Electoral Competition with Rationally Inattentive Voters*

Filip Matějka[†] and Guido Tabellini[‡]

November 2019

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, that divisive issues attract most attention and that 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.

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 has 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'

*We are grateful for comments from Bard Harstad, three anonymous referees, Avi Acharya, Michal Bauer, Nicola Gennaioli, David Levine, Alessandro Lizzeri, Massimo Morelli, Salvo Nunnari, Jakub Steiner, Jim Snyder, David Stromberg, Stephane Wolton, Leeat Yariv, Jan Zápal, and seminar and conference participants at Barcelona GSE, Bocconi University, CSEF-IGIER, CIFAR, Columbia University, , Ecole Polytechnique, Mannheim, NBER, NYU BRIC, NYU Abu Dhabi, NYU, Royal Holloway, Stanford GSB, University of Chicago, University of Oxford, EIEF, CEU, CEPR, ECARES, ETH Zurich, the Political Economy Conference at the University of Bath and the CPB Netherlands Bureau for Economic Policy Analysis.

[†]CERGE-EI, a joint workplace of Charles University in Prague and the Economics Institute of the Czech Academy of Sciences, Politickych veznu 7, 111 21 Prague, Czech Republic; CEPR. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No 678081).

[‡]Department of Economics and IGIER, Bocconi University; CEPR; CES-Ifo; CIFAR

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. The provision of information has been disintermediated (Prior 2007, Sunstein 2017). What we know is determined not so much by what information the media have chosen to provide, but by what and where we search of our own initiative. 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. In contrast, when network television and newspapers were the main sources of political information, instead, it was more difficult to become very well informed about narrow and specific issues; moreover, 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 the 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 maximize the probability of winning and 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. Their 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 the internet), if the cost of information drops, or if the economy is hit by shocks.

We 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 or with more polarized policy views are more informed about the policy positions of presidential candidates (Palfrey and Poole 1987) and of members of Congress (Lauderdale 2013). 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), Ansolobehere et al. (2001) and Stone and Simas (2010).

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 are thus very attentive to detecting possible deviations on these instruments.

We illustrate this point in a model similar to Gavazza and Lizzeri (2009), and show

[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.

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, and we can do comparative statics.

Some features of the equilibria we study are similar to those of models of lobbying, where organized groups with high stakes exert a disproportionate influence over policy, directly through bribes, or indirectly by reducing the cost of information over some issues. Nevertheless there are some important differences. First, lobbying equilibria are generally Pareto efficient (Grossmand and Helpman 2001). With multidimensional policy instruments, instead, equilibria with rationally inattentive voters are generally not Pareto efficient. Second, the empirical implications of who is more influential (groups with the ability to get organized vs groups with low information gathering costs) are obviously different. Third, even models of lobbying through information provision, such as Coate (2004), do not explain why voters pay attention to information provided by the lobbies. Fourth, as discussed in Subsection 4.2, the specific normative implications also differ.

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.[2]

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 by-product 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 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

[2]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. Hu and Li (2018) study electoral competition with rationally inattentive voters in one dimension and with a flexible choice of signals.

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 by-product 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. In our paper, instead, political candidates only care about the probability of winning, and policy divergence occurs if candidates differ in their visibility. 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. In our paper, instead, informational asymmetries emerge as an outcome of the political equilibrium. Voters have rational beliefs of the candidates' equilibrium policy choices, and this determines their demand for specific types of information. As discussed in Subsection 4.2, this also leads to different normative implications.

Our result on policy divergence due to differences in transparency between candidates is related to Glaeser et al (2005). That paper too takes information as exogenous, and assumes that core party supporters are more informed about their own party than about the opponent, in a setting with endogenous turnout. Our model, instead, has the opposite feature: since in equilibrium all voters have a positive probability of voting for either candidate, they pay more attention to the more distant candidate. The specific predictions of our paper are also quite different.[3]

A large theoretical literature studies voters' incentives to collect information and /or vote, 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.

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 despite very small attention costs.

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. 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.[4]

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 concludes. The appendix contains the proofs and shows that our results generalize to a setting where voters' uncertainty reflects learning by the

[3] 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.

[4] 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).

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 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 their 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. Alternatively, one can think of the main political actors over which voters form rational expectations as parties. But policies are implemented by candidates, and parties can make unobserved errors in selecting candidates (Coate 2004 takes a related approach to study campaign advertising). The implementation errors $e_{C,i}$ are independent across candidates C and policy instruments i, and their variance $\sigma^2_{C,i}$ is given exogenously.[5]

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.[6]

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' equilibrium 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

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

[6] The assumption that $\tilde{x}^v$ is realized at the last stage is made just to simplify notation, so that attention strategies of voters are the same within each group.

$\sigma_C^2 = [\sigma_{C,1}^2, ..., \sigma_{C,M}^2]$ the vector of prior variances. Note that, to ensure consistency, the prior variances coincide with the variance of the implementation errors e_C in (2).[7]

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,

$$s_{C,i}^{v,J} = q_{C,i} + \epsilon_{C,i}^{v,J},$$

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.[8]

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.[9] 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

[7] 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 $\bar{\Sigma}$.

[8] 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.

[9] 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.

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 $\lambda_{C,i}^{J}\in R_{+}$ 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.

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.

This formulation 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. Since in a large election the actual probability of being pivotal is close to zero, 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 higher 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 satisfactory theory of voter behavior, this seems the most natural and least arbitrary 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. [10]

Note that, in line with our assumption that voters are motivated by the desire to cast 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.[11]

We further discuss these assumptions on voters' behavior in the next subsection.

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^2_C 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^2_C.*

[10] 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 (9) below would still hold exactly.

[11] 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.

(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 $e_{C,i} \sim N(0, \sigma^2_{C,i})$, which have an exogenously given distribution reflecting the process governing each electoral campaign. And the noise in the policy signals observed by the voters, $\epsilon^{v,J}_{C,i} \sim N(0, \gamma^J_{C,i})$, whose variance $\gamma^J_{C,i}$ corresponds to the chosen level of attention, $\xi^J_{C,i}$. 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. A previous version also considered 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 previous version 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 the noisy signal of the actual policy platforms and vote.[12] The results discussed in the next section generalize almost identically to this setting, with one difference. If policy uncertainty is due to implementation errors, as in the baseline model, then it is entirely exogenous. If instead policy uncertainty reflects shocks to the preferences of voters, then policy volatility is endogenous: it is determined

[12]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).

in equilibrium by how candidates react to such preference shocks.

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 they bother to acquire political information in large elections has been neglected. The standard approach views political information as a by-product 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.

As mentioned above, we assume that voters seek information because they are intrinsically motivated to cast a meaningful vote. 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.

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 exist a unique solution to the voter's attention problem, and attention would be increasing in both stakes and uncertainty. Also, the

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

cost of information should be interpreted as net of its entertainment value, which can explain why petty news often receives more attention than important issues.[14]

We also assume that the cost of information does not depend on whether the candidate announces a policy that is close to or far away from the voter's bliss point. This makes sense with opportunistic candidates who choose similar policies. But with partisan candidates, voters may face lower cost to acquire information about the closer candidate (social media could have this effect). A previous version discussed how this could increase policy divergence in equilibrium.

Finally, we assume throughout that voters' utility is strictly concave. This is important, because it implies that individual welfare is more sensitive to policy changes when the candidate is further away from the voter's bliss point. Although this assumption is reasonable in many contexts, it need not always hold. If voters' utility was linear or convex in some component of the policy vector, some of the specific predictions of the model would be different. We discuss this point in Section 4 with regard to specific applications. Note however that if the utility function was strictly convex, existence of the equilibrium could be problematic or it could entail corner solutions.

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, $\lambda^J_{C,i} = 0$. Then our model boils down to standard probabilistic voting

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

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. Thus, under full information the equilibrium corresponds to the utilitarian optimum.

