Electoral Competition with Rationally Inattentive Voters[*]

Filip Matějka[†]and Guido Tabellini[‡]

September 2018

Abstract

This paper studies how voters' selective ignorance interacts with policy design by political candidates. It shows that the selectivity empowers voters with extreme preferences and small groups, divisive issues attract most attention and public goods are underfunded. Finer granularity of information increases these inefficiencies. Rational inattention can also explain why competing opportunistic candidates do not always converge on the same policy issues, and under what conditions internet may lead to increased polarization by partisan candidates.

1 Introduction

As a result of the digital revolution, the supply of political information has become virtually unlimited and almost free. One would think that this greatly increased voters' information and awareness of political processes. Yet, the major observed changes have been compositional. As emphasized by Prior (2007), some individuals have become much more informed, others less. Informational asymmetries across issues (what one is informed about) have also become more prominent. On average, however, Americans'

public knowledge did not increase relative to the late 1980s (The Pew Research Center 2007).

A plausible explanation of these patterns is that the availability and granularity of information has vastly increased, but at the same time it has become easier to avoid being informed (Prior 2007). Anyone can easily collect very detailed information on a narrow issue, while remaining uninformed about everything else. Because information remains costly to absorb and process, individuals can be very selective in the information that they acquire. When network television was the main source of political information, instead, individuals could not avoid being exposed to general news while searching for specific bits of information or seeking entertainment. As a result, political information was more uniform across individuals and issues.

In other words, the digital revolution had the following important implication. The patterns of information that bear on political process (who is informed and over what) are now largely determined by the individual demand for information, while the packaging of information by the media has become less important.

What effect does the possibility of selective ignorance have on political and policy outcomes? In particular, who is informed and over what, in a world in which information is easy to obtain but remains costly to absorb? And how do these informational patterns interact with and affect policy choices in a representative democracy? Could better information technology have adverse effects on the functioning of representative democracies, as many commentators suggest?

The goal of this paper is to address these questions. We study a general and unified theoretical framework where rationally inattentive voters allocate costly attention to political news, and politicians take this into account in setting policies. An important advantage of our framework is that voters' information is derived directly from first principles, i.e., from voters' preferences and their rational expectations of political outcomes. Thus, our results are applicable to a broad range of issues and do not require additional assumptions on voters' information when a new situation is studied.

Policy is set in the course of electoral competition by two candidates, who commit to policy platforms ahead of elections. As in standard probabilistic voting, voters trade off their policy preferences against their (random) preferences for one candidate or the other - see Persson and Tabellini (2000) and Lindbeck and Weibull (1987). The novelty is that here rational but uninformed voters also decide how to allocate costly attention. Voters cannot perfectly predict equilibrium policies, either because candidates make random implementation errors, or because candidates have private information over their type. Attention is modelled as the precision of the noisy signals that voters receive about the candidates' policies. More precise signals are more costly, and voters optimally choose

the precision of the signals they receive.

Voters' attention and public policies are jointly determined. Since attention is scarce, voters optimally allocate it to what is most important to them. Voters' priorities are not exogenously given, however, but depend on expected policy choices. In turn, voters' attention affects the incentives of political candidates, who design their policies taking into account who is informed about what. This interaction between candidates and optimally inattentive voters gives rise to systematic patterns of information acquisitions and deviations of equilibrium policies from the full information benchmark. These patterns are endogenous, and we study how they react if the granularity of available information increases (e.g., because of the diffusion of internet), if the cost of information drops, or if the economy is hit by shocks.

We first assume that candidates are opportunistic and maximize the probability of winning, and derive two general results. First, attention is not uniform, but differs across voters and policy issues. Voters are more attentive if they have higher stakes from observing a deviation from the expected equilibrium policy. Second, the equilibrium maximizes a modified *"perceived"* social welfare function that reflects voters' attention strategies. Thus, perceived welfare reacts to policy announcements in ways that differ across voters and policy issues. Where attention is higher, perceived welfare is more responsive to policy changes, and political candidates take this into account by catering more to the more attentive voters.

We then illustrate the general implications of these results with two examples. First, we study conflict over a single policy dimension. Here the focus is on which voters are more attentive and hence more influential. The main point is that rational inattention amplifies the effects of preference intensity and dampens the effects of group size on equilibrium outcomes, relative to full information. A group can have high policy stakes (and hence high attention) at the expected equilibrium policy for one of two reasons: because its preferences are very different from the rest of the population - it is an extremist group; or because it is small in size, so that political candidates can afford to neglect it. Thus, minorities and extremists tend to be more attentive and more influential in the political process, compared to full information. If the distribution of voters' policy preferences is not symmetric, this moves the equilibrium policy away from the full (or uniform) information benchmark.

The prediction that extremists and minorities are more informed and attentive is consistent with evidence from survey data. First, voters with more extreme partisan preferences are more informed about the policy positions of presidential candidates - Palfrey and Poole (1987). Second, they also consume more media (blogs, TV, radio and newspapers) - Ortoleva and Snowberg (2015). Third, ethnic minorities generally are more

informed about racial issues - Carpini and Keeter (1996).

Rational inattention also implies that the equilibrium can display policy divergence, even if candidates only care about winning the election and not about the policy per se, and they are equally popular.[1] Suppose that candidates differ in their informational attributes (e.g., one candidate has more media coverage and hence a lower cost of attention). Then the candidate with less media coverage caters to the relatively more attentive voters, namely those at one of the extremes, while his opponent chooses more centrist policies and is thus favored at the elections. An implication is that political entrants, who are likely to have less media coverage, tend to choose more extreme policies, and are less likely to win the election. This effect is weaker when policy stakes are particularly high, i.e., when a new important issue comes up or in unusual times such as in a crisis. Such times provide windows of opportunity for the less established candidates. The prediction that weaker candidates choose more extremist policies is consistent with the evidence from the US Congress in Fiorina (1973) and Ansolobehere et al. (2001).

We then consider a second example, where policy is multi-dimensional. We show that availability of fine-grained information can have perverse effects. Rational inattention implies that voters are more attentive to the policy dimensions over which they have higher stakes. These are typically the most controversial policies, because it is here that the political equilibrium cannot please everyone. On the issues on which everyone agrees, instead, voters expect an equilibrium policy close to their bliss point, and thus they have low stakes and low attention. Thus, attention to, say, spending on the justice system or on defense is predicted to be low. On the other hand, information about targeted transfers will be high, particularly amongst the potential beneficiaries of these policies. The reason is not only that these policies provide significant benefits to specific groups, but also that they are opposed by everyone else. This widespread opposition implies that in equilibrium these targeted policies are always insufficient from the perspective of the beneficiaries, who thus are very attentive to detect possible deviations on these instruments.

We illustrate this point in a model similar to Gavazza and Lizzeri (2009), and show that the equilibrium is Pareto inefficient: public goods that benefit all are under-provided, general tax distortions affecting everyone are too high, while there is excessive targeting to specific groups through tax credits or transfers. The final policy distortion is similar to that in Gavazza and Lizzeri (2009), but here informational asymmetries are endogenously determined in equilibrium, rather than assumed at the start. In addition, we can do comparative statics. For instance, we find that recessions reduce policy distortions, which

---

[1]Groselcose (2001) explains policy divergence as due to differences in valence, In our model valence can be captured by average popularity, which is assumed to be the same for the two candidates.

confirms the benefit of crisis for economic reforms (e.g., Kingdon 1984, OECD 2012).[2]

Finally, we show that our results extend to a setting where politicians are "citizen candidates" with policy preferences that are unknown to voters. This framework yields two additional insights. First, equilibrium policy divergence between the two candidates reflects the cost of attention by voters. A uniform drop in the cost of attention leads to more policy convergence, because voters are more responsive and candidates are less free to pursue their pereferred policies. But compositional effects also matter. If the cost of information drops only for the more extremist voters, then this has the opposite effect and policy divergence increases in equilibrium. This clarifies some of the mechanisms through which the new media technologies may lead to increased political polarization by elected representatives. Internet can lead to more polarization to the extent that it reduces the cost of information for voters with extreme partisan views, but not if it brings about a generalized and uniform improvement in political information. Second, in this setting policy uncertainty is endogenous and interacts with voters' attention. In particular, candidates who are expected to be more extreme receive more attention; this makes them respond less to their own preference shocks. That is, more extreme candidates are also more predictable (for a given variance of the preference shocks), which in turn partially attenuates voters' attention.

Our paper borrows analytical tools from the recent literature on rational inattention in other areas of economics, e.g., Sims (2003), Mackowiak and Wiederholt (2009), Van Nieuwerburgh and Veldkamp (2009), Woodford (2009), Matějka and McKay (2015), and Caplin and Dean (2015). This approach popularized and reinvented for economics the idea that attention is a scarce resource, and thus information can be imperfect even if it is freely available, such as on the internet or in financial journals.[3]

The notion that voters are very poorly informed is widespread in political economy (e.g., Carpini and Keeter 1996, Lupia and McCubbins 1998), yet the traditional approach views political information as a byproduct of other activities (Downs 1957). Trade policy is natural example, studied by Ponzetto (2011). In his model, workers acquire heterogeneous information about the positive effects of trade protection on their employment sector, and remain less informed about the cost of protection for their consumption. This asymmetry in information leads to a political bias against free trade. Thus information is endogenous but, unlike in our setting, it is not collected by citizens in order to cast a vote and this is reflected in the properties of the equilibrium. Moreover, such endogeneity

---

[2]Rahm Emmanuel (President Obama's first Chief of Staff): "Never want a serious crisis to go to waste." November 2008.

[3]Bordalo, Gennaioli and Shleifer (2013) provide an alternative theoretical framework to study how salience affects choices made by consumers with limited attention. Nunnari and Zapal (2017) introduce a closely related framework to a model of electoral competition.

requires a different model outside of electoral competition for each new issue studied.

A large literature has explored the political effects of information supplied by the media (see Stromberg 2001, Enikolopov et al. (2011), Gentzkow 2006, Gentzkow et al 2011, Perego and Yuksel 2016 and the surveys by Stromberg 2015, Prat and Stromberg 2013 and Della Vigna 2010). In terms of our theoretical framework, all these contributions endogenize the cost of acquiring political information, and their results are complementary to ours. One difference is that we look at how individuals process information, thus the source of the friction is different. A second important difference is that we look at voters' demand of information for purely political reasons. The media literature instead studies how the supply of information responds to demand, but information demand is a byproduct of other private activities, the utility of which depends on government policy. Thus, this literature concludes that large groups are more informed, because they are more relevant for profit maximizing media. We reach the opposite conclusion. Moreover, our approach allows us to study the effects of changes in the availability of information, when demand for political information responds endogenously to its cost or its degree of granularity.

Yuksel (2014) studies a model where policy is perfectly observed but voters seek costly information over the state of nature (a shock to voters' policy preferences). When policy is multidimensional, voters specialize in gathering information over the state of nature in the policy dimension that is most important to them, and remain less informed about the other states of nature. This makes them less responsive to the remaining policy dimensions, and partisan candidates are free to set policy closer to their bliss points, compared to a setting where the state of nature is perfectly observed by voters. Levy and Razin (2015) also study the implications of voters' cognitive limitations on equilibrium polarization, but their focus on correlation neglect is quite different from this paper.

Several papers study the effects of exogenously given imperfect information on policy outcomes. As already discussed, our second example is related to Gavazza and Lizzeri (2009), who study electoral competition when voters' information varies across policy instruments. The main difference is that they assume a given pattern of information, and their analysis relies on specific out of equilibrium beliefs. Our result on policy divergence due to differences in transparency between candidates is related to Glaeser et al (2005). That paper too assumes a specific pattern of exogenous information asymmetries, however.[4]

---

[4]In particular, they assume that core party supporters are more likely to observe a deviation from the expected equilibrium, compared to other voters, in a model with endogenous turnout. In our framework, instead, informational asymmetries are endogenous and everyone votes, and the specific predictions are quite different. As already mentioned, Groseclose (2001) also predicts policy divergence, but based on differences in valence between candidates. Finally, Alesina and Cukierman (1990) study how partisan candidates may have an incentive to hide their true ideological preferences.

A large theoretical literature studies voters' incentives to bear the cost of collecting information and /or voting, starting with the seminal contribution by Ledyard (1984). Most research on costly information focuses on the welfare properties of the equilibrium (Martinelli 2006) or on small committees (Persico 2003), however, and does not ask how voters' endogenous information shapes equilibrium policies. The literature on endogenous participation studies the equilibrium interaction of voting and policy design, but without an explicit focus on information acquisition.[5]

Finally, our paper is also related to a rapidly growing empirical literature on the economic and political effects of policy instruments with different degrees of visibility (see Congdon et al. 2011 for a general discussion of behavioral public finance). The findings in that literature confirm that policy instruments with different degrees of transparency are not politically equivalent, and directly or indirectly support the theoretical results of our paper.[6]

The outline of the paper is as follows. In section 2 we describe the general theoretical framework, where policy is ex-ante uncertain because of implementation errors by candidates. Section 3 presents some general results. Section 4 illustrates two applications to specific policy issues. Section 5 extends the framework to citizen candidates with random policy preferences. Section 6 concludes. The appendix contains the proofs and shows that our results generalize to a setting where voters' uncertainty reflects learning by the candidates about the state of nature, rather than implementation errors.

2   The general framework

This section presents a general model of electoral competition with rationally inattentive voters. Two opportunistic political candidates $C \in \{A, B\}$ maximize the probability of

---

[5]Regarding empirical evidence of limited and endogenous attention, Gabaix et al.(2006), and many others, explore endogenous attention allocation in a laboratory setting. Bartoš et al. (2016) explore attention to applicants in the field in rental and labor markets. They show that employers' and landlords' attention is endogenous to market conditions, it is selective, and it affects their decisions although the costs of attention (such as of reading applicant's CV) seem small.

[6]Chetty et al. (2009) show that consumer purchases reflect the visibility of indirect taxes. Finkelstein (2009) shows that demand is more elastic to toll increases when customers pay in cash rather than by means of a transponder, and toll increases are more likely to occur during election years in localities where transponders are more diffuse. Cabral and Hoxby (2012) compare the effects of two alternative methods of paying local property tax: directly by homeowners, vs indirectly by the lender servicing the mortgage, who then bills the homeowner through monthly automatic installments, combining all amounts due (for mortgage, insurance and taxes). Households paying indirectly are less likely to know the true tax rate (although they have no systematic bias). Moreover, in areas where indirect payment is (randomly) more prevalent, property tax rates are significantly higher. Bordignon et al. (2010) study the effects of a tax reform in Italy that allowed municipalities to partially replace a (highly visible) property tax with a (much less visible) surcharge added to the national income tax. Mayors in their first term switched to the less visible surcharge to a significantly greater extent than mayors who were reaching the limits of their terms. See also the earlier literature on fiscal illusion surveyed by Dollery and Worthington (1996).

winning the election and set a policy vector $q_C = [q_{C,1}, ..., q_{C,M}]$ of M elements. The elements may be targeted transfers to particular groups, tax rates, levels of public good, etc.

There are N distinct groups of voters, indexed by $J = 1, 2, ..., N$. Each group has a continuum of voters with a mass m^J, indexed by the superscript v. Voters' preferences have two additive components, as in standard probabilistic voting models (Persson and Tabellini, 2000). The first component $U^J(q_C)$ is a concave and differentiable function of the policy and is common to all voters in J. The second component is a preference shock x^v in favor of candidate B. Thus, the utility function of a voter of type $\{v, J\}$ from voting for candidate A or B is respectively:

$$U_A^{v,J}(q_A) = U^J(q_A), \qquad U_B^{v,J}(q_B) = U^J(q_B) + x^v. \tag{1}$$

The preference shock x^v in favor of candidate B is the sum of two random variables: $x^v = \tilde{x} + \tilde{x}^v$, where $\tilde{x}^v$ is a voter specific preference shock, while $\tilde{x}$ is a shock common to all voters. We assume that $\tilde{x}^v$ is uniformly distributed on $[-\frac{1}{2\phi}, \frac{1}{2\phi}]$, i.e., it has mean zero and density ϕ and is iid across voters. The common shock $\tilde{x}$ is distributed uniformly in $[-\frac{1}{2\psi}, \frac{1}{2\psi}]$. In what follows we refer to $\tilde{x}^v$ as an idiosyncratic preference shock and to $\tilde{x}$ as a popularity shock.

The distinguishing feature of the model is that voters are uninformed about the candidates' policies, but they can choose how much of costly attention to devote to these policies and their elements. To generate some voters' ex-ante uncertainty, we assume that candidates target a policy of their choice (which in equilibrium can be perfectly predicted by voters), but the policy platform actually set by each candidate is drawn by nature from the neighborhood of the targeted policy. Specifically, each candidate commits to a *target* policy platform $\hat{q}_C = [\hat{q}_{C,1}, ..., \hat{q}_{C,M}]$. The *actual* policy platform on which candidate C runs, however, is

$$q_{C,i} = \hat{q}_{C,i} + e_{C,i} \tag{2}$$

where $e_{C,i} \sim N(0, \sigma_{C,i}^2)$ is a random variable that reflects implementation errors in the course of the campaign. For instance, the candidate announces a specific target tax rate on real estate, $\hat{q}_{C,i}$, but when all details are spelled out and implemented during the electoral campaign, the actual tax rate to which each candidate commits may contain additional provisions such as homestead exemptions, or for assessment of market value. The implementation errors $e_{C,i}$ are independent across candidates C and policy instruments i, and their variance $\sigma_{C,i}^2$ is given exogenously.[7]

---

[7]The assumption of independence could easily be dropped, and then e_C would be multivariate normal with a variance-covariance matrix Σ - see below.

The sequence of events is as follows.

1. Voters form prior beliefs about the policy platforms of each candidate and choose attention strategies.

2. Candidates set policy (i.e. they choose *target* platforms, and *actual* policy platforms are determined as in (2)).

3. Voters observe noisy signals of the actual platforms.

4. The ideological bias x^v is realized and elections are held. Whoever wins the election enacts their announced actual policies.

