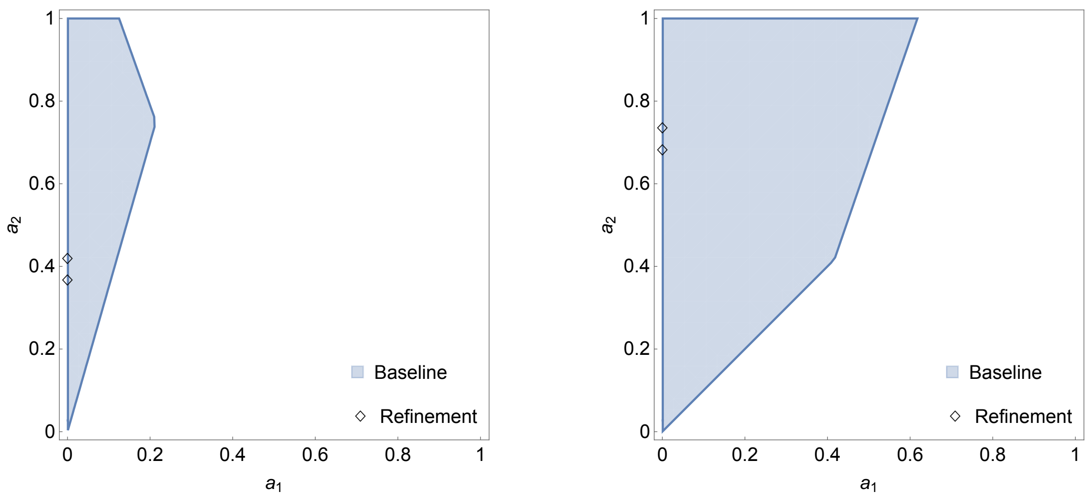

Figure 5: Equilibrium policies before and after refinement: $u_+(a,t) = 6 - 9|t-a|$ on the left panel; $u_+(a,t) = 6 - 8|t-a|$ if $a < t$ and $6 - 40|t-a|$ if $a > t$ on the right panel; $u_-(a,t) = -|t-a|$; t_c's equal $\pm 1/4$ and $\pm 3/4$ with equal probability; each A_c consists of 20 evenly spaced policies on Θ_c.

Limited commitment In Appendix B.3, we allow the winning candiate to renege and implement his ideal policy position probabilistically, and regard the policy proposal in Stage 2 of the election game as a campaign message. The main result thereof, namely campaign messages are extreme and exaggerated, is a mere understatement of reality. As Roger Ailes described his role as a campaign strategist: "Every single thing I did from issues to speeches to ads, etc. was designed to push the varying kinds of the candidates further apart" (Vavreck (2009)).

5 Extension: Noisy News

5.1 Setup

In the baseline model, now suppose, instead, that news is a profile $\boldsymbol{\omega} = (\omega_\alpha, \omega_\beta)$ of reports, each drawn independently from a finite support $\Omega_c \subset \Theta_c$ according to a probability mass function $f_c(\cdot \mid a_c)$. Define the *news technology* by the joint probability function $f = f_\alpha \times f_\beta$, and name each $\boldsymbol{\omega} \in \Omega$ a *news state* drawn from support $\Omega = \Omega_\alpha \times \Omega_\beta$.

In this new setting, redefine the attention rule of any arbitrary voter t by a mapping $m_t : \Omega \to [0,1]$, where each $m_t(\boldsymbol{\omega})$ specifies the probability that the voter chooses candidate β in news state $\boldsymbol{\omega}$. For any given profile $x = (f, \sigma)$ of news technology and candidates' strategies, let

$$\nu_x(\omega, t) = \mathbb{E}_x\left[v(\tilde{\mathbf{a}}, t) \mid \boldsymbol{\omega}\right]$$

be voter t's expected utility difference from choosing candidate β over candidate α in news state $\boldsymbol{\omega}$, and let

$$V_t(m_t, x) = \mathbb{E}_x\left[m_t(\tilde{\boldsymbol{\omega}})\, \nu_x(\tilde{\boldsymbol{\omega}}, t)\right].$$

be the expected utility difference under (m_t, x). Voter t's expected payoff is then

$$V_t(m_t, x) - \mu_t \cdot I(m_t, x),$$

where $I(m_t, x)$ is the mutual information between his decision and the news state. Under any profile (m, x) of attention rules, candidates' strategies and news technology, the expected payoff of candidate c is equal to

$$V_c(m, x) = \mathbb{E}_{m,x}\left[w_c(\tilde{\boldsymbol{\omega}})\, u_+\left(\tilde{a}_c, \tilde{t}_c\right) + (1 - w_c(\tilde{\boldsymbol{\omega}}))\, u_-\left(\tilde{a}_{-c}, \tilde{t}_c\right)\right]$$

In particular, $\omega_c(\boldsymbol{\omega})$ represents the candidate's winning probability in news state $\boldsymbol{\omega}$ and can be obtained by replacing $\mathbf{a}$ with $\boldsymbol{\omega}$ in Equation (2.1).

5.2 Equilibrium Analysis

A strategy profile (m^*, σ^*) constitutes a *Bayesian Nash equilibrium* of the election game under news technology f if each player maximizes his expected utility, taking f and other players' strategies as given. As before, we assume that the environment, which now includes the news technology, is symmetric around the median voter:

Assumption 5. *$f_\alpha(-\omega \mid -a) = f_\beta(\omega \mid a)$ for all $a \in A_\beta$ and $\omega \in \Omega_\beta$.*

As before, we focus on symmetric equilibrium where $\sigma_c(a \mid t) = \sigma_c(-a \mid -t)$ for all $a \in A_c$ and $t \in T_c$, and $m_t(-\omega, \omega') = 1 - m_{-t}(-\omega', \omega)$ for all $t \in \Theta$ and $(-\omega, \omega') \in \Omega$.

5.2.1 Matrix Representation

Let $-\omega_K < \cdots < -\omega_1 < 0 < \omega_1 < \cdots < \omega_K$ denote the realizations of news reports and write $\boldsymbol{\omega}_{mn} = (-\omega_m, \omega_n)$ for $m, n = 1, \cdots, K$, where K is an exogenous number greater than one. Let $\mathbf{A}$ and $\mathbf{\Sigma}$ be as above, and let $\mathbf{W}$ be a $K \times K$ matrix whose mn^{th} entry $w_{mn} \in \{0, 1/2, 1\}$ represents candidate β's winning probability in news state $\boldsymbol{\omega}_{mn}$.

A tuple $\langle \mathbf{A}, \mathbf{\Sigma}, \mathbf{W} \rangle$ can be attained in a symmetric equilibrium if (1) $[\mathbf{A}, \mathbf{\Sigma}]$ is *incentive compatible for candidates under* $\langle f, \mathbf{W} \rangle$ and if (2) $\mathbf{W}$ *can be rationalized by optimal attention allocation under* $\langle f, \mathbf{A}, \mathbf{\Sigma} \rangle$ (hereafter, $\langle f, \mathbf{W} \rangle$-IC and $\langle f, \mathbf{A}, \mathbf{\Sigma} \rangle$-rationalizable). Formally,

Definition 3. $[\mathbf{A}, \mathbf{\Sigma}]$ *is* $\langle f, \mathbf{W} \rangle$-IC *if in Definition 1, the winning probability matrix is replaced with an* $N \times N$ *matrix whose* ij^{th} *entry is*

$$\sum_{m,n=1}^{K} w_{mn} f \left(\boldsymbol{\omega}_{mn} \mid \mathbf{a}_{ij} \right) \sigma_{ij}.$$

Definition 4. $\mathbf{W}$ *is* $\langle f, \mathbf{A}, \mathbf{\Sigma} \rangle$-rationalizable *if*

$$w_{mn} = w \left(\boldsymbol{\omega}_{mn} \right) \ \forall m, n,$$

where $w \left(\boldsymbol{\omega}_{mn} \right)$ *can be obtained from plugging* $m_t^* \left(\boldsymbol{\omega}_{mn} \right)$, $t \in \Theta$ *under* $\langle f, \mathbf{A}, \mathbf{\Sigma} \rangle$ *into Equation (2.1).*

5.2.2 Equilibrium Policies

For any probability matrix $\mathbf{\Sigma}$, define $\mathcal{E} \left(\mathbf{\Sigma}, f \right)$ as the set of policy matrices which, when paired with $\mathbf{\Sigma}$, can be attained in equilibrium under f:

$$\mathcal{E} \left(\mathbf{\Sigma}, f \right) = \left\{ \mathbf{A} : \exists \mathbf{W} \text{ s.t. } \begin{array}{c} [\mathbf{A}, \mathbf{\Sigma}] \text{ is } \langle f, \mathbf{W} \rangle - \text{IC} \\ \mathbf{W} \text{ is } \langle f, \mathbf{A}, \mathbf{\Sigma} \rangle - \text{rationalizable} \end{array} \right\}.$$

The upcoming analysis assumes that the news technology is log-supermodular, meaning that we are more likely to observe extreme rather than centrist news reports as the underlying policy becomes more extreme (Milgrom (1981)):

Assumption 6. *For all c, $a, a' \in A_c$ and $\omega, \omega' \in \Omega_c$ such that $a < a'$ and $\omega < \omega'$, we have*

$$\frac{f_c(\omega' \mid a)}{f_c(\omega \mid a)} < \frac{f_c(\omega' \mid a')}{f_c(\omega \mid a')}.$$

The next theorem gives a full characterization of the set $\mathcal{E}(\boldsymbol{\Sigma}, f)$:

Theorem 3. *Assume Assumptions 1, 2, 5 and 6. Then for any probability matrix $\boldsymbol{\Sigma}$ of order N,*

(i) if $N = 1$, then

$$\mathcal{E}(\boldsymbol{\Sigma}, f) = \left\{\mathbf{A} : [\mathbf{A}, \boldsymbol{\Sigma}] \text{ is } \langle f, \frac{1}{2}\mathbf{J}_K\rangle - \text{IC}\right\};$$

(ii) if $N \geq 2$, then

$$\mathcal{E}(\boldsymbol{\Sigma}, f) = \left\{\mathbf{A} : [\mathbf{A}, \boldsymbol{\Sigma}] \text{ is } \langle f, \widehat{\mathbf{W}}_K\rangle - \text{IC}\right\}.$$

Theorem 3 shows that in every symmetric equilibrium that yields a non-degenerate policy state, *winner is determined as if news were fully revealing*, meaning that in every news state, the candidate who earns the most centrist news report wins the election for sure, whereas candidates with equally distant news reports from the median voter split the votes evenly.

The implication of this result is twofold. First, the key insight of Theorem 1, namely voter characteristics such as marginal attention cost are irrelevant in the determination of equilibrium policies, carries over to the current setting. Second, news technology clearly matters for candidates' incentives, raising new questions as to how news quality affects equilibrium levels of voter attention and policy extremism and exaggeration.

