ELECTORAL COMPETITION WITH RATIONALLY INATTENTIVE VOTERS*

Filip Matějka[†] and Guido Tabellini[‡]

First version: September 2015; This version: January 2016

Abstract

How do voters allocate costly attention to alternative political issues? And how does selective ignorance of voters interact with policy design by politicians? We address these questions by developing a model of electoral competition with rationally inattentive voters. Rational inattention amplifies the effects of preference intensity, because voters pay more attention where stakes are higher. The model has many potential applications, and those that we discuss in more detail imply that extremist voters are more attentive and influential, public goods are under-provided, divisive issues receive more attention, and less transparent candidates choose more extreme policies. Endogenous attention can also lead to multiple equilibria, explaining how poor voters in developing countries can be politically empowered by welfare programs.

Keywords: electoral competition, policy design, rational inattention.

JEL codes: D83, D72.

*We are grateful for comments from Michal Bauer, David Levine, Alessandro Lizzeri, Nicola Gennaioli, Massimo Morelli, Salvo Nunnari, Jakub Steiner, Stephane Wolton, Leet Yariv, Jan Zápal, and seminar and conference participants at Barcelona GSE, Bocconi University, CIFAR, Columbia University, CSEF-IGIER, Ecole Polytechnique, Mannheim, NBER, NYU BRIC, NYU Abu Dhabi, Royal Holloway and University of Oxford.

†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.

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

1 Introduction

Voters are typically very poorly informed about public policies. This is a well known fact, documented by extensive research in political science (eg. Carpini and Keeter 1996, Bartels 1996) and emphasized by classic works like Mill (1861), Schumpeter (1943) and Downs (1957). Nevertheless, voters' ignorance is not uniform nor entirely random. Some voters are more informed than others about many issues, and citizens are generally more informed about what is more important to them. For instance, blacks are generally less informed than whites in the US, but they tend to be relatively more informed about racial policies; women are more informed about education policies than men - see Carpini and Keeter (1996). Moreover, although voters miss a lot of specific details and are affected by seemingly irrelevant events (Achen and Bartels 2004), there is also evidence that they grasp the essentials of major issues (Page and Shapiro 1992). In other words, although voters are uninformed, there are regularities in what they know and don't know, and this is reflected in their views about public policy.

How does this selective ignorance of voters interact with policy formation by politicians? In particular, how can the observed patterns of what voters know be explained, and how does their knowledge depend on the political process? Conversely, how do the endogenous patterns in voters' information influence policy choices by elected representatives? These are the general questions addressed in this paper.

We study a theoretical model in which voters optimally choose how to allocate costly attention, and politicians take this into account in setting policies. In equilibrium, voters' attention to specific issues and public policies are jointly determined and influence each other. We first formulate a general theoretical framework, which we then use to study a number of more specific applications. Policy is set in the course of electoral competition by two vote maximizing candidates, who commit to policy platforms in advance of the 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). The novelty is that here rational but uninformed voters also decide how to allocate costly attention to alternative candidates and to alternative policy issues. We don't study how politicians seek to grab attention, but rather how scarce attention is allocated by voters, and how this influences electoral platforms. Since attention is costly for the voters, they optimally allocate it to what is most important to them - i.e. where their stakes are higher - and to those issues or candidates where the cost of information is lower (because of media coverage or transparency of policies). This in turn affects the incentives of the political candidates, who design their policies so as to increase the visibility of policy benefits and to hide the costs, taking voters' attention as given but also taking into account that different groups of voters may be differently informed. This interaction between optimally inattentive voters and opportunistic candidates gives rise to systematic policy distortions and to other predictions.