In Section 2.2 we define the equilibrium, which is a pair of targeted policy vectors chosen by the candidates, and a set of attention strategies chosen by each voter. The attention strategies are optimal for each voter, given their prior beliefs about policies, and policy vectors maximize the probability of winning for each candidate, given the voters' attention strategies. Moreover, voters' prior beliefs are consistent with the candidates' policy targets.

2.1 Voters' behavior

The voters' decision process has two stages: information acquisition and voting.

2.1.1 Imperfect information and attention

All voters have identical prior beliefs about the policy vectors q_C of the two candidates. In the beliefs, elements of the policy vector are independent, and so are the policy vectors of the two candidates. Let each element of the vector of prior beliefs be drawn from $N(\bar{q}_{C,i}, \sigma^2_{C,i})$, where $\bar{q}_C = [\bar{q}_{C,1}, ..., \bar{q}_{C,M}]$ is the vector of prior means, and $\sigma^2_C = [\sigma^2_{C,1}, ..., \sigma^2_{C,M}]$ the vector of prior variances. Note that, to insure consistency, the prior variances coincide with the variance of the implementation errors e_C in (2).[8]

In the first stage voters choose attention, that is they choose how much information about each element of each policy vector to acquire. We model this as the choice of the level of noise in signals that the voters receive. Each voter (v, J) receives a vector $s^{v,J}$ of independent signals on all the elements $\{1, ..., M\}$ of both candidates, A and B,

$$s^{v,J}_{C,i} = q_{C,i} + \epsilon^{v,J}_{C,i},$$

---

[8]Like for the implementation errors, the assumption of independence could easily be dropped, and then $\tilde{q}_C$ would be multivariate normal with a variance-covariance matrix $\bar{\Sigma}$.

where the noise $\epsilon_{C,i}^{v,J}$ is drawn from a normal distribution $N(0, \gamma_{C,i}^J)$, and is iid across voters.[9]

It is convenient to define the following vector $\xi^J \in [0,1]^{2M}$, which is the decision variable for attention in our model: $\xi^J = \left\{ [\xi_{A,1}^J ... \xi_{A,M}^J], [\xi_{B,1}^J ... \xi_{B,M}^J] \right\}$, where

$$\xi_{C,i}^J = \frac{\sigma_{C,i}^2}{\sigma_{C,i}^2 + \gamma_{C,i}^J} \in [0,1].$$

The more attention is paid by the voter to $q_{C,i}$, the closer is $\xi_{C,i}^J$ to 1. This is reflected by the noise level $\gamma_{C,i}^J$ being closer to zero, and also by a smaller variance $\rho_{C,i}^J$ of posterior beliefs.[10] Naturally, higher attention is more costly; see below. We also allow for some given level $\xi_0 \in [0,1)$ of minimal attention paid to each instrument, which is forced upon the voter exogenously, i.e., the choice variables must satisfy $\xi_{C,i}^J \geq \xi_0$.

Higher levels of precision of signals are more costly. Here we employ the standard cost function in rational inattention (Sims, 2003), but this choice is not crucial. We assume that the cost of attention is proportional to the relative reduction of uncertainty upon observing the signal, measured by entropy. For uni-variate normal distributions of variance σ^2, entropy is proportional to $log(\pi e \sigma^2)$. Thus, the reduction in uncertainty that results from conditioning on a normally distributed signal s is given by $\log(\pi e \sigma^2) - \log(\pi e \rho)$, where σ^2 is the prior variance and ρ denotes the posterior variance. Since in a multivariate case of independent uncorrelated elements, the total entropy equals the sum of entropies of single elements, the cost of information in our model is:

$$\sum_{C \in \{A,B\}, i \leq M} \lambda_{C,i}^J \log\left(\sigma_{C,i}^2 / \rho_{C,i}^J\right) = - \sum_{C \in \{A,B\}, i \leq M} \lambda_{C,i}^J log\left(1 - \xi_{C,i}^J\right).$$

The term $-log(1 - \xi_{C,i}^J)$ measures the relative reduction of uncertainty about the policy element $q_{C,i}$, and it is increasing and convex in the level of attention $\xi_{C,i}$. The parameter $\lambda_{C,i}^J \in \mathbb{R}_+$ scales the unit cost of information of voter J about $q_{C,i}$. It can reflect the supply of information from the media or other sources, the transparency of the policy instrument $q_{C,i}$, or the ability of voter J to process information.

---

[9]All voters belonging to the same group choose the same attention strategies, since ex-ante (i.e., before the realization of x^v and $\epsilon_{C,i}^{v,J}$) they are identical.

[10]The posterior variance equals $\rho_{C,i}^J = \gamma_{C,i}^J \sigma_{C,i}^2 / (\sigma_{C,i}^2 + \gamma_{C,i}^J)$. Thus, the variable $\xi_{C,i}^J$ also measures the relative reduction of uncertainty about $q_{C,i}$; $\xi_{C,i}^J = 1 - \frac{\rho_{C,i}^J}{\sigma_{C,i}^2}$. The more attention is paid, the closer is $\xi_{C,i}^J$ to 1 and hence the lower is the posterior variance.

2.1.2   Voting

The second stage is a standard voting decision under uncertainty. After voters receive additional information of the selected form, and knowing the realization of the candidate bias x^v, they choose which candidate to vote for. Specifically, after a voter receives signals $s^{v,J}$, he forms posterior beliefs about utilities from policies that will be implemented by each candidate, and he votes for A if and only if:

$$E[U^J(q_A)|s_A^{v,J}] - E[U^J(q_B)|s_B^{v,J}] \geq x^v. \tag{3}$$

where the expectations operator refers to the posterior beliefs about the unobserved policy vectors q_C, conditional on the signals received.

2.1.3   Voter's objective

In the first stage the voter chooses an attention strategy to maximize expected utility in the second stage, considering what posterior beliefs and preference shocks can be realized, less the cost of information. Thus, voters in each group J choose attention strategy ξ^J that solves the following maximization problem:

$$\max_{\xi^J \in [\xi_0, 1]^{2M}} \quad E\left[max_{C \in \{A,B\}} E[U_C^{v,J}(q_C)|s_C^{v,J}]\right] + \sum_{C \in \{A,B\}, i \leq M} \lambda_{C,i}^J log\left(1 - \xi_{C,i}^J\right). \tag{4}$$

The first term is the expected utility from the selected candidate (inclusive of the candidate bias x^v), i.e., it is the maximal expected utility from either candidate conditional on the received signals. The inner expectation is over a realized posterior belief. The outer expectation is determined by prior beliefs; it is over realizations of $\epsilon_C^{v,J}$ and x^v. The second term is minus the cost of information.

2.2   Equilibrium

In equilibrium, neither candidates nor voters have an incentive to deviate from their strategies. In particular, voters' prior beliefs are consistent with the equilibrium choice of *targeted* policy vectors of the candidates, and candidates select a best response to the attention strategies of voters and to each other's policies. Specifically:

Definition 1 *Given the level of noise σ_C^2 in candidates' policies, the equilibrium is a set of targeted policy vectors chosen by each candidate, $\hat{q}_A, \hat{q}_B$, and of attention strategies ξ^J chosen by each group of voters, such that:*

(a) *The attention strategies ξ^J solve the voters' problem (4) for prior beliefs with means $\bar{q}_C = \hat{q}_C$ and noise σ_C^2.*

(b) *The targeted policy vector $\hat{q}_C$ maximizes the probability of winning for each candidate C, taking as given the attention strategies chosen by the voters and the policy platforms chosen by his opponent.*

2.3 Discussion

Here we briefly discuss some of the previous modeling assumptions. Most of our findings are robust to slight variations in these assumptions, however, since the results that follow are based on intuitive monotonicity arguments only.

Noise in prior beliefs. There are two primitive random variables in this set up: the campaign implementation errors $e_{C,i} \sim N(0, \sigma_{C,i}^2)$, which have an exogenously given distribution reflecting the process governing each electoral campaign. And the noise in the policy signals observed by the voters, $\epsilon_{C,i}^{v,J} \sim N(0, \gamma_{C,i}^J)$, whose variance $\gamma_{C,i}^J$ corresponds to the chosen level of attention, $\xi_{C,i}^J$. The distribution of voters' prior beliefs then reflects the distribution of the implementation errors, $e_{C,i}$.

The assumption that candidates make random mistakes or imprecisions in announcing the policies is used to generate uncertainty in prior beliefs. This assumption follows the well known notion of a trembling hand from game theory (Selten 1975, McKelvey and Palfrey 1995). There needs to be a source of uncertainty in the model, otherwise limited attention would play no role, but there are other ways of introducing it. Section 5 considers a model with no implementation errors, but where candidates have policy preferences that are unknown to voters. The main difference is that candidates maximize expected utility, rather then the probability of winning. This yields additional implications, but the main insights of section 3 extend to that environment. The online appendix considers yet another setting, where opportunistic candidates maximize the probability of winning and there are no implementation errors. Policy is ex-ante uncertain because candidates observe private signals of the environment. In particular, voters' policy preferences take the form: $U^J(q - \eta)$, where η is a random variable. Candidates observe a noisy private signal of η and set policy. Voters observe the realization of η and set attention strategies. They then observe noisy signal of the actual policy platforms and vote. The interpretation is that candidates have to commit to non-state-contingent platforms before the state of nature is fully revealed to voters, and candidates have different views about the state of nature (or equivalently about the welfare consequences of alternative policies). The results discussed in the next section generalize almost identically to this setting, with one difference. If policy uncertainty is due implementation errors, as in the baseline model, then it is entirely exogenous. If instead policy uncertainty reflects shocks to the preferences of candidates (as in section 5) or of voters (as in the appendix), then policy

volatility is endogenous: it is determined in equilibrium by how candidates react to such preference shocks. This yields additional implications that are discussed below.

Another feature of prior beliefs that is worth discussing is the assumed independence of all shocks across policy instruments. We make this assumption for the sake of simplicity. If we allowed for correlated shocks across policy instruments, the main implications of our model would not change in a fundamental way, but expressions for Bayesian updating would become more complicated, and thus also some analytical results in Section 3 would be less elegant. Similarly, we could also extend beyond the iid noise in signals and, for instance, model the effect of media, which generates correlated noise in information for many voters. We leave this for future research.

The introduction of a minimal level of attention $\xi_0 > 0$ is useful to simplify the discussion of the example in Section 4.2. If $\xi_0 = 0$, voters would pay no attention at all to some policy instruments within some range of their level, and there would be multiple equilibria with similar properties. Any positive ξ_0 pins down the solution uniquely. The minimal level of attention $\xi_0 > 0$ could be derived (with more complicated notation) from the plausible assumption that all voters receive a costless signal about policy (such as when they turn on the radio or open their internet browser).

Voters' objectives. Why do individuals bother to vote and pay costly attention? With a continuum of voters, the probability of being pivotal is zero, and selfish voters should not be willing to pay any positive cost of information or of voting. Even with a finite number of voters, in a large election the probability of being pivotal is so small that it cannot be taken as a the main motivation for voting or paying costly attention. This is the same issue faced by many papers in the field of political economy, and we do not aspire to solve it.

While a large literature has sought to explain why individuals vote in large elections, the question of why individuals bother to acquire political information in large elections has been neglected. The standard approach views political information as a byproduct of other activities (Downs 1957). While there is no doubt that political information is also acquired in this indirect way, it is also obvious that political information is sought purposefully from the media or from political sources.

Our formulation of the voters' objective, (4), literally states that the voter chooses how much and what form of information to acquire *as if* he were pivotal in his subsequent voting decision. This can be interpreted as saying that voters are motivated by "sincere attention" and want to cast a meaningful vote. That is, they draw utility from voting for the right candidate (i.e., the one that is associated with his highest expected utility), because they consider it their duty (cf. Feddersen and Sandroni 2006) or because they

want to tell others (as in Della Vigna et al. 2015). In other words, individuals are motivated to acquire political information by exactly the same considerations that induce them to vote one way or the other in the ballot. In the absence of a complete and satisfactrory theory of voter behavior, this seems the most natural and least arbitray assumption. In this interpretation, the parameter $\lambda^J_{C,i}$ captures the cost of attention relative to the psychological benefit of voting for the right candidate. [11]

An alternative and more ambitious formulation would have been to derive both turnout and information acquisition from a group-utilitarian model of voters' behavior, adapting the approach of Coate and Conlin (2004) to our setting. In this alternative formulation, the demand of political information would continue to reflect the importance of the policy stakes for the group, as in our setting, but the closeness of the election could also matter (in Coate and Conlin (2004), a closer election induces more people to vote). This aspect of the demand of political information is missing from our model.

In line with our assumption that voters are motivated by the desire of casting a meaningful vote and not by the expectation of being pivotal, we also assume that voters do not condition their beliefs on being pivotal when they vote. This is the standard approach in the literature on electoral competition, and it is consistent with the fact that with a continuum of voters the probability of being pivotal is zero.[12]

The cost of information need not be entropy-based. We just use this form since it is standard in the literature. However, almost any function that is globally convex, and increasing in elements of ξ^J, would generate qualitatively the same results; see a note under Proposition 2 below.[13] There would exists a unique solution to the voter's attention problem, and attention would be increasing in both stakes and uncertainty. Also, the cost of information should be interpreted as net of its entertainment value, which can explain why petty news often receive more attention than important issues.[14]

Finally, the assumption that voters care about both policies and candidates, as in probabilistic voting models, is made to insure existence of the equilibrium when the policy space is multidimensional. The preferences for candidates could reflect their personal

---

[11]An alternative interpretation is that voters expect to be pivotal with an exogenously given probability, say $\delta > 0$. Then the first term in (4), the expected utility from the selected policy, would be pre-multiplied by δ. Such a modification would be equivalent to rescaling the cost of information λ by the factor $1/\delta$, with no substantive change in any result. If the probability of being pivotal was endogenous and part of the equilibrium, the model would become more complicated, but most qualitative implications discussed below would again remain unchanged. The first order condition (8) below would still hold exactly.

[12]If we allowed for learning from being pivotal, then under some assumptions voters could learn the policy exactly, and limited attention would have no effect. This assumption is more restrictive in asymmetric equilibria, where a lot of information may be revealed by discovering that the race is close.

[13]"Almost any" here denotes functions with sufficient regularity and symmetry across its arguments.

[14]In surveys run by Carpini and Keeter (1996), during the Bush vs Clinton campaign, 85% of respondents knew that the dog of President Bush was named Millie, while only 15% knew that both candidates supported the death penalty. Deriving the entertainment value of information from primitive assumption of voters' behavior goes beyond the scope of this paper.

attributes, or non-pliable policy issues that will be chosen after the election on the basis of candidates' ideological beliefs or partisan preferences. The specific timing, that the idiosyncratic preference shock $\tilde{x}^v$ is realized only at the voting stage, implies that the attention strategies of voters are the same within each group. This assumption could be relaxed at the price of notational complexity. Since these candidate features are fixed and do not interact with their pre-electoral policy choices, we neglect the issue of how much attention is devoted to the candidates (as distinct from their policies).

3   Preliminary results

In this section we first describe how the equilibrium policy is influenced by voters' attention, and then we describe the equilibrium attention strategies. The equilibrium policy solves a specific modified social welfare function which can be compared with that of standard probabilistic voting models. If noise in candidates' policies and thus in voters' prior uncertainty is small, the equilibrium can be approximated by a convenient first order condition. This result is useful when discussing particular examples and applications of the general model.

3.1   A "perceived" social welfare function

To characterize the equilibrium, we need to express the probability of winning the election as a function of the candidate's announced policies. In this, we follow the standard approach in probabilistic voting models (Persson and Tabellini, 2000).

Let p_C be the probability that C wins the elections. Suppose first that the cost of information is 0, $\lambda_{C,i}^J = 0$. Then our model boils down to standard probabilistic voting with full information. The distributional assumptions and the additivity of the preference shocks $x^v = \tilde{x} + \tilde{x}^v$ then imply:

$$p_A = \frac{1}{2} + \psi \left(\sum_J m^J \left[U^J(q_A) - U^J(q_B) \right] \right). \tag{5}$$

The probability that C wins is increasing in the social welfare $\sum_J m^J U^J(q_C)$ that C provides.[15]

In our model, however, voters do not base their voting decisions on the true utilities they derive from policies, but on expected utilities only. Appendix 7.1 shows that with

---

[15]This holds when the support of the popularity shock $\tilde{x}$ is sufficiently large.

inattentive voters and $\lambda_{C,i}^J > 0$, the probability that candidate A wins is:

$$p_A = \frac{1}{2} + \psi \left(\sum_J m^J E_{\epsilon,q_A,q_B}^J \left[E[U^J(q_A)|s_A^{v,J}] - E[U^J(q_B)|s_B^{v,J}] \right] \right) \qquad (6)$$

where the outer expectations operator is indexed by J because voters' attention differ across groups. Obviously, $p_B = 1 - p_A$. For a particular realization of policies, in our model the probability of winning is analogous to (5), except that the voting decision is not based on $U^J(q_C)$, but on $E[U^J(q_A)|s_A^{v,J}]$.[16] The overall probability of winning is then an expectation of this quantity over all realizations of policies and of noise in signals.

Given an attention strategy, candidate A cannot affect $E[U^J(q_B)|s_B^{v,J}]$, and vice versa for candidate B. Thus we have:

Lemma 1 *In equilibrium, each candidate C solves the following maximization problem.*

$$\max_{\hat{q}_C \in \mathbb{R}^M} \quad \sum_J m^J E_{\epsilon,e}^J \left[E[U^J(q_C)|s_C^{v,J}] \Big| \hat{q}_C \right] \qquad (7)$$

In equilibrium, candidate C maximizes the "perceived social welfare" provided by his policies. It is the weighted average of utilities from policy q_C expected by voters in each group (weighted by the mass of voters, and pdf of realizations of errors e in announced policies and observation noise ϵ). Under perfect information this quantity equals the social welfare provided by q_C. Here instead different groups will generally select different attention strategies, resulting in perceptions of welfare that also differ between groups or across policy issues.

Lemma 1 thus reveals the main difference between this framework and standard probabilistic voting models. For instance, if some voters pay more attention to some policy deviations, then their expected utilities vary more with such policy changes compared to other voters. Therefore, perceived welfare can systematically differ from actual welfare, and rational inattention can lead politicians to select distorted policies.[17]

Finally, note that the candidates' objective (7) is a concave function of the realized policy vector q_C.[18] Thus, the equilibrium can be characterized by the first order conditions

---

[16]Again, this holds if the support of the popularity shock $\tilde{x}$ is sufficiently large relative to the RHS of (6).