5.2.3 Comparative Statics

For any probability matrix $\boldsymbol{\Sigma}$ and any news technology f, define the *attention set* of an arbitrary voter $t < 0$ by the set of policy matrices which, when paired with $\boldsymbol{\Sigma}$, induces the voter to attend to politics under f:

$$\mathcal{A}_t(\boldsymbol{\Sigma}, f) = \left\{\mathbf{A} : \mathbb{E}_{\langle f, \mathbf{A}, \boldsymbol{\Sigma}\rangle}\left[\exp\left(\mu_t^{-1}\nu_{\langle f, \mathbf{A}, \boldsymbol{\Sigma}\rangle}(\boldsymbol{\omega}, t)\right)\right] \geq 1\right\}.$$

Intersecting this set with $\mathcal{E}(\mathbf{\Sigma}, f)$, we obtain the set of policy matrices that draws the voter's attention to politics in equilibrium:

$$\mathcal{E}\mathcal{A}_t(\mathbf{\Sigma}, f) = \mathcal{E}(\mathbf{\Sigma}, f) \cap \mathcal{A}_t(\mathbf{\Sigma}, f),$$

as well as the set of policy matrices that leads the voter to act based on ideology in equilibrium:

$$\mathcal{E}\mathcal{I}_t(\mathbf{\Sigma}, f) = \mathcal{E}(\mathbf{\Sigma}, f) \cap \mathcal{A}_t^c(\mathbf{\Sigma}, f).$$

In Appendix B.2 Corollary 1, we show that the above described sets vary with the marginal attention cost the exact same way as described in Theorem 2. Here we consider instead how they depend on the informativeness of the news technology.

Our notion of informativeness is adopted from Blackwell (1953):

Definition 5. *f is more* Blackwell-informative *than f' (f'* is a garble of *f, $f \succeq f'$) if there exists a Markov kenel ρ such that for all $\mathbf{a}$ and $\boldsymbol{\omega'}$,*

$$f'(\boldsymbol{\omega}' \mid \mathbf{a}) = \sum_{\boldsymbol{\omega} \in \Omega} f(\boldsymbol{\omega} \mid \mathbf{a}) \rho(\boldsymbol{\omega}' \mid \boldsymbol{\omega}).$$

It is well known that garbling destroys information and makes decision makers worse-off in all decision problems. A leading example of de-garbling occurred between 1870 and 1920, a period during which advances in printing technology and increasing competition made news papers less partisan and more information (Gentzkow et al. (2004)). Recently, political scientists, journalists and managers of social media platforms have voiced concerns for the rise of partisan media and fake news (Levendusky (2013); Barthel et al. (2016); Lee and Kent (2017); Issac (2018)). We seek to understand how these concerns, when combined with limited voter attention, can have real impacts for equilibrium policy settings and welfare.

We first examine the effect of garbling on the attention set:

Theorem 4. *Assume Assumptions 1, 4, 5 and 6, and fix any probability matrix $\mathbf{\Sigma}$ of order $N \geq 2$ and any $f \succeq f'$. Then for any camp α voter $t < 0$ such that $\mathcal{A}_t(\mathbf{\Sigma}, f)$ and $\mathcal{A}_t(\mathbf{\Sigma}, f')$ are different and non-empty,*

(i) $\mathcal{A}_t(\mathbf{\Sigma}, f') \subset \mathcal{A}_t(\mathbf{\Sigma}, f)$;

(ii) $\min_{\mathbf{A} \in \mathcal{A}_t(\mathbf{\Sigma}, f')} \nu_{\langle f'', \mathbf{A}, \mathbf{\Sigma} \rangle}(\boldsymbol{\omega}_{K1}, 0) > \min_{\mathbf{A} \in \mathcal{A}_t(\mathbf{\Sigma}, f)} \nu_{\langle f'', \mathbf{A}, \mathbf{\Sigma} \rangle}(\boldsymbol{\omega}_{K1}, 0)$ *for* $f'' = f, f'$.

Theorem 4 shows that as the news technology becomes less Blackwell-informative, retaining voter's attention becomes harder and requires that the varying types of candidates adopt increasingly different policies. The degree of policy difference is measured in terms of the median voter's utility gain from choosing candidate β over candidate α upon hearing the most centrist news report about the former and the most extreme news report about the latter.

To build intuition, notice that

$$\nu_{\langle f,\mathbf{A},\mathbf{\Sigma}\rangle}\left(\boldsymbol{\omega}_{K1},0\right)=\underbrace{\mathbb{E}_{\langle f,\mathbf{A},\mathbf{\Sigma}\rangle}\left[u\left(\tilde{a}_{\beta},0\right)\mid\omega_{\beta}=\omega_{1}\right]}_{(1)}-\underbrace{\mathbb{E}_{\langle f,\mathbf{A},\mathbf{\Sigma}\rangle}\left[u\left(\tilde{a}_{\beta},0\right)\mid\omega_{\beta}=\omega_{K}\right]}_{(2)},$$

where (1) and (2) is the median voter's expected valuation of policies given the most centrist and the most extreme news report, respectively. Expanding these terms yields

$$(1)=\frac{\sum_{i=1}^{N}f_{\beta}\left(\omega_{1}\mid a_{i}\right)u(a_{i},0)\sigma_{i}}{\sum_{i=1}^{N}f_{\beta}\left(\omega_{1}\mid a_{i}\right)\sigma_{i}},$$

and

$$(2)=\frac{\sum_{i=1}^{N}f_{\beta}\left(\omega_{K}\mid a_{i}\right)u(a_{i},0)\sigma_{i}}{\sum_{i=1}^{N}f_{\beta}\left(\omega_{K}\mid a_{i}\right)\sigma_{i}},$$

where $\sigma_i=\sum_j\sigma_{ji}$ is the marginal probability that candidate β's policy is a_i. By log-supermodularity, we know that (1) weighs centrist policies heavily and (2) extreme policies heavily. Thus, an enlarging difference between (1) and (2) means that centrist and extreme policies have become more different on average, holding probabilities and news technology fixed.

The overall impact of garbling on equilibrium outcomes is subtle and is illustrated by the next example:

Example 2. In Example 1, now suppose, instead, that news report can be either *centrist* ($\omega=\pm\omega_1$) or *extreme* ($\omega=\pm\omega_2$), where $0<\omega_1<\omega_2<1$. The news technology is $f_\xi=f_{\alpha,\xi}\times f_{\beta,\xi}$, where $f_{\beta,\xi}\left(\omega_2\mid a\right)=a+\xi(1-a)$. The parameter $\xi\in(0,1)$ represents the degree of slanting as well as the Blackwell-order of the news technology: as ξ increases, we are more likely to hear extreme news reports, ceteris paribus, and the quality of news deteriorates.

Figure 6 depicts equilibrium outcomes for three levels of ξ's. In particular, the diamonds represent policy profiles that arise in equilibrium where candidates adopt pure symmetric strategies, whereas the shaded areas above colored lines represent the attention sets of an arbitrary voter $t = -.001$. As ξ increases, the attention set shrinks and moves northwest, and the remaining policy profiles exhibit a greater degree of policy extremism. Meanwhile, equilibrium policies converge to candidates' ideological bliss points $(1/4, 3/4)$, but the trajectory can depend subtly on properties of candidates' utility functions and the news technology.

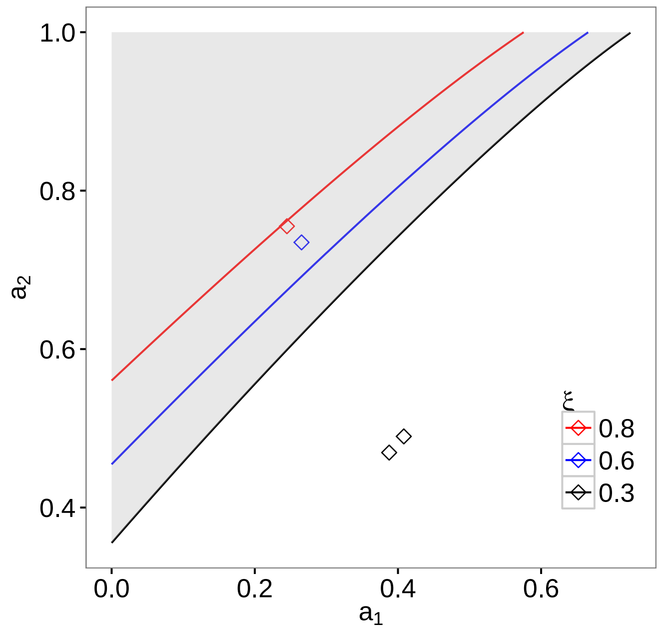

Figure 6: Equilibrium outcomes: $R = 8$, $\delta_+ = 3$, $\delta_- = 1$, $t = -.001$ and $\mu = 1$; diamonds represent policy profiles that arise in equilibrium where candidates adopt pure symmetric strategies, whereas shaded areas above the colored lines represent attention sets.

Unlike in the baseline model, here we don't know whether garbling makes it harder or easier to arouse and attract voter's attention in equilibrium. What we do know is that any equilibrium that achieves this goal after garbling must exhibit a greater degree of policy extremism than those that fail to do so prior to garbling — a phenomenon we shall hereafter call *media-driven extremism*.

One should not confuse noisy news with the aggregate shock to voters' preferences in probabilistic voting models. In our model, a deterioration of the Blackwell-informativeness of the news signals can have subtle impact on the equilibrium degree of policy extremism. In probabilistic voting models, greater preference shocks lead to higher levels of policy divergence, which itself is a different concept than the policy extremism defined above.

5.2.4 Proof sketch

The proof of Theorem 4 exploits basic properties of garbling. We begin by proving the following lemma:

Lemma 3. *Fix any t, σ and $f \succeq f'$, and write $x = (f, \sigma)$ and $x' = (f', \sigma)$. Then,*

(i) $\mathbb{E}_x [\nu_x (\boldsymbol{\omega}, t)] = \mathbb{E}_{x'} [\nu_x (\boldsymbol{\omega}', t)]$*;*

(ii) for each $\boldsymbol{\omega}' \in \Omega$, there exist probability weights $\{\pi (\boldsymbol{\omega}', \boldsymbol{\omega})\}_{\boldsymbol{\omega} \in \Omega}$ such that

$$\nu_{x'}(\boldsymbol{\omega}', t) = \sum_{\boldsymbol{\omega}} \pi (\boldsymbol{\omega}', \boldsymbol{\omega}) \nu_x (\boldsymbol{\omega}, t) .$$

Lemma 3 shows that voter's utility difference from choosing one candidate over another after garbling constitutes a mean-preserving spread of the same quantity prior to garbling. Then from the fact that the exponential function is convex, it follows that garbling causes contractions of the attention set, and this completes the proof of Part (i).

We next argue that a policy matrix $\mathbf{A}$ belongs to the attention set only if

$$\nu_{\langle f,\mathbf{A},\boldsymbol{\Sigma}\rangle} (\boldsymbol{\omega}_{K1}, 0) \geq \text{ a constant independent of } f, \tag{5.1}$$

and that garbling reduces the left-hand side of the above inequality, as extreme (resp. centrist) news reports no longer lead to extreme (resp. centrist) beliefs about candidates' policies:

Lemma 4. *For any $f \succeq f'$ and $[\mathbf{A}, \boldsymbol{\Sigma}]$, $\nu_{\langle f,\mathbf{A},\boldsymbol{\Sigma}\rangle} (\boldsymbol{\omega}_{K1}, 0) \geq \nu_{\langle f',\mathbf{A},\boldsymbol{\Sigma}\rangle} (\boldsymbol{\omega}_{K1}, 0)$.*

Combining these results with Part (i), we find that the inequality in (5.1) must hold true strictly prior to garbling in order for the policy matrix to remain in the attention set after garbling, and this completes the proof of Part (ii).