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 6.1 shows that with inattentive voters and $\lambda^J_{C,i} > 0$, the probability that candidate A wins is:

$$p_A = \frac{1}{2} + \psi \left(\sum_J m^J E^J_{\epsilon,q_A,q_B} \left[E[U^J(q_A)|s^{v,J}_A] - E[U^J(q_B)|s^{v,J}_B] \right] \right) \tag{6}$$

where the outer expectations operator is indexed by J because voters' attention differs 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 $E[U^J(q_A)|s^{v,J}_A]$. The overall probability of winning is then an expectation of this quantity over all realizations of policies and of noise in signals.

In deriving (6), we have assumed that the support of the shocks $\tilde{x}$ and $\tilde{x}^v$ are sufficiently large, so that in a neighborhood of the equilibrium all voters have a positive probability of voting for either candidate. Specifically, we assume throughout that, if an equilibrium exists, then in a neighborhood of the equilibrium policies the following condition is satisfied:

$$\left| E^J_{\epsilon,q_A,q_B} \left[E[U^J(q_A)|s^{v,J}_A] - E[U^J(q_B)|s^{v,J}_B] \right] \right| < min(\frac{1}{2\psi}, \frac{1}{2\phi}) \tag{7}$$

This assumption requires the variance of the implementation shocks $\sigma^2_{C,i}$ to be sufficiently small, and it is more restrictive in the equilibria where there is policy divergence - see subsection 4.1. Note that a similar assumption is also needed under full information: in standard probabilistic voting models, preference shocks need to be sufficiently large so that all groups include voters that in equilibrium vote for either candidate. The assumption that noise in beliefs is small is consistent with the approach taken below. Throughout the rest of the paper we thus assume that condition (7) holds.

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

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

$$\max_{\hat{q}_C \in \mathbb{R}^M} \quad \sum_J m^J E^J_{\epsilon,e}\Big[E[U^J(q_C)|s_C^{v,J}]\Big|\hat{q}_C\Big] \tag{8}$$

In equilibrium, candidate C maximizes the "perceived social welfare" provided by his policies. This 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.

Finally, note that the candidates' objective (8) is a concave function of the realized policy vector q_C. Thus, the equilibrium can be characterized by the first order conditions of the objective (8), since they are necessary and sufficient for an optimum.[15]

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 the utility function is quadratic, in which case the approximation is exact, or if prior uncertainty in beliefs is small, in which case we can use a local first order approximation to the utility function. The distinctive feature of our model is that it studies implications of endogenous imperfect information for outcomes of electoral competition. These approximations emphasize the first-order effects of information imperfection. As shown here, these effects can be highly relevant even if information imperfections are small.

[15] Concavity is implied by the following arguments: 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: $\check{q}_{C,i} = \xi^J_{C,i} s^{v,J}_{C,i} + (1-\xi^J_{C,i})\bar{q}_{C,i}$, where $s^{v,J}_{C,i} = q_{C,i} + \epsilon^{v,J}_{C,i}$ and $\bar{q}_{C,i}$ denotes the prior mean. Thus, $\check{q}_{C,i} = \xi^J_{C,i}(\hat{q}_{C,i} + e_{C,i} + \epsilon^{v,J}_{C,i}) + (1-\xi^J_{C,i})\bar{q}_{C,i}$. 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.

Let us denote by

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

the marginal utility of a change in the i^{th} component of the policy vector for a voter in group J, evaluated at the expected policies. Thus, $u^J_{C,i}$ measures intensity of preferences about $q_{C,i}$ in a neighborhood of the equilibrium.

3.2.1 Candidates' problem.

The Appendix proves the following result:

Proposition 1 *Suppose that (7) holds. For small noise σ^2_C, the first order conditions that define equilibrium policies can be approximated by:*

$$\sum_{J=1}^{N} m^J \xi^J_{C,i} u^J_{C,i} = 0 \qquad \forall i, \tag{9}$$

The proof in fact shows that (9) holds for both first and second order approximations of U, and thus it also holds exactly for quadratic utility functions. The magnitude of the contribution of the higher order terms of the approximation is a function of the higher-order derivatives and scales with variance σ^2_C, and thus vanishes if σ^2_C approaches zero. From now on, we suppose that the noise σ^2_C is sufficiently small so that equilibrium policies are implicitly defined by (9).

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 the expected posterior mean about $q_{C,i}$ in group J changes by $\xi^J_{C,i}\Delta$ only. Thus, while the effect on voters' utility is $\Delta u^J_{C,i}$, the effect on expected, i.e., perceived, utility is only $\xi^J_{C,i}\Delta u^J_{C,i}$.

Several remarks are in order. First, with only one policy instrument, equation (9) is the first order condition for the maximum of a modified social planner's problem, where each group J is weighted by its attention, ξ^J_C. Thus, if all voters paid the same attention, so that $\xi^J_{C,i} = \xi$ for all J, C, i, then the equilibrium coincides with that under full information and hence 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. This implication is similar to those found in the literature that has studied exogenous informational asymmetries between groups, such as Grossman and Helpman (2001).

Second, if policy is multi-dimensional, the attention weights $\xi^J_{C,i}$ in (9) generally also vary by policy instrument i. In this case, equation (9) does not correspond to the first order condition for the maximum of a modified social planner problem. Hence, the equilibrium is not even Pareto efficient. The public good example in subsection 4.2 illustrates this point.

Third, if attention weights also differ by candidate C (besides differing by group J), that is if different voters pay more attention to one candidate and less to the other, then in general the two candidates choose different equilibrium policies.

We summarize this discussion in the following:

Corollary 1 *(i) If $\xi^J_{C,i}$'s are the same for all voters and candidates, i.e., $\xi^J_{C,i} = \xi_i$, then the equilibrium corresponds to the utilitarian optimum. (ii) If $\xi^J_{C,i}$'s vary only across voters, i.e. $\xi^J_{C,i} = \xi^J$, then the equilibrium is Pareto efficient but it does not correspond to the utilitarian optimum. (iii) If $\xi^J_{C,i}$'s vary across J and i, i.e. $\xi^J_{C,i} = \xi^J_i$, then the equilibrium is not Pareto efficient. (iv) If $\xi^J_{C,i}$'s vary across J and C, then in equilibrium the two candidates choose different policies.*

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

3.2.2 Voters' problem.

Let us now focus on the voter's problem. How should costly attention be allocated to alternative components of the policy vector? Consider a first order approximation of U in the voters' optimization problem stated in (4). The Appendix proves:

Lemma 2 *Suppose that (7) holds. For small noise σ^2_i, voters choose attention vectors $\xi^J \in [\xi_0, 1]^M$ to maximize the following (approximated) objective function:*

$$\left(\sum_{C\in\{A,B\},i=1}^{M} \xi^J_{C,i} (u^J_{C,i})^2 \sigma^2_{C,i} \right) + \sum_{C\in\{A,B\},i\leq M} \hat{\lambda}^J_{C,i} log \left(1 - \xi^J_{C,i}\right) \tag{10}$$

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

In the proof we use condition (7) above, saying that the noise σ^2_C and potential a priori divergence between equilibrium policies q_A and q_B are small relative to the support of preference shocks x. If this condition is satisfied then, as shown in the Appendix, the first term in the voters objective function in (10) is exactly $\sum_{C\in\{A,B\},i=1}^{M} \xi_{C,i}^{J} \sigma_{C,i}^{2}((u_{C,i}^{J})^2 + \hat{\delta}_i(\sigma_C^2))$, where $|\hat{\delta}_i|$ is decreasing in the scale of σ_C^2 and equal to zero for $\sigma_C^2 = 0$.

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 (an approximation of) 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.

3.2.3 Implications.

An immediate implication of (10) is the next proposition.

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

$$\xi_{C,i}^{J} = \max\left(\xi_0, 1 - \frac{\hat{\lambda}_{C,i}^{J}}{(u_{C,i}^{J})^2\sigma_{C,i}^2}\right). \tag{11}$$

Thus, the voter pays weakly 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 $q_{C,i}$. Strict monotonicity holds for any positive cost $\hat{\lambda}_{C,i}^{J}$ that is sufficiently large that the constraint $\xi_{C,i}^{J} \geq \xi_0$ does not bind.[16]

This also implies that the attention weights $\xi_{C,i}^{J}$ may differ across candidates, because the cost of information $\lambda_{C,i}^{J}$ or prior uncertainty $\sigma_{C,i}^2$ could differ between the two candidates. If so, the two candidates in equilibrium may 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, which lead

[16]Note that for any convex information-cost function $\Gamma(\xi^J)$, the objective (10) 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. For instance, the effects hold for any cost function that is symmetric across policy elements, i.e., invariant to permutations in ξ^J.