First, if policy is one-dimensional, voters with stronger and more extreme policy preferences are more influential in the political process. The reason is that they are more attentive to policy deviations, because they care more about them. Thus, rational inattention amplifies the effects of preference intensity. If the distribution of voters' policy preferences is not symmetric, this entails systematic distortions. In equilibrium, opportunistic politicians aim to please the more extremist voters (who have higher stakes) compared to a standard probabilistic voting model, moving the equilibrium away from the utilitarian optimum. This mechanism can also explain why policy can over-react to novel policy issues, or when the economic environment suddenly changes (eg. after a large financial shock), or to issues where there is genuine uncertainty about the urgency of policy intervention (eg. global warming). This is because, if the policy is also imperfectly observed, the political process is influenced by voters who received more extreme signals about the state of the world or the urgency of the issue, and hence have more extreme policy preferences.

Second, if candidates differ in their informational attributes, voters take this into account. They pay more attention to candidates whose policies are less costly to get information about. Thus, candidates with greater media coverage (typically those favored in the polls or who are more established) attract more attention from all voters, compared to less transparent or less visible candidates. This effect is not uniform across voters, however. Voters with higher stakes find it optimal to pay relatively more attention to the less visible or less transparent candidates, compared to voters with lower stakes. This interaction between voters' attention and candidates' informational attributes implies that the equilibrium displays policy divergence: even if candidates only care about winning the election, and not about the policy per se, different candidates select different equilibrium policies, and in equilibrium have different probabilities of winning. In general, candidates receiving less media attention enact policies that are more favorable to extremist voters, while the more established candidates, who receive more attention from the media and from all voters (and from the centrist voters in particular), choose policies preferred by average voters. Therefore, in equilibrium the more visible candidates have a higher probability of winning the election. This result also implies that both candidates would like to grab more attention, if they could, since this allows them to better explain their policies to the average voter.

Third, if policy is multidimensional, additional distortions arise from selective attention to different policy instruments. Voters pay more attention to the policy instruments that are more important to them, neglecting those instruments where policy deviations are expected to have only marginal effects. This implies that equilibrium public goods that provide benefits to all are under-provided, and general tax distortions affecting everyone are too high, while there is an excessive amount of targeted redistribution (through tax credits or transfers) that only benefits specific groups. The reason is that voters optimally select to pay more attention to targeted instruments compared to general public goods or general taxes. This in turn induces competing candidates to tilt their equilibrium policies away from general public goods and towards targeted transfers, and to rely on general tax instruments even if they are highly distorting. Unlike in other models of electoral competition, this behavior does not result from the asymmetric influence of one group of voters over another. Instead, it reflects the optimal behavior of all voters who choose to pay more attention to some public policies than to others.

Fourth, this framework yields predictions about the pattern of information amongst voters. In equilibrium, voters allocate attention where the stakes are expected to be higher. Thus, voters tend to be more informed about policy instruments on which there is more heterogeneity of preferences, such as targeted redistribution. This is because, if everyone agrees on a policy issue, voters expect politicians to enact optimal policies, they face small stakes from policy deviations around the optimum, and hence they have no incentive to be informed.

Thus, information about, say, defense policy or other general public goods will be very low. On the other hand, information about targeted transfers will be higher, 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 will always be insufficient from the perspective of the beneficiaries. Hence special interest groups are very attentive to possible deviations on these targeted instruments. For the same reason, in a one-dimensional conflict, voters in the middle of the ideological divide will be less informed than those at the extremes (given the same cost of information), because they expect the policy to be about right from their perspective. This is consistent with evidence on US survey data: first, voters with more extreme policy preferences choose to pay more attention to the media (blogs, TV, radio and newspapers) - Ortoleva and Snowberg (2015); second, they are also more informed about the policy positions of presidential candidates - Palfrey and Poole (1987).

Finally, political attention also reflects the opportunity cost of time or psychological stress from poverty, which in turn is directly affected by some public policies. We illustrate this with reference to welfare programs in developing countries. Poor relief programs in Latin America have been found to increase poor voters' participation and attention to politics (Manacorda et al. 2009). Motivated by this finding, we study a simple model of poverty alleviation, where pro-poor policies enable the poor to be more attentive and hence more influential in the political process. This in turn induces politicians to enact more pro-poor policies, giving rise to multiple equilibria that can explain some stylized facts on the political effects of welfare programs in developing countries.