6 Extension: Multiple Issues

In the baseline model, now suppose, instead, that a policy consists of two issues a and b (e.g., inflation and unemployment, defense and economy), both take values in Θ. The Pareto frontier $\mathcal{B}(a)$, defined as a function of a, is strictly decreasing, strictly concave and smooth, and satisfies the Inada condition $\lim_{a \to -1} \mathcal{B}'(a) = 0$ and $\lim_{a \to 1} \mathcal{B}'(a) = -\infty$. The electorate consists of a continuum of voters and two candidates named a and b. Each player is either pro-a or pro-b, depending on whether his type — a real number in Θ that represents his preference weight on issue b — belongs to $\Theta_a = [-1, 0]$ or $\Theta_b = [0, 1]$. Players' types are drawn from the same probability distributions as described in Section 2, subject to minor relabeling. Their utility functions are $u(a, b, t)$, $u_+(a, b, t)$ and $u_-(a, b, t)$, all strictly increasing and smooth in (a, b). Voters' utility functions are strictly concave in issues and satisfy the Spence-Mirrlees' generalized single-crossing property:

Assumption 7. *$u(a, b, t)$ is strictly concave in (a, b) and $-\frac{u_a(a,b,t)}{u_b(a,b,t)}$ is strictly increasing in t for all (a, b).*

Under the above described assumptions, each voter t's indifference curve is tangent to the Pareto frontier at a unique point $(a^\circ(t), \mathcal{B}(a^\circ(t)))$ defined by

$$\frac{u_a(a^\circ(t), \mathcal{B}(a^\circ(t)))}{u_b(x^\circ(t), \mathcal{B}(a^\circ(t)))} + \mathcal{B}'(a^\circ(t), \mathcal{B}(a^\circ(t))) = 0,$$

where $a^\circ(t)$ is decreasing in t. Suppose the image of a°, denoted by $a^\circ(\Theta)$, coincides with Θ, and hence each $a \in \Theta$ is associated with a unique type $(a^\circ)^{-1}(a)$ of voter whose indifference curve is tangent to the Pareto frontier at $(a, \mathcal{B}(a))$.

Under the assumption that utilities are increasing in both issues, all equilibria of the game must lie on the Pareto frontier, or else a Pareto improvement can be easily constructed. Thus it suffices to consider an augmented economy with a single issue a and the following augmented utility functions: $\widehat{u}(a, t) = u(a, \mathcal{B}(a), t)$, $\widehat{u}_+(a, t) = u_+(a, \mathcal{B}(a), t)$ and $\widehat{u}_-(a, t) = u_-(a, \mathcal{B}(a), t)$, where the resemblance between the function $\widehat{u}$ and voter's utility function u in the baseline model is noteworthy:

Lemma 5. *The function $\widehat{u}$ satisfies the following properties:*

(i) $\widehat{u}(\cdot, t)$ is strictly increasing on $[-1, a^\circ(t)]$ and is strictly decreasing on $[a^\circ(t), 1]$ for all t;

(ii) $\widehat{u}(\cdot, t)$ *is strictly concave for all* t*;*

(iii) $\widehat{u}$ *has strict increasing differences if* $u_{at} \geq 0$ *and* $u_{bt} \leq 0$ *and one of these inequalities is strict.*

The next theorem generalizes Theorems 1 and 2 to encompass multiple issues:

Theorem 5. *In the augmented economy, Theorem 1 holds under Assumptions 1 and 7, and Theorem 2 holds under Assumptions 1, 3, 4 and 7.*

Theorem 5 provides a formal justification for one of the most sung folklores in political science, namely salient issue positions go hand in hand with high levels of public attention. A textbook example in this regard concerns Democratic candidates' issue positions during eras of high inflation. As documented in Popkin (1994), candidates who paradoxically "owned" the issue of unemployment contrary to their own senses of urgency successfully were highly visible and triggered discussions and debates among the public.

7 Conclusion

The analysis so far has left several questions open. For instance, can we enrich our model with the self-interested media as those considered in Gentzkow and Shapiro (2016) and Puglisi and Snyder (2016)? Can the joint effect of limited voter attention and the market conditions we identify partially explain the growing polarization in the U.S. as documented in Poole and Rosenthal (2001), Abramowitz and Saunders (2008) and Fiorina and Abrams (2008)? And can our framework be applied to the study of related problems, such as optimal product differentiation against rationally inattentive consumers? We hope that someone, maybe ourselves, will pursue these research agendas in the future.

A Omitted Proofs

A.1 Proofs of Section 4

Throughout this section, we will write $\Delta_{ij}(t) = v\left(\mathbf{a}_{ij}, t\right)$ and suppress the notation for t in the case of $t = 0$. By symmetry and our assumption on $u(\cdot, 0)$, we know that (1) $\Delta_{ij} = -\Delta_{ij}$, (2) $\Delta_{ij} > 0$ for all $i > j$ and (3) $\Delta_{N1} = \max_{i,j} \Delta_{ij}$. Define the function $\gamma : \mathbb{R} \to \mathbb{R}$ by

$$\gamma(x) = \exp(x) + \exp(-x). \tag{A.1}$$

Notice the following properties: (1) $\gamma \geq 2$ and the equality holds at $x = 0$, (2) $\gamma'' > 0$ and (3) $\gamma' > 0$ on $\mathbb{R}_+$.

Proof of Lemma 2

Proof. Take any profile of candidate strategies that is symmetric and non-degenerate. Let $[\mathbf{A}, \mathbf{G}]$ be the corresponding matrix representation, where $\mathbf{A}$ and $\mathbf{G}$ are square matrices of order $N \geq 2$. By symmetry,

$$\begin{aligned}\mathbb{E}_{[\mathbf{A},\mathbf{\Sigma}]}\left[\exp\left(\mu_0^{-1}\tilde{\Delta}_{ij}\right)\right] &= \sum_{i=1}^{N} \sigma_{ii} \exp\left(\mu_0^{-1}\Delta_{ii}\right) + \sum_{i=1}^{N}\sum_{j=1}^{i-1} \sigma_{ij}\gamma\left(\mu_0^{-1}\Delta_{ij}\right) \\ &> \sum_{i=1}^{N} \sigma_{ii} \cdot 1 + \sum_{i=1}^{N}\sum_{j=1}^{i-1} \sigma_{ij} \cdot 2 = 1,\end{aligned}$$

and hence $\mathbb{E}_{[\mathbf{A},\mathbf{\Sigma}]}\left[\mu_0^{-1}\exp\left(-\tilde{\Delta}_{ij}\right)\right] > 1$, where the inequalities are strict because $N \geq 2$. Then by continuity, the above inequalities hold true in a neighborhood I of $t = 0$, and the result can be easily read from Lemma 1. □

Proof of Theorem 1

Proof. By symmetry, the following holds true for all i and j:

$$\begin{aligned}\int m_t^*\left(\mathbf{a}_{ij}\right) dP(t) &= \int_{t<0} 1 - m_{-t}^*\left(\mathbf{a}_{ji}\right) dP(t) + \int_{t>0} m_t^*\left(\mathbf{a}_{ij}\right) dP(t) \\ &= \int_{t>0} m_t^*\left(\mathbf{a}_{ij}\right) - m_t^*\left(\mathbf{a}_{ji}\right) dP(t) + \frac{1}{2}.\end{aligned}$$

Thus in the case of $N = 1$, the result follows because the first term of the above line is equal to zero.

Consider the case of $N \geq 2$. Below we argue in two steps that for all $i > j$, (1) $m_t^*\left(\mathbf{a}_{ij}\right) \geq m_t^*\left(\mathbf{a}_{ji}\right)$ for all t, and (2) the inequality is strict in a neighborhood of $t = 0$:

Step 1 Since the function $u(\cdot, t)$ is concave and hence the function $u(\cdot, t) + u(-\cdot, t)$ is decreasing in its argument, we can compare any voter t's utility difference from choosing one candidate over another across symmetric policy states as follows:

$$\Delta_{ij}(t) - \Delta_{ji}(t) = u\left(a_j, t\right) + u\left(-a_j, t\right) - \left[u\left(a_i, t\right) + u\left(-a_i, t\right)\right] \geq 0.$$

Then the result follows from Lemma 1, which shows that m_t^* is increasing in v.

Step 2 A standard continuity argument shows that $\Delta_{ij}(t) > \Delta_{ji}(t)$ for all $i > j$ in a neighborhood I' of $t = 0$, whereas Lemmas 1 and 2 imply that m_t^* is strictly increasing in v in a neighborhood I of $t = 0$. Intersecting these neighborhoods yields $m_t^*\left(\mathbf{a}_{ij}\right) > m_t^*\left(\mathbf{a}_{ji}\right)$ for all $t \in I \cap I'$, and this completes the proof.

□

Proof of Theorem 2

Proof. Let $t < 0$ and $\mathbf{\Sigma}$ be as described in Theorem 2. By Theorem 1, it suffices to show that as μ increases from zero to infinity, (i) the set $\mathcal{A}_t\left(\mathbf{\Sigma}, \mu\right)$ shrinks, (ii) $\min\left\{u(a_1, 0) - u(a_N, 0) : \mathbf{A} \in \mathcal{A}_t(\mathbf{\Sigma}, \mu)\right\}$ increases, (iii) the above described objects do not always stay constant.

Part (i): It suffices to show that the mutual information between voter's optimal decision and the policy state under any given profile of candidate strategies is decreasing in μ, which is the case by revealed preference.

Part (ii): Let $\mathbf{A}$ be any policy matrix of order N. By Assumption 4,

$$\mathbb{E}_{[\mathbf{A},\mathbf{\Sigma}]}\left[\exp\left(\mu^{-1}\tilde{\Delta}_{ij}\left(t\right)\right)\right] \leq \exp\left(\kappa|t|/\mu\right)\mathbb{E}_{[\mathbf{A},\mathbf{\Sigma}]}\left[\exp\left(\mu^{-1}\tilde{\Delta}_{ij}\right)\right],$$

where the last term of the above inequality can be bounded above as follows:

$$\begin{aligned}\mathbb{E}_{[\mathbf{A},\mathbf{\Sigma}]}\left[\exp\left(\mu^{-1}\tilde{\Delta}_{ij}\right)\right] &= \sum_{i=1}^{N}\sigma_{ii}\exp\left(\mu^{-1}\Delta_{ii}\right) + \sum_{i=1}^{N}\sum_{j=1}^{i-1}\sigma_{ij}\gamma\left(\mu^{-1}\Delta_{ij}\right) \\ &\leq \sum_{i=1}^{N}\sigma_{ii}\cdot 1 + \frac{1}{2}\left(1 - \sum_{i=1}^{N}\sigma_{ii}\right)\gamma\left(\mu^{-1}\Delta_{N1}\right).\end{aligned}$$

Thus, a necessary condition for $\mathbb{E}_{[\mathbf{A},\mathbf{\Sigma}]}\left[\exp\left(\mu^{-1}\tilde{\Delta}_{ij}(t)\right)\right] \geq 1$ to hold true is

$$\exp(-\kappa|t|/\mu)\left[\sum_{i=1}^{N}\sigma_{ii} + \frac{1}{2}\left(1 - \sum_{i=1}^{N}\sigma_{ii}\right)\gamma(x)\right] \geq 1,$$

where

$$x \triangleq \mu^{-1}\Delta_{N1} = \mu^{-1}[u(a_1, 0) - u(a_N, 0)].$$

This is equivalent to

$$\gamma(x) \geq 2b,$$

where

$$b \triangleq \frac{\exp(\kappa|t|/\mu) - \sum_{i=1}^{N}\sigma_{ii}}{1 - \sum_{i=1}^{N}\sigma_{ii}} > 1.$$