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.

Finally, Lemma 2 also implies that prior knowledge about one candidate does not affect the choice of attention to the other ($\xi^J_{A,i}$ does not depend on the voter's belief about what B does in equilibrium, and vice versa). More generally, the voter's prior belief about his ranking of the two candidates has no influence on how he allocates attention. The reason is that the marginal value of better information about the policy shocks $e_{C,i}$ does not depend on the prior probability of voting for one candidate or the other. This in turn follows from our assumption that shocks to preferences over candidates, $\tilde{x}$ and $\tilde{x}^v$, have a uniform distribution and with a sufficiently large support (so that condition (7) above holds). Also recall that the cost of attention is additively separable.

We summarize this discussion in the following:

Corollary 2 *The attention weights $\xi^J_{C,i}$ that voters in group J give to the policy instrument $q_{C,i}$ set by candidate C have the following properties:*
(i) $\xi^J_{C,i}$ is weakly increasing in voters' stakes $|u^J_{C,i}|$ and decreasing in their unit cost of information $\hat{\lambda}^J_{C,i}$.
(ii) $\xi^J_{C,i}$ does not depend on the policies set by the other candidate $C' \neq C$.

Combining this result with Proposition 1, voters with higher stakes on policy instrument $q_{C,i}$ (or lower cost of observing $q_{C,i}$) are more attentive, and hence more responsive to changes in this policy instrument. As a consequence, in setting $q_{C,i}$, candidate C has stronger incentives to provide policy favors to these high-stake (or low information cost) voters, compared to less attentive voters. In the next section we explore the implications of these insights for equilibrium policy.

The appendix also solves a second order (rather than first order) approximation of the voters' optimization problem, see (35), which is of course exact for quadratic utilities. The main qualitative difference from (11) 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 the variance of realized utility - i.e., variance of the maximal utility from the two offered policy vectors. Paying more attention decreases the likelihood of selecting a candidate that provides a very low realized utility, which in turn decreases volatility of realized utilities of the selected candidate. This also implies that more risk-averse voters are relatively more attentive and hence relatively more influential than under perfect information, the more so the greater is the prior uncertainty and their 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 other issues. In particular, these examples illustrate why an increase in the granularity of information can be welfare deteriorating, and why new and lesser known candidates often cater to minorities or political extremists.

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 (Corollary 2). 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.

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

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

$q \in R$ and $U(.)$ is concave and symmetric about its maximum at 0. With a one dimensional policy, by (9) 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 (11), 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). 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. These features of the equilibrium become more relevant the higher the cost of attention is. Moroever, groups with a lower cost $\hat{\lambda}^J$ also receive a greater weight, for the same reason.

These properties of the equilibrium can be illustrated with two examples. Suppose first that there are three groups of the same size, and with a distribution of bliss points skewed to the right, for instance $t^1 < t^2 < t^3$ and $t^2 - t^1 < t^3 - t^2$. Thus, group 3 has more extreme policy preferences than the other groups. The cost of information is the same for all groups. Then, the following holds:

Corollary 3 *Let $t^1 < t^2 < t^3$ such that $t^2 - t^1 < t^3 - t^2$. There exists $\lambda_0 > 0$ such that, for any $\hat{\lambda} < \lambda_0$, the equilibrium policy $\hat{q}^*$ is strictly increasing in $\hat{\lambda}$. That is, as the cost of attention rises, the equilibrium moves closer to the bliss point of the group with more extreme preferences (here group 3).*

Voters with more extreme preferences, i.e., location of their bliss-points, pay more attention. This is an immediate implication of (11) and the concavity of $U^J(q)$. The stakes are increasing in $|\hat{q}^* - t^J|$, which in turn drives the attention level up. The monotonicity is weak, because when the stakes are low enough, then voters pay only the minimal attention ξ_0; but strict monotonicity applies anytime the cost of information is sufficiently low that the minimal level of attention is not binding.

Group size has a similar effect, because it also implies that in equilibrium some groups are further away from their bliss points than others. To show this, consider an example with two groups. Now group 2 is the smaller group. Here in equilibrium group 2 is further away from the equilibrium policy and thus pays more attention. Specifically:

Corollary 4 *Let $J \in \{1, 2\}$, $m^1 > m^2$, and $\hat{\lambda}$ be the cost of information of both groups. Then the distance between equilibrium policy and bliss-point of the smaller group $|\hat{q}^* - t^2|$*

is weakly lower for $\hat{\lambda} > 0$ than under perfect information. Moreover, there exists $\lambda_0 > 0$ such that $|\hat{q}^* - t^2|$ is strictly decreasing in $\hat{\lambda}$ for all $\hat{\lambda} \in (0, \lambda_0)$.

Unless voters are constrained by the lower bound on attention ξ_0, increases in $\hat{\lambda}$ imply that the small group becomes even more influential, because attention of the larger group decreases by more. Again, strict monotonicity applies anytime the cost of information is sufficiently low that the minimal level of attention is not binding.

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, 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 prevail 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.

Clearly, these results reflect the assumption that utility is strictly concave. If preferences were linear - eg. $U^J(q) = -\alpha^J|q - t^J|$), then the more attentive groups would still be those with higher stakes, but here these would be the groups with high α^J, and not necessarily those with extreme bliss points. Nevertheless, the prediction that extremist voters pay more attention is in line with results from some 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. Similarly, Lauderdale (2013) find that citizens who are better informed about policy positions of members of Congress are more polarized in their policy views. Moreover, 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.[17]

Policy divergence and new candidates. If the two candidates differ in how visible they are to voters, we obtain a new implication. Suppose that the cost of collecting information is lower, say, for candidate A. For instance, B could be a less established candidate to which the media pay less attention. Then all voters pay more attention to the more established or transparent candidate, here A ($\xi_A^J > \xi_B^J$ for all J). But this effect is not the same across groups of voters. By (11), the difference in attention given by voters to 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 B, while the centrist voters pay relatively more attention to the more established or transparent candidate A (this can be seen by evaluating the derivative of ξ^J with respect to $\hat{\lambda}$ in (11)).[18] This in turn affects the incentives of both candidates and leads to policy divergence if the distribution of bliss points is asymmetric.

This result is best illustrated in the example of Corollary 3 discussed above. Specifically, suppose that there are three groups of the same size, and group 3 has more extreme preferences: $t^2 - t^1 < t^3 - t^2$. We have:

Corollary 5 *Let $\hat{\lambda}_A < \hat{\lambda}_B$. There exists $\lambda_0 > 0$ such that, for any $\hat{\lambda}_B < \lambda_0$, the equilibrium policy $\hat{q}_B^*$ is strictly closer to the bliss point of the group with more extreme preferences (here group 3), compared to $\hat{q}_A^*$. Moreover, in equilibrium A has a higher probability of winning the election than B.*

The more extremist voters pay more attention overall, but they also pay relatively more attention to the less established candidate B, because their stakes are higher. The centrist voters pay relatively more attention to candidate A, who has greater media coverage. Thus, more established candidates tend to cater to the average voter, while

[17] Yet another interpretation of these findings, not relying on concavity of preferences, is that more extremist voters have lower cost of seeking information (particularly on the candidate they feel closer too), because of motivated beliefs or through social interactions (see Sunstein 2017).

[18] The derivative of ξ^J with respect to $\hat{\lambda}$ is $-\frac{1}{u^J(q)\sigma^2}$ if and only if $1 - \frac{\hat{\lambda}}{(u^J)^2\sigma^2} > \xi_0$, otherwise it equals zero. Thus, the change in attention when $\hat{\lambda}$ changes is larger for voters with lower $|u^J|$.

candidates receiving less media coverage go after extremist voters. A similar result applies if one group has higher stakes because it is smaller, as in Corollary 4.