[17]This can happen even if all groups are equally influential in the sense of having the same distribution of ideological preference shocks $\tilde{x}^v$.

[18]This is because: i) For Gaussian beliefs and signals, posterior means depend linearly on the target policy $\hat{q}_C$ set by each candidate, and their variance as well as variances of posterior beliefs are independent of $\hat{q}_C$. Variance of posterior belief can be expressed in terms of prior variance and the attention vector: $\rho_{J,i} = (1 - \xi_i^J)\sigma_i^2$. Upon acquisition of a signal $s_{C,i}^{v,J}$, the posterior mean is: $\check{q}_{C,i} = \xi_{C,i}^J s_{C,i}^{v,J} + (1 - \xi_{C,i}^J)\bar{q}_{C,i}$, where $s_{C,i}^{v,J} = q_{C,i} + \epsilon_{C,i}^{v,J}$ and $\bar{q}_{C,i}$ denotes the prior mean. Thus, $\check{q}_{C,i} = \xi_{C,i}^J(\hat{q}_{C,i} + e_{C,i} + \epsilon_{C,i}^{v,J}) + (1 - \xi_{C,i}^J)\bar{q}_{C,i}$. ii) For a given vector of posterior variances, the term $E[U^J(q_C)|s_C^{v,J}]$ is a concave function of the vector of posterior means of the belief about the policy vector q_C.

of the objective (7), since they are necessary and sufficient for an optimum.

3.2   Small noise approximations or quadratic utility

In this subsection we introduce an approach that can be used to determine the exact form of the equilibrium. This can be done if utility function is quadratic or if prior uncertainty in beliefs is small, and we can use a local approximation to the utility function. The distinctive feature of our model is that it studies implications of imperfect information for outcomes of electoral competition. Thus, these approximations emphasize the first-order effects of such information imperfection. As shown here, these effects can be highly relevant even if information imperfections are small.

Let us denote by

$$u_{C,i}^J = \left(\frac{\partial U^J(q_{C,i})}{\partial q_{C,i}} \right) \Big|_{q_C = \bar{q}_C}$$

the marginal utility for a voter in group J of a change in the ith component of the policy vector, evaluated at the expected policies. Thus, $u_{C,i}^J$ measures intensity of preferences about $q_{C,i}$ in a neighborhood of the equilibrium. Suppose that the noise σ_C^2 is small. Then Appendix 7.2 proves:

Proposition 1 *The equilibrium policies satisfy the following first order conditions:*

$$\sum_{J=1}^N m^J \xi_{C,i}^J u_{C,i}^J = 0 \qquad \forall i, \tag{8}$$

where $\xi_{C,i}^J$ are the equilibrium attention weights.

The proof in fact shows that (8) holds for both first and second order approximations of U, and thus it also holds exactly for quadratic utility functions, which we use in the example in Section 4.1.

This proposition emphasizes the main forces in electoral competition with inattentive voters. For a policy change to have an effect on voting, it needs to be paid attention to and observed. If $q_{C,i}$ changes by an infinitesimal Δ, then expected posterior mean in group J about $q_{C,i}$ changes by $\xi_{C,i}^J \Delta$ only. Thus, while the effect on voters' utility is $\Delta u_{C,i}^J$, the effect on expected, i.e., perceived, utility is only $\xi_{C,i}^J \Delta u_{C,i}^J$.

Several remarks are in order. First, with only one policy instrument, equation (8) is the first order condition for the maximum of a modified social planner's problem, where each group J is weighted by its attention, $\xi_{C,i}^J$. Thus, if all voters paid the same attention, so that $\xi_{C,i}^J = \xi$ for all J, C, i, then the equilibrium coincides with the utilitarian optimum. If some groups pay more attention, however, then they are assigned a greater weight by

both candidates. That is, more attentive voters are more influential, because they are more responsive to any policy change.

Second, if policy is multi-dimensional, the attention weights $\xi_{C,i}^J$ in (8) generally vary by policy instrument i. If they do, then equation (8) does not correspond to the first order condition for the maximum of a modified social planner problem, and hence the equilibrium is not constrained Pareto efficient. The public good example in subsection 4.2 below illustrates this point.

Third, these results hold for any attention weights, and not just for those that are optimal from the voters' perspectives. In other words, Proposition 1 characterizes equilibrium policy with imperfectly attentive voters, irrespective of how voters' attention is determined.

Let us now focus on the voter's problem. How should costly attention be allocated to alternative components of the policy vector? We start with a first order approximation of U in the voters' optimization problem stated in (4). Thus, suppose again that the noise in prior beliefs σ_C^2 is small.[19] Then Appendix proves:

Lemma 2 *The voter chooses the attention vector* $\xi^J \in [\xi_0, 1]^M$ *that maximizes the following objective.*

$$\left(\sum_{C \in \{A,B\}, i=1}^{M} \xi_{C,i}^J (u_{C,i}^J)^2 \sigma_{C,i}^2 \right) + \sum_{C \in \{A,B\}, i \leq M} \hat{\lambda}_{C,i}^J log \left(1 - \xi_{C,i}^J \right), \tag{9}$$

where $\hat{\lambda}_{C,i}^J = 2\lambda_{C,i}^J / Min(\psi, \phi)$.

The form of (9) for second order approximations is presented in (39) in the Appendix.

The benefit of information for voters reflects the expected difference in utilities from the two candidates. If both candidates provide the same expected utility, then there is no gain from information. Specifically, the term $\sum_{C \in \{A,B\}, i=1}^{M} \xi_{C,i}^J (u_{C,i}^J)^2 \sigma_{C,i}^2$ is the variance of the difference in expected utilities under each of the two candidates, conditional on posterior beliefs. The larger is the discovered difference in utilities, the larger is the gain, since then the voter can choose the candidate that provides higher utility.

Note also that $\xi_{C,i}^J \sigma_{C,i}^2 = (\sigma_{C,i}^2 - \rho_{C,i})$ measures the reduction of uncertainty between prior and posterior beliefs. Thus, net of the cost of attention, the voter maximizes a weighted average of the reduction in uncertainty, where the weights correspond to the (squared) marginal utilities from deviations in $q_{C,i}$. That is, the voter aims to achieve a greater reduction in uncertainty where the instrument-specific stakes are higher.

---

[19]Again, analogously to probabilistic voting, we also assume that the support of the preference shock is large relatively to the difference in expected utilities from the two candidates.

An immediate implication of (9) is the next proposition.[20]

Proposition 2 *The solution to the voter's attention allocation problem is:*

$$\xi_{C,i}^J = \max\left(\xi_0, 1 - \frac{\hat{\lambda}_{C,i}^J}{(u_{C,i}^J)^2 \sigma_{C,i}^2}\right). \tag{10}$$

Quite intuitively, the solution (10) implies that, for a given cost of information $\hat{\lambda}^J$, the voter pays more attention to those elements $q_{C,i}$ for which the unit cost of information $\lambda_{C,i}^J$ is lower, i.e. are more transparent, prior uncertainty $\sigma_{C,i}^2$ is higher, and which have higher utility-stakes $|u_{C,i}^J|$ from changes in $q_{C,i}$. Note that for any convex information-cost function $\Gamma(\xi^J)$, the objective (9) would be concave, and thus there would exist a unique maximum, which would solve $\partial\Gamma(\xi^J)/\partial\xi_{C,i}^J = Min(\psi, \phi)(u_{C,i}^J)^2\sigma_{C,i}^2/2$. The effect of stakes and uncertainty also holds more generally.[21]

Putting implications of (8) and (10) together, we infer that in our model voters with higher stakes have relatively more impact on equilibrium policies than under perfect information. To summarize, voter's higher stakes imply higher attention, which in turn implies stronger voting response to a policy change. Therefore, candidates have stronger incentives to appeal to these high-stake voters than if all voters were equally attentive. These results are very intuitive, and since they are mostly based on monotonicity, we believe that they are robust to slight changes of its assumptions.

Finally, the attention weights $\xi_{C,i}^J$ also depend on the identity of the candidate, because the cost of information or prior uncertainty $\sigma_{C,i}^2$, could differ between the two candidates. If so, the two candidates in equilibrium end up choosing different policy vectors. Thus, rational inattention can lead to policy divergence if candidates differ in their informational attributes, even though both candidates only care about winning the elections. This contrasts with other existing models of electoral competition, that lead to policy divergence in pure strategies only if candidates have policy preferences themselves (see Persson and Tabellini 2000). Subsection 4.1 below illustrates this result with an example.

For lack of better alternatives, we have assumed that voters' beliefs are rational (i.e. their priors described by the mean $\bar{q}_C$ and variance σ_C^2 are consistent with equilibrium policy choices). What are the implications of systematic biases in beliefs? Specifically, what happens if the mean $\bar{q}_C$ is viewed to be different from the true equilibrium q_C, or perceived policy uncertainty σ_C^2 is higher? The effect of σ_C^2 on beliefs is clear: if q_C is

---

[20]The solution for second order approximation is in (40).

[21]For instance, the effects hold for any cost function that is symmetric across policy elements, i.e., invariant to permutations in ξ^J.

believed to be more volatile, then it has the same effect as if λ_C were lower - voters pay more attention to such a candidate, and the equilibrium policies of C move closer to the perfect information case. The opposite is implied by lower σ_C^2. The effect of $\bar{q}_C$ is to move the equilibrium policy in the opposite direction than the error in beliefs. If C is believed to be further from the bliss point of voter J, then the stakes $|u_C^J|$ increase and candidate C is paid more attention by J, which pushes equilibrium policy closer to J (assuming that candidates are aware of this distortion in beliefs). Therefore, voters who incorrectly expect policy to be unfavorable for them will get a better policy, and viceversa.

The appendix also solves a second order (rather than first order) approximation of the voters' optimization problem, which is of course exact for quadratic utilities. In this case, the optimal attention ξ^J is given by (40), only a slightly more complicated formula than in (10), and its qualitative properties remain almost the same. The difference is that if voters are not risk-neutral, then they acquire information not just to make a better choice of which candidate to vote for, but also to decrease uncertainty conditional on a chosen candidate. This also implies that more risk-averse voters are relatively more influential than under perfect information. The voters' optimality condition then contains an additional term, which implies that voters' attention is higher than stated in (10). This additional term is larger the greater is prior uncertainty and risk aversion.

4   Applications

In this section we present two examples to illustrate some basic implications of inattentive voters. Besides explaining what voters know and don't know and predicting specific policy distortions relative to the full information equilibrium, rational inattention also sheds light on several other issues. In particular, these examples illustrate why an increase in the granularity of information can be welfare deteriorating, why new and lesser known candidates often cater to minorities or political extremists, the role of parties as labels that reduce information costs, the benefit of crisis for economic reforms.

We start with electoral competition on a one-dimensional policy. Here the focus is on how different voters allocate attention to the same policy issue, with resulting differences in political influence. Then we turn to multi-dimensional policies, in a symmetric model. Here the focus is on how voters allocate attention to different policy issues and the resulting policy distortions.

4.1 One dimensional conflict

This example explores the effects of rational inattention on equilibrium policy outcomes in a simple setting. We study how electoral competition resolves heterogeneity in preferences regarding a single policy dimension. Rational inattention amplifies the effects of preference intensity and dampens the effects of group size. The reason is that voters with higher stakes pay more attention and hence are more influential. Who has the higher stakes is endogenous, however, since it depends on expected policy platforms. This leads to equilibrium policies that favor smaller and more extremist groups, relative to full information. Moreover, as the cost of information drops, attention becomes more uniform amongst voters, and extremist voters become less influential.

Let voters differ in their preferences for a one dimensional policy q. Voters in group J have a bliss-point t^J and their marginal cost of information is $\hat{\lambda}^J$, for now assumed to be the same for all candidates C. The voters' utility function is

$$U^J(q) = U(q - t^J),$$

$q \in \mathbb{R}$ and $U(.)$ is concave and symmetric about its maximum at 0.[22] With a one dimensional policy, by Proposition 1 the equilibrium with rational inattention can be computed as the solution to a modified social planning problem, where each candidate C maximizes $\sum_J m^J \xi_C^J U^J(q_C)$.

Who is more attentive and influential? By (10), voters' attention increases with the distance $|\hat{q}^* - t^J|$, where $\hat{q}^*$ denotes the equilibrium policy target. The reason is that the utility stakes, $|u^J(q_C)|$, increase in this distance, due to concavity of U^J. The distance $|\hat{q}^* - t^J|$, in turn, reflects two features of a group: its size m^J and the location of its bliss point t^J in the overall distribution of voters' preference. Clearly groups with extreme preferences tend to have high stakes, since the equilibrium policy is generally far away from their bliss point. Smaller groups also have higher stakes, because the equilibrium policy treats them less favorably than larger groups. Hence, if the cost of collecting information $\hat{\lambda}^J$ is the same for all groups of voters, then groups with extreme policy preferences and of small size pay more attention to q_C and are politically more influential (i.e. they receive a higher weight ξ_C^J in the modified planner's problem). Groups with a lower cost $\hat{\lambda}^J$ also receive a greater weight, for the same reason.

The effect of group size on attention can easily be illustrated if there are only two

---

[22]Political disagreement is often one-dimensional, as policy preferences tend to be aligned along left-to-right ideological positions (see Poole and Rosenthal 1997).

groups, $J = 1, 2$, with $m^1 \neq m^2$ and $t^1 \neq t^2$. Proposition 1 implies that in equilibrium:

$$\frac{\xi_C^1 u^1(q_C)}{\xi_C^2 u^2(q_C)} = -\frac{m^2}{m^1}. \tag{11}$$

Inserting (10) in (11), it can be verified that in equilibrium the smaller group has larger stakes and hence is more attentive: $m^J < m^K$ implies that $|u^J| > |u^K|$ and hence that $\xi_C^J > \xi_C^K$.

More generally, rational inattention amplifies the effect of preference intensity (i.e. the intensive margin) and dampens the effect of group size (the extensive margin) on the equilibrium policy. Consider a group with a bliss point above the equilibrium policy target: $t^J > \hat{q}^*$. If t^J increases further, then both the policy stakes u^J and attention ξ_C^J increase, and thus the overall effect of higher stakes is super-proportional. On the other hand, the effect of an increase in group size is less than proportional. If the mass of voters m^J increases, then for given attention the weight of group J increases proportionately. However, (11) implies that larger groups pay less attention (ξ_C^J drops as m^J rises), with a partially offsetting effect on the equilibrium policy.

This implication of rational inattention, that smaller groups are more informed and hence more influential compared to full information, contrasts with the opposite result in the literature on the political effects of the media. Profit maximizing media typically target larger groups, who are thus predicted to be better informed and more influential (Stromberg 2001, Prat and Stromberg 2013). If one interprets the cost $\hat{\lambda}^J$ as influenced by the media, then the media literature predicts that larger groups have smaller $\hat{\lambda}^J$, while rational inattention predicts that smaller groups have higher stakes u^J. Which effects prevails on attention ξ_C^J is a priori ambiguous. Nevertheless, the evidence in Carpini and Keeter (1996) quoted in the introduction suggests that minorities are generally more informed about the issues that are relevant to them, compared to the rest of the population.

The prediction that extremist voters pay more attention is also in line with results from two previous empirical studies. Using the survey data of U.S. presidential elections held in 1980, Palfrey and Poole (1987) find that voters who are highly informed about the candidate policy location tend to be significantly more polarized in their ideological views compared to uninformed voters. Using data from the 2010 Cooperative Congressional Election Survey and the American National Election Survey, Ortoleva and Snowberg (2015) find that voters with more extreme policy preferences consume more media such as newspapers, TV, radio and internet blogs. Ortoleva and Snowberg interpret this finding as suggesting that greater media exposure enhances overconfidence and extremism, because of correlation neglect (voters don't take into account that signals are correlated and overestimate the accuracy of the information that they acquired). But an alternative

interpretation, consistent with rational inattention, is that voters with more extreme policy preferences deliberately seek more information, because they have greater stakes in political outcomes.

The specific implications for how the equilibrium differs from that with full information depend on the shape of the distribution of bliss-points t^J. If the distribution is asymmetric, then voters in the longer tail pay relatively more attention, and thus the equilibrium under rational inattention is closer to them relative to the perfect information equilibrium. In other words, the equilibrium policy lies on the other side of social optimum than the median voter's bliss point. Similarly, asymmetries in group size could move the equilibrium away from the majority's preferences. Two competing groups of unequal size is a special example of this point. In this case, the implemented policy could be relatively suboptimal for the majority and beneficial for a small number of insiders.

These effects are stronger if it is more difficult to observe the policy, i.e., if the issue is more complicated, for instance as with financial regulation. The size of this deviation from the utilitarian optimum increases with the size of the information cost. Specifically, suppose that $\hat{\lambda}^J = \hat{\lambda}$ for all J. The derivative of the first order condition (8) that characterizes the equilibrium with inattentive voters with respect to $\hat{\lambda}$ is $-\frac{1}{\sigma^2} \sum_{J \in P} \frac{m^J}{u^J(q)}$, where $P = \{J : 1 - \frac{\hat{\lambda}}{(u^J)^2 \sigma^2} > \xi_0\}$. If this derivative is negative, then the equilibrium value of q drops if $\hat{\lambda}$ rises. Notice that this holds for negatively skewed distributions of t^J.

Policy divergence and new candidates. This example also sheds light on the implications of differences in information costs between the two candidates. Suppose that the cost of collecting information is lower, say, for candidate B, so that $\lambda_B < \lambda_A$. For instance, A could be a less established candidate to which the media pay less attention. Then all voters pay more attention to the more established candidate, here B ($\xi_B^J > \xi_A^J$ for all J). But this effect is not the same across groups of voters. By (10), the difference in attention given by voters between the two candidates depends on u^J, and it is higher in the center, i.e., for t^J closer to q, than at the extremes of the voters' distribution. Specifically, the more extremist voters pay relatively more attention to the less established candidate A, while the centrist voters pay relatively more attention to candidate B who has greater media coverage (this can be seen by evaluating the derivative of ξ^J with respect to λ in (10)). This in turn affects the incentives of both candidates and leads to policy divergence if the distribution of bliss points is asymmetric.