Since the equation $\gamma(x) = 2b$ has a unique positive root

$$\gamma^{-1}(2b) = \log\left(b + \sqrt{b^2 - 1}\right),$$

the above inequality is equivalent to

$$u(a_1, 0) - u(a_N, 0) \geq \mu\gamma^{-1}(2b) \triangleq \delta(\mu).$$

It is easy to verify that $\delta(\mu)$ is strictly positive for all $\mu > 0$. Below we show that this term is strictly increasing in μ:

Step 1 We first show that $\frac{d\delta(\mu)}{d\mu} > 0$ for $\mu > \frac{\kappa}{2\log 2}$. For ease, write $y = \exp(\kappa|t|/\mu)$, $z = b + \sqrt{b^2 - 1}$ and $\Sigma = \sum_{i=1}^{N}\sigma_{ii}$, and notice that $y, z > 1$ and $\Sigma < 1$. Differentiating $\delta(\mu)$ with respect to μ yields

$$\frac{d\delta(\mu)}{d\mu} = \log z - \frac{y\log y}{(1 - \Sigma)\sqrt{b^2 - 1}}.$$

Since

$$\log z = \log[1 + (z - 1)] \geq \frac{2(z - 1)}{2 + (z - 1)} = \frac{2(z - 1)}{z + 1},$$

and

$$\log y = \log[1 + (y - 1)] \leq \frac{(y - 1)}{\sqrt{1 + (y - 1)}} = \frac{y - 1}{\sqrt{y}},$$

a sufficient condition for $\frac{d\delta(\mu)}{d\mu} > 0$ to hold true is that

$$\frac{2(z-1)}{z+1} - \frac{y \cdot \frac{y-1}{\sqrt{y}}}{(1-\Sigma)\sqrt{b^2-1}} > 0.$$

Plugging in $z = b + \sqrt{b^2-1}$ and $b = \frac{y-\Sigma}{1-\Sigma}$ into the above inequality and rearranging, we obtain

$$\begin{aligned} 2\left(y - 1 + \sqrt{(y-1)(y+1-2\Sigma)}\right)\sqrt{y+1-2\Sigma} \\ - \left(y+1-2\Sigma) + \sqrt{(y-1)(y+1-2\Sigma}\right)\sqrt{y(y-1)} > 0. \end{aligned}$$

This is equivalent to

$$\sqrt{(y-1)(y+1-2\Sigma)}\left(\sqrt{y-1} + \sqrt{y+1-2\Sigma}\right)(2-\sqrt{y}) > 0,$$

which holds true if and only if $2 - \sqrt{y} > 0$, or equivalently $\exp(\kappa|t|/2\mu) < 2$. A sufficient condition is that $\mu > \frac{\kappa}{2\log 2}$.

Step 2 We next show that $\frac{d^2\delta(\mu)}{d\mu^2} < 0$. Since

$$\frac{d^2\delta(\mu)}{d\mu^2} = -\frac{\log y}{\sqrt{y+1-2\Sigma}}\left(1 - \frac{y(y-\Sigma)}{y+1-2\Sigma}\right)\frac{\partial y}{\partial \mu},$$

a sufficient condition for $\frac{d^2\delta(\mu)}{d\mu^2} < 0$ to hold true is

$$1 - \frac{y(y-\Sigma)}{y+1-2\Sigma} < 0,$$

or equivalently $y > 2 - \frac{1}{\Sigma}$. This is true because $y > 1$ and $\Sigma < 1$.

Part (iii): When μ is small, $\mathcal{A}_t(\boldsymbol{\Sigma}, \mu) \neq \emptyset$ because of the assumption that $v(\mathbf{a}, t) > 0$ for some $\mathbf{a} \in \mathbf{A} \in \mathcal{E}(\boldsymbol{\Sigma})$. When μ is large, the following holds true for all policy matrices $\mathbf{A}$ of order N:

$$\mathbb{E}_{[\mathbf{A},\boldsymbol{\Sigma}]}\left[\exp\left(\mu^{-1}\tilde{\Delta}_{ij}(t)\right)\right] \underbrace{\approx}_{(1)} \mathbb{E}_{[\mathbf{A},\boldsymbol{\Sigma}]}\left[1 + \mu^{-1}\tilde{\Delta}_{ij}(t)\right] \underbrace{<}_{(2)} \mathbb{E}_{[\mathbf{A},\boldsymbol{\Sigma}]}\left[1 + \mu^{-1}\tilde{\Delta}_{ij}\right] \underbrace{=}_{(3)} 0,$$

where (1) uses the fact that $\exp(z) \approx 1 + z$ when $z \approx 0$, (2) follows from Assumption 3 and (3) properties of Δ_{ij}'s. Thus, $\mathcal{A}_t(\boldsymbol{\Sigma}, \mu) = \emptyset$ in this case, and this completes the proof. $\square$

A.2 Proofs of Section 5

Throughout this section, we will write $x = (f, \sigma)$ for notational ease, where σ can be any profile of strategies that the analysis below takes as given.

A.2.1 Useful Lemmas

The next lemma generalizes Lemma 1 to encompass noisy news:

Lemma 6. *Fix any $x = (f, \sigma)$. Then for all t and $\boldsymbol{\omega}$,*

(i) $m_t^*(\boldsymbol{\omega}) = 0$ if and only if $\mathbb{E}_x\left[\exp\left(\mu_t^{-1}\nu_x(\tilde{\boldsymbol{\omega}}, t)\right)\right] < 1$;

(ii) $m_t^*(\boldsymbol{\omega}) = 1$ if and only if $\mathbb{E}_x\left[\exp\left(-\mu_t^{-1}\nu_x(\tilde{\boldsymbol{\omega}}, t)\right)\right] < 1$;

(iii) otherwise $m_t^*(\boldsymbol{\omega}) \in (0, 1)$ and satisfies the following first-order condition:

$$\nu_x(\boldsymbol{\omega}, t) = \mu_t \cdot \log\left(\frac{m_t^*(\boldsymbol{\omega})}{1 - m_t^*(\boldsymbol{\omega})} \cdot \frac{1 - \mathbb{E}_x[m_t^*(\tilde{\boldsymbol{\omega}})]}{\mathbb{E}_x[m_t^*(\tilde{\boldsymbol{\omega}})]}\right).$$

Proof. The proof is analogous to that of Lemma 1 and thus is omitted. $\square$

The next lemma is adapted from Milgrom (1981):

Lemma 7 (Milgrom (1981))**.** *Assume Assumption 6. Then for any σ that is symmetric and non-degenerate, $\mathbb{E}_x[h(\tilde{a}_\beta) \mid \omega_\beta = \omega]$ is increasing (resp. strictly increasing) in ω if h is an increasing (resp. strictly increasing) function.*

Proof. See Milgrom (1981). $\square$

The next lemma summarizes the key properties of $\nu_x(\boldsymbol{\omega}, t)$'s:

Lemma 8. *Assume Assumptions 1, 5 and 6, and take any σ that is symmetric and and non-degenerate. Then,*

(i) $\nu_x(\boldsymbol{\omega}_{mn}, t) \geq \nu_x(\boldsymbol{\omega}_{nm}, t)$ for all t and $m > n$;

(ii) For the case of $t = 0$,

(a) $\nu_x(\boldsymbol{\omega}_{mn}, 0) = -\nu_x(\boldsymbol{\omega}_{nm}, 0)$ *for all* m *and* n*;*

(b) $\nu_x(\boldsymbol{\omega}_{mn}, 0) > \nu_x(\boldsymbol{\omega}_{nm}, 0)$ *for all* $m > n$*;*

(c) $\nu_x(\boldsymbol{\omega}_{K1}, 0) = \max_{\boldsymbol{\omega}} \nu_x(\boldsymbol{\omega}, 0)$.

Proof. Part (i): By symmetry,

$$\nu_x(\boldsymbol{\omega}_{mn}, t) = \mathbb{E}_x[u(\tilde{a}_\beta, t) \mid \omega_\beta = \omega_n] - \mathbb{E}_x[u(-\tilde{a}_\beta, t) \mid \omega_\beta = \omega_m],$$

and hence

$$\begin{aligned}\nu_x(\boldsymbol{\omega}_{mn}, t) - \nu_x(\boldsymbol{\omega}_{nm}, t) = {} & \mathbb{E}_x[u(\tilde{a}_\beta, t) + u(-\tilde{a}_\beta, t) \mid \omega_\beta = \omega_n] \\ & - \mathbb{E}_x[u(\tilde{a}_\beta, t) + u(-\tilde{a}_\beta, t) \mid \omega_\beta = \omega_m].\end{aligned}$$

In the above derivation, exploiting the concavity of $u(\cdot, t)$ and Lemma 7 shows that the term $\mathbb{E}_x[u(\tilde{a}_\beta, t) + u(-\tilde{a}_\beta, t) \mid \omega_\beta = \omega]$ is decreasing in ω, and straightforward rearranging gives the desired result.

Part (ii): Using the fact that $u(a, 0) = u(-a, 0)$ in the above derivation, we obtain

$$\begin{aligned}\nu_x(\boldsymbol{\omega}, 0) &= \mathbb{E}_x[u(\tilde{a}_\beta, 0) \mid \omega_\beta = \omega'] - \mathbb{E}_x[u(-\tilde{a}_\beta, 0) \mid \omega_\beta = \omega] \\ &= \mathbb{E}_x[u(\tilde{a}_\beta, 0) \mid \omega_\beta = \omega'] - \mathbb{E}_x[u(\tilde{a}_\beta, 0) \mid \omega_\beta = \omega],\end{aligned}$$

where the term $\mathbb{E}_x[u(\tilde{a}_\beta, 0) \mid \omega_\beta = \omega]$ is strictly decreasing in ω by Lemma 7 and the assumption that $u(\cdot, 0)$ is strictly decreasing on $\mathbb{R}_+$. The rest of the proof is immediate and thus is omitted. □

A.2.2 Proofs of Main Results

For ease, we will write $\Delta_{x,mn}(t) = \nu_x(\boldsymbol{\omega}_{mn}, t)$ and suppress the notation for t in the case of $t = 0$. We will normalize μ_t's to one.

Proof of Theorem 3

Proof. The case of degenerate policies is straightforward, so assume non-degenerate policies from now on. For the same reason as discussed in the proof of Theorem 1, it suffices to show that $m_t^*(\boldsymbol{\omega}_{mn}) \geq m_t^*(\boldsymbol{\omega}_{nm})$ for all $t \geq 0$ and $m > n$, and that the inequality is strict for a positive measure of voters.

The proof below fixes any $m > n$:

Step 1 First, combining Lemma 6 and Lemma 8 (i) shows that $m_t^*\left(\boldsymbol{\omega}_{mn}\right) \geq m_t^*\left(\boldsymbol{\omega}_{nm}\right)$ for all $t \geq 0$.