With policy divergence and different attention weights, the probability of victory differs from 1/2. The less established candidate B (who receives less attention by all voters and by the centrist voters in particular) is less likely to win. Equation (6) shows that the candidate who provides higher perceived social welfare has a higher probability of winning. Perceived social welfare provided by each candidate C equals social welfare at the expected equilibrium policy $\bar{q}_C$, plus terms that depend on the noise in policies.[19] If noise is small, then the sign of the difference in the expectations determines the sign of differences in perceived social welfare. Thus, if A is closer to the social optimum, then she also has a higher probability of winning. 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), Ansolabehere et al. (2001) and Stone and Simas (2010).

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 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 vice versa.

[19] See for instance the expansion of perceived utility in (19) in the Appendix.

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 the equilibrium is Pareto inefficient: 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.

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:

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

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 seigniorage, 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 seigniorage 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: $\sigma^J_{C,i} = \sigma$ and $\lambda^J_{C,i} = \lambda$ for all C, J, i.

Equilibrium policy with rational inattention. It is easy to verify that the socially optimal policy vector q° (i.e, the policy that maximizes $\sum_J m^J U^J(q)$) satisfies $s^\circ = \tau^\circ = 0$, i.e., distorting taxes are not used, achieves equal consumption for all groups, $c^J = c^\circ$ for all J, and sets the public good so as to satisfy Samuelson optimality condition, namely $H'(g^\circ) = V'(c^\circ)/N$. 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.[20]

However, with rationally inattentive voters, candidates are motivated to distort the policies away from the social optimum. Let $\xi^J_g, \xi^J_\tau, \xi^J_J, \xi^J_{-J}$ denote the level of attention that voters in group J denote to g, τ, b^J, and b^K for $K \neq J$ respectively (by symmetry, the targeted taxes paid by all other groups receive the same level of attention in equilibrium). The Appendix proves:

Proposition 3 *Under costly attention the equilibrium policy vector $\hat{q}^*$ and allocation of attention have the following features: (i) there is under-provision of the public good relative to the social optimum, $\hat{g}^* \leq g^o$, and the government relies on distorting (observable and unobservable) sources of revenues: $\hat{s}^*, \hat{\tau}^* \geq 0$. (ii) All voters pay only the minimal attention to non-divisive issues, $\xi^J_g = \xi^J_\tau = \xi_0$, and they pay weakly more attention to their own targeted taxes (or transfers) than to the targeted instruments affecting others, $\xi^J_J \geq \xi^J_{-J}$. There exists $\lambda_0 > 0$ such that for all $\hat{\lambda} > \hat{\lambda}_0$ all inequalities are strict.*

[20] See the beginning of the proof of Proposition 3, and in particular the note following equation (40) in the Appendix.

Strict under-provision and inefficient use of taxes occurs anytime the agent pays higher than minimal attention to his own targeted taxes. 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).

What drives equilibria away from the social optimum is heterogeneity in ξ_i^J across different voters. Stakes regarding g and τ are uniform across all voters, and thus all voters pay the same attention to these instruments, which yields the same FOCs as for the social optimum. However, when it comes to targeted taxes b^J, stakes differ across voters. Voters of type J choose to pay more attention to b^J than other voters do. This incentivizes candidates to decrease b^J away from the social optimum, so as to please voters who pay most attention to their own group specific taxes. By symmetry, in equilibrium all targeted taxes b^J are decreased, public goods are underprovided, and uniform distorting taxes are over-used.

Note that in equilibrium, although $\hat{\tau}^*$ and $\hat{g}^*$ differ from the social optimum, all groups J have $u_\tau^J = 0$ and $u_g^J = 0$. That is, the policy instruments that have a uniform effect on all groups are set at the (identical) bliss point of each group. These bliss points differ from the social optimum, because each group wishes to under-provide the public good and raise distorting taxes so as to direct targeted transfers to itself. By (11) 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, (9) implies that the equilibrium attention is at the minimal level ξ_0. On the other hand, divisive issues are paid more attention to. Since these policy instruments are not set optimally from the perspective of each individual voter, then voters' stakes are positive, and they pay attention to such issues.[21]

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

[21]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 (9) 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.

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, consistent 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. We distinguish between two types of transparency. One that affects the cost of information, and the other that affects the flexibility of choice. The diffusion of the Internet is a case in point. The Internet provides not only very cheap information , but also information on very fine issues. Agents can now choose to be informed on very narrow issues of their choice, and also what information to avoid. Such granular information was not available before at all as agents had to get information on broader issues presented by TV, for instance. The Appendix proves:

Corollary 6 *The equilibrium becomes less distorted, i.e., $\hat{b}^{J*}$ and $\hat{g}^*$ increase while $\hat{\tau}^*$ and $\hat{s}^*$ fall, if*

(i) the cost of information on instruments targeted at others (λ_{-J}^J) falls or if

(ii) the cost of information on instruments targeted at themselves (λ_J^J) increases.

The potential inefficiencies are driven by too much attention to instruments targeted at themselves relative to those targeted at others. Intuitively, inducing voters to pay more attention to benefits targeted at other groups, raises the political costs of targeting.[22] On the other hand, while Proposition 3 above states that welfare is weakly lower under rational inattention than under perfect information, lowering the cost of information $\hat{\lambda}$ uniformly on all instruments has a non-monotonic effect. If voters pay higher than minimal attention to both b^J and b^{-J}, then efficiency increases, while if the cost is sufficiently high that voters pay the minimal attention to b^{-J}, then the effect is the opposite.

Now we illustrate the second point that transparency in the form of packaging of information can also affect welfare. 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

[22] 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 oneself 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.

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 information is: lower k means more granular information.

Corollary 7 *As k increases (i.e., granularity decreases) the equilibrium becomes less distorted, i.e., $\hat{b}^{J*}$ and $\hat{g}^*$ increase while $\hat{\tau}^*$ and $\hat{s}^*$ fall. The equilibrium reaches the social optimum 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. An important implication of the model is that more information can have adverse effects on social welfare, because it can enhance endogenous informational asymmetries.

Another welfare improving information repackaging would be to also give voters information 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.[23]

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.

[23]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.

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 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.[24]

5 Concluding remarks

Digital technologies provide an almost unlimited and easily accessible supply of very detailed information. Yet, because of limited attention, information remains costly to absorb and process. This has raised the relevance of informational asymmetries between voters and across policy issues. Such asymmetries are not random or unexplainable, however. What we know and don't know about the political process is largely determined by information that we purposefully seek of our own initiative, or to which we are exposed through social media. This paper has studied how voters allocate costly attention to political information, and how this interacts with the behavior of office seeking politicians.

Our analysis delivers two general and intuitive insights. First, voters pay more attention to policy issues where they have higher stakes. Second, the political process rewards attention with policy favors. In equilibrium attention and policy are jointly determined, since policy stakes are endogenous. We have then illustrated the implications of these general insights with two examples. If policy is one-dimensional, endogenous attention acts as an amplifier of preference intensity, and can lead to policy divergence even if politicians only care about winning the election. If policy is multidimensional, attention is devoted to the more controversial policy issues. As a result, the equilibrium is generally Pareto inefficient and policies that provide uniform benefits to all are under-provided. These results have normative implications for how to structure the cost of acquiring information over public policies (eg., distortions get worse if the granularity of information

[24]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).

increases).

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. A previous version of this paper showed that our results extend to a setting where politicians have partisan policy preferences that are unknown to voters. This framework yields an additional insight. 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 preferred 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. The 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.

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 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.[25]

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.

[25] 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.

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

Grossman, Gene and Elhanan Helpman (2001) "Special Interest Politics", MIT Press, Cambridge

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.

Hu, Li and Li, Anqi (2018), "The Politics of Attention", mimeo.

Lauderdale, Benjamin E. (2013) "Does Inattention to Political Debate Explain the Polarization Gap between the US Congress and Public?", Public Opinion Quarterly, Vol 77, S1, p2-23

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). "-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.

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.