The policy divergence emerges because candidate A assigns a greater weight to the more extreme voters compared to candidate B, since these voters are more attentive to his policies given their higher stakes. In other words, more established candidates tend to cater to the average voter, while candidates receiving less media coverage go after ex-
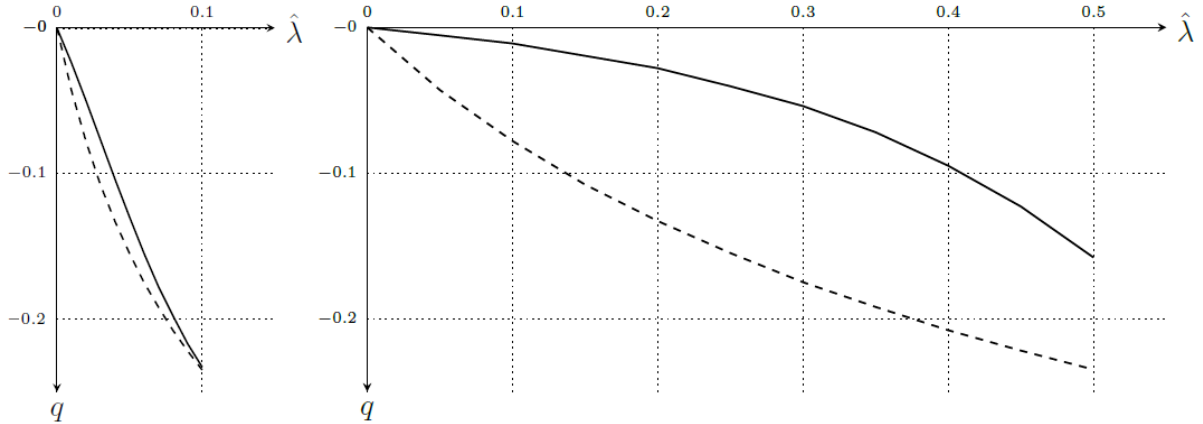
Figure 1: Effect of the cost of information, left: $\sigma_C^2 = 0.05$, right: $\sigma_C^2 = 0.25$, solid: exact solution, dashed: first-order approximation.

tremist voters. With policy divergence and different attention weights, the probability of victory differs from $1/2$, and the less established candidate A (who receives less attention by all voters and by the centrist voters in particular) is less likely to win (since $\xi_B^J > \xi_A^J$ for all J, the value of the objective function $\sum_J m^J \xi_C^J U^J(q_C)$ at the optimum will be larger for B than for A). This effect is weaker when the policy stakes u^J are scaled up, however. This implies that in unusual times,e.g., in a crisis when policy stakes are particularly high, or when a new important issue comes up, then less established candidates have a higher chance of winning the elections. Such situations provide windows of opportunity for new challengers.

The prediction that electorally advantaged candidates pursue more centrist policies, while weak candidates cater to the extremes, is consistent with evidence from US Congressional elections discussed by Fiorina (1973) and Ansolabehere et al. (2001)

Numerical example. To illustrate these findings, let there be three types of voters of equal mass such that $t^1 = t^2 = \frac{1}{2}$ and $t^3 = -1$. Let us also assume $U^J(q) = -(q - t^J)^2$. Information costs are the same for both candidates, who thus set the same policy targets. Under perfect information, $\hat{\lambda} = 0$, the equilibrium policy coincides with the social optimum, $q = 0$. It is the average of the bliss-points in the population. The median voters' bliss point is $1/2$. However, when the cost of information increases, the equilibrium q becomes negative, i.e., moves in the direction of the smaller group.

Figure 1 presents the equilibrium q as a function of $\hat{\lambda}$. The solid curve represents the exact solution using (40) in the Appendix, and the dashed curve is based on the first order approximation, (10). The left panel shows results for $\sigma_C^2 = 0.05$. There, when $\hat{\lambda} = 0.01$, then $q \doteq -0.02$, when $\hat{\lambda} = 0.05$, then $q \doteq -0.13$, and when $\hat{\lambda} = 0.1$, then $q \doteq -0.23$. For positive costs of information, the smaller group $J = 3$, i.e., the extreme voters, pay

relatively more attention than $J = 1$ and $J = 2$ when q is in the neighborhood of zero, and thus the equilibrium policy moves in their direction.[23] Note that here the variance of prior uncertainty about policies is of moderate size: it is one tenth of the total variance of bliss points in the population. We can see that the first order approximation works quite well here.

The right panel in Figure 1 presents equilibrium policies for $\sigma_C^2 = 0.25$. In this case, the variance of policies is somewhat extreme - it is as large as half of the variance of bliss points in the population. Due to the much larger uncertainty, voters choose to pay closer attention, and for the same $\hat{\lambda}$ the equilibrium departs less from the social optimum $q = 0$ than in the left panel, both in the first order approximation and in the exact solution. The distance between the first order approximation and the exact solution increases with a larger variance, however. The reason is that with a large variance, the risk aversion effect (which is present only in the exact solution) induces voters to pay even more attention as σ_C^2 increases.

The equilibrium policies are represented by Figure 1 also when candidates differ in their media coverage, i.e., in the costs $\hat{\lambda}$ associated with processing information about their policy instruments. In such a case, the policies of the two candidates diverge, and the candidate with less media coverage (higher $\hat{\lambda}$) chooses a lower q.

If the cost of attention is heterogeneous across voters, then the equilibrium policy reflects that, too. Preferences of voters with a lower marginal cost weigh more in equilibrium. For instance for $\sigma_C^2 = 0.05$, if $\hat{\lambda}^3 = 0.01$ and $\hat{\lambda}^1 = \hat{\lambda}^2 = 0.1$, then in equilibrium $q = -0.34$, policy is closer to the more attentive voters $J = 3$.

Political parties as labels. This model can also shed light on the role of parties, as ideological labels that save voters' attention.[24] By consistently taking positions in defense of specific economic interests, or according to specific ideological views, political parties can save voters the cost of collecting information on different issues or over time. This role of parties as labels can be illustrated by a simple extension of the one-dimensional policy application. Suppose that there is one national electoral district and two regional districts. A one dimensional policy has to be chosen at each level of government, and voters care about both the national and regional policies. The three elections are run simultaneously. Each voter participates in two elections, in his region and in the nation. There are two political parties, each running in all three elections. But now suppose

---

[23]When the cost of information increases beyond a certain level, then attention becomes uniform again since all voters are at the lower bound for attention, ξ_0. Once this lower bound is reached, policy is again at the social optimum since all voters are weighted equally.

[24]This insight is emphasized by Downs (1957). See also Snyder and Ting (2002), where voters get information about the ideological preferences of individual candidates by observing the party label. In our approach, instead, the label also affects the subsequent choice of learning about policies.

that, before voters choose attention, each party chooses whether to coordinate policy across elections, or to let the policy be set independently at the regional vs national level. Coordination amounts to a commitment to run on the same electoral platform at the national and regional level.

The important piece here is that voters know whether polices are set nationally, or independently across regions. The presence of a party organization allows for such labeling across electoral districts. The advantage of a coordinated policy is twofold. First, by increasing the voters' stakes, it increases their attention. Second, it reduces in half the cost of attention, since attention devoted to this policy is useful in two elections (regional and national) rather than in one only. If voters draw the same utility from the national and the regional policy, coordination has the same effect as a four-fold reduction in λ (see (10), where stakes enter squared). As a result, the equilibrium policy gets closer to the social optimum and this increases the party's probability of winning both elections. This benefit of a single coordinated policy is offset by the cost of a worse local fit; the cost is higher the more districts differ in terms of voters' policy preferences. Under perfect information, both parties would always prefer full decentralization, rather than a single coordinated policy. But if heterogeneity is not too large and the cost of attention is high, then both parties may prefer to coordinate national and regional policies, so as to grab more attention.

For a numerical example, let $U^J(q) = -(q - t^J)^2$, $t^1 = 1, t^2 = -1$, $\hat{\lambda} = 0.3$ and $\sigma_C^2 = 0.1$. There are two regions, with different combinations of voters type. The national population is the sum of the two regions. Let there be 75% of type 1 voters in region 1. A centralized policy is then optimal if the share of type 1 voters in region 2 is higher than 60%. In this case, if set independently, optimal policies in the two regions and nationally would be similar, and it is thus better to set one common policy, which would then allow voters to be more informed about it. On the other hand, if the share of type 1 voters in region 2 is less than 60%, i.e., the two regions are less alike, then it is better to target independent policies across the two regional elections. When the cost of information is lower, then the two regions have to differ even more for the coordination to emerge.

This reasoning may explain why often political parties are reluctant to adapt their policy platforms to local conditions in elections for State or local government. In many US states, elections are uncompetitive, with one party always winning. Why doesn't the losing party adjust its policies locally, deviating from the national position? E.g., why do not Midwest democrats support the "pro life" stand? Our explanation is that voters would then have a harder time finding out what both the national and state policies are, which in equilibrium would hurt the party's chances of winning a higher sum of chairs.[25]

---

[25]Strictly speaking, addressing this puzzle would require a richer model, where candidates also have

Similar forces may be at work in a dynamic setting, where electoral platforms could be coordinated over time and across policy issues. In this case, parties of political candidates would be motivated to choose persistent policies. If the costs of information were sufficiently high, then they would prefer stable policies over policies responding to current needs of voters. Announcing a policy persistent over a long horizon would increase voters' stakes in the same way as a policy common across regions.

4.2 Multidimensional policy: Targeted transfers and public good provision

When the policy is multi-dimensional, rational inattention has additional implications, because voters also choose how to allocate attention amongst policy instruments. As discussed above, equilibrium attention is higher on the policy instruments where the stakes for the voter are more important. Typically these are the most divisive policy issues, on which there is sharp disagreement amongst voters. The reason is that voters realize that the equilibrium will not deliver their preferred policies on the more controversial issues, while they expect to be pleased (and hence have low stakes - i.e. low marginal utility from observing a policy deviation) on the issues where they all have the same preferences.

We illustrate this result in a model of public good provision and targeted redistribution. The model is symmetric and all voters behave identically. The framework is similar to Gavazza and Lizzeri (2009), except that there information is given exogenously. Our agents instead choose what to get informed about. They all choose to pay minimal attention to the public good and to uniform taxes (on which they all agree), and focus their attention on the targeted policy instruments, with highest attention on those instruments that are more relevant for them. As a result in equilibrium there is under-provision of the public good and over-reliance on uniform but distorting taxes in order to finance targeted redistribution. Equilibrium distortions are worse if the granularity of information increases, if policy instruments allow for finer redistribution and if resources are more abundant.

A simple model. Consider an economy where $N > 2$ groups of voters indexed by J derive utility from private consumption c^J and a public good g:

$$U^J = V(c^J) + H(g),$$

---

partisan preferences. In a setting with opportunistic candidates, there is policy convergence and both parties always make identical choices of coordination vs non coordination, so that the probability of winning is 1/2 for both.

where $V(.)$ and $H(.)$ are strictly increasing and strictly concave functions. Each group has a unit size. Government spending can be financed through alternative policy instruments: a non distorting lump sum tax targeted to each group, b^J, with negative values of b^J corresponding to targeted transfers; a uniform tax, τ, that cannot be targeted and that entails tax distortions; and a non observable source of revenue, s for seignorage, also distorting and non targetable. Thus, the government and private budget constraints can be written respectively as:

$$
\begin{aligned}
g &= \sum_J b^J + N\tau + s \\
c^J &= y - b^J - T(\tau) - S(s)/N.
\end{aligned}
$$

where y is personal income and the functions $T(\cdot)$ and $S(\cdot)$ capture the distorting effects of these two sources of revenues. Specifically, we assume that both $S(\cdot)$ and $T(\cdot)$ are increasing, differentiable, and convex functions. Moreover, $S(0) = T(0) = 0$ and for derivatives $S'(0) = T'(0) = 1$. From a technical point of view, the non observable tax has the role of a shock absorber and allows us to retain the assumption of independent noise shocks to all observable policy instruments. Its distorting effects capture the idea that any excess of public spending over tax revenues must be covered through inefficient sources of finance, such as seignorage or costly borrowing. Putting these pieces together, we get:

$$ U^J(q) = V[y - b^J - T(\tau) - S(g - \sum_K b^K - N\tau)/N] + H(g) \tag{12} $$

The observable policy vector is $q = [b^1, ..., b^N, g, \tau]$, and the non observable tax can be inferred by voters from information on the observable policy vector. For simplicity, we assume that prior uncertainty is the same for all voters, all candidates and all policy instruments, and all voters have the same information costs: $\sigma_{C,i}^J = \sigma$ and $\lambda_{C,i}^J = \lambda$ for all C, J, i.

It is easy to verify that the socially optimal policy vector satisfies $\hat{s} = \hat{\tau} = 0$, i.e., eliminates all distorting taxes, achieves equal consumption for all groups, $c^J = \hat{c}$ for all J, and sets the public good so as to satisfy Samuelson optimality condition, namely $H'(\hat{g}) = V'(\hat{c})/N$. Thus the optimal level of the public good is financed through an equal targeted lump sum tax on all groups. Under full information, electoral competition would deliver this outcome.

Equilibrium policy with rational inattention. Next consider the equilibrium with rational inattention. To express the first order conditions (8), we use: $u_J^J = (-1 + S'/N)V'(c^J)$, $u_{-J}^J = V'(c^J)S'/N$, $u_\tau^J = (T' - S')V'(c^J)$ and $u_g^J = H' - V'(c^J)S'/N$, where the J and $-J$ subscripts refer to partial derivatives of U^J with respect to a voters'

own taxes b^J, and taxes targeted at others, b^K for $K \neq J$, respectively; and the g and τ subscripts refer to partial derivatives of U^J with respect to g and τ respectively; all derivatives are evaluated at the equilibrium policy targets. By symmetry, in equilibrium all groups are treated in the same way, so that $c^J = \hat{c}^*$, where a * denotes the equilibrium. The first order conditions with respect to $\hat{g}$ and $\hat{\tau}$, as long as attention to these instruments is positive, are the same as for the social planner's problem, respectively:

$$-V'S'/N + H' = 0 \tag{13}$$
$$-T' + S' = 0 \tag{14}$$

The reason is that all types J pay the same level of attention to g and τ, and thus ξ_g^J and ξ_τ^J do not enter these expressions.[26] What could drive equilibria away from the social optimum is heterogeneity in ξ_i^J across different voters, only, which does not arise with these uniform tax instruments and given the symmetry of the model.

The first order condition (8) with respect to $\hat{b}^J$ can be written as:

$$\xi_J^J V'(c^J)(-1 + S'/N) + \sum_{K \neq J} \xi_J^K V'(c^K) S'/N = 0 \tag{15}$$

Exploiting symmetry again and simplifying, this can be written as:

$$[1 + (N-1)\frac{\xi_J^{-J}}{\xi_J^J}]S'/N = 1 \tag{16}$$

Equation (16) is inconsistent with social optimum. At the social optimum, $S' = 1$ (since $s = 0$), which in turn implies that $\xi_J^{-J} < \xi_J^J$, since $N > 2$ - cf (10). Namely, at the socially optimal policy, all groups pay more attention to their own taxes than to taxes paid by other groups. But if $\xi_J^{-J} < \xi_J^J$, then equation (16) implies $S' > 1$, a contradiction. Hence in equilibrium, it must be that $S' > 1$, and hence that $\hat{s} > 0$. Equations (13)-(14) then imply that $H' > V'/N$ and that $T' > 1$. There is under-provision of the public good relative to the social optimum, and the government relies on distorting (observable and unobservable) sources of revenues, despite the availability of lump sum taxes. In fact, if the marginal tax distortions T' and S' do not rise too rapidly, it is even possible that the equilibrium entails negative values of $\hat{b}^J$. That is, both candidates collect revenue through distorting taxes from all citizens, and then give it back to each group in the form of targeted transfers (i.e. there is fiscal churning).

---

[26]This can be seen from (10) and from the fact that u_τ^J and u_g^J are common to all voters.

The allocation of attention. The source of these distortions is the asymmetry in attention: voters pay more attention to the targeted instruments, because (in equilibrium) the stakes are higher, and they neglect the instruments that have the same effects on all citizens, for the same reason. Moreover, they pay more attention to their own targeted taxes (or transfers) than to the targeted instruments affecting others. This in turn induces both candidates to deviate from the efficient allocation, in order to appear to please each group. The higher is the cost of information λ and the larger is N, the larger is the distortion.

Note that in equilibrium $u_\tau^J = T' - S' = 0$ and $u_g^J = H' - V'S'/N = 0$. By (10) this in turn implies that $\xi_g^J = \xi_\tau^J = \xi_0$. Namely, in equilibrium all voters pay minimal attention to public goods and to the uniform distorting tax, as if they were non-observable. This point applies generally, beyond this specific example. If there is no disagreement amongst voters regarding a policy instrument, then all voters expect both candidates to set these general instruments at their optimal values (from the individual voter's selfish perspective). Marginal utility from policy deviations is then zero, and voters have no incentive to devote costly attention to these items. For issues that are non-divisive, (8) implies that the equilibrium attention is at the minimal level ξ_0. On the other hand, divisive issues are paid more attention to. Since the policy is not set optimally from the perspective of each individual voter, then voters' stakes are positive, and they pay attention to such issues.[27]

The result that in equilibrium voters are inattentive to policies on which everyone agrees (such as g and τ in the model) while they pay attention to divisive issues (such as targeted instruments), is consistent with existing evidence on the content of Congressional debates and on the focus of US electoral campaigns. Ash et al. (2015) construct indicators of divisiveness in the floor speeches of US congressmen. Exploiting within-legislator variation, they show that the speeches of US senators become more divisive during election years, consistently with the idea that voters' attention is greater on the more divisive issues. Moreover, Hillygus and Shields (2008) show that divisive issues figure prominently in US presidential campaigns, contrary to the expectation that candidates instead try to avoid divisive policy positions in order to win more widespread support.

The effects of fiscal transparency. An important implication of the model is that more information can have adverse effects on social welfare, because it can enhance endogenous informational asymmetries. The diffusion of the Internet is a case in point.