Step 2 From Parts (ii (a)) and (ii (b)) of Lemma 8, it follows that

$$\begin{aligned}
\mathbb{E}_x\left[\exp\left(\tilde{\Delta}_{x,mn}\right)\right] =& \sum_{m=1}^{K} \mathbb{P}_x\left(\boldsymbol{\omega}_{mm}\right)\exp\left(\Delta_{x,mm}\right) + \sum_{m=1}^{K}\sum_{n=1}^{m-1} \mathbb{P}_x\left(\boldsymbol{\omega}_{mn}\right)\gamma\left(\Delta_{x,mn}\right) \\
>& \sum_{m=1}^{K} \mathbb{P}_x\left(\boldsymbol{\omega}_{mm}\right)\cdot 1 + \sum_{m=1}^{K}\sum_{n=1}^{m-1} \mathbb{P}_x\left(\boldsymbol{\omega}_{mn}\right)\cdot 2 \\
=& 1,
\end{aligned}$$

and hence that $\mathbb{E}_x\left[\exp\left(-\tilde{\Delta}_{x,mn}\right)\right] > 1$. By continuity, the above inequalities hold true in a neighborhood I'' of $t = 0$, and so $m_t^* \in (0,1)$ for $t \in I''$ by Lemma 6.

Step 3 By continuity and Lemma 8 (ii (b)), there exists a neighborhood I''' of $t = 0$ where $\Delta_{x,mn}(t) > \Delta_{x,nm}(t)$. Intersecting this neighborhood with that described in Step 2, we obtain $m_t^*\left(\boldsymbol{\omega}_{mn}\right) > m_t^*\left(\boldsymbol{\omega}_{nm}\right)$ for all $t \in I'' \cap I'''$, and this completes the proof.

$\square$

Proof of Lemma 3

Proof. Take any $f \succeq f'$ with ρ being the Markov kernel, and write $x = (f, \sigma)$ and $x' = (f', \sigma)$ for notational ease. For any $\boldsymbol{\omega}, \boldsymbol{\omega}'$, define

$$\pi\left(\boldsymbol{\omega}', \boldsymbol{\omega}\right) = \frac{\mathbb{P}_x\left(\boldsymbol{\omega}\right)\rho\left(\boldsymbol{\omega}' \mid \boldsymbol{\omega}\right)}{\sum_{\tilde{\boldsymbol{\omega}}} \mathbb{P}_x\left(\boldsymbol{\omega}\right)\rho\left(\boldsymbol{\omega}' \mid \tilde{\boldsymbol{\omega}}\right)} = \frac{\mathbb{P}_x\left(\boldsymbol{\omega}\right)\rho\left(\boldsymbol{\omega}' \mid \boldsymbol{\omega}\right)}{\mathbb{P}_{x'}\left(\boldsymbol{\omega}'\right)},$$

and notice that $\pi\left(\boldsymbol{\omega}', \boldsymbol{\omega}\right) \geq 0$ for all $\boldsymbol{\omega}, \boldsymbol{\omega}'$ and that $\sum_{\boldsymbol{\omega}} \pi\left(\boldsymbol{\omega}', \boldsymbol{\omega}\right) = 1$ for all $\boldsymbol{\omega}$.

By the definition of garbling,

$$\mathbb{P}_{x'}\left(\boldsymbol{\omega}'\right) = \sum_{\boldsymbol{\omega}, \mathbf{a} \in \operatorname{supp}(\sigma)} \rho\left(\boldsymbol{\omega}' \mid \boldsymbol{\omega}\right) f\left(\boldsymbol{\omega} \mid \mathbf{a}\right)\sigma(\mathbf{a}) = \sum_{\boldsymbol{\omega}} \rho\left(\boldsymbol{\omega}' \mid \boldsymbol{\omega}\right)\mathbb{P}_x\left(\boldsymbol{\omega}\right),$$

and

$$
\begin{aligned}
\nu_{x'}(\boldsymbol{\omega}', t) =& \mathbb{E}_{x'}\left[v(\mathbf{a}, t) \mid \boldsymbol{\omega}'\right] \\
=& \frac{\sum\limits_{\mathbf{a} \in \operatorname{supp}(\sigma)} f'(\boldsymbol{\omega}' \mid \mathbf{a})\, \sigma(\mathbf{a})\, v(\mathbf{a}, t)}{\sum\limits_{\mathbf{a} \in \operatorname{supp}(\sigma)} f'(\boldsymbol{\omega}' \mid \mathbf{a})\, \sigma(\mathbf{a})} \\
=& \frac{\sum\limits_{\mathbf{a} \in \operatorname{supp}(\sigma)} \sum\limits_{\boldsymbol{\omega}} \rho(\boldsymbol{\omega}' \mid \boldsymbol{\omega})\, f(\boldsymbol{\omega} \mid \mathbf{a})\, \sigma(\mathbf{a})\, v(\mathbf{a}, t)}{\sum\limits_{\mathbf{a} \in \operatorname{supp}(\sigma)} \sum\limits_{\boldsymbol{\omega}} \rho(\boldsymbol{\omega}' \mid \boldsymbol{\omega})\, f(\boldsymbol{\omega} \mid \mathbf{a})\, \sigma(\mathbf{a})} \\
=& \frac{\sum\limits_{\boldsymbol{\omega}} \rho(\boldsymbol{\omega}' \mid \boldsymbol{\omega})\, \mathbb{P}_x(\boldsymbol{\omega})\, \nu_x(\boldsymbol{\omega}, t)}{\sum\limits_{\boldsymbol{\omega}} \rho(\boldsymbol{\omega}' \mid \boldsymbol{\omega})\, \mathbb{P}_x(\boldsymbol{\omega})} \\
=& \sum_{\boldsymbol{\omega}} \pi(\boldsymbol{\omega}', \boldsymbol{\omega})\, \nu_x(\boldsymbol{\omega}, t).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathbb{E}_{x'}\left[\nu_{x'}(\boldsymbol{\omega}', t)\right] = \sum_{\boldsymbol{\omega}'} \left(\sum_{\boldsymbol{\omega}} \pi(\boldsymbol{\omega}', \boldsymbol{\omega})\, \nu_x(\boldsymbol{\omega}, t) \right) \cdot \mathbb{P}_{x'}(\boldsymbol{\omega}') \\
= \sum_{\boldsymbol{\omega}} \nu_x(\boldsymbol{\omega}, t)\, \mathbb{P}_x(\boldsymbol{\omega}) \cdot \sum_{\boldsymbol{\omega}'} \rho(\boldsymbol{\omega}' \mid \boldsymbol{\omega}) = \mathbb{E}_x\left[\nu_x(\boldsymbol{\omega}, t)\right],
\end{aligned}
$$

and this completes the proof. □

Proof of Lemma 4

Proof. Take any $f \succeq f'$ with ρ being the Markov kernel, and write $x = (f, \sigma)$ and $x' = (f', \sigma)$ for notational ease. Using the proven fact by Lemma 8 (ii (c)) in the derivation of Lemma 3, we obtain that for all $\boldsymbol{\omega}'$,

$$\begin{aligned}
\nu_{x'}\left(\boldsymbol{\omega}', 0\right) =& \frac{\sum_{\boldsymbol{\omega}} \sum_{\mathbf{a} \in \operatorname{supp}(\sigma)} v\left(\mathbf{a}, 0\right) \rho\left(\boldsymbol{\omega}' \mid \boldsymbol{\omega}\right) f\left(\boldsymbol{\omega} \mid \mathbf{a}\right) \sigma(\mathbf{a})}{\sum_{\boldsymbol{\omega}} \sum_{\mathbf{a} \in \operatorname{supp}(\sigma)} \rho\left(\boldsymbol{\omega}' \mid \boldsymbol{\omega}\right) f\left(\boldsymbol{\omega} \mid \mathbf{a}\right) \sigma\left(\mathbf{a}\right)} \\
\leq& \max_{\boldsymbol{\omega}} \frac{\sum_{\mathbf{a} \in \operatorname{supp}(\sigma)} v\left(\mathbf{a}, 0\right) f\left(\boldsymbol{\omega} \mid \mathbf{a}\right) \sigma(\mathbf{a})}{\sum_{\mathbf{a} \in \operatorname{supp}(\sigma)} f\left(\boldsymbol{\omega} \mid \mathbf{a}\right) \sigma\left(\mathbf{a}\right)} \\
=& \max_{\boldsymbol{\omega}} \nu_x\left(\boldsymbol{\omega}, 0\right) \\
=& \nu_x\left(\boldsymbol{\omega}_{K1}, 0\right).
\end{aligned}$$

Using the same fact again yields $\nu_{x'}(\boldsymbol{\omega}_{K1}, 0) = \max_{\boldsymbol{\omega}'} \nu_{x'}\left(\boldsymbol{\omega}', t\right) \leq \nu_x\left(\boldsymbol{\omega}_{K1}, 0\right)$, and this completes the proof. □

Proof of Theorem 4

Proof. Part (i): Exploiting Lemma 3 and the convexity of exponential function, we obtain

$$\begin{aligned}
& \mathbb{E}_{x'}\left[\exp\left(\nu_{x'}\left(\boldsymbol{\omega}', t\right)\right)\right] \\
=& \sum_{\boldsymbol{\omega}'} \mathbb{P}_{x'}\left(\boldsymbol{\omega}'\right) \exp\left(\nu_{x'}\left(\boldsymbol{\omega}', t\right)\right) \\
=& \sum_{\boldsymbol{\omega}'} \mathbb{P}_{x'}\left(\boldsymbol{\omega}'\right) \exp\left(\sum_{\boldsymbol{\omega}} \pi\left(\boldsymbol{\omega}', \boldsymbol{\omega}\right) \nu_x(\boldsymbol{\omega}, t)\right) \\
\leq& \sum_{\boldsymbol{\omega}', \boldsymbol{\omega}} \mathbb{P}_{x'}\left(\boldsymbol{\omega}'\right) \pi\left(\boldsymbol{\omega}', \boldsymbol{\omega}\right) \exp\left(\nu_x(\boldsymbol{\omega}, t)\right) \\
=& \sum_{\boldsymbol{\omega}} \left(\sum_{\boldsymbol{\omega}'} \rho(\boldsymbol{\omega}' \mid \boldsymbol{\omega})\right) \mathbb{P}_x\left(\boldsymbol{\omega}\right) \exp\left(\nu_x\left(\boldsymbol{\omega}, t\right)\right) \\
=& \mathbb{E}_x\left[\exp\left(\nu_x\left(\boldsymbol{\omega}, t\right)\right)\right],
\end{aligned}$$

and likewise $\mathbb{E}_{x'}\left[\exp\left(-\nu_{x'}\left(\boldsymbol{\omega}', t\right)\right)\right] \leq \mathbb{E}_{x'}\left[\exp\left(-\nu_{x'}\left(\boldsymbol{\omega}', t\right)\right)\right]$. The result can then be read from Lemma 6.