Stone, Walter J, and Elisabeth N. Simas (2010) "Candidate Valence and Ideological Positions in US House Elections". American Journal of Political Science, 54(2), pp371-388

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

Sunstein, Cass (2017) "#Republic: Divided Democracy in the Age of Social Media", Princeton University Press, Princeton

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 Santa Barbara

6 Appendix

6.1 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:

$$\tilde{x}^v = E[U^J(q_A)|s_A^{v,J}] - E[U^J(q_B)|s_B^{v,J}] - \tilde{x} \equiv \tilde{x}_T^{v,J} \tag{13}$$

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:

$$\pi_A = \sum_J m^J E_\epsilon^J[\Pr(\tilde{x}^v \leq \tilde{x}_T^{v,J})] = \frac{1}{2} + \phi \sum_J m^J E_\epsilon^J[\tilde{x}_T^{v,J}] \tag{14}$$

Note that (14) holds when the noise in the ideological preference shocks $\tilde{x}^v$ is sufficiently large to affect the vote with positive probability.[26]

By (13)-(14), 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:

$$p_A = \frac{1}{2} + \psi \left(\sum_J m^J E_{\epsilon,q_A,q_B}^J \left[E[U^J(q_A)|s_A^{v,J}] - E[U^J(q_B)|s_B^{v,J}] \right] \right) \tag{15}$$

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.

[26]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.

6.2 Small noise approximations or quadratic utility

We reformulate Proposition 1 in the following more general way and then we prove it:

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

$$\sum_{J=1}^{N} m^J \xi^J_{C,i}[u^J_{C,i} + \delta_i(\sigma^2_C)] = 0 \qquad \forall i, \tag{16}$$

where $\xi^J_{C,i}$ are the equilibrium attention weights. The function $|\delta_i(\sigma^2_C)|$ is decreasing in the scale of elements of σ^2_C and equal zero for $\sigma^2_C = 0$. For small noise σ^2_C equation (16) can be approximated by:

$$\sum_{J=1}^{N} m^J \xi^J_{C,i} u^J_{C,i} = 0 \qquad \forall i, \tag{17}$$

Proof: We will express derivatives of the candidate's objective (8) 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 $\bar{q}_C$.

$$\tilde{U}^J(q_C) \simeq U^J(\bar{q}_C) + \sum_{i=1}^{M} u^J_{C,i}(q_{C,i} - \bar{q}_{C,i}) + \frac{1}{2}\sum_{i,j=1}^{M,M} u^J_{C,i,j}(q_{C,i} - \bar{q}_{C,i})(q_{C,j} - \bar{q}_{C,j}), \tag{18}$$

where $u^J_{C,i}$ and $u^J_{C,i,j}$ 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 equals $E[\tilde{U}^J(q_C)|s^{v,J}_C]$ plus expectation of the higher order terms. Terms of n^{th} order equal products of $\frac{1}{n!}$, of n^{th} order derivatives of $U^J(\bar{q}_C)$, and of n terms of the form of $(q_{C,i} - \bar{q}_{C,i})$.

$$\begin{aligned} E[\tilde{U}^J(q_C)|s^{v,J}_C] \quad &= \quad U^J(\bar{q}_C) + \sum_{i=1}^{M} u^J_{C,i}(\check{q}_{C,i} - \bar{q}_{C,i}) \\ &\quad + \frac{1}{2}\sum_{i,j=1}^{M,M} u^J_{C,i,j} E\Big[(q_{C,i} - \bar{q}_{C,i})(q_{C,j} - \bar{q}_{C,j})|s^{v,J}_C\Big], \end{aligned} \tag{19}$$

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

$$\begin{aligned} &\frac{1}{2}\sum_{i,j=1}^{M,M} u^J_{C,i,j} E\Big[\Big((q_{C,i} - \check{q}_{C,i}) - (\bar{q}_{C,i} - \check{q}_{C,i})\Big)\Big((q_{C,j} - \check{q}_{C,j}) - (\bar{q}_{C,j} - \check{q}_{C,j})\Big)|s^{v,J}_C\Big] \\ = \quad &\frac{1}{2}\sum_{i,j=1}^{M,M} u^J_{C,i,j}(\check{q}_{C,i} - \bar{q}_{C,i})(\check{q}_{C,j} - \bar{q}_{C,j}) + \frac{1}{2}\sum_{i=1}^{M} u^J_{C,i,i}(1 - \xi_{C,i})\sigma^2_{C,i}. \end{aligned} \tag{20}$$

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 $(1 - \xi_{C,i})\sigma^2_{C,i}$.

The expectation of the higher order terms in $E[U^J(q_C)|s^{v,J}_C]$ takes an analogous form to the RHS of (20). It is a sum of products of the higher order derivatives and elements of $(q_C - \check{q}_C)$ and $(\bar{q}_C - \check{q}_C)$, where elements of $(q_C - \check{q}_C)$ are in an even order in each term.

We use $\check{q}_{C,i} = \xi^J_{C,i} s^{v,J}_{C,i} + (1 - \xi^J_{C,i})\bar{q}_{C,i}$ to express $E_{\epsilon,e}[\cdot]$ of the first term on the RHS of (20), which is

$$\begin{aligned} & \frac{1}{2} E_{\epsilon,e}\Big[\sum_{i,j=1}^{M,M} u^J_{C,i,j} \xi^J_{C,i} \xi^J_{C,j} (\hat{q}_{C,i} + e_i + \epsilon^J_{C,i} - \bar{q}_{C,i})(\hat{q}_{C,j} + e_j + \epsilon^J_{C,j} - \bar{q}_{C,j})\Big] \\ = \; & \frac{1}{2} \sum_{i=1}^{M} u^J_{C,i} (\xi^J_{C,i})^2 (\sigma^2_{C,i} + \frac{1 - \xi^J_{C,i}}{\xi^J_{C,i}} \sigma^2_{C,i}) \\ & + \frac{1}{2} \sum_{i,j=1}^{M,M} u^J_{C,i,j} \xi^J_{C,i} \xi^J_{C,j} (\hat{q}_{C,i} - \bar{q}_{C,i})(\hat{q}_{C,j} - \bar{q}_{C,j}), \end{aligned} \tag{21}$$

where $\frac{1-\xi^J_{C,i}}{\xi^J_{C,i}} \sigma^2_{C,i}$ is the variance of $\epsilon^J_{C,i}$. Putting (19)-(21) together, we get

$$\begin{aligned} E_{\epsilon,e}\Big[E[\tilde{U}^J(q_C)|s^{v,J}_C]\Big|\hat{q}_C\Big] \; = \; & U^J(\bar{q}_C) + \sum_{i=1}^{M} \xi^J_{C,i} u^J_{C,i} (\hat{q}_{C,i} - \bar{q}_{C,i}) + \frac{1}{2} \sum_{i=1}^{M} u^J_{C,i,i} \sigma^2_{C,i} \\ & + \frac{1}{2} \sum_{i,j=1}^{M,M} u^J_{C,i,j} \xi^J_{C,i} \xi^J_{C,j} (\hat{q}_{C,i} - \bar{q}_{C,i})(\hat{q}_{C,j} - \bar{q}_{C,j}). \end{aligned} \tag{22}$$

Expectation $E_{\epsilon,e}\Big[E[U^J(q_C)|s^{v,J}_C]\Big|\hat{q}_C\Big]$ equals the RHS of (22) plus $E_{\epsilon,e}\Big[\cdot\Big]$ of the higher order terms. All the n^{th}-order terms would then be a product of $1/n!$, an n^{th}-order derivative, a constant given by a product of attention weights, and of k terms of prior variances $\sigma^2_{C,i}$ and $n - 2k$ terms of $(\hat{q}_{C,j} - \bar{q}_{C,j})$.

Therefore, the derivative of the RHS of (22) with respect to $\hat{q}_{C,i}$, evaluated at the equilibrium $\hat{q}_C = \bar{q}_C$, is

$$\frac{\partial E^J_{\epsilon,e}\Big[E[\tilde{U}^J(q_C)|s^{v,J}_C]\Big|\hat{q}_C\Big]}{\partial \hat{q}_{C,i}}\Big|_{\hat{q}_C = \bar{q}_C} = \xi^J_{C,i} u^J_{C,i}. \tag{23}$$

Weighting this by m^J, we get (8).