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), or Matějka and McKay (2015). This approach presumes that attention is a scarce resource, even if information is freely available, such as on the internet or in financial journals. Rationally inattentive agents choose how much and what pieces of information to pay attention to. Regarding empirical evidence of endogenous attention, Gabaix et al.(2006), for instance, explore attention allocation in a laboratory setting, and Bartoš et al. (2014) explore attention to applicants in rental and labor markets. Bordalo, Gennaioli and Shleifer (2013, 2015) provide an alternative theoretical framework to study how salience affects choices made by consumers with limited attention.

Although the notion that voters are very poorly informed is widespread (cf. Carpini and Keeter 1996, Lupia and Mc Cubbins 1998), not many papers have attempted to explore the policy implications of this in large elections where voters' information is endogenous and results from the optimal behavior of voters. A closely related contribution is the interesting paper by Gavazza and Lizzeri (2009) on electoral competition with partially uninformed voters. They show that specific patterns of information asymmetries give rise to intertemporal distortions, to under-provision of public goods, and to "churning" (i.e. the same groups receive targeted transfers and pay general taxes, so that net transfers are smaller than gross transfers). The pattern of imperfect information is exogenously given, however, and their equilibrium is supported by particular out of equilibrium beliefs. Our result on policy divergence due to differences in transparency between candidates is related to Glaeser et al (2005). That paper too assumes a specific pattern of exogenous information asymmetries, however. In particular, they assume that core party supporters are more likely to observe a deviation from the expected equilibrium, compared to other voters, in a model with endogenous turnout. In our framework, informational asymmetries are instead endogenous, and everyone votes.[1] Ponzetto (2011) studies a model of trade policy in which 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

[1]Alesina and Cukierman (1990) study the incentives of partisan politicians to hide their ideological preferences from voters.

consumption. This asymmetry in information leads to a political bias against free trade. Ansolabehere et al. (2014) provide evidence that voters' views are biased by the information to which they are exposed as economic agents. Although information is endogenous in these two papers, it is a byproduct of other economic activities, and unlike in our paper, it does not result from a deliberate allocation of attention to the political process. Also, a large literature has explored the political effects of information supplied by the media (see 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.

Our paper is also related to a rapidly growing empirical literature on the economic and political effects of policy instruments with different degrees of visibility (see Congdon et al. 2011 for a general discussion of behavioral public finance). 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. All these findings confirm that policy instruments with different degrees of transparency are not politically equivalent, and directly or indirectly support the theoretical results of our paper.[2]

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

The outline of the paper is as follows. In section 2 we describe the general theoretical framework. Section 3 presents some general results. Section 4 illustrates several applications to specific policy issues. Section 5 concludes. The appendix contains the main proofs.

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:

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

[2]See also the earlier literature on fiscal illusion surveyed by Dollery and Worthington (1996.

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

The distinguishing feature of the model is that voters are uninformed about the candidates' policies, but they can choose how much of costly attention to devote to these policies and their elements. To generate some voters' uncertainty, we assume that candidates target a policy of their choice (which in equilibrium will be known 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 $e_{C,i} \sim N(0, \sigma^2_{C,i})$ is a random variable that reflects implementation errors in the course of the campaign. For instance, the candidate announces a specific target tax rate on real estate, $\hat{q}_{C,i}$, but when all details are spelled out and implemented during the electoral campaign, the actual tax rate to which each candidate commits may contain additional provisions such as homestead exemptions, or for assessment of market value. The implementation errors $e_{C,i}$ are independent across candidates C and policy instruments i, and their variance $\sigma^2_{C,i}$ is given exogenously.[3]

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

3. Voters observe noisy signals of the actual platforms.
4. The ideological bias x^v is realized and elections are held. Whoever wins the election enacts their announced actual policies.