---

[27] For any $\xi_0 > 0$ the equilibrium is unique. However, when $\xi_0 = 0$, there is an interval of equilibria about the unique equilibrium for a positive ξ_0. This is because, when attention to g and τ is zero, then the first order conditions (8) with respect to these instruments are satisfied trivially. At the social optimum, u_g^J and u_τ^J equal zero, and thus attention is zero, and it is zero in its neighborhood as well.

Internet provided very cheap information on very fine issues. Such granular information was not available before at all. Voters can now pay attention to very narrow issues, yet their total span of attention is still limited. This can lead to large differences in attention across voters and policy issues.

To illustrate this point, suppose that agents cannot choose attention to each targeted transfer independently, but that information about several such targetable instruments is packaged together in M information bins. Specifically, the number N of targetable instruments is decomposed as: $N = kM$, where k and M are both integers and k denotes the size of each information bin (all bins are of equal size to preserve symmetry). Voters are constrained to pay uniform attention to the objects inside each bin. That is, they observe b^J and b^I separately for $J \neq I$. But they can only vary attention across the M information bins, not across the N targetable instruments. Thus k is a measure of how coarse is information: lower k means more granular information.

Denote by ξ_J^J the attention paid by J to the information bin that contains b^J, and by ξ_{-J}^J the attention paid by J to the information bins that does not contain b^J. Using the first order condition for b^J (15) together with the constraint on bins of ξ, we get the following instead of (16):

$$[k + (N - k)\frac{\xi_{-J}^J}{\xi_J^J}]\frac{S'}{N} = 1. \tag{17}$$

As k increases (more coarse information), S' monotonically decreases towards 1 (the social optimum). This is because, for fixed $\frac{\xi_{-J}^J}{\xi_J^J} < 1$, the term $[k + (N - k)\frac{\xi_{-J}^J}{\xi_J^J}]$ is increasing in k. In addition, it can be shown that in equilibrium $\frac{\xi_{-J}^J}{\xi_J^J}$ is increasing in k too. Intuitively, as k makes incentives to pay attention to different targetable instruments more uniform, different voters pay more similar attention to the same instrument.[28]

Thus, as k increases, welfare improves, and reaches Pareto efficiency when $k = N$ (i.e. information is the least granular). In other words, more granular information leads to more distorted policies and is welfare deteriorating.

More generally, the result highlights how to package information for voters so as to reduce political distortions. The equilibrium would become less distorted if the cost of

---

[28]By contradiction: for increasing k to be associated with increasing S', the ratio $\frac{\xi_{-J}^J}{\xi_J^J}$ would according to (17) need to be decreasing in k. Following (10):

$\xi_J^{-J} = \max\left(\xi_0, 1 - \frac{\hat{\lambda}}{(V'S')^2\sigma^2}\right)$ and $\xi_J^J = \max\left(\xi_0, 1 - \frac{\hat{\lambda}}{\frac{1}{k}(V'(1-S'))^2\sigma^2 + \frac{k-1}{k}(V'S')^2\sigma^2}\right).$

The denominator in ξ_J^J is a weighted average of equilibrium squared marginal utilities from taxes targeted at the agent's group (decreasing in S') and at other groups (increasing in S'). And thus for the ratio to be increasing in k, equilibrium V' would need to be decreasing in k, because both increasing k and the assumed associated increasing S' increase the ratio $\frac{\xi_{-J}^J}{\xi_J^J}$, too. But that is not possible, since decreasing V' would imply higher equilibrium consumption at higher inefficiencies S.

information on instruments targeted at others (λ^J_{-J}) fell, while the cost of information on instruments targeted at themselves (λ^J_J) increased. This can be seen from (16): a higher λ^J_J and a lower λ^J_{-J} would raise the ratio $\frac{\xi^J_{-J}}{\xi^J_J}$, leading to less seignorage, more public good provision and less distorting taxation. Intuitively, voters would pay more attention to benefits targeted at other groups, raising the political costs of targeting.[29]

Another welfare improving information repackaging would be to also give voters on the net taxes that they pay, $b^J + \tau$, besides on b^J and τ separately. Then voters would pay some attention to it, and candidates would be less tempted to raise τ and reduce b^J, because voters would be less likely to detect a direct welfare improvement. If information is separately provided on b^J and τ, instead, such a deviation would be more profitable for the candidates, because voters would be attentive to b^J while paying only minimal attention to τ.[30]

The more general normative lesson is that more information is not necessarily better, but information should be packaged so that the value of attention is similar across policy dimensions and groups of voters. This is different from Gavazza and Lizzeri (2009), who emphasize the distorting effects of asymmetric information in a setting where voters' information is exogenous. They argue that more information on aggregate spending is welfare improving, while information on aggregate taxes is counter-productive in an intertemporal setting. Our model instead highlights the distinction between targeted vs general instruments. Changing the cost of information on general taxation (τ) or general public goods (g) has no effect in our framework, because voters choose to pay no attention irrespective of the cost. What matters instead is the cost of collecting information on instruments targeted at them vs. those targeted at others.

Finally, and almost trivially, the model could be extended to capture the evidence in Cabral and Hoxby (2012), or Bordignon et al. (2010). These empirical papers find that policymakers tend to charge lower tax rates when the visibility of taxation is higher, shifting the tax burden on less visible sources of revenue. This prediction would follow almost immediately from a modified version of this example, where the cost of information λ^J varies across policy instruments. From a normative perspective, this implies that

---

[29]Of course, there is a limit to how much these costs can be exogenously changed by the government, since the cost of observing instruments targeted at one-self will generally be lower than the cost of instruments targeted at others (see Ponzetto (2011) for a specific example of this point with regard to trade policy). Moreover, transparency is also a policy choice, and it is not clear that politicians would always benefit from it.

[30]Note that the incentive to under-provide the public good would not be affected by this repackaging of information, since candidates would still have the possibility of reducing g (to which voters only pay minimal attention) so as to reduce targeted taxes on all groups. For this reason, it would not be optimal to *only* provide information on $b^J + \tau$, since the attention paid to targeted taxes paid by others dampens the incentive to under-provide g. Deriving these results formally would entail additional complications, because now the error terms would be correlated across observable variables, and the expressions in Propositions 1 and 2 and in Lemma 2 would have to be modified accordingly.

more transparency of taxation is not always unambiguously welfare improving. Suppose, in particular, that there are differences in transparency across policy instruments, and for technological reasons some policy instruments cannot become more transparent (for instance because income tax withholding is preferable due to economies of scale or for other administrative reasons). Then, it may be optimal to reduce the transparency of other sources of revenues, so as to put them on an even footing in terms of political costs.[31]

Comparative statics: Policy fragmentation and income shocks. Distortions get worse if the policy vector is more fragmented. While this point is about the policy space, it is related to the discussion above about granularity of information. The detrimental effect of fragmented policy can easily be seen if the number of groups increases. As N goes up, the cost of a targeted transfer to a particular group is spread out over many other groups. As a result $\frac{\xi^J_{-J}}{\xi^J_J}$ decreases, i.e, each transfer targeted at a particular group is paid relatively less attention by others. Thus, as N increases, targeting increases, the level of public good decreases and uniform distorting taxes (τ and s) increase.

This point applies more generally, and provides a rationale for simplicity of policy. Fine and complex policies are not paid detailed attention across all dimensions, which induces distortions. While fine policy vectors might address finer issues, they also typically imply larger differences in stakes across different voters and policy issues. Hence, as the heterogeneity of stakes and of attention increases, the departure from the social optimum is larger. In the limit, the distortions would disappear if the government was forced to treat all groups in the same way. Proposition 1 implies that the social optimum emerges if and only if attention weights are uniform across policy instruments and voters.

This reasoning also has implications for the optimal task allocation to different levels of government. A traditional argument in favor of fiscal decentralization is that local governments have better information about local preferences. Here we are led to a similar conclusion, but through a different logic. To the extent that groups are determined by geography (i.e they are residents of different localities), fiscal decentralization reduces conflict between voters, because it reduces opportunities for redistribution. As such, fiscal decentralization reduces the scope of informational asymmetries across different policy issues associated with a single election, and this leads to more efficient outcomes.

Finally, equilibrium distortions are also affected by income shocks. As income y falls, the marginal utility of consumption rises, which induces voters to become generally more attentive. But attention to targeted transfers to others rises more than the attention

---

[31]Inattention also changes the behavioral implications of how economic agents respond to tax policy or other instruments, including the deadweight losses of taxation. Here we neglect these issues, discussed at length for instance in Congdon et al. (2011).

paid to own taxes, i.e., it can be shown that in equilibrium the ratio $\frac{\xi^J_{-J}}{\xi^J_J}$ is an increasing function of y. In other words, the effect of a negative income shock is similar to that of a drop in the cost of information. As a result, distortions due to inattention are reduced by a drop in income and the equilibrium gets closer to the full information benchmark. This result is in line with empirical findings that efficiency enhancing reforms are more likely to occur during large recessions - see for instance Høj et al. (2006) and OECD (2012).

5 Partisan politicians

There is ample evidence that US elected representatives have become more polarized over recent decades, arguably more so than their electorate (Fiorina 2009). Several alternative explanations have been put forward, some of which emphasize the radical changes that occurred in the media environment with the diffusion of the internet while others are more sceptical on the role of the internet (see Barber and McCarty 2013 but also Boxell et al. 2017). Can a drop in the cost of political information lead to more policy divergence? We now show that the answer is not straightforward, and depends on how different types of voters are affected by the new media environment. The section also provides an alternative microfoundation for the sources of political uncertainty faced by voters.

Shocks to candidates' preferences   Suppose that candidates care about the policy implemented, just like voters, and not about the probability of winning per se. In other words, they are "citizen candidates", and candidate C has utility $V^C(q) = V(q - t^C)$, for $C = A, B$. Policy is one-dimensional ($q \in \mathbb{R}$) and t^C is the candidate's bliss-point (i.e., V is maximized at $q = t^C$). Politicians do not make implementation errors, and the policies they choose are the actual policies they enter elections with. However, the candidates' bliss points t^C are random and only candidate C observes the realization of t^C. This is the source of voters' uncertainty: since the preference shocks t^C are ex ante unknown by voters, so are the policies chosen by candidates.

The sequence of events is as follows:

1. Voters and candidates form priors, nature draws iid preference shocks $t^C \sim N(\bar{t}^C \sigma_t^2)$.
2. Each candidate observes his own preference shock t^C, and sets policy q^C.
3. Voters set attention startegies and then they observe noisy signals of the chosen policies q^A and q^B.
4. The ideological bias x^v is realized and elections are held. Whoever wins the election enacts his announced policy.

The voters' problem is exactly as in the baseline model, where attention solves problem (4). The candidates' problem is now different. Candidate A chooses policy to maximize expected utility, which is:

$$p_A V^A(q_A) + (1 - p_A)E[V^A(q_B)], \tag{18}$$

The expectation operator in the second term is over q_B (which depends on the unobserved preference shock t^B). The probability that A wins the elections, as perceived by A, is:

$$p_A(q_A, q_B) = \frac{1}{2} + \psi \left(\sum_J m^J E_\epsilon^J \left[E[U^J(q_A)|s_A^{v,J}] - E[U^J(q_B)|s_B^{v,J}] \right] \right) \tag{19}$$

The only difference from (6) is that here the expectation is not also over implementation errors, but only over noise in voters' signals.[32]

The definition of equilibrium below is different from the benchmark Definition 1 because here it must include the whole profile of candidates' policies (policies for all preference shocks), not just one target policy for each candidate. This is because the profile determines the distribution of policies, and thus voters' prior, which in turn affects voters' attention. We look for candidates' strategies in the class of policies that are linear in preference shocks:

$$q_C(t^C) = q_C^0 + \beta_C(t^C - \bar{t}^C), \tag{20}$$

where $q_C^0 \in \mathbb{R}$ is the policy vector selected when the mean preference $t^C = \bar{t}^C$ is realized and $\beta_C \in \mathbb{R}$ is a linear coefficient. This is in line with our focus on small-noise approximations, as these policies are local approximations to any continuous policy. Linearity is important for tractability too, since it translates Gaussian preferences into Gaussian policies. If voters' priors were not Gaussian, then the machinery of Bayesian updating would be much more complicated.

Definition 2 Equilibrium with partisan candidates. *The equilibrium is a profile of policy vectors $q_C(t^C)$ given by $\{q_C^0, \beta_C\}$ in (20) and of attention strategies ξ^J such that:*

(a) *The attention strategies ξ^J solve the voters' problem (4) for prior beliefs that are consistent with candidates' strategies (namely prior beliefs q_C are normally distributed with mean q_C^0 and variance $\beta_C^2 \sigma_t^2$).*

(b) *For any t^C, the policy vector $q_C(t^C)$ maximizes the objective (18), taking as given*

---

[32]Note that, even for symmetric equilibria where the two candidates have the same policy profiles, the probability of winning perceived by candidate A conditional on the realized shock t^A differs from $1/2$ anytime t^A differs from its mean $\bar{t}^A$.

voters' equilibrium attention and the equilibrium distribution of policies of the opponent.

No change in the voters' problem   Let us emphasize that there is nothing new on the voters' side, regardless of the drivers of candidates' choices. The resulting attention strategies are thus given by (10). Voters choose how much attention to pay based on the marginal utility around the expected policies and on the policy variance. The difference from the baseline model is that noise in policies is driven by candidates' preference shocks, and not by implementation errors - yet, voters' attention strategies are determined by their prior beliefs only. The insights of earlier sections fully apply.

What information do voters choose to get? Voters get imperfect signals on the realized policies. In the baseline model this is equivalent to information on implementation errors, while here it is equivalent to signals on candidates' preferences. If a candidate is expected to be more extreme or more volatile, voters choose to pay more attention to him/her.

Equilibrium policies   The properties of equilibrium policies, instead, differ from those in the baseline model of section 2 in two respects. First, here the equilibrium displays policy divergence. This is an immediate implication of the candidates' first order conditions. Similarly to (8), we have that $\partial p_A / \partial q_A = \psi \sum_J m^J \xi_A^J u_A^J$. Hence we can write candidate $A's$ first order conditions for each realized shock t^A as:

$$\left(V^A(q_A) - E[V^A(q_B)] \right) \psi \left(\sum_J m^J \xi_A^J u_A^J \right) + v^A(q_A) p_A = 0, \tag{21}$$

where v^C is the marginal utility of V^C. Thus, each candidate trades off the expected electoral gain of pleasing the voters (the first term on the LHS of (21)) against the marginal welfare loss of being far from its bliss point (the second term on the LHS of (21)). Since generally $1 > p_A > 0$, equation (21) implies that $q_A^* \neq q_B^*$ in equilibrium.

Second, the variance of policies is endogenous, unlike when it was driven purely by implementation error. By (20), the variance of policies is $\sigma_C^2 = \beta_C^2 \sigma_t^2$, and it depends on β_C, the chosen responsiveness to preference shocks. It can be shown that, with costly information ($\lambda > 0$), the optimal β_C falls as candidates become more extreme (i.e. with $\bar{t}^C$ further from social optimum). In other words, more extreme candidates are more stable (assuming the variance of preference shocks is the same for both candidates). The reason is that in equilibrium these extreme candidates receive more attention. Hence additional deviation towards their bliss-point, and perhaps away from the bliss-point of voters, is more costly in terms of votes. A more extreme candidate thus responds less to his own preference shocks.

More extreme candidates are thus less volatile (they have lower σ_C^2), which delivers a partially attenuating effect on the voters' attention. This attenuation effect was absent from the baseline model with implementation error, where the volatility of policies was given exogenously.

Rational inattention and polarization Does costly attention amplify or moderate the extent of policy divergence, relative to a full information equilibrium? At the equilibrium policies, a majority of voters wishes more moderate policies by each candidate. That is, enacting more moderate policies would be rewarded by voters and the probability of winning would rise (by (21), $v^A(q_A)$ and $\partial p_A/\partial q_A = \psi \sum_J m^J \xi_A^J u_A^J$ have opposite signs). Costly attention dampens voters' reactions, because voters are less responsive to policy choices than under full information. This reduces the incentive to enact moderate policies. As a result, the equilibrium with costly attention displays more policy divergence compared to full information.

More generally, suppose that λ rises uniformly for all voters. Then it can be shown that equilibrium policy divergence rises (q_C^0 gets closer to $\bar{t}^C$, and β^C gets closer to one), because voters are less responsive to deviations (they only observe a noisy signal of actual policies).[33]

Numerical solution These properties of the equilibrium are illustrated in the following numerical examples, where we assume (20) and use a small-noise approximations to (21).[34]

Figure 2 presents numerical solutions for two different cases. For the numerics, we assume symmetric voters $t^1 = 1/2, t^2 = -1/2, m^1 = m^2 = 1/2, \psi = 0.1$ and quadratic utilities of both voters and candidates, i.e, $U^J = -(q - t^J)^2, V^C = -(q - t^C)^2$. The social optimum is thus $q = 0$. The cost of information is $\lambda_C = 0.01$ and candidates' bliss-points have variance $\sigma_t^2 = 0.01$.

The left panel in Figure 2 presents solutions q_A^0 and q_B^0 for symmetric candidates ($\bar{t}^B = -\bar{t}^A$) as both mean bliss-points are varied, $\bar{t}^A \in (0,1)$. The dashed curves are

---

[33]Recall that $\frac{\partial p_A}{\partial q_A} = \psi \frac{1}{N} \sum_J \xi_A^J u^J(q_A)$. By (10), differentiating this expression w.r. to λ around the equilibrium $\frac{\partial^2 p_A}{\partial q_A \partial \lambda} = -\psi \frac{1}{N} \frac{1}{\sigma^2} \sum_{J \in P} \frac{1}{u^J(q_A^*)}$, where $P = \{J : 1 - \frac{\hat{\lambda}}{(u^J(q_A^*))^2 \sigma^2} > \xi_0\}$. Suppose that the distribution of voters' bliss points is symmetric around $1/2$, and consider a symmetric equilibrium with $0 < t^A < 1/2$ and $t^A = -t^B$. Since $q_A^* < 1/2$, $u^J(q_A^*) \gtreqqless 0$ as $t^J \gtreqqless q_A^*$ and the distribution of voters is symmetric around $1/2$, we have that $\sum_{J \in P} \frac{1}{u^J(q_A^*)} > 0$. Thus, $\frac{\partial^2 p_A}{\partial q_A \partial \lambda} < 0$ around the equilibrium. In words, if the cost of political information rises for all voters, then the probability of victory becomes less sensitive to policy choices. By (21), then, partisan politicians can afford to enact more divergent policies.