The derivative of expectation of the original non-approximated $U^J(q_C)$ in addition includes terms where each is a product of $\xi^J_{C,i}$, of a higher-order derivative evaluated at zero, and of at least one variance $\sigma^2_{C,j}$. The only terms that do not drop out after differentiating with respect to $\hat{q}_{C,i}$ are those where $(\hat{q}_{C,i} - \bar{q}_{C,i})$ appears in exactly the order of one, and thus each remaining term is a product of $\sigma^2_{C,j}$ for some j, too. If the utility is smooth at zero, then the whole sum is finite and scales down weakly super-proportionally if σ_C^2 is scaled down, the limit of the contribution as σ_C^2 goes to zero is zero. ∎

Proof of Corollary 1: If for each i: ξ_i^J's are positive and equal for all J, then dividing (9) by ξ_i^J yields the FOCs for the maximization of welfare. ∎

We reformulate Lemma 2 in the following more general way and then we prove it.

Lemma 2: Suppose that (7) holds. Then the voters chooses the attention vector $\xi^J \in [\xi_0, 1]^M$ to maximize the following objective function:

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

where $\hat{\lambda}_{C,i}^J = 2\lambda_{C,i}^J/Min(\psi,\phi)$, and $|\hat{\delta}_i(\sigma_C^2)|$ is decreasing in the scale of σ_C^2 and equal to zero for $\sigma_C^2 = 0$. For small noise σ_i^2 this objective function approximates to:

$$\left(\sum_{C\in\{A,B\},i=1}^{M} \xi_{C,i}^J (u_{C,i}^J)^2 \sigma_{C,i}^2\right) + \sum_{C\in\{A,B\},i\leq M} \hat{\lambda}_{C,i}^J log\left(1-\xi_{C,i}^J\right) \tag{24}$$

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

$$\begin{aligned} E\left[\max_{C\in\{A,B\}} E[U_C^{v,J}(q_C)|s_C^{v,J}]\right] - \text{cost of info} \;&=\; \frac{1}{2}E\left[E[U_A^{v,J}(q_A)|s_A^{v,J}] + E[U_B^{v,J}(q_B)|s_B^{v,J}]\right] + \\ &+\frac{1}{2}E\left[\left|E[U_A^{v,J}(q_A)|s_A^{v,J}] - E[U_B^{v,J}(q_B)|s_B^{v,J}]\right|\right] - \\ &-\text{cost of info.} \end{aligned} \tag{25}$$

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 as 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 (25) yields:

$$\begin{aligned}
&\frac{1}{2}E\Big[\sum_{C\in\{A,B\}} E[U_C^{v,J}(q_C)|s_C^{v,J}]\Big] \\
\simeq\;& \frac{1}{2}E\Big[\sum_{C\in\{A,B\}} E[U_C^{v,J}(\bar{q}_C)+\sum_{i=1}^{M}u_{C,i}^{J}(q_{C,i}-\bar{q}_{C,i})+\frac{1}{2}\sum_{i,j=1}^{M,M}u_{C,i,j}^{J}(q_{C,i}-\bar{q}_{C,i})(q_{C,j}-\bar{q}_{C,j})|s_C^{v,J}]\Big] \\
=\;& \frac{1}{2}\sum_{C\in\{A,B\}}\Big(U^J(\bar{q}_C)+\frac{1}{2}\sum_{i,j=1}^{M,M}u_{C,i,j}^{J}E\Big[E\Big[\Big((q_{C,i}-\check{q}_{C,i})-(\bar{q}_{C,i}-\check{q}_{C,i})\Big) \\
&\qquad\Big((q_{C,j}-\check{q}_{C,j})-(\bar{q}_{C,j}-\check{q}_{C,j})\Big)|s_C^{v,J}\Big]\Big]\Big) \\
=\;& \frac{1}{2}\sum_{C\in\{A,B\}}\Big(U^J(\bar{q}_C)+\frac{1}{2}\sum_{i=1}^{M}(u_{C,i,i}^{J}\xi_{C,i}\sigma_{C,i}^2+u_{C,i,i}^{J}(1-\xi_{C,i})\sigma_{C,i}^2)\Big) \\
=\;& \frac{1}{2}\sum_{C\in\{A,B\}}\Big(U^J(\bar{q}_C)+\frac{M}{2}u_{C,i,i}^{J}\sigma_{C,i}^2\Big) \qquad (26)
\end{aligned}$$

In the first steps we omitted third and higher order terms, which would result in additive terms with higher powers of σ_C^2. 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 $(1-\xi_{C,i})\sigma_{C,i}^2$, and also that variance of posterior means, $(\check{q}_{C,i}-\bar{q}_{C,i})$, is $\xi_{C,i}\sigma_{C,i}^2$ (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.

The RHS of (26) is independent of ξ^J, and thus the voter's choice of attention is given by the maximization of expectation of the absolute value of:

$$\frac{1}{2}\Delta^v=\frac{1}{2}\Big(E[U_A^{v,J}(q_A)|s_A^{v,J}]-E[U_B^{v,J}(q_B)|s_B^{v,J}]\Big) \qquad (27)$$

less the cost of information. Let

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

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

$\Delta > 0$):

$$
\begin{aligned}
\int_{-\infty}^{\infty} f(x)|\Delta - x|dx &= \int_{-\infty}^{\Delta} f(x)(\Delta - x)dx - \int_{\Delta}^{\infty} f(x)(\Delta - x)dx \\
&= \Delta\Big(\int_{-\infty}^{\Delta} f(x)dx - \int_{\Delta}^{\infty} f(x)dx\Big) + \\
&\quad +\Big(-\int_{-\infty}^{\Delta} f(x)xdx + \int_{\Delta}^{\infty} f(x)xdx\Big) \\
&= \Delta \int_{-\Delta}^{\Delta} f(x)dx + 2\int_{\Delta}^{\infty} f(x)xdx. \qquad (28)
\end{aligned}
$$

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, we use the assumption that the noise σ_C^2 is very small relative to the support of preference shocks x. When Δ, the perceived difference in utilities from the two candidates, is very small relative to the support of x, then we can assume that $f(x)$ is constant on $(-\Delta, \Delta)$:

$$
\begin{aligned}
\Delta \int_{-\Delta}^{\Delta} f(x)dx &\simeq 2f(0)\Delta^2, \\
2\int_{\Delta}^{\infty} f(x)xdx &= 2\int_{0}^{\infty} f(x)xdx - 2\int_{0}^{\Delta} f(x)xdx \simeq E_f[|x|] - f(0)\Delta^2. \qquad (29)
\end{aligned}
$$

Therefore, conditional on Δ, the expectation of $|\Delta^v|$ equals $(E_f[|x|] + f(0)\Delta^2)$. The approximation is exact if the support of x is sufficiently large. In fact, $f(x)$ is constant on $(-a, a)$, where $a = min(\frac{1}{2\psi}, \frac{1}{2\phi})$. Therefore, it is exact if $|\Delta| < a$. 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 (20), we get:

$$
\begin{aligned}
\Delta &\simeq U^J(\bar{q}_A) - U^J(\bar{q}_B) + \sum_{i=1}^{M}\Big(u_{A,i}^J(\check{q}_{A,i} - \bar{q}_{A,i}) - u_{B,i}^J(\check{q}_{B,i} - \bar{q}_{B,i})\Big) \\
&\quad +\frac{1}{2}\sum_{i=1}^{M}\Big(u_{A,i,i}^J((\check{q}_{A,i} - \bar{q}_{A,i})^2 + (1 - \xi_{A,i}^J)\sigma_{A,i}^2) - u_{B,i,i}^J((\check{q}_{B,i} - \bar{q}_{B,i})^2 \qquad (30) \\
&\quad +(1 - \xi_{B,i}^J)\sigma_{B,i}^2)\Big). \qquad (31)
\end{aligned}
$$

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 (30) squared, the last term is expectation of a product of the first and the third terms.