Part (ii): By Assumption 4,

$$\mathbb{E}_x\left[\exp\left(\tilde{\Delta}_{x,mn}(t)\right)\right] \leq \exp(\kappa|t|)\,\mathbb{E}_x\left[\exp\left(\tilde{\Delta}_{x,mn}\right)\right],$$

where the last term of the above inequality can be bounded above as follows:

$$\begin{aligned}\mathbb{E}_x\left[\exp\left(\tilde{\Delta}_{x,mn}\right)\right] &= \sum_{m=1}^{K}\mathbb{P}_x\left(\boldsymbol{\omega}_{mm}\right)\exp\left(\Delta_{x,mm}\right) + \sum_{m=1}^{K}\sum_{n=1}^{m-1}\mathbb{P}_x\left(\boldsymbol{\omega}_{mm}\right)\gamma\left(\Delta_{x,mn}\right)\\ &\leq \sum_{m=1}^{K}\mathbb{P}_x\left(\boldsymbol{\omega}_{mm}\right) + \frac{1}{2}\left(1-\sum_{m=1}^{K}\mathbb{P}_x\left(\boldsymbol{\omega}_{mm}\right)\right)\gamma\left(\Delta_{x,K1}\right).\end{aligned}$$

Thus, a necessary condition for $\mathbb{E}_x\left[\exp\left(\tilde{\Delta}_{x,mn}(t)\right)\right] \geq 1$ to hold true is

$$\gamma\left(\Delta_{x,K1}\right) \geq 2\left[\frac{\exp(\kappa|t|) - \sum_{m=1}^{K}\mathbb{P}_x(\boldsymbol{\omega}_{mm})}{1-\sum_{m=1}^{K}\mathbb{P}_x(\boldsymbol{\omega}_{mm})}\right]. \tag{A.2}$$

Two things are noteworthy. First, the right-hand side of (A.2) is strictly increasing in $\sum_{m=1}^{K}\mathbb{P}_x(\boldsymbol{\omega}_{mm})$. Second, for all x,

$$\sum_{m=1}^{K}\mathbb{P}_x\left(\boldsymbol{\omega}_{mm}\right) = \sum_{i,j=1}^{N}\sum_{m=1}^{K} f\left(\omega_m \mid a_i\right) f\left(\omega_m \mid a_j\right)\sigma_{ij} \geq \sum_{i,j=1}^{N}\sum_{m=1}^{K}\frac{1}{K}\cdot\frac{1}{K}\cdot\sigma_{ij} = \frac{1}{K},$$

where the inequality follows from solving the optimization problem below:

$$\min_{z_m,z_m'}\sum_{m=1}^{K} z_m z_m' \text{ s.t. } z_m, z_m' \geq 0\ \forall m \text{ and } \sum_{m=1}^{K} z_m = \sum_{m=1}^{K} z_m' = 1,$$

which attains its minimum value at $z_m = z_m' = \frac{1}{K}$, $m = 1, \cdots, K$. Thus, (A.2) holds true only if

$$\Delta_{x,K1} \geq \gamma^{-1}\left(2\left[\frac{K\exp(\kappa|t|)-1}{K-1}\right]\right), \tag{A.3}$$

where the right-hand side of the above condition is independent of x. The remainder of the proof follows closely the argument sketched in Section 5.2.2 and thus is omitted.

□

A.3 Proofs of Section 6

Proof. Part (i): Differentiating $\widehat{u}(a,t)$ with respect to a and using the fact that

$$\frac{u_a\left(a^\circ(t), \mathcal{B}\left(a^\circ(t)\right), t\right)}{u_b\left(a^\circ(t), \mathcal{B}\left(a^\circ(t)\right) t\right)} + \mathcal{B}'\left(a^\circ(t)\right) = 0,$$

we obtain

$$\frac{d}{da}\widehat{u}(a,t) = u_a\left(a, \mathcal{B}(a), t\right) + u_b\left(a, \mathcal{B}(a), t\right)\mathcal{B}'(a),$$

and hence

$$\left.\frac{d}{da}\widehat{u}(a,t)\right|_{a=a^\circ(t)} = 0.$$

Meanwhile, since $a^\circ(t)$ is decreasing in t and $a^\circ(I) = I$, it follows that for any $a < a^\circ(t)$, there exists $t' > t$ such that $a^\circ\left(t'\right) = a$ and hence

$$\frac{u_a\left(a, \mathcal{B}\left(a\right), t'\right)}{u_b\left(a, \mathcal{B}\left(a\right), t'\right)} + \mathcal{B}'\left(a\right) = 0.$$

Therefore,

$$\begin{aligned}
\left.\frac{d}{da}\widehat{u}(a,t)\right|_{a<a^\circ(t)} &= u_a\left(a, \mathcal{B}(a), t\right) - u_b\left(a, \mathcal{B}(a), t\right)\frac{u_a\left(a, \mathcal{B}\left(a\right), t'\right)}{u_b\left(a, \mathcal{B}(a), t'\right)} \\
&= u_b\left(a, \mathcal{B}(a), t\right)\left[\frac{u_a\left(a, \mathcal{B}\left(a\right), t\right)}{u_b\left(a, \mathcal{B}(a), t\right)} - \frac{u_a\left(a, \mathcal{B}\left(a\right), t'\right)}{u_b\left(a, \mathcal{B}(a), t'\right)}\right] \\
&< 0,
\end{aligned}$$

where the inequality holds true because $u_b < 0$ and $-\frac{u_a}{u_b}$ is strictly increasing in t. The proof of $\left.\frac{d}{da}\widehat{u}(a,t)\right|_{a>a^\circ(t)} > 0$ is analogous and thus is omitted.

Part (ii): Differentiating $\widehat{u}(a,t)$ with respect to a and rearranging, we obtain

$$\begin{aligned}
\frac{d^2}{da^2}\widehat{u}(a,t) =& u_{aa}\left(a, \mathcal{B}(a), t\right) + \left[u_{ab}\left(a, \mathcal{B}(a), t\right) + u_{ba}\left(a, \mathcal{B}(a), t\right)\right]\mathcal{B}'(a) \\
& + u_{bb}\left(a, \mathcal{B}(a), t\right)\left(\mathcal{B}'(a)\right)^2 + u_b\left(a, \mathcal{B}(a), t\right)\mathcal{B}''(a) \\
=& \left[1, \mathcal{B}'(a)\right]\begin{bmatrix} u_{aa} & u_{ab} \\ u_{ba} & u_{bb} \end{bmatrix}\begin{bmatrix} 1 \\ \mathcal{B}'(a) \end{bmatrix} + u_b\left(a, \mathcal{B}(a), t\right)\mathcal{B}''(a) \\
<& 0,
\end{aligned}$$

where the inequality holds true because $u(a,b,t)$ is strictly concave in (a,b) and $\mathcal{B}(a)$ is strictly concave in a.

Part (iii): Differentiating $\widehat{u}(a,t)$ with respect to a and t yields

$$\frac{\partial^2 \widehat{u}(a,t)}{\partial a \partial t} = u_{at}\left(a, \mathcal{B}(a), t\right) + u_{bt}\left(a, \mathcal{B}(a), t\right) \mathcal{B}'(a).$$

Thus, $\frac{\partial^2 \widehat{u}(a,t)}{\partial a \partial t} > 0$ if $u_{bt} \geq 0$ and $u_{at} \leq 0$ and one of these inequalities is strict. □

B Online Appendix (For Online Publication Only)

B.1 Omitted Details From Examples 1 and 2

Example 1 (Continued)**.** The outcomes of our interest feature $\mathbf{\Sigma} = \frac{1}{4}\mathbf{J}_2$, where $\mathbf{J}_2$ denotes the 2×2 matrix of all ones.

Equilibrium policies By Theorem 1, the set of policy matrices that can be attained in equilibrium is

$$\mathcal{E}\left(\frac{1}{4}\mathbf{J}_2\right) = \left\{\mathbf{A} : \left[\mathbf{A}, \frac{1}{4}\mathbf{J}_2\right] \text{ is } \widehat{\mathbf{W}}_2 - \text{ IC}\right\}.$$

Under the assumption that policies outside the support of equilibrium strategies can be perfectly discerned and penalized by voters, incentive compatibility means that each type of candidate prefers his own policy to that of the opposing type. That is, either $(t,t') = \left(-\frac{1}{4}, -\frac{3}{4}\right)$ or $\left(-\frac{3}{4}, -\frac{1}{4}\right)$ satisfies (i)

$$\frac{3}{4}\left(R - \delta_+|t+a_1|\right) - \frac{1}{4}\delta_-|t-a_1| \geq \frac{1}{4}\left(R - \alpha_+|t+a_2|\right) - \frac{1}{4}\delta_-|t-a_1| - \frac{1}{4}\delta_-|t-a_2|,$$

and (ii)

$$\frac{1}{4}\left(R - \delta_+|t'+a_2|\right) - \frac{1}{4}\delta_-|t'-a_1| - \frac{1}{4}\delta_-|t'-a_2| \geq \frac{3}{4}\left(R - \delta_+|t'+a_1|\right) - \frac{1}{4}\delta_-|t'-a_1|.$$

Solving this system of inequalities yields the policy profiles depicted in the blue and yellow areas of Figure 3.

Attention set The attention set of voter $t \approx 0$ is

$$\mathcal{A}_t\left(\frac{1}{4}\mathbf{J}_2, \mu\right) = \left\{\mathbf{A} : a_2 - a_1 \geq \mu\gamma^{-1}\left(4\exp(2|t|/\mu) - 2\right)\right\},$$

where tedious but straightforward algebra as in the proof of Theorem 2 shows that the term $\mu\gamma^{-1}\left(4\exp(2|t|/\mu) - 2\right)$ is strictly increasing in μ. Combining this result with the characterization of the set $\mathcal{E}\left(\frac{1}{4}\mathbf{J}_2\right)$ yields the patterns delineated at the beginning of Section 4.

Example 2 (Continued)**.** Let $-a_2 \leq -a_1 \leq 0 \leq a_1 \leq a_2$ denote the policies adopted by the various types of the candidates. By assumption, $\mathbf{\Sigma} = 1$ if $a_1 = a_2$ and $\frac{1}{4}\mathbf{J}_2$ if $a_1 < a_2$.

Equilibrium policies For any $\mathbf{a} \in A_\alpha \times A_\beta$, define

$$w_\xi\left(\mathbf{a}\right) = \frac{1}{2}\left[f_\xi\left(\boldsymbol{\omega}_{11} \mid \mathbf{a}\right) + f_\xi\left(\boldsymbol{\omega}_{22} \mid \mathbf{a}\right)\right] + f_\xi\left(\boldsymbol{\omega}_{21} \mid \mathbf{a}\right)$$

as candidate β's expected winning probability in policy state $\mathbf{a}$. In equilibrium, candidate β's expected utility conditional on candidate α's policy being of $a \in A_\alpha$ is

$$\overline{w}_\xi(a) = \frac{1}{2}w_\xi\left(a, a_1\right) + \frac{1}{2}w_\xi\left(a, a_2\right),$$

and the problem faced by any type t of candidate α is

$$\max_{a \in A_\alpha}\left(1 - \overline{w}_\xi(a)\right)\left(R - \delta_+|t - a|\right) + \overline{w}_\xi\left(a\right)u_{-,\xi}\left(a, t\right),$$

where

$$u_{-,\xi}(a,t) = \frac{-\delta_-}{\overline{w}_\xi(a)}\left[\frac{1}{2}w_\xi\left(a, a_1\right)|t - a_1| + \frac{1}{2}w_\xi\left(a, a_2\right)|t - a_2|\right]$$

denotes the conditional expected utility in the case the candidate proposes policy a and loses the election to candidate β. Solving this optimization problem yields the policy profiles depicted in Figure 6.