[34]We find $\{q_A^0, \beta_A, q_B^0, \beta_B\}$ such that the approximation to (21) is satisfied for all $\{t_A, t_B\}$. This amounts to solving for four equations (requirements on constant and linear coefficients for the FOCs of A and B) and four unknowns.
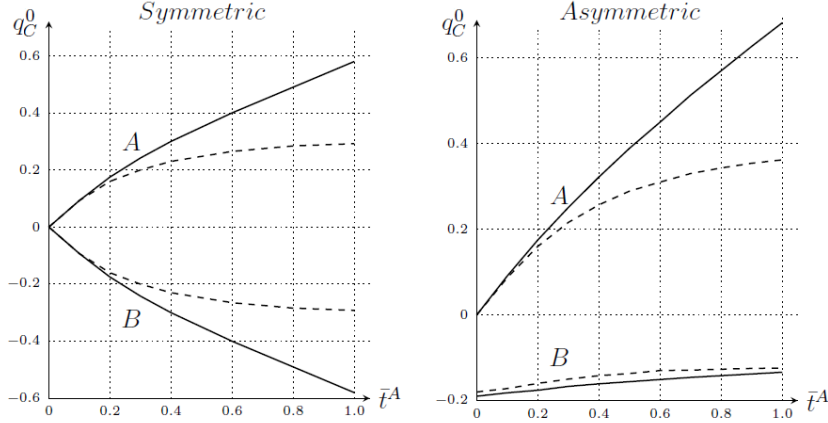
Figure 2: Effect of the bliss-point $\bar{t}^A$. Solid: rational inattention ($\lambda_C = 0.01$), dashed: perfect information.

solutions under perfect information, and the solid ones are under rational inattention with $\lambda_C = 0.01$.

As candidates become more extreme, the gap between full information and costly attention increases (i.e. the distance between solid and dashed curves increases). This is the result of three forces. First, for given but imperfect attention (i.e. if $\xi_C^J < 1$ but fixed), more extreme candidates get closer to their bliss point compared to full information, because $\xi_C^J < 1$ implies that voters are less responsive to policy choice (compared to full information) when candidates become more extreme. Second, with endogenous attention but exogenous volatility (as in the baseline model with implementation errors), this effect is attenuated because attention increases as candidates become more extreme. Third, if policy volatility is also endogenous (through β_A), more extreme candidates are less volatile, which partly reduces the increase in attention they receive and increases the gap between costly attention and full information.[35] On net, the first and third force prevail, and polarization increases more under costly attention than under full information as candidates become more extreme.

The right panel presents the asymmetric case: the bliss-point of candidate B is fixed at $\bar{t}^B = -0.2$, and we vary $\bar{t}^A \in (0,1)$. The pattern of increased polarization emerges again. While the responsiveness β_A drops from 1 all the way to 0.47 for $\bar{t}^A = 1$, for B it decreases from 0.91 to 0.86. The more extreme candidates choose lower β_C, which decreases the attention paid to them.

Internet and polarization   Can internet diffusion explain why polarization increased in the US Congress? If internet brought about a uniform drop in the cost of

---

[35]For the solution with rational inattention, β_A is equal to 1 for $\bar{t}^A = 0$ and then gradually decreases all the way down to 0.5 for $\bar{t}^A = 1$.

political information for all voters, then the answer suggested by this model is no. As explained above, a uniform drop in the cost of information strengthens the incentive for policy moderation.

It is not clear, however, that the new digital technologies reduced the cost of political information for all voters. The internet makes it easier to find the information one is interested in. Individuals who are interested in politics can now become more easily informed about the policy positions of different candidates. But many others use internet for entertainment purposes, or to seek other kinds of information, which affects their opportunity cost of attention. Most likely, internet and cable news increased the difference in information between individuals who are more vs less interested in politics (Prior 2007 and Gavazza et al. 2016 emphasize this point). Since individuals more interested in politics often tend to have stronger and more extreme partisan views (an implication of our model too), this probably amplified the difference in information between extremist vs moderate voters. This effect could be captured in our model with the assumption that λ, the cost of acquiring political information, drops only for individuals with more extreme bliss points (the opportunity cost of being politically informed may even have risen for those who primarily use internet for entertainment purposes). How would such compositional changes affect policy divergence?

To answer, suppose that the cost of information λ drops only for individuals with more extreme policy preferences, namely those with bliss points to the left of $t < 1/2$ and to the right of $1 - t > 1/2$. Individuals with $t^J \in [t, 1-t]$ are not affected. Then we can show that in a symmetric equilibrium policy divergence $(q_B^* - q_A^*)$ rises as λ drops, as long as $t \leq q_A^*$. In, fact by continuity of preferences, equilibrium policy divergence rises even if $t > q_A^*$, as long as t is not too close to $1/2$. In other words, our model predicts that the diffusion of internet leads partisan candidates to enact more divergent policies only if it reduces the cost of political information for voters with sufficiently extreme policy preferences.[36]

The intuition behind this result is subtle. Consider voters with more extreme bliss points than the equilibrium policy platforms, i.e. with $t^J \notin [q_A^*, q_B^*]$. Since the equilibrium displays policy divergence, any such voter pays more attention to the more distant candidate (since the stakes $u^J(q_C)$ are higher - cf. (10)). This asymmetry in attention to the two candidates is a moderating force against policy divergence. An attentive voter

---

[36]To prove this result, repeat the same steps as in footnote 33 above, except that now the set P is defined as $P = \{J : t^J \notin [t, 1-t]\}$. Again, we have that $u^J(q_A^*) \gtreqless 0$ as $t^J \gtreqless q_A^*$. Note also that for any $t^I = 1 - t^J$ we also have $u^J(q_A^*) < |u^I(q_A^*)|$ if $|t^J - q_A^*| < |t^I - q_A^*|$, by concavity of $U(.)$. By symmetry of the distribution of voters' bliss points, we then have that $\frac{\partial^2 p_A}{\partial q_A \partial \lambda} = -\psi \frac{1}{N} \frac{1}{\sigma^2} \sum_{J \in P} \frac{1}{u^J(q_A^*)} > 0$. Thus, as λ drops, p_A becomes less sensitive to policy choices ($\frac{\partial p_A}{\partial q_A}$ drops too), and by (21) equilibrium policy divergence increases. Exploiting continuity, this results holds also for t sufficiently close to but larger than q_A^*.

is someone who is considering whether to vote for A or B. Voters whose bliss point is closer to, say, A would like B to set a more moderate (centrist) policy. Their greater attention to B than to A thus induces B to choose a more moderate policy, compared to a full information equilibrium, or to an environment in which attention is uniform across candidates. Now suppose that, as a result of the new media technologies, the cost of information λ drops for all these extreme voters. Their attention increases, but not uniformly: attention increases more on the closer candidate (where the stakes are lower). In other words, the (positive) gap in attention between the more distant and the closer candidate shrinks, as λ drops. This reduces the electoral penalty for choosing extreme policies, and thus equilibrium policy divergence increases.[37]

This mechanism is quite general, since it relies on the simple idea that attention is higher where stakes are higher. This idea in turn follows from the assumption that attention is allocated instrumentally, to maximize the expected political benefit of the relevant information. Thus, we conjecture that the results discussed above would continue to hold in different behavioral models of voting, as long as attention is allocated instrumenally. Suppose for instance that, unlike in our model, some voters are attached to a specific candidate, and they either vote for him/her or abstain, but they never vote for the more distant candidate. Suppose further that the decision to vote (rather than abstain) is motivated by the perceived difference in expected utility under candidates A vs B. In this case too, attached voters have to allocate attention to the policies chosen by A vs B, and the stakes (and hence costly attention) continue to be higher for the policies of the more distant candidate. If B chose a more divergent policy and this was observed by $A\prime s$ supporters, more of them would turn out to vote for A, reducing $B\prime s$ chances of winning. This electoral loss is higher the larger is the difference in attention to the distant vs close candidate.Thus, a higher gap in the attention to close vs distant candidates moderates policy divergence by partisan candidates, just as in our model. A drop in the cost of information for the more extremist voters would reduce this attention gap, encouraging more policy divergence in equilibrium.[38]

6 Concluding remarks

Voters tend to be poorly informed about policy issues raised during an electoral campaign, and about the political process in general. This fact is well known and undisputed.

---

[37]Throughout we assume that the preference shock towards candidates, x^v, has large enough support relative to the extent of the equilibrium policy divergence between candidates, so that, for all realizations of the preference shock x^v, voters expect to be trading off policy preferences against candidate preferences.

[38]Glaeser et al. (2005) study a model with similar behavioral features by voters. However, they assume (without deriving it from a demand for information) that attached voters are more informed about the closer candidate.

Nevertheless, not much is known about what explains the specific patterns of voters' lack of information, and how this interacts with the behavior of politicians. This paper seeks to fill this gap, studying how voters allocate costly attention in a simple model of electoral competition.

The model is highly portable across applications, since attention allocation is derived from first principles, i.e., directly from preferences in a general setup. It can thus be applied to study a large variety of questions within electoral competition, and it could also be extended to study several other aspects of the political process.

In future research, it would be fruitful to integrate our political demand for information in a framework where the cost of information is affected by equilibrium behavior of others, such as media, interest groups, or politicians. This would entail studying the incentives of whoever provides this information, and how this interacts with rational inattention. The literature on lobbying has studied the role of organized groups in providing information to voters, but much of this literature makes very demanding assumptions on the voters' ability to process information (e.g., Coate 2004, Prat 2006). Perego and Yuksel (2016) show that competition may induce media to differentiate from each other by specializing on different policy dimensions, and the more granular information that they provide has counterproductive effects on voters' behavior. This insight is related to our results in Subsection 4.2, but they only consider the supply side of the media market, and policy platforms are taken as given. Studying how individuals choose to pay attention to information provided by others (media, lobbies or political parties), how the suppliers of information compete for attention, and how this interacts with electoral competition, is an important area for future research.

In this paper we have focused on forward looking voting, in the course of electoral campaigns. Voters also vote retrospectively, however, reacting ex post to the incumbent's behavior. A large theoretical and empirical literature on electoral accountability has focused on this aspect of elections (see Persson and Tabellini 2000, Besley 2007). These contributions generally assume that voters' information, although incomplete, is exogenous. Endogenizing what voters pay attention to, in a framework of retrospective voting and where policy is manipulated by the incumbent so as to hide or attract attention, is likely to yield other novel insights. More generally, rational inattention could shed light on when voters behave retrospectively, when they pay attention to proposed new policies, and when to candidates' valence. This could help integrate several strands of literature in political economy.[39]

---

[39]Diermeier and Li (2015) study electoral control by behavioral and non-strategic voters. Prato and Wolton (2017) study a signalling model where voters' attention can endogenously be high or low. Bandyopadhyay et al. (2018) study under what conditions a challenger chooses an extremist policy in order to attract the attention of a profit maximizing media.

Finally, in this paper, attention influences policy but the reverse channel (from policy to attention) only occurs through voters' expectations. A previous version considered a setting where implemented policies influence attention by changing the opportunity cost of time. In particular, poverty alleviation programs allow the poor to engage in activities other than mere survival. This makes the poor more attentive and hence more influential, which increases the likelihood of pro-poor policies. This complementarity is consistent with empirical findings about the consequences of welfare programs in Latin America (eg. Manacorda et al. 2009), and can give rise to multiple equilibria.

References

Achen, C. and L. Bartels (2004), "Blind Retrospection: Electoral Responses to Drought, Flu and Shark Attacks", mimeo, Princeton University.

Alesina, Alberto and Alex Cukierman (1990), "The Politics of Ambiguity", *Quarterly Journal of Economics,* 105, 829-850

Ansolabehere, Stephen D., James M. Snyder, Jr., and Charles Stewart III. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45:136-159.

Ash Elliott, Massimo Morelli and Richard van Weelden (2015), "Election and Divisiveness : Theory and Evidence", Bococni University, mimeo.

Bandyopadhyay, Siddharta, Kalyan Chatterjee and Jaideep Roy (2018), "Triggering Extremist Platforms: Political Consequences of Profit-Seeking Media", mimeo, University of Bath

Bartels, Larry (1996), "Uninformed Voters: Information Effects in Presidential Elections", *American Journal of Political Science,* Vpl. 40 N. 1, February, 194-230

Bartoš, Vojtěch, Michal Bauer, Julie Chytilova, and Filip Matějka (2016), "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review* (2016), 106(6): 1437-75.

Besley, Timothy (2007), "Principled Agents? The Political Economy of Good Government," *The Lindahl Lectures*, Oxford University Press.

Bordalo, P., N. Gennaioli and A. Shleifer (2013), "Salience and Consumer Choice", *Journal of Political Economy*, October

Bordignon, Massimo, Veronica Grembi, and Santino Piazza (2010), "Who do you blame in local finance? Analysis of municipal financing in Italy," CESifo Working Paper N. 3100.

Boxell, Levi, Matthew Gentzkow and Jesse M. Shapiro (2017) "Greater Internet Use is Not Associated with Faster Growth in Political Polarization among US Demographic Groups", *Proceedings of the National Academy of Sciences.* 114(40). October 3

Cabral, Marika, and Caroline Hoxby (2012), "The hated property tax: Salience, tax rates, and tax revolts," NBER Working Paper 18514.

Caplin, Andrew and Mark Dean (2015), "Revealed Preference, Rational Inattention, and Costly Information Acquisition", *American Economic Review*, 105 (7), 2183-2203.

Carpini, Delli, Michael X., and Scott Keeter (1996), "What Americans Know about Politics and Why It Matters," Yale University Press.

Chetty, Ray, Adam Looney, and Kory Kroft (2009), "Salience and Taxation: Theory and Evidence," *American Economic Review*, 99(4), 1145-1177.

Coate, Stephen (2004), "Political Competition with Campaign Contributions and Informative Advertising," *Journal of the European Economic Association*, 2(5), 772-804.

Coate, Stephen and Michael Conlin (2004), "A Group Rule: Utilitarian Approach to Voter Turnout: Theory and Evidence " *The American Economic Review*, 94(5) pp. 1476-504

Congdon, William J., Jeffrey R. Kling, and Sendhil Mullainathan (2011), "*Policy and Choice: Public Finance through the Lens of Behavioral Economics*," Brookings Institution Press.

Della Vigna, Stefano (2010), "Persuasion: Empirical Evidence", *Annual Review of Economics*, 2:643–69

Della Vigna, Stefano, John List, Ulrike Malmendier and Gautam Rao (2015), "Voting to Tell Others", Berkeley, mimeo

Diermeier, Daniel and Christopher Li (2015), "Electoral Control with Behavioral Voters", University of Chicago, mimeo

Dollery, Brian E., and Andrew C .Worthington (1996), " The Empirical Analysis of Fiscal Illusion," *Journal of Economic Surveys*, 10(3), 261-97.

Downs, Anthony (1957), "*An Economic Theory of Democracy*", New York, Harper

Enikolopov, R., M. Petrova and E. Zhuravskaya (2011), "Media and Political Persuasion: Evidence from Russia"; *American Economic Review*, 101: 3253-85

Feddersen, Timothy, and Alvaro Sandroni (2006) "A Theory of Participation in Elections." *American Economic Review*, 96(4): 1271-1282.

Finkelstein, Amy (2009), "EZ Tax: Tax Salience and Tax Rates," *Quarterly Journal of Economics*, 124(3), 969-1010.

Fiorina, Morris (1973). "Electoral Margins, Constituency Influence, and Policy Moderation: A Critical Assessment." *American Politics Quarterly* 1:479-498.

Fiorina, Morris with Samuels Abrahms (2009) "*Disconnect - The Breakdown of Representation in American Politics*" , University of Oklhaoma Press, Norman

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg (2006):, "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model," *The American*

Economic Review, 96, 1043–1068.

Gavazza, Alessandro, and Alessandro Lizzeri (2009), "Transparency and Economic Policy," *Review of Economic Studies Limited*, 76, 1023–1048.

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

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson (2011), "The effect of newspaper entry and exit on electoral politics." *The American Economic Review 101*, no. 7 (2011): 2980-3018.

Glaeser, Edward L and Ponzetto, Giacomo AM and Shapiro, Jesse M (2005), "Strategic Extremism: Why Republicans and Democrats Divide on Religious Values," *Quarterly Journal of Economics* , 120(4), 1283-1330.

Groseclose, Tim (2001), "A Model of Candidate Location When One Candidate Has a Valence Advantage", *American Journal of Political Science*, 45 (4), 862-886

Hillygus D. Sunshine and Todd G. Shields (2008), "*The Persuadable Voter: Wedge Issues in Presidential Campaigns*", Princeton University Press, Princeton

Høj Jens, Vincenzo Galasso, Giuseppe Nicoletti, Thai-Thanh Dang (2006), "The Political Economy of Structural Reform: Empirical Evidence from OECD Countries", *OECD*, Working Paper No. 501.

Kingdon, J.W. (1984), "Agendas." *Alternatives, and Public Policies* 45.

Ledyard, John O. (1984), "The Pure Theory of Large Two Candidate Elections." *Public Choice*, 44, 7-41.

Levy, Gilat, and Ronny Razin (2015) "Does Polarization of Opinions Lead to Polarization of Platforms? The case of Correlation Neglect", *Quarterly Journal of Political Science* 10(3): 321-55, September

Lindbeck, A., & Weibull, J. W. (1987). Balanced-budget redistribution as the outcome of political competition. *Public choice*, 52(3), 273-297.

Lupia, Arthur, and Mathew D. McCubbins (1998), "The Democratic Dilemma. Can Citizens Learn What They Need to Know?", Cambridge University Press.

Mackowiak, Bartosz, and Mirko Wiederholt (2009), "Optimal Sticky Prices under Rational Inattention," *The American Economic Review*, 99(3), 769-803.