$$\begin{aligned}
E[\Delta^2] \;\simeq\;& \Big(U^J(\bar{q}_A) - U^J(\bar{q}_B)\Big)^2 + \sum_{i=1, C\in\{A,B\}}^{M} \xi^J_{C,i}(u^J_{C,i})^2\sigma^2_{C,i} \\
&+\frac{1}{4}E\Big[\Big(\sum_{i=1}^{M} u^J_{A,i,i}((\check{q}_{A,i}-\bar{q}_{A,i})^2+(1-\xi^J_{A,i})\sigma^2_{A,i}) - u^J_{B,i,i}((\check{q}_{B,i}-\bar{q}_{B,i})^2+(1-\xi^J_{B,i})\sigma^2_{B,i})\Big)^2\Big] \\
&+\Big(U^J(\bar{q}_A) - U^J(\bar{q}_B)\Big)\Big(\sum_{i=1}^{M} u^J_{A,i,i}\sigma^2_{A,i} - u^J_{B,i,i}\sigma^2_{B,i}\Big). \qquad (32)
\end{aligned}$$

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

$$\begin{aligned}
-\quad & 2\sum_{i,j=1}^{M,M} u^J_{A,i,i}u^J_{B,j,j}\sigma^2_{A,i}\sigma^2_{B,j} + 2\sum_{i,j=1, C\in\{A,B\}}^{M,M} u^J_{C,i,i}u^J_{C,j,j}\xi^J_{C,i}(1-\xi^J_{C,j})\sigma^2_{C,i}\sigma^2_{C,j} \\
& + \sum_{i,j=1, C\in\{A,B\}}^{M,M} u^J_{C,i,i}u^J_{C,j,j}(1-\xi^J_{C,i})(1-\xi^J_{C,j})\sigma^2_{C,i}\sigma^2_{C,j} \qquad (33) \\
& + \sum_{i,j=1, C\in\{A,B\}}^{M,M} u^J_{C,i,i}u^J_{C,j,j}\xi^J_{C,i}\xi^J_{C,j}\sigma^2_{C,i}\sigma^2_{C,j} + 2\sum_{i=1, C\in\{A,B\}}^{M} (u^J_{C,i,i})^2(\xi^J_{C,i})^2(\sigma^2_{C,i})^2 \\
= \quad & -2\sum_{i,j=1}^{M,M} u^J_{A,i,i}u^J_{B,j,j}\sigma^2_{A,i}\sigma^2_{B,j} + \sum_{i,j=1, C\in\{A,B\}}^{M,M} u^J_{C,i,i}u^J_{C,j,j}\sigma^2_{C,i}\sigma^2_{C,j} \\
& +2\sum_{i=1, C\in\{A,B\}}^{M} (u^J_{C,i,i})^2(\xi^J_{C,i})^2(\sigma^2_{C,i})^2. \qquad (34)
\end{aligned}$$

The first term on the LHS of (33) 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 $(1-\xi^J_{C,i})\sigma^2_{C,i}$, the third is product of between terms with $(1-\xi^J_{C,i})\sigma^2_{C,i}$, the fourth 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 fuorth 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 equivalent to (25) 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

$$F(\xi^J) = \sum_{i=1, C\in\{A,B\}}^{M} \Big(\xi^J_{C,i}\sigma^2_{C,i}(u^J_{C,i})^2 + 2(\xi^J_{C,i})^2(\sigma^2_{C,i})^2(u^J_{C,i,i})^2\Big). \qquad (35)$$

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 (35); and we also denote $\hat{\lambda}^J_{C,i} = 2\lambda^J_{C,i}/Min(\psi, \phi)$. If we did not omit any higher order terms, then the RHS of (35) takes the form of

$$\sum_{C\in\{A,B\},i=1}^{M} \xi^J_{C,i}\sigma^2_{C,i}((u^J_{C,i})^2 + \hat{\delta}_i(\sigma^2_C)),$$

where $\hat{\delta}_i$ is a sum of terms that are products of $\sigma^2_{C,i}$ of a power of at least one, and of higher-order derivatives of U and f. The function $|\hat{\delta}_i(\sigma^2_C)|$ is hence decreasing in the scale of σ^2_C and equal to zero for $\sigma^2_C = 0$.

The solution to the voter's maximization problem for the second-order approximation is then:

$$\xi^J_{C,i} = \max\left(\xi_0, \frac{4\sigma^2_{C,i}(u^J_{C,i,i})^2 - (u^J_{C,i})^2 + \sqrt{(4\sigma^2_{C,i}(u^J_{C,i,i})^2 + (u^J_{C,i})^2)^2 - 16\hat{\lambda}^J_{C,i}(u^J_{C,i,i})^2}}{8\sigma^2_{C,i}(u^J_{C,i,i})^2}\right). \tag{36}$$

∎

Proof of Proposition 2: Derivative of the objective (10) with respect to $\xi^J_{C,i}$ equals $(u^J_{C,i})^2\sigma^2_{C,i} - \lambda^J_{C,i}/(1-\xi^J_{C,i})$. The first order condition together with the constraint $\xi^J_{C,i} \geq \xi_0$ then takes the form of (11). ∎

6.3 Applications

Proof of Corollary 3: We first establish that the equilibrium policy $q^* \in (t^2, t^3)$. By contradiction, if $q^* \geq t^3$, then LHS of FOC (9) is negative (all terms are negative, except perhaps for $J = 3$, which can be non-positive), while for $q^* \leq t^2$ the LHS would be positive (the term with $J = 2$ is positive and $\xi^1 u^1 + \xi^3 u^3$ is positive because $|q^* - t^1| < |q^* - t^3|$).

Second, we show that ξ^3 is weakly greater than ξ^1 and ξ^2. Since $q^* \in (t^2, t^3)$, then $\xi^2 u^2$ must be negative, which means that $\xi^1 u^1 + \xi^3 u^3$ is positive. Therefore, $|u^3|$ is greater than both $|u^1|$ and $|u^2|$, which implies also that ξ^3 is weakly greater than ξ^1 and ξ^2.

Third, we show that for $\hat{\lambda} > 0$, the equilibrium policy is weakly closer to t^3 than under perfect information, i.e., $q^* \geq q^P$, where both equilibria are in (t^2, t^3). Under perfect information, the FOC (9) implies $u^1 + u^2 = -u^3$. But as we showed above, since $\xi^3 \geq \max(\xi^1, \xi^2)$, then under imperfect information (9) implies $u^1 + u^2 \geq -u^3$. By contradiction, if $q^* < q^P$, then relative to perfect information both u^1 and u^2 would decrease while u^3 would increase. However, then $u^1 + u^2 \geq -u^3$ stated above could not hold, and thus $q^* \geq q^P$.

Finally, we prove the finer dependence on $\hat{\lambda}$. Note that as $\hat{\lambda}$ increases, so does the radius R_0 such that for $|t^J - q^*| \leq R_0$ the voters J pay only the minimal attention ξ_0. The radius is due to (11) given by $|u(R_0)|^2 = (1-\xi_0)\hat{\lambda}/\sigma^2$, and hence it approaches zero as $\hat{\lambda}$ does. Since $q^* \geq q^P$, then if $\hat{\lambda}$ is small enough such that q^P is outside of the radius of the minimal attention for $J = 2$, $|t^2 - q^P| > R_0$, then the equilibrium with rational inattention is outside of the radius too, and all voters pay higher attention than ξ_0. Given q^P, there always exists λ_0 such that for $\hat{\lambda} < \lambda_0$ all voters pay higher attention than ξ_0.

We now plug the formula for the choice of attention (11) for $\xi^J > \xi_0$ into the FOC (9). The derivative of the FOC with respect to $\hat{\lambda}$ takes a form:

$$-\frac{\hat{\lambda}}{\sigma^2}(1/u^1 + 1/u^2 + 1/u^3).$$

Since $|u^1| < |u^3|$, then $(1/u^1 + 1/u^3)$ is negative, and so is $1/u^2$. The derivative of the FOC above is thus positive. Increasing $\hat{\lambda}$ for $\hat{\lambda} < \lambda_0$ pushes the equilibrium policy closer to t^3.