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

2.1 Voters' behavior

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

2.1.1 Imperfect information and attention

All voters have identical prior beliefs about the policy vectors q_C of the two candidates. In the beliefs, elements of the policy vector are independent, and so are the policy vectors of the two candidates. Let each element of the vector of prior beliefs be drawn from $N(\bar{q}_{C,i}, \sigma^2_{C,i})$, where $\bar{q}_C = [\bar{q}_{C,1}, ..., \bar{q}_{C,M}]$ is the vector of prior means, and $\sigma^2_C = [\sigma^2_{C,1}, ..., \sigma^2_{C,M}]$ the vector of prior variances. Note that, to insure consistency, the prior variances coincide with the variance of the implementation errors e_C in (2).[4]

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

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

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

where the noise $\epsilon^{v,J}_{C,i}$ is drawn from a normal distribution $N(0, \gamma^J_{C,i})$, and is iid across voters.[5]

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

$$\xi^J_{C,i} = \frac{\sigma^2_{C,i}}{\sigma^2_{C,i} + \gamma^J_{C,i}} \in [0,1].$$

The more attention is paid by the voter to $q_{C,i}$, the closer is $\xi^J_{C,i}$ to 1. This is reflected by the noise level $\gamma^J_{C,i}$ being closer to zero, and also by a smaller variance $\rho^J_{C,i}$ of posterior beliefs.[6] 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 $\xi^J_{C,i} \geq \xi_0$.

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

$$\sum_{C \in \{A,B\}, i \leq M} \lambda^J_{C,i} \log\left(\sigma^2_{C,i}/\rho^J_{C,i}\right) = - \sum_{C \in \{A,B\}, i \leq M} \lambda^J_{C,i} log\left(1 - \xi^J_{C,i}\right).$$

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

[6] The posterior variance equals $\rho^J_{C,i} = \gamma^J_{C,i}\sigma^2_{C,i}/(\sigma^2_{C,i} + \gamma^J_{C,i})$. Thus, the variable $\xi^J_{C,i}$ also measures the relative reduction of uncertainty about $q_{C,i}$; $\xi^J_{C,i} = 1 - \frac{\rho^J_{C,i}}{\sigma^2_{C,i}}$. The more attention is paid, the closer is $\xi^J_{C,i}$ to 1 and hence the lower is the posterior variance.

The term $-log(1 - \xi^J_{C,i})$ 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 $\lambda^J_{C,i} \in \mathbb{R}_+$ scales the unit cost of information of voter J about $q_{C,i}$. It can reflect the supply of information from the media or other sources, the transparency of the policy instrument $q_{C,i}$, or the ability of voter J to process information.

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:

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

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:

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

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

2.2 Equilibrium

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

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

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

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

2.2.1 Discussion

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

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

The assumption that candidates make random mistakes or imprecisions in announcing the policies is used to generate some 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 could also be other ways of introducing uncertainty, however. For instance, candidates could have unknown partisan or ideological preferences favoring some groups or some policy instruments, or they could have idiosyncratic information about the environment (e.g., the composition of the population of voters). And obviously, voters' uncertainty can also be a behavioral assumption. Most of the qualitative implications of the model would stay unchanged in all of these cases.

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

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

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

Our formulation of the voters' objective, (4), literally states that the voter chooses how much and what form of information to acquire *as if* he were pivotal in his subsequent voting decision. This can be interpreted as saying that voters are motivated by "sincere attention" and want to cast a meaningful vote. That is, they draw utility from voting for the right candidate (i.e., the one that is associated with his highest expected utility), because they consider it their duty (cf. Feddersen and Sandroni 2006) or because they want to tell others (as in Della Vigna et al. 2015). In this interpretation, the parameter $\lambda^J_{C,i}$ captures the cost of attention relative to the psychological benefit of voting for the right candidate.[7]

In line with this interpretation, that voters are motivated by the desire of casting a meaningful vote and not by the expectation of being pivotal, we also assume that voters do not condition their beliefs on being pivotal when they vote. This is the standard approach in the literature on electoral competition, and it is consistent with the fact that in our model the probability of being pivotal is zero (or would be negligible with a large but finite number of voters).[8]

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.[9] There would exists a unique solution

[7] 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 by the factor $1/\delta$, with no substantive change in any result. If the probability of being pivotal was endogenous and part of the equilibrium, the model would become more complicated, but most qualitative implications discussed below would again remain unchanged. The first order condition (8) below would still hold exactly. See however the next paragraph, on how individuals vote without conditioning on being pivotal.