Manacorda, Marco, Edward Miguel, and Andrea Vigorito (2011), "Government Transfers and Political Support," *American Economic Journal: Applied Economics*, 3(3), 1-28.

Martinelli, César (2006) "Would Rational Voters Acquire Costly Information?," *Journal of Economic Theory*, 129(1), 225–251.

Matějka, Filip, and Alisdair McKay (2015), "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *The American Economic Review*, 105(1), 272-98.

McKelvey, Richard D., and Thomas R. Palfrey (1995), "Quantal response equilibria for normal form games," *Games and economic behavior*, 10(1), 6–38.

Van Nieuwerburgh, Stijn, and Laura Veldkamp (2009), "Information immobility and the home bias puzzle," *The Journal of Finance*, 64(3), 1187-1215.

Nunnari, Salvatore, and Jan Zapal (2017), "A Model of Focusing in Political Choice ", mimeo.

OECD (2012), "Economic Policy Reforms 2012: Going for Growth", *OECD*, Paris.

Ortoleva, Pietro and Eric Snowberg (2015), "Overconfidence in Political Behavior", *American Economic Review* 105(2): 504-35

Page, Benjamin I., and Robert Y. Shapiro (1992), "*The rational public,*" The university of Chicago Press.

Palfrey, Thomas R., and Keith T. Poole (1987), "The Relationship between Information, Ideology, and Voting Behavior," *American Journal of Political Science*, 31(3), 511-530.

Perego,Jacopo, and Sevgi Yuksel (2016), "Media Competition and the Source of Disagreement", mimeo, NYU

Persico, Nicola (2003), "Committee Design with Endogenous Information," *Review of Economic Studies*, 70, 1–27.

Persson, Torsten, and Guido Tabellini (2000), "Political economics – Explaining economic policy," MIT Press.

Ponzetto, Giacomo A. M. (2011), "Heterogeneous Information and Trade Policy", CEPR Discussion Papers n. 8726.

Prat, Andrea (2006), "Rational Voters and Political Advertising," *Oxford Handbook of Political Economy* (eds. Barry Weingast and Donald Wittman), Oxford University Press.

Prat, Andrea and David Stromberg (2013), "The Political Economy of Mass Media", in: *Advances in Economics and Econometrics*, edited by Daron Acemoglu, Manuel Arellano and Eddie Dekel, Cambridge University Press

Prato, Carlo and Wolton, Stephane (2017), "Rational ignorance, populism, and reform", *European Journal of Political Economy.* ISSN 0176-2680

Prior, Markus. (2007) "*Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*" Cambridge University Press, Cambridge.

Poole, Keith T., and Howard Rosenthal (2001) "D-Nominate after 10 Years: A Comparative Update to Congress: A Political-Economic History of Roll-Call Voting" *Legislative Studies Quarterly* 26 (1):5–29.

Selten, Reinhard. (1975), "Reexamination of the perfectness concept for equilibrium

points in extensive games," *International journal of game theory*, 4(1), 25–55.

Sims, Christopher A (2003), "Implications of rational inattention," *Journal of monetary Economics*, 50.3 (2003): 665-690.

Snyder Jr, James M. and Ting, Michael M. (2002), "An informational rationale for political parties," *American Journal of Political Science*, 90–110.

Stromberg, David (2001), "Mass Media and Public Policy", *European Economic Review* 45: 652-663

Stromberg, David (2015), "Media and Politics", *Annual Review of Economics*, 7: 173-205

The Pew Research Center (2007), *"What Americans Know: 1989-2007",* Washington DC 2007

Van Nieuwerburgh, Stijn, and Laura Veldkamp (2009), "Information immobility and the home bias puzzle," *The Journal of Finance*, 64(3), 1187-1215.

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

Yuksel, Sevgi (2014), "Specialized Learning and Political Polarization", mimeo, University of California t Santa Barbara

For Online Publication

7 Appendix

7.1 Perceived welfare

Consider those voters in group J who receive signals with realization of noise $\epsilon^{v,J} = \{\epsilon_A^{v,J}, \epsilon_B^{v,J}\}$. By (3), they are just indifferent between candidates A and B if:

$$\tilde{x}^v = E[U^J(q_A)|s_A^{v,J}] - E[U^J(q_B)|s_B^{v,J}] - \tilde{x} \equiv \tilde{x}_T^{v,J} \tag{22}$$

Thus, $\tilde{x}_T^{v,J}$ is the threshold preference shock in favor of candidate B that defines the "swing voters" in group J. Any voter receiving signals with noise $\epsilon^{v,J}$ votes for A if and only if $\tilde{x}^v \leq \tilde{x}_T^v$. Note that each group has a distribution of swing voters, corresponding to the distribution of the noise $\epsilon^{v,J}$. Define the "average swing voter" in group J as $E_\epsilon^J[\tilde{x}_T^{v,J}]$, where the expectation $E_\epsilon^J[\cdot]$ is over realizations of noise $\epsilon^{v,J}$. Then, for given announced policies q_A and q_B, exploiting the assumption that $\tilde{x}^v$ has the same uniform distribution in each group, we can express the vote share of candidate A as:

$$\pi_A = \sum_J m^J E_\epsilon^J[\Pr(\tilde{x}^v \leq \tilde{x}_T^{v,J})] = \frac{1}{2} + \phi \sum_J m^J E_\epsilon^J[\tilde{x}_T^{v,J}] \tag{23}$$

Note that (23) holds when the noise in the ideological preference shocks $\tilde{x}^v$ is sufficiently large to affect the vote with positive probability.[40]

By (22)-(23), the vote share π_A is a linear function of the popularity shock $\tilde{x}$. Since the latter is also uniformly distributed, the probability of winning for candidate A is then:

$$p_A = \frac{1}{2} + \psi \left(\sum_J m^J E_{\epsilon,q_A,q_B}^J \left[E[U^J(q_A)|s_A^{v,J}] - E[U^J(q_B)|s_B^{v,J}] \right] \right) \tag{24}$$

Obviously, $p_B = 1 - p_A$. Again, this holds if the support of the popularity shock $\tilde{x}$ is sufficiently large relative to the RHS of (6), which in a symmetric equilibrium will always be true.

---

[40]This holds for all $\{J, \epsilon^{v,J}, q_A, q_B\}$ and $\tilde{x}$ for which

$$\left(E[U^J(q_A)|\epsilon_A^{v,J}] - E[U^J(q_B)|\epsilon_B^{v,J}] - x^v \right)$$

can be both positive and negative depending on $\tilde{x}^v$, i.e., for which the support of uniformly distributed preference shocks is sufficiently large to affect the vote of v with positive probability. With increasing support of this noise the measure of such cases potentially affected by $\tilde{x}^v$ approaches one.

7.2 Small noise approximations or quadratic utility

Proof of Proposition 1: We will express derivatives of the candidate's objective (7) with respect to $\hat{q}_C$, which are then weighted by masses m^J.

Let $\tilde{U}^J$ denote the second-order approximation to U^J around $\bar{q}_C$.

$$\tilde{U}^J(q_C) \simeq U^J(\bar{q}_C) + \sum_{i=1}^{M} u_{C,i}^J(q_{C,i} - \bar{q}_{C,i}) + \frac{1}{2} \sum_{i,j=1}^{M,M} u_{C,i,j}^J(q_{C,i} - \bar{q}_{C,i})(q_{C,j} - \bar{q}_{C,j}),$$

where $u_{C,i}^J$ and $u_{C,i,j}^J$ are the first and second derivatives of $U^J(q_C)$; both evaluated at $\bar{q}_C$. Voter's expected utility conditional on posterior beliefs is:

$$
\begin{aligned}
E[U^J(q_C)|s_C^{v,J}] &\simeq E[\tilde{U}^J(q_C)|s_C^{v,J}] = \\
&= U^J(\bar{q}_C) + \sum_{i=1}^{M} u_{C,i}^J(\check{q}_{C,i} - \bar{q}_{C,i}) \\
&\quad + \frac{1}{2} \sum_{i,j=1}^{M,M} u_{C,i,j}^J E\Big[(q_{C,i} - \bar{q}_{C,i})(q_{C,j} - \bar{q}_{C,j})|s_C^{v,J}\Big],
\end{aligned}
\tag{25}
$$

where $\check{q}_c$ is the vector of posterior means $E[q_C|s_C^{v,J}]$. The last term can be written as:

$$
\begin{aligned}
&\frac{1}{2} \sum_{i,j=1}^{M,M} u_{C,i,j}^J E\Big[\big((q_{C,i} - \check{q}_{C,i}) - (\bar{q}_{C,i} - \check{q}_{C,i})\big)\big((q_{C,j} - \check{q}_{C,j}) - (\bar{q}_{C,j} - \check{q}_{C,j})\big)|s_C^{v,J}\Big] \\
&= \frac{1}{2} \sum_{i,j=1}^{M,M} u_{C,i,j}^J(\check{q}_{C,i} - \bar{q}_{C,i})(\check{q}_{C,j} - \bar{q}_{C,j}) + \frac{1}{2} \sum_{i=1}^{M} u_{C,i,i}^J(1 - \xi_{C,i})\sigma_{C,i}^2.
\end{aligned}
\tag{26}
$$

This is because elements of noise in beliefs $(q_{C,i} - \check{q}_{C,i})$ about the posterior means are independent from each other as well as from anything else. The second term on the RHS is variance of $(q_{C,i} - \check{q}_{C,i})$, i.e., posterior variance, which equals $(1 - \xi_{C,i})\sigma_{C,i}^2$.

We use $\check{q}_{C,i} = \xi_{C,i}^J s_{C,i}^{v,J} + (1 - \xi_{C,i}^J)\bar{q}_{C,i}$ to express $E_{\epsilon,e}[\cdot]$ of the first term on the RHS of (26), which is

$$
\begin{aligned}
&\frac{1}{2} E_{\epsilon,e}\Big[\sum_{i,j=1}^{M,M} u_{C,i,j}^J \xi_{C,i}^J \xi_{C,j}^J(\hat{q}_{C,i} + e_i + \epsilon_{C,i}^J - \bar{q}_{C,i})(\hat{q}_{C,j} + e_j + \epsilon_{C,j}^J - \bar{q}_{C,j})\Big] \\
&= \frac{1}{2} \sum_{i=1}^{M} u_{C,i}^J(\xi_{C,i}^J)^2\Big(\sigma_{C,i}^2 + \frac{1 - \xi_{C,i}^J}{\xi_{C,i}^J}\sigma_{C,i}^2\Big) \\
&\quad + \frac{1}{2} \sum_{i,j=1}^{M,M} u_{C,i,j}^J \xi_{C,i}^J \xi_{C,j}^J(\hat{q}_{C,i} - \bar{q}_{C,i})(\hat{q}_{C,j} - \bar{q}_{C,j}),
\end{aligned}
\tag{27}
$$

where $\frac{1-\xi^J_{C,i}}{\xi^J_{C,i}}\sigma^2_{C,i}$ is the variance of $\epsilon^J_{C,i}$. Putting (25)-(27) together, we get

$$
\begin{aligned}
E_{\epsilon,e}\Big[E[U^J(q_C)|s^{v,J}_C]\Big|\hat{q}_C\Big] &\simeq U^J(\bar{q}_C) + \sum_{i=1}^M \xi^J_{C,i} u^J_{C,i}(\hat{q}_{C,i}-\bar{q}_{C,i}) + \frac{1}{2}\sum_{i=1}^M u^J_{C,i,i}\sigma^2_{C,i} \\
&\quad + \frac{1}{2}\sum_{i,j=1}^{M,M} u^J_{C,i,j}\xi^J_{C,i}\xi^J_{C,j}(\hat{q}_{C,i}-\bar{q}_{C,i})(\hat{q}_{C,j}-\bar{q}_{C,j}). \qquad (28)
\end{aligned}
$$

Therefore, derivative of the RHS of (28) with respect to $\hat{q}_{C,i}$, evaluated at the equilibrium $\hat{q}_C = \bar{q}_C$, is

$$
\frac{\partial E^J_{\epsilon,e}\Big[E[U^J(q_C)|s^{v,J}_C]\Big|\hat{q}_C\Big]}{\partial\hat{q}_{C,i}}\Bigg|_{\hat{q}_C=\bar{q}_C} \simeq \xi^J_{C,i} u^J_{C,i}.
$$

Weighting this by m^J, we get (7) ∎

Proof of Lemma 2: The voter maximizes the expectation of $\max_{C\in\{A,B\}} E[U^{v,J}_C(q_C)|s^{v,J}_C]$ less the cost of information, see (4). The objective can be rewritten:

$$
\begin{aligned}
E\Big[\max_{C\in\{A,B\}} E[U^{v,J}_C(q_C)|s^{v,J}_C]\Big] - \text{cost of info} &= \frac{1}{2}E\Big[E[U^{v,J}_A(q_A)|s^{v,J}_A] + E[U^{v,J}_B(q_B)|s^{v,J}_B]\Big] + \\
&\quad + \frac{1}{2}E\Big[\Big|E[U^{v,J}_A(q_A)|s^{v,J}_A] - E[U^{v,J}_B(q_B)|s^{v,J}_B]\Big|\Big] - \\
&\quad - \text{cost of info}. \qquad (29)
\end{aligned}
$$

The inner expectations are over realized posterior beliefs. The outer expectations are over all realizations of q_C, noise in signals and preference shocks.

Using similar steps in the proof of Proposition 1 and imposing $\hat{q}_C = \bar{q}_C$, the second-order approximation of the first term on the RHS of (29) yields:

$$
\begin{aligned}
&\frac{1}{2}E\Big[\sum_{C\in\{A,B\}} E[U^{v,J}_C(q_C)|s^{v,J}_C]\Big] \\
&\simeq \frac{1}{2}E\Big[\sum_{C\in\{A,B\}} E[U^{v,J}_C(\bar{q}_C) + \sum_{i=1}^M u^J_{C,i}(q_{C,i}-\bar{q}_{C,i}) + \frac{1}{2}\sum_{i,j=1}^{M,M} u^J_{C,i,j}(q_{C,i}-\bar{q}_{C,i})(q_{C,j}-\bar{q}_{C,j})|s^{v,J}_C]\Big] \\
&= \frac{1}{2}\sum_{C\in\{A,B\}} \Big(U^J(\bar{q}_C) + \frac{1}{2}\sum_{i,j=1}^{M,M} u^J_{C,i,j}E\Big[E\Big[\Big((q_{C,i}-\check{q}_{C,i})-(\bar{q}_{C,i}-\check{q}_{C,i})\Big) \\
&\qquad\qquad \Big((q_{C,j}-\check{q}_{C,j})-(\bar{q}_{C,j}-\check{q}_{C,j})\Big)|s^{v,J}_C\Big]\Big]\Big) \\
&= \frac{1}{2}\sum_{C\in\{A,B\}} \Big(U^J(\bar{q}_C) + \frac{1}{2}\sum_{i=1}^M (u^J_{C,i,i}\xi_{C,i}\sigma^2_{C,i} + u^J_{C,i,i}(1-\xi_{C,i})\sigma^2_{C,i})\Big) \\
&= \frac{1}{2}\sum_{C\in\{A,B\}} \Big(U^J(\bar{q}_C) + \frac{M}{2}u^J_{C,i,i}\sigma^2_{C,i}\Big) \qquad (30)
\end{aligned}
$$

49

In the second to last step we use the fact that variance of $(q_{C,i} - \check{q}_{C,i})$, i.e., posterior variance, equals $(1 - \xi_{C,i})\sigma^2_{C,i}$, and also that variance of posterior means, $(\check{q}_{C,i} - \bar{q}_{C,i})$, is $\xi_{C,i}\sigma^2_{C,i}$ (also see footnotes 6 and 12). We also use independence of noise across instruments. Note that unlike in the proof of Proposition 1, $\hat{q}_C$ does not enter these expressions, since voters condition on their beliefs only.

(30) is independent of ξ^J, and thus the voter's choice of attention is thus given by the maximization of the expectation of only:

$$\frac{1}{2}\Delta^v = \frac{1}{2}\Big(E[U^{v,J}_A(q_A)|s^{v,J}_A] - E[U^{v,J}_B(q_B)|s^{v,J}_B] \Big) \tag{31}$$

less the cost of information. Let

$$\Delta = E[U^J(q_A)|s^{v,J}_A] - E[U^J(q_B)|s^{v,J}_B] = \Delta^v + x^v$$

denote the difference in expected utilities after signals are received, but before the preference and popularity shocks are realized.

Since x^v is the sum of two independent and uniformly distributed random variables, its p.d.f $f(x)$ is continuous and symmetric. Conditional on Δ, expectation of $|\Delta^v|$ is (with $\Delta > 0$):

$$
\begin{aligned}
\int_{-\infty}^{\infty} f(x)|\Delta - x|dx &= \int_{-\infty}^{\Delta} f(x)(\Delta - x)dx - \int_{\Delta}^{\infty} f(x)(\Delta - x)dx \\
&= \Delta\Big(\int_{-\infty}^{\Delta} f(x)dx - \int_{\Delta}^{\infty} f(x)dx \Big) + \\
&\quad + \Big(-\int_{-\infty}^{\Delta} f(x)xdx + \int_{\Delta}^{\infty} f(x)xdx \Big) \\
&= \Delta\int_{-\Delta}^{\Delta} f(x)dx + 2\int_{\Delta}^{\infty} f(x)xdx. \tag{32}
\end{aligned}
$$

In the last step we use symmetry of $f(x)$, which also implies $\int_{-\Delta}^{\Delta} f(x)xdx = 0$ and $\int_{-\infty}^{-\Delta} f(x)xdx = -\int_{\Delta}^{\infty} f(x)xdx$.

Now, when Δ is very small relative to the size of the bulk of the support of x:

$$
\begin{aligned}
\Delta\int_{-\Delta}^{\Delta} f(x)dx &\simeq 2f(0)\Delta^2, \\
2\int_{\Delta}^{\infty} f(x)xdx &= 2\int_{0}^{\infty} f(x)xdx - 2\int_{0}^{\Delta} f(x)xdx \simeq E_f[|x|] - f(0)\Delta^2. \tag{33}
\end{aligned}
$$

Therefore, conditional on Δ, the expectation of $|\Delta^v|$ equals $(E_f[|x|] + f(0)\Delta^2)$. Now we just need to express the unconditional expectation of Δ^2, i.e., of the square of difference

between expected utilities from the two candidates after signals are acquired, evaluated at $\hat{q}_C = \bar{q}_C$.