∎

Proof of Corollary 4: We first establish that in equilibrium the smaller group $J = 2$ faces higher stakes and pays weakly more attention. Since $m^1 > m^2$, then the FOC (9) implies that $|\xi^1_C u^1(q_C)/\xi^2_C u^2(q_C)| < 1$. Because ξ^J_C is weakly increasing in stakes $|u^J|$, then $|u^2| > |u^1|$, which also implies that $\xi^2_C \geq \xi^1_C$. Under perfect information $\frac{|u^2|}{|u^1|}$ equals $\frac{m^1}{m^2}$, while with costly information it equals $\frac{m^1\xi^1_C}{m^2\xi^2_C}$, and thus the ratio $\frac{|u^2|}{|u^1|}$ is weakly lower under rational inattention, because $J = 1$ pays weakly lower attention than $J = 2$. Because there is one-to-one correspondence between u^1 and u^2 for q between t^1 and t^2, one is decreasing in q and the other is increasing, then a weakly lower $\frac{|u^2|}{|u^1|}$ also implies weakly lower $|u^2|$ than under perfect information. Concavity of utility then implies the first part of the statement.

To prove the finer dependence on $\hat{\lambda}$, note that as $\hat{\lambda}$ increases, so does the radius R_0 such that for $|t^J - \hat{q}*| \leq R_0$ the voters J pay only the minimal attention. The radius is due to (11) given by $|u(R_0)|^2 = (1-\xi_0)\hat{\lambda}/\sigma^2_C$, and hence it approaches zero as $\hat{\lambda}$ does. WLOG, $t^1 < t^2$. Now, take the perfect information equilibrium, q^P, which lies in $(t_1, \frac{t_1+t_2}{2}))$. Using the arguments above, any equilibrium with rationally inattentive voters must satisfy $\hat{q}^* \geq q^P$. Thus if $\hat{\lambda}$ is small enough such that q^P is outside of the radius of the minimal attention, $|t^J - q^P| > R_0$, then the equilibrium with rational inattention is too, and all voters pay higher attention than ξ_0.

Finally, we plug the formula for the choice of attention (11) for $\xi^J_C > \xi_0$ into the FOC

(9). The derivative of the FOC with respect to $\hat{\lambda}$ takes a form:

$$-\frac{\hat{\lambda}}{\sigma_C^2}(m^1/u^1 + m^2/u^2).$$

As long as $m^2 > 0$, which we assume, $u^1 < 0 < u^2$. Since $|u^1| < |u^2|$, then $(m^1/u^1 + m^2/u^2)$ is negative, and the derivative of the FOC above is thus positive. Increasing $\hat{\lambda}$ pushes the equilibrium policy closer to t^2.

∎

Proof of Corollary 5: The first part is an immediate implication of Corollary 3 applied to the policy of candidate B.

The second part is implied by (6) and (22). The probability of winning is increasing in the expected perceived welfare. According to (22), the expectation equals the social welfare given by the expected policy plus higher order terms, which vanish for small deviations of the policy q_C from the expected policy $\bar{q}_C$. For small noise in policies, the candidate that provides a target policy that would deliver higher welfare thus has a higher probability of winning. ∎

Proof of Proposition 3: To express the first order conditions (9), we denote: $u_J^J = (-1 + S'/N)V'(c^J)$, $u_{-J}^J = V'(c^J)S'/N$, $u_\tau^J = (T' - S')V'(c^J)$ 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:

$$-V'S'/N + H' = 0 \tag{37}$$

$$-T' + S' = 0 \tag{38}$$

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.[27] 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.

[27]This can be seen from (11) and from the fact that u_τ^J and u_g^J are common to all voters.

The first order condition (9) with respect to $\hat{b}^J$ can be written as:

$$\xi_J^J V'(c^J)(-1+S'/N) + \sum_{K\neq J} \xi_J^K V'(c^K) S'/N = 0. \tag{39}$$

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

$$[1+(N-1)\frac{\xi_J^{-J}}{\xi_J^J}]S'/N = 1. \tag{40}$$

For the social optimum, where $\frac{\xi_{-J}^J}{\xi_J^J} = 1$, FOC (40) simplifies to $S' = 1$. This implies that $s^o = 0$, and from (38) we get $\tau^o = 0$.

We now show that with rational inattention, $\frac{\xi_{-J}^J}{\xi_J^J} \leq 1$. By contradiction, $\frac{\xi_{-J}^J}{\xi_J^J} > 1$ cannot be an equilibrium, because (40) then implies that $S' < 1$. However, $S' < 1$ implies $|u_J^J| > \frac{V'}{2}$ and $|u_{-J}^J| < \frac{V'}{2}$. Since attention ξ_i^J is a weakly increasing function of the stakes $|u_i^J|$, then this is a contradiction.

Finally, $\frac{\xi_{-J}^J}{\xi_J^J} \leq 1$ together with FOC (40) implies that in equilibrium $S' \geq 1$ and hence that $\hat{s} \geq 0$. Equations (37)-(38) then imply that $H' \geq V'/N$, which implies $\hat{g} \leq g^o$, and that $T' \geq 1$.

When the cost of information is sufficiently low, then the inequalities are strict. Using (37), we find that $u_{-J}^J = H'$ which is greater or equal to $H'^o) > 0$, since $\hat{g} \leq g^o$ and H is concave and increasing. Therefore, there exists a cost of information for which $\xi_{-J}^J > \xi_0$, and thus also $\xi_J^J > \xi_0$, which implies also that for all lower costs of information the agent pays higher than minimal attention, and the inequalities are strict.

∎

Proof of Corollary 6: The statement is an implication of (39). First, let us assume that the cost of information is sufficiently low that the agent pays attention to $\hat{b}_J^J$ as well as to $\hat{b}_{-J}^J$. We show in the proof of Proposition 3 that such $\lambda_0 > 0$ exists. Strict monotonicities then hold for all $\hat{\lambda} < \lambda_0$.

Differentiating the LHS of the FOC (39) for $\hat{b}^J$ with respect to $\hat{\lambda}$, we get:

$$\frac{1}{|u_J^J|} - \frac{N-1}{|u_{-J}^J|}.$$

This derivative is negative because $|u_{-J}^J| \leq |u_J^J|$ and $N > 2$, therefore b^J decreases if $\hat{\lambda}$ increases.

Finally, if either $\hat{\lambda}_J^J$ increases, or if $\hat{\lambda}_{-J}^J$ decreases, then weight on the second term in (39) weakly increases, which must weakly decrease equilibrium $\hat{s}$ and increase $\hat{b}^J$. ∎

Proof of Corollary 7: 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 do not contain b^J. Using the first order condition for b^J (39) together with the constraint on bins of ξ, we get the following instead of (40):

$$[k + (N-k)\frac{\xi^J_{-J}}{\xi^J_J}]\frac{S'}{N} = 1. \tag{41}$$

We will now show that if k increases (information is coarser), then S' decreases. For fixed $\frac{\xi^J_{-J}}{\xi^J_J} \leq 1$, the term $[k + (N-k)\frac{\xi^J_{-J}}{\xi^J_J}]$ is increasing in k. Therefore, if the equilibrium $\frac{\xi^J_{-J}}{\xi^J_J}$ is increasing in k, then (41) implies that S' must be decreasing in k.

Therefore, if S' were strictly increasing in k (for some k), then $\frac{\xi^J_{-J}}{\xi^J_J}$ would be strictly decreasing. Note that increasing S' also implies increasing V' since consumption must be lower at higher inefficiencies S.

We will now prove by contradiction that S' is decreasing in k. Let us assume that S' is strictly increasing in k (for some k). The above implies that $\frac{\xi^J_{-J}}{\xi^J_J}$ must be strictly decreasing for such k and V' increasing. Following (11), we have:

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

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'). For V' increasing in k, the ratio $\frac{\xi^J_{-J}}{\xi^J_J}$ thus must be increasing in k, because both increasing k and the assumed associated increasing S' increase the ratio $\frac{\xi^J_{-J}}{\xi^J_J}$. This is a contradiction since the assumed strictly increasing S' must be associated with strictly decreasing $\frac{\xi^J_{-J}}{\xi^J_J}$.

Therefore, S' is monotonically decreasing in k. Finally, for $k = N$, we have $\frac{\xi^J_{-J}}{\xi^J_J} = 1$. Equation (41) then implies S', and hence $S = 0$. ∎