Attention set Normalize marginal attention cost to one. When $t \approx 0$, the voter's attention set is given by the following inequality:

$$\nu_{\xi}\left(\boldsymbol{\omega}_{21}, 0\right) \geq \gamma^{-1}\left(\frac{2\left[\exp\left(2|t|\right) - \sum_{m=1}^{2} \mathbb{P}_{\xi}\left(\boldsymbol{\omega}_{mm}\right)\right]}{1 - \sum_{m=1}^{2} \mathbb{P}_{\xi}\left(\boldsymbol{\omega}_{mm}\right)}\right),$$

where

$$\nu_{\xi}\left(\boldsymbol{\omega}_{21}, 0\right) = \frac{\left(a_2 - a_1\right)^2}{\left(2 - a_1 - a_2\right)\left[2\xi + \left(1 - \xi\right)\left(a_1 + a_2\right)\right]},$$

and

$$\sum_{m=1}^{2} \mathbb{P}_{\xi}\left(\boldsymbol{\omega}_{mm}\right) = \frac{1}{4}\left[\left(1 - \xi\right)\left(2 - a_1 - a_2\right)\right]^2 + \frac{1}{4}\left[2\xi + \left(1 - \xi\right)\left(a_1 + a_2\right)\right]^2.$$

B.2 Omitted Results From Section 5

In the model presented in Section 5, write the attention set as $\mathcal{A}_t\left(\boldsymbol{\Sigma}, f, \mu\right)$ in order to make its dependence on μ explicit. The next corollary generalizes Theorem 2 to encompass noisy news:

Corollary 1. *Assume Assumptions 1-3 and 5-6, and let $\boldsymbol{\Sigma}$ be any probability matrix of order $N \geq 2$ for which the set $\mathcal{E}\left(\boldsymbol{\Sigma}, f\right)$ is non-empty. Then for any $t < 0$ that satisfies Assumption 4 and where $\nu_{\langle f, \mathbf{A}, \boldsymbol{\Sigma}\rangle}\left(\boldsymbol{\omega}, t\right) > 0$ for some $\mathbf{A} \in \mathcal{E}\left(\boldsymbol{\Sigma}, f\right)$ and $\boldsymbol{\omega} \in \Omega$, the following hold true as we increase μ from zero to infinity:*

(i) the set $\mathcal{EA}_t\left(\boldsymbol{\Sigma}, f, \mu\right)$ shrinks and the set $\mathcal{EI}_t\left(\boldsymbol{\Sigma}, f, \mu\right)$ expands;

(ii) $\min\left\{u\left(a_1, 0\right) - u\left(a_N, 0\right) : \mathbf{A} \in \mathcal{EA}_t\left(\boldsymbol{\Sigma}, f, \mu\right)\right\}$ increases;

(iii) the above described objects do not always stay constant.

Proof. For the same reason as discussed in the proof of Theorem 3, it suffices to show that as μ increases from zero to infinity, (i) the set $\mathcal{A}_t\left(\boldsymbol{\Sigma}, f, \mu\right)$ shrinks, (ii) $\min\left\{u(a_1, 0) - u(a_N, 0) : \mathbf{A} \in \mathcal{A}_t(\boldsymbol{\Sigma}, f, \mu)\right\}$ increases, (iii) the above described objects do not always stay constant.

Part (i): the proof is analogous to that of Theorem 2 (i) and thus is omitted.

Part (ii): Take any policy matrix $\mathbf{A}$ of order N and write $x = \langle f, \mathbf{A}, \boldsymbol{\Sigma} \rangle$ for convenience. By Assumption 4,

$$\mathbb{E}_x \left[\exp\left(\mu^{-1} \tilde{\Delta}_{x,mn}(t) \right) \right] \leq \exp\left(\kappa |t| / \mu \right) \mathbb{E}_x \left[\exp\left(\mu^{-1} \tilde{\Delta}_{x,mn} \right) \right],$$

where the last term of the above inequality can be bounded above as follows:

$$\begin{aligned} \mathbb{E}_x \left[\exp\left(\mu^{-1} \tilde{\Delta}_{x,mn} \right) \right] &= \sum_{m=1}^{K} \mathbb{P}_x \left(\boldsymbol{\omega}_{mm} \right) \exp\left(\mu^{-1} \Delta_{x,mm} \right) + \sum_{m=1}^{K} \sum_{n=1}^{m-1} \mathbb{P}_x \left(\boldsymbol{\omega}_{mm} \right) \gamma\left(\mu^{-1} \Delta_{x,mn} \right) \\ &\leq \sum_{m=1}^{K} \mathbb{P}_x \left(\boldsymbol{\omega}_{mm} \right) + \frac{1}{2} \left(1 - \sum_{m=1}^{K} \mathbb{P}_x \left(\boldsymbol{\omega}_{mm} \right) \right) \gamma\left(z \right), \\ \text{where } z &\triangleq \mu^{-1} \left(u\left(a_1, 0 \right) - u\left(a_N, 0 \right) \right). \end{aligned}$$

Thus, a necessary condition for $\mathbb{E}_x \left[\exp\left(\mu^{-1} \tilde{\Delta}_{x,mn}(t) \right) \right] \geq 1$ to hold true is

$$\gamma(z) \geq 2 \left[\frac{\exp(\kappa |t| / \mu) - \sum_{m=1}^{K} \mathbb{P}_x(\boldsymbol{\omega}_{mm})}{1 - \sum_{m=1}^{K} \mathbb{P}_x(\boldsymbol{\omega}_{mm})} \right],$$

where the right-hand side of the above inequality is bounded below by the following term that is independent of x (the derivation is exact same as that of (A.3)):

$$2 \left[\frac{K \exp(\kappa |t| / \mu) - 1}{K - 1} \right].$$

The remainder of the proof follows closely that of Theorem 2 (ii) and thus is omitted.

Part (iii): When μ is small, $\mathcal{A}_t \left(\boldsymbol{\Sigma}, f, \mu \right) \neq \emptyset$ by the assumption that $\nu_{\langle f, \mathbf{A}, \boldsymbol{\Sigma} \rangle} \left(\boldsymbol{\omega}, t \right) > 0$ for some $\mathbf{A} \in \mathcal{E} \left(\boldsymbol{\Sigma}, f \right)$ and $\boldsymbol{\omega} \in \Omega$. When μ is large, the following holds true for any policy matrix of order N:

$$\mathbb{E}_x \left[\exp\left(\mu^{-1} \tilde{\Delta}_{x,mn}(t) \right) \right] \approx \mathbb{E}_x \left[1 + \mu^{-1} \tilde{\Delta}_{x,mn}(t) \right] \underbrace{<}_{(1)} \mathbb{E}_x \left[1 + \mu^{-1} \tilde{\Delta}_{x,mn} \right] \underbrace{=}_{(2)} 0,$$

where (1) follows by Assumption 3 and (2) Lemma 8. Thus $\mathcal{A} \left(\boldsymbol{\Sigma}, f, \mu \right) = \emptyset$ in this case, and this completes the proof. $\square$

B.3 Limited Commitment and Campaign Message

In the baseline model, now suppose, instead, that the winning candidate honors his policy proposal in Stage 2 of the election game with probability η and adopts his most preferred position equal to his preference type otherwise. The parameter $\eta \in [0, 1]$ captures the winner's level of commitment power and is taken as exogenously given throughout the analysis. The policy proposal, now regarded as a campaign promise, serves a new role beyond what we have seen in the baseline model: it enables inferences about the winner's type in the case the latter reneges and does what pleases himself the most.

The upcoming analysis limits candidates to using *monotone pure symmetric strategies*, that is, pure symmetric strategies that are strictly increasing in the candidate's type. As the reader will soon realize, this assumption helps preserve the monotone relationship between policy positions and voters' decisions that the analysis so far has utilized a lot. We could alternatively consider mixed strategies that satisfy the monotone likelihood ratio property, whereby high type candidates are more likely to propose high policies than low type candidates do. We choose not to pursue this route in order to keep the analysis simple and tractable.

Let $-t_N < \cdots < -t_1 \leq 0 \leq t_1 < \cdots < t_N$ denote candidates' types and $-a_N < \cdots < -a_1 \leq 0 \leq a_1 < \cdots < a_N$ their policies, where N is now an exogenous integer assumed to be greater than one. For $i, j = 1, \cdots, N$, write $\mathbf{t}_{ij} = (-t_i, t_j)$, and define

$$\tilde{v}(\mathbf{a}_{ij}, t) = \eta v\left(\mathbf{a}_{ij}, t\right) + (1 - \eta) v\left(\mathbf{t}_{ij}, t\right)$$

as voter t's utility difference from choosing candidate β over candidate α in state $\mathbf{a}_{ij}$. Let

$$\tilde{u}_+\left(a_c, t_c\right) = \eta u_+\left(a_c, t_c\right) + (1 - \eta) u_+\left(t_c, t_c\right)$$

and

$$\tilde{u}_-\left(a_c, t_{-c}\right) = \eta u_-\left(a_c, t_{-c}\right) + (1 - \eta) u_-\left(t_c, t_{-c}\right)$$

be the utility of the winner and the loser, respectively, when the winning policy proposal is a_c and candidates' types are t_c and t_{-c}. Replacing v, u_+ and u_- with $\tilde{v}, \tilde{u}_+$ and $\tilde{u}_-$, respectively, leads us to define players' value functions analogously to what we did in the baseline model.

Under the aforementioned restrictions, the probability matrix $\mathbf{\Sigma}$ is fully determined by candidates' type distribution and will thus be taken as exogenously given. Let $\mathbf{T}$ be the square matrix whose ij^{th} entry is $\mathbf{t}_{ij}$, and let $\mathbf{A}$ and $\mathbf{W}$ be as above. For any given level η of commitment power, a tuple $\langle \mathbf{A}, \mathbf{W} \rangle$ can be attained in equilibrium if (1) $\mathbf{A}$ is *incentive compatible for candidates under* $\mathbf{W}$ (hereafter, $\langle \eta, \mathbf{W} \rangle$-IC), and if (2) $\mathbf{W}$ is *can be rationalized by optimal attention allocation under* $\mathbf{A}$ (hereafter, $\langle \eta, \mathbf{A} \rangle$-rationalizable). In particular, (1) can be obtained by limiting attention to monotone pure symmetric strategies in Definition 1, and (2) by treating $\mathbf{\Sigma}$ as exogenously given in Definition 2.

For any $\eta \in [0,1]$, define

$$\tilde{\mathcal{E}}(\eta) = \left\{ \mathbf{A} : \exists \mathbf{W} \text{ s.t. } \begin{array}{c} \mathbf{A} \text{ is } \langle \eta, \mathbf{W} \rangle - \text{IC} \\ \mathbf{W} \text{ is } \langle \eta, \mathbf{A} \rangle - \text{rationalizable} \end{array} \right\}$$

as the set of policy matrices that can be attained in the equilibrium of our interest, and define

$$\tilde{\mathcal{A}}_t(\eta, \mu) = \left\{ \mathbf{A} : \mathbb{E}_{[\mathbf{A}, \mathbf{T}, \mathbf{\Sigma}]} \left[\exp\left(\mu^{-1} \tilde{v}(\tilde{\mathbf{a}}, t) \right) \right] > 1 \right\}$$

as the attention set of any camp α voter $t < 0$ at any level $\mu > 0$ of marginal attention cost. Taking intersections yields the set $\widetilde{\mathcal{EA}}_t(\eta, \mu)$ of policy matrices that arouses and attracts the voter's attention in equilibrium, as well as the set $\widetilde{\mathcal{EI}}_t(\eta, \mu)$ of policy matrices that leads to ideology-based voting in equilibrium.