[8] 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.

[9] "Almost any" here denotes functions with sufficient regularity and symmetry across its ar-

to the voter's attention problem, and attention would be increasing in both stakes and uncertainty.

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

3 Preliminary results

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

3.1 A "perceived" social welfare function

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

Let p_C be the probability that C wins the elections. Suppose first that the cost of information is 0, $\lambda^J_{C,i} = 0$. Then our model boils down to standard probabilistic

guments.

voting with full information. The distributional assumptions and the additivity of the preference shocks $x^v = \tilde{x} + \tilde{x}^v$ then imply:

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

The probability that C wins is increasing in the social welfare $\sum_J m^J U^J(q_C)$ that C provides.[10]

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:

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

where the outer expectations operator is indexed by J because voters' attention differ across groups. Obviously, $p_B = 1 - p_A$. For a particular realization of policies, in our model the probability of winning is analogous to (5), except that the voting decision is not based on $U^J(q_C)$, but on $E[U^J(q_A)|s_A^{v,J}]$.[11] The overall probability of winning is then an expectation of this quantity over all realizations of policies and of noise in signals.

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

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

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

In equilibrium, candidate C maximizes the "perceived social welfare" provided by his policies. It is the weighted average of utilities from policy q_C expected by

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

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

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.[12]

Finally, note that the candidates' objective (7) is a concave function of the realized policy vector q_C. This is because: i) For Gaussian beliefs and signals, posterior means depend linearly on the target policy $\hat{q}_C$ set by each candidate, and their variance as well as variances of posterior beliefs are independent of $\hat{q}_C$.[13] 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. Thus, the equilibrium can be characterized by the first order conditions of the objective (7), since they are necessary and sufficient for an optimum.

3.2 Small noise approximations or quadratic utility

In this subsection we introduce an approach that can be used to determine the exact form of the equilibrium. This can be done if utility function is quadratic or if prior uncertainty in beliefs is small, and we can use a local approximation to the utility function. The distinctive feature of our model is that it studies implications of imperfect information for outcomes of electoral competition. Thus,

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

[13]Variance of posterior belief can be expressed in terms of prior variance and the attention vector: $\rho_{J,i} = (1 - \xi_i^J)\sigma_i^2$. Upon acquisition of a signal $s_{C,i}^{v,J}$, the posterior mean is: $\check{q}_{C,i} = \xi_{C,i}^J s_{C,i}^{v,J} + (1 - \xi_{C,i}^J)\bar{q}_{C,i}$, where $s_{C,i}^{v,J} = q_{C,i} + \epsilon_{C,i}^{v,J}$ and $\bar{q}_{C,i}$ denotes the prior mean. Thus, $\check{q}_{C,i} = \xi_{C,i}^J(\hat{q}_{C,i} + e_{C,i} + \epsilon_{C,i}^{v,J}) + (1 - \xi_{C,i}^J)\bar{q}_{C,i}$.

these approximations emphasize the first-order effects of such information imperfection. As shown here, these effects can be highly relevant even if information imperfections are small.

Let us denote by

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

the marginal utility for a voter in group J of a change in the i^{th} component of the policy vector, evaluated at the expected policies. Thus, $u_{C,i}^{J}$ measures intensity of preferences about $q_{C,i}$ in a neighborhood of the equilibrium. Suppose that the noise σ_C^2 is small. Then Appendix 6.2 proves:

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

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

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

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

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

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

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

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

Let us now focus on the voter's problem. How should costly attention be allocated to alternative components of the policy vector? We start with a first order approximation of U in the voters' optimization problem stated in (4). Thus, suppose again that the noise in prior beliefs σ^2_C is small.[14] Then Appendix proves:

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

$$\left(\sum_{C\in\{A,B\},i=1}^{M} \xi^J_{C,i} (u^J_{C,i})^2 \sigma^2_{C,i} \right) + \sum_{C\in\{A,B\},i\leq M} \hat{\lambda}^J_{C,i} log \left(1 - \xi^J_{C,i}\right), \tag{9}$$

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

The form of (9) for second order approximations is presented in (37) in the Appendix.