Using the second order approximation, and manipulations similar to those in (26), we get:

$$
\begin{aligned}
\Delta \;\simeq\;& U^J(\bar{q}_A) - U^J(\bar{q}_B) + \sum_{i=1}^{M}\Big(u_{A,i}^J(\check{q}_{A,i} - \bar{q}_{A,i}) - u_{B,i}^J(\check{q}_{B,i} - \bar{q}_{B,i})\Big) \\
&+ \frac{1}{2}\sum_{i=1}^{M}\Big(u_{A,i,i}^J((\check{q}_{A,i} - \bar{q}_{A,i})^2 + (1 - \xi_{A,i}^J)\sigma_{A,i}^2) - u_{B,i,i}^J((\check{q}_{B,i} - \bar{q}_{B,i})^2 \qquad (34) \\
&+ (1 - \xi_{B,i}^J)\sigma_{B,i}^2)\Big). \qquad (35)
\end{aligned}
$$

Finally, to express $E[\Delta^2]$, we get to more tedious algebra. The first three terms of the following are expectations of the terms in (34) squared, the last term is expectation of a product of the first and the third terms.

$$
\begin{aligned}
E[\Delta^2] \;\simeq\;& \Big(U^J(\bar{q}_A) - U^J(\bar{q}_B)\Big)^2 + \sum_{i=1,C\in\{A,B\}}^{M} \xi_{C,i}^J(u_{C,i}^J)^2\sigma_{C,i}^2 \\
&+ \frac{1}{4}E\Bigg[\Big(\sum_{i=1}^{M}u_{A,i,i}^J((\check{q}_{A,i} - \bar{q}_{A,i})^2 + (1 - \xi_{A,i}^J)\sigma_{A,i}^2) - u_{B,i,i}^J((\check{q}_{B,i} - \bar{q}_{B,i})^2 + (1 - \xi_{B,i}^J)\sigma_{B,i}^2)\Big)^2\Bigg] \\
&+ \Big(U^J(\bar{q}_A) - U^J(\bar{q}_B)\Big)\Big(\sum_{i=1}^{M}u_{A,i,i}^J\sigma_{A,i}^2 - u_{B,i,i}^J\sigma_{B,i}^2\Big). \qquad (36)
\end{aligned}
$$

The term with expectation equals $\frac{1}{4}$ times

$$
\begin{aligned}
&-\; 2\sum_{i,j=1}^{M,M} u_{A,i,i}^J u_{B,j,j}^J \sigma_{A,i}^2\sigma_{B,j}^2 + 2\sum_{i,j=1,C\in\{A,B\}}^{M,M} u_{C,i,i}^J u_{C,j,j}^J \xi_{C,i}^J(1 - \xi_{C,j}^J)\sigma_{C,i}^2\sigma_{C,j}^2 \\
&+ \sum_{i,j=1,C\in\{A,B\}}^{M,M} u_{C,i,i}^J u_{C,j,j}^J(1 - \xi_{C,i}^J)(1 - \xi_{C,j}^J)\sigma_{C,i}^2\sigma_{C,j}^2 \qquad (37) \\
&+ \sum_{i,j=1,C\in\{A,B\}}^{M,M} u_{C,i,i}^J u_{C,j,j}^J \xi_{C,i}^J\xi_{C,j}^J\sigma_{C,i}^2\sigma_{C,j}^2 + 2\sum_{i=1,C\in\{A,B\}}^{M} (u_{C,i,i}^J)^2(\xi_{C,i}^J)^2(\sigma_{C,i}^2)^2 \\
=\;& -2\sum_{i,j=1}^{M,M} u_{A,i,i}^J u_{B,j,j}^J \sigma_{A,i}^2\sigma_{B,j}^2 + \sum_{i,j=1,C\in\{A,B\}}^{M,M} u_{C,i,i}^J u_{C,j,j}^J \sigma_{C,i}^2\sigma_{C,j}^2 \\
&+ 2\sum_{i=1,C\in\{A,B\}}^{M} (u_{C,i,i}^J)^2(\xi_{C,i}^J)^2(\sigma_{C,i}^2)^2. \qquad (38)
\end{aligned}
$$

The first term on the LHS of (37) is the product of all terms associated with A and all

associated with B, the second is a product of terms with $(\check{q}_{C,i} - \bar{q}_{C,i})^2$ and those with $(1 - \xi_{C,i}^J)\sigma_{C,i}^2$, the third is product of between terms with $(1 - \xi_{C,i}^J)\sigma_{C,i}^2$, the forth and fifth are product of the terms including $(\check{q}_{C,i} - \bar{q}_{C,i})^2$ and $(\check{q}_{C,j} - \bar{q}_{C,j})^2$, and the last term being a correction of the forth one for $i = j$, since if $x \sim N(0, \sigma^2)$, then $E[x^4] = 3(\sigma^2)^2$.

Therefore, putting everything together and omitting constants independent of ξ^J, the objective equivalent to (29) is

$$\frac{f(0)}{2}F(\xi^J) - \text{cost of info},$$

where $f(0) = Min(\psi, \phi)$ given the distributional assumption on $x^v = \tilde{x} + \tilde{x}^\nu$, and

$$F(\xi^J) = \sum_{i=1, C \in \{A,B\}}^{M} \left(\xi_{C,i}^J \sigma_{C,i}^2 (u_{C,i}^J)^2 + 2(\xi_{C,i}^J)^2 (\sigma_{C,i}^2)^2 (u_{C,i,i}^J)^2 \right). \tag{39}$$

For simplicity, in the statement of this Lemma in the text we report the first-order approximation only, and thus include only the first-order term from (39); and we also denote $\hat{\lambda}_{C,i}^J = 2\lambda_{C,i}^J / Min(\psi, \phi)$.

The solution to the voter's maximization problem is then:

$$\xi_{C,i}^J = \max\left(\xi_0, \frac{4\sigma_{C,i}^2 (u_{C,i,i}^J)^2 - (u_{C,i}^J)^2 + \sqrt{(4\sigma_{C,i}^2 (u_{C,i,i}^J)^2 + (u_{C,i}^J)^2)^2 - 16\hat{\lambda}_{C,i}^J (u_{C,i,i}^J)^2}}{8\sigma_{C,i}^2 (u_{C,i,i}^J)^2} \right). \tag{40}$$

∎

8 Voters' preference shocks

In this section we study an alternative foundation to voters' uncertainty about policies. Here, voters are subject to preference shocks which are only imperfectly observed by politicians when they make the policy decision. There is no policy implementation error and candidates are opportunistic and maximize the probability of winning. Nevertheless policies are random, because they depend on candidates' noisy perceptions of voters' preferences. Voters observe their own preferences, but do not observe the candidates' noisy perceptions of such preferences.

The sequence of events is as follows:

1. Nature draws an unobservable preference shock η. Voters' utility then takes the form $U^J(q - \eta)$.

2. Candidates receive i.i.d. signals on the preference shocks η, and set policy.

3. Voters observe preference shocks η, form prior beliefs about the policy platforms of each candidate, and choose attention strategies.

4. Voters observe noisy signals of the actual platforms.

5. The ideological bias x^v is realized and elections are held. Whoever wins the election enacts their announced actual policies.

The preference shocks are drawn from a normal distribution, $\eta_i \sim N(0, \sigma_\eta^2)$, and candidates receive i.i.d. signals on them:

$$\varsigma_{C,i} = \eta_i + \varepsilon_{C,i},$$

where $\varepsilon_{C,i}$ is drawn from $N(0, (1-\alpha_C)/\alpha_C \sigma_\eta^2)$. The precision of candidates' signals is given by $\alpha_C \in (0,1)$, which is fixed and plays the same role as ξ^J for voters' attention. Candidate's belief about η_i upon acquisition of the signal $\varsigma_{C,i}$ (before choosing the policy) is:

$$\eta_i \sim N(\alpha_C \varsigma_{C,i}, (1-\alpha)\sigma_\eta^2).$$

The voters' problem is now only slightly different from (4). The only novelty is that now preference shocks η enter the utility function and beliefs are conditioned on η:

$$\max_{\xi^J \in [\xi_0, 1]^{2M}} E\left[max_{C \in \{A,B\}} E[U_C^{v,J}(q_C - \eta)|s_C^{v,J}, \eta]\Big|\eta \right] + \sum_{C \in \{A,B\}, i \leq M} \lambda_{C,i}^J log\left(1 - \xi_{C,i}^J\right). \tag{41}$$

The candidates' problem is also almost the same as in the baseline model. Maximizing the probability of winning the elections is equivalent to maximizing the perceived social welfare.

$$\max_{q_C \in \mathbb{R}^M} \sum_J m^J E\left[E[U^J(q_C - \eta)|s_C^{v,J}, \eta]\Big|q_C, \varsigma_C \right]. \tag{42}$$

The difference from the problem (7) is that here the candidate conditions on the signal on η, and the resulting uncertainty about perceived welfare is not driven by policy implementation errors, but by uncertainty about η. Conditional on politicians' beliefs about η, given by the signal ς_C, the politician chooses directly the policy vector q_C (there is no implementsation error).

We will again study the small noise approximation. Below we describe the first order approximation. The second order approximation is available upo request. We look for candidates' strategies in the class of policies that are linear in the expected preference shocks:

$$q_C = q_C^0 + \beta_C \cdot E[\eta|\varsigma_C], \tag{43}$$

where $q_C^0 \in \mathbb{R}^M$ is the policy vector selected when $E[\eta|\varsigma_C] = 0$ and $\beta_C \in \mathbb{R}^M$ is a vector of linear coefficients. These policies are local approximations to any continuous policy.

Definition 3 Equilibrium with preference shocks. *The equilibrium is a profile of policy vectors $q_C(\varsigma_C)$, given by q_C^0 and β_C, and of attention strategies $\xi^J(\eta)$ such that: (a) The attention strategies $\xi^J(\eta)$ solve the voters' problem (41), where their prior beliefs are consistent with candidates' strategies, namely, conditional on the observed η, the prior $q_{C,i}$ is normally distributed with mean $q_{C,i}^0 + \beta_{C,i}\alpha_C\eta$ and variance $\beta_{C,i}^2 \alpha_C(1-\alpha_C)\sigma_\eta^2$. (b) For any ς_C, the policy vector $q_C(\varsigma_C)$ maximizes the probability of winning of candidate C, taking as given the attention profiles chosen by the voters and the policy profiles chosen by his opponent, i.e., solves (42).*

Proposition 3 *The equilibrium attention strategies $\xi^J(\eta)$ are given by (10) in Proposition 2. The equilibrium policies of the form of (43) are such that $\beta_{C,i} = 1/\alpha_C$ and the policy q_C^0 for $E[\eta|\varsigma^C] = 0$ solves*

$$\sum_{J=1}^{N} m^J \xi_{C,i}^J u_{C,i}^J = 0 \qquad \forall i, \tag{44}$$

where $\xi_{C,i}^J$ are the equilibrium attention weights at $\eta = 0$.

The proposition implies for the first order approximations (i.e., very small noise):

$$q_C = q_C^0 + E[\eta|\varsigma^C]/\alpha_C.$$

Candidates that are highly uninformed (e.g., don't work with experts) move their equilibrium policies with their beliefs more (i.e., $\beta_{C,i}$ is high). On the other hand, their beliefs respond to the preference shocks less, because they do not put much weight on the noisy signals that they receive (i.e., α_C is low). In expectation, the equilibrium policy change due to η is independent of the level of candidate's precision of signals.

$$E[q_C|\eta] = q_C^0 + \eta.$$

However, this does not mean that the distribution of movements, or the noise in policies are independent of the precision. Conditional on the realized η, voters' prior beliefs are:

$$\bar{q}_{C,i} = q_{C,i}^0 + \eta, \qquad \sigma_{C,i}^2 = \frac{1-\alpha_C}{\alpha_C}\sigma_\eta^2. \tag{45}$$

Thus, the variance of policies of the less informed candidates is larger. They respond strongly to their posterior expectations, which are driven also by the noise in signals.

Voters thus face higher uncertainty about these candidates, and pay more attention to their policies.

The FOC for $\xi^J(\eta)$ and for q_C^0 are the same as in the baseline model, but here the variance of prior beliefs is endogenous since it depends on β_C. Notice that the policy response to beliefs about preference shocks, $\beta_{C,i}$, is greater than one, and so can $\beta_{C,i}^2 \alpha_C(1-\alpha_C)$, i.e., policies can have higher variance than the fundamental shocks.

8.1 Discussion

While the median policy is determined similarly to the baseline model, (44), there are also some deviations from it. The main additional insight is that, as in the model with shocks to candidates' preferences of section 5, policy volatility is endogenous. Uninformed politicians respond to beliefs about η more, they are more volatile, and are then paid more attention.

More specifically, preferences shocks η do not translate one-to-one into policy shocks, $\beta \neq 1$, unlike under perfect information where instead $\beta = 1$. This is because voters know that politicians on average underestimate the preference shocks by a factor of α_C. If the policy response to the belief about η_i were one-to-one ($\beta_{C,i} = 1$), then voters would adjust their attention as a response to the observed η as if the policy shock were $\alpha_C \eta_i$ and the candidate's FOCs wouldn't hold (the resulting effect on preferences - the $q - \eta$ term - would be as if policy changed by $(\alpha_C - 1)\eta_i$ and there were no preference shock). The politician thus knows that on average attention adjusts as if voters expected a policy change of the opposite sign, and thus the politician is more motivated to move in the direction of the shock η. As a result:

$$q = q^0 + \frac{E[\eta]}{\alpha_C}.$$

Candidates who receive more imprecise signals (lower α_C) respond to the changes in their beliefs more. While posterior beliefs of such candidates have lower variance, the resulting noise in their policy choices is larger (because they scale up the already high noise in their signals); it is $\frac{1-\alpha_C}{\alpha_C}$. Uninformed politicians choose to move more, are less predictable, and draw more attention.

Proof of Proposition 3: We will now derive the FOCs for (42) assuming small noise approximations. First, let us express candidate's expectation of voter's perceived utility,

where $\check{q}_C$ is the vector of voter's posterior means.

$$
\begin{aligned}
E\Big[E[U^J(q_C - \eta)|s_C^{v;J}, \eta]\Big|q_C, \varsigma_C\Big] &\cong E\Big[U^J(\check{q}_C - \eta)\Big|q_C, \varsigma_C\Big] \cong U^J\Big(E[\check{q}_C - \eta|q_C, \varsigma_C]\Big) \cong \\
&\cong U^J(q_C^0) + \sum_{i=1}^{M} u_{C,i}^J \cdot \Big(E[\check{q}_{C,i} - \eta_i - q_{C,i}^0|q_C, \varsigma_C]\Big) = \\
&= U^J(q_C^0) + \sum_{i=1}^{M} u_{C,i}^J \cdot \Big(E[\xi_{C,i}^J(\eta) \cdot (q_{C,i} + \epsilon_{C,i}^J) + (1 - \xi_{C,i}^J(\eta))\bar{q}_{C,i} - \eta_i \\
&= \sum_{i=1}^{M} u_{C,i}^J q_{C,i} E[\xi_{C,i}^J(\eta)|\varsigma_C] + \text{terms independent of } q_C \cong \\
&\cong \sum_{i=1}^{M} u_{C,i}^J q_{C,i} \xi_{C,i}^J(E[\eta|\varsigma_C]) + \text{terms independent of } q_C.
\end{aligned}
$$

In the first two steps we apply the insight that expectation of $U(x)$ is approximately equal to $U(\cdot)$ evaluated at the expected x. The third step is the first order approximation of $U(\cdot)$ around q_C^0, where $u_{C,i}^J$ is the first derivative evaluated at this point. In the next step, we express the posterior mean $\check{q}_C$ in terms of the prior mean $\bar{q}_{C,i}$ and the imperfect signal on q_C. Here we emphasize that the voter's attention $\xi_{C,i}^J(\eta)$ depends on the shock η, since it affects voters' prior beliefs about chosen policies and thus the marginal utility at the prior mean $(\bar{q}_C - \eta)$.

The last step uses the same approximation as in the first two steps: expectation of a function is approximated by the function evaluated at the expectation of its argument. Choice of attention $\xi_{C,i}^{\eta,J}$, as in Proposition 2, depends on the voter's prior mean. Voter's prior mean expected by the candidate is:

$$
E[\eta|\varsigma_C] = E[\bar{q}_C - \eta|\varsigma_C] = E[q_C^0 + \alpha_C \cdot \beta_C \eta - \eta|\varsigma_C] = q_C^0 + (\alpha_C \cdot \beta_C - 1)\alpha_C \varsigma].
$$

Differentiating the RHS of (46) by $q_{C,i}$ we get the candidate's FOC for i and ς_C:

$$
\sum_{J=1}^{N} u_{C,i}^J \xi_{C,i}^J(E[\eta|\varsigma_C]) = \sum_{J=1}^{N} u_{C,i}^J \xi_{C,i}^J\Big(q_C^0 + (\alpha_C \cdot \beta_C - 1)\alpha_C \varsigma_C\Big) = 0. \tag{47}
$$

Let $q_{C,i}^0$ be a solution for $E[\eta|\varsigma_C] = \varsigma_C = 0$, which is the analog to the equilibrium target of the baseline model. The marginal utilities $u_{C,i}^J$ are evaluated at q_C^0, where q_C^0 is the unconditional prior mean and thus also the solution for $\varsigma_C = 0$.

Setting $\beta_{C,i} = 1/\alpha_C$ thus implies $(\alpha_C \cdot \beta_C - 1) = 0$. Therefore, because $q_{C,i}^0$ solves (47) for $\varsigma_C = 0$, then policies (43) with $\beta_{C,i} = 1/\alpha_C$ solve the FOC for all ς_C. Conversely, since the FOC must hold for all ς_C, these are the only linear coefficients that solve the FOCs, otherwise $q_{C,i}^0$ would not be locally unique. ∎