Armed with these definitions and notations, we can now state the main result of this section:

Corollary 2. *Assume Assumptions 1 and 2. Then for any η,*

$$\tilde{\mathcal{E}}(\eta) = \left\{ \mathbf{A} : \mathbf{A} \text{ is } \langle \eta, \widehat{\mathbf{W}}_N \rangle - \text{IC} \right\}.$$

Assume, in addition, Assumption 3. In the case where $\tilde{\mathcal{E}}(\eta)$ is non-empty, for any $t < 0$ that satisfies Assumption 4 and where $\tilde{v}(\mathbf{a}, t) > 0$ for some $\mathbf{a} \in \mathbf{A} \in \tilde{\mathcal{E}}(\eta)$, the following hold true as μ increases from zero to infinity:

(i) the set $\widetilde{\mathcal{EA}}_t(\eta, \mu)$ shrinks and the set $\widetilde{\mathcal{EI}}_t(\eta, \mu)$ expands;

(ii) $\min \left\{ u(a_1, 0) - u(a_N, 0) : \mathbf{A} \in \widetilde{\mathcal{EA}}_t(\eta, \mu) \right\}$ increases;

(iii) the above described objects do not always stay constant.

The main prediction of Corollary 2, namely extreme and exaggerated campaign messages are instrumental for arousing and attracting voters' attention in equilibrium, is a mere understatement of reality. In the first presidential election in the post-watergate era, Jimmy Carter campaigned as an honest outsider with fresh leadership, although subsequent conversations with Hubert Humphrey revealed that he was not as far away from the core values of the Democratic Party as he was portrayed to be (Popkin (1994)). As Roger Ailes famously described his role as a campaign strategist: "Every thing I did from issue selection to speeches to ads, etc. was designed to push the different kinds of the candidates further apart" (Vavreck (2009)).

Proof. By definition,

$$\begin{aligned}\tilde{v}\left(\mathbf{a}_{ij},t\right)=&\eta\left[u\left(a_j,t\right)+u\left(-a_j,t\right)-\left[u\left(a_i,t\right)+u\left(-a_i,t\right)\right]\right]\\&+\left(1-\eta\right)\left[u\left(t_j,t\right)+u\left(-t_j,t\right)-\left[u\left(t_i,t\right)+u\left(-t_i,t\right)\right]\right],\end{aligned}$$

and thus inherit all the desired properties of $v\left(\mathbf{a}_{ij},t\right)$'s. Replacing the latter with the former in the proofs of Theorems 1 and 2 gives the desired result. □

The impact of limited commitment on equilibrium outcomes is subtle. Below we give some intuition through an example:

Example 3. In Example 1, now suppose, instead, that the winning candidate honors his campaign promise with probability η and reneges with probability $1-\eta$. Tedious but straightforward algebra shows that in this case, the attention set of any voter $t \approx 0$ is defined by the following inequality:

$$\eta \cdot (a_2 - a_1) + (1-\eta) \cdot \underbrace{1/2}_{\text{inference effect}} \geq \underbrace{\mu\gamma^{-1}\left(4\exp(2|t|/\mu) - 2\right)}_{\text{hurdle of indifference}}.$$

On the right-hand side of this inequality is the same *hurdle of indifference* as in the baseline model. This term represents the hurdle that candidates need to bypass in order to capture the voter's attention, and it is this exact term that generates the phenomenon of policy extremism and exaggeration as we increase the marginal attention cost from zero to infinity in the baseline model. On the left-hand side of this inequality is a new term called the *inference effect.* This term represents the inference drawn by the voter about the winner's type, and it serves as a powerful motivator for the voter to be attentive rather than indifferent.

Which of the above described effects is dominant matters for our (rather partial) assessment of how limited commitment affects equilibrium outcomes. There are two cases to consider:

Case 1 $\mu\gamma^{-1}\left(4\exp(2|t|/\mu)-2\right)\leq\frac{1}{2}$. In this case, the inference effect itself is strong enough to overcome the hurdle of indifference. As the level of commitment power increases, the need for inference diminishes, and the logic behind Theorem 2 kicks in, suggesting that a greater degree of policy extremism is needed for arousing and attracting the voter's attention.

Case 2 $\mu\gamma^{-1}\left(4\exp(2\epsilon/\mu)-2\right)>\frac{1}{2}$. In this case, the inference effect is weak compared to the hurdle of indifference, and the opposite of the above happens as we bestow the winner with higher levels of commitment power.

References

Abramowitz, A. I., and K. L. Saunders (2008): "Is Polarization a Myth?" *Journal of Politics*, 70(2), 542-555.

Aragones, E., and T. R. Palfrey (2002): "Mixed Equilibrium in a Downsian Model with a Favored Candidate," *Journal of Economic Theory*, 103, 131-161.

Banks, J. S. (1990): "A Model of Electoral Competition with Incomplete Information," *Journal of Economic Theory*, 50, 309-325.

Barthel, M., A., Mitchell and J. Holcomb (2016): "Many Americans Believe Fake News Is Sowing Confusion," *Pew Research Center*, December 15.

Baum, M. A., and S. Kernell (1999): "Has Cable Ended the Golden Age of Presidential Television," *American Political Science Review*, 93(1), 99-114.

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

Callander, S. (2008): "Political Motivations," *Review of Economic Studies*, 75(3), 671-697.

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

——— (1985b): "Robustness of the Multidimensional Voting Model: Candidate Motivations, Uncertainty, and Convergence," *American Journal of Political Science*, 29(1), 69-95.

Campbell, A., P. Converse, W. E. Miller and D. E. Stokes (1960): *The American Voter.* Chicago, IL: The University of Chicago Press.

Chen, J., and M. Yang (2011): "Common Value Voting with Rational Inattention," *Working Paper*, Princeton University.

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

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

Duggan, J. (2017): "A Survey of Equilibrium Analysis in Spatial Model of Elections," *Working Paper*, University of Rochester.

Dunaway, J. (2016): "Mobile vs. Computer: Implications for News Audiences and Outlets," *Shorenstein Center on Media, Politics and Public Policy*, August 30.

Feldman, N. (2017): *The Three Lives of James Madison: Genius, Partisan, President.* New York, NY: Random House, 1st ed.

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

Gentzkow, M., and J. M. Shapiro (2011): "Ideological Segregation Online and Offline," *Quarterly Journal of Economics*, 126(4), 1799-1839.

——— (2016): "Media Bias in the Marketplace: Theory," in *Handbook of Media Economics*, ed. by S. Anderson, J. Waldfogel and D. Stromberg, Chapter 14, pp. 624-645. Elsevier.

Gentzkow, M., E. L. Glaeser and C. Goldin (2004): "How Newspapers Became Informative and Why It Mattered," *NBER Working Paper No. 10791.*

GERARDI, D., AND L. YARIV (2008): "Information Acquisition in Committees," *Games and Economic Behavior*, 62(2), 436-459.

GERSHKOV, A., AND B. SZENTES (2009): "Optimal Voting Schemes with Costly Information Acquisition," *Journal of Economic Theory*, 144(1), 36-68.

ISSAC, M. (2018): "FaceBook Overhauls News Feed to Focus on What Friends and Family Share," *The New York Times*, January 11, 2018.

KARTIK, N., AND R. P. MCAFEE (2007): "Signaling Character in Electoral Competition," *American Economic Review*, 97(3), 852-870.

LEE, C., AND J. L. KENT (2017): "Facebook Says Russian-Backed Election Content Reached 126 Million Americans," *NBCNews*, October 30.

LEVENDUSKY, M. (2013): *How Partisan Media Polarize America.* Chicago, IL: The University of Chicago Press.

MAĆKOWIAK, B., AND M. WIEDERHOLT (2009): "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 99(3), 769-803.

MARTIN, D. (2017): "Strategic Pricing with Rational Inattention to Quality," *Games and Economic Behavior*, 104, 131-145.

MARTINELLI, C. (2006): "Would Rational Voters Acquire Costly Information?," *Journal of Economic Theory*, 129(1), 225-251.

MATÉJKA, F., AND A. MCKAY (2012): "Simple Market Equilibria with Rationally Inattentive Consumers," *American Economic Review: Papers and Proceedings*, 102(3), 24-29.

——— (2015): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105(1), 272-298.

MATĚJKA, F., AND G. TABELLINI (2016): "Electoral Competition with Rationally Inattentive Voters," *Working Paper*, CERGE-EI and Bocconi University.

MILGROM, P. (1981): "Good News and Bad News: Representation Theorems and Applications," *The Bell Journal of Economics*, 12(2), 380-391.

Milgrom, P., and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62(1), 157-180.

Mitchell, A., J. Gottfried, J. Kiley and K. E. Matsa (2014): "Political Polarization and Media Habits," *Pew Research Center*, October 21.

Perez, S. (2017): "U.S. Consumers Now Spend 5 Hours Per Day on Mobile Devices," *TechCrunch*, March 3.

Persson, T., and G. Tabellini (2000): *Political Economics: Explaining Economic Policy.* Cambridge, MA: The MIT Press.

Peterson, E., S. Goel and S. Iyengar (2018): "Echo Chambers and Partisan Polarization: Evidence from the 2016 Presidential Campaign," *Working Paper*, Dartmouth College and Stanford University.

Persico, N. (2004): "Committee Design with Endogenous Information," *Review of Economic Studies*, 71(1), 165-191.

Petrocik, J. R. (1996): "Issue Ownership in Presidential Elections, with a 1980 Case Study," *American Journal of Political Science*, 40(3), 825-850.

Poole, K. T., and H. L. Rosenthal (2011): *Ideology and Congress.* New Brunswick (U.S.A.) and London (U.K.) : Transaction Publishers.

Popkin, S. L. (1994): *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns.* Chicago, IL: The University of Chicago Press.

Prato, C., and S. Wolton (2016): "The Voters' Curses: Why We Need Goldilocks Voters," *American Journal of Political Science*, 60(3), 726-737.

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

Puglisi, R., and J. M Snyder (2016): "Empirical Studies of Media Bias," in *Handbook of Media Economics*, ed. by S. Anderson, J. Waldfogel and D. Stromberg, Chapter 15, pp. 647-667. Elsevier.

Ravid, D. (2017): "Bargaining with Rational Inattention," *Working Paper*, University of Chicago.

Shannon, C. (1948): "A Mathematical Theory of Communication," *Bell Labs Technical Journal*, 27(3), 379-423.

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

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

Teixeira, T. S. (2014): "The Rising Cost of Consumer Attention: Why You Should Care, and What You Can Do about It," *Working Paper*, Harvard Business School.

Vavreck, L. (2009): *The Message Matters: The Economy and Presidential Campaigns.* Princeton, NJ: Princeton University Press.

Wittman, D. (1983): "Candidate Motivation: A Synthesis of Alternative Theories," *American Political Science Review*, 77(1), 142-157.

——— (2009): "Information-Constrained State-Dependent Pricing," *Journal of Monetary Economics*, 56(1), 100-124.

Yang, M. (2015): "Coordination with Flexible Information Acquisition," *Journal of Economic Theory*, 158, 721-738.

——— (2016): "Optimality of Debt under Flexible Information Acquisition," *Working Paper*, Duke University.

Yuksel, S. (2014): "Specialized Learning and Political Polarization," *Working Paper*, UCSB.