The benefit of information for voters reflects the expected difference in utilities from the two candidates. If both candidates provide the same expected utility, then there is no gain from information. Specifically, the term $\sum^M_{C\in\{A,B\},i=1} \xi^J_{C,i}(u^J_{C,i})^2\sigma^2_{C,i}$ is the variance of the difference in expected utilities under each of the two candidates, conditional on posterior beliefs. The larger is the discovered difference in

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

utilities, the larger is the gain is, since then the voter can choose the candidate that provides higher utility.

Note also that $\xi^J_{C,i}\sigma^2_{C,i} = (\sigma^2_{C,i} - \rho_{C,i})$ 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.

An immediate implication of (9) is the next proposition.[15]

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

$$\xi^J_{C,i} = \max\left(\xi_0, 1 - \frac{\hat{\lambda}^J_{C,i}}{(u^J_{C,i})^2\sigma^2_{C,i}}\right). \tag{10}$$

Quite intuitively, the solution (10) implies that, for a given cost of information $\hat{\lambda}^J$, the voter pays more attention to those elements $q_{C,i}$ for which the unit cost of information $\lambda^J_{C,i}$ is lower, i.e. are more transparent, prior uncertainty $\sigma^2_{C,i}$ is higher, and which have higher utility-stakes $|u^J_{C,i}|$ from changes in $q_{C,i}$. Note that for any convex information-cost function $\Gamma(\xi^J)$, the objective (9) would be concave, and thus there would exist a unique maximum, which would solve $\partial\Gamma(\xi^J)/\partial\xi^J_{C,i} = Min(\psi, \phi)(u^J_{C,i})^2\sigma^2_{C,i}/2$. The effect of stakes and uncertainty also holds more generally.[16]

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

[15]The solution for second order approximation is in (38).

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

mostly based on monotonicity, we believe that they are robust to slight changes of its assumptions.

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

The appendix also solves a second order (rather than first order) approximation of the voters' optimization problem, which is of course exact for quadratic utilities. In this case, the optimal attention ξ^J is given by (38), only a slightly more complicated formula than in (10), and its qualitative properties remain almost the same. The difference is that if voters are not risk-neutral, then they acquire information not just to make a better choice of which candidate to vote for, but also to decrease uncertainty conditional on a chosen candidate. The voters' optimality condition then contains an additional term, which implies that voters' attention is higher than stated in (10). This additional term is larger the greater is prior uncertainty, $\sigma^2_{C,i}$.

4 Applications

In this section we present three examples to illustrate some basic implications of inattentive voters. Throughout, we compare the equilibrium with rational inattention and the equilibrium with fully informed voters, which, as stated above, coincides with the utilitarian optimum. We start with electoral competition on a one-dimensional policy, then turn to the choice of multi-dimensional policies, and finally show that rational inattention can lead to multiple equilibria.

4.1 One dimensional conflict

This example explores the effects of rational inattention on equilibrium policy outcomes in a simple setting. 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 λ^J, for now assumed to be the same for all candidates C. The voters' utility function is

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

$q \in \mathbb{R}$ and $U(.)$ is concave and symmetric about its maximum at 0. Political disagreement is often one-dimensional, as policy preferences tend to be aligned along left-to-right ideological positions (see Poole and Rosenthal 1997).

With a one dimensional policy, by Proposition 1 the equilibrium with rational inattention can be computed as the solution to a modified social planning problem, where each candidate C maximizes $\sum_J m^J \xi_C^J U^J(q_C)$.

By (10), voters' attention increases with the distance $|\hat{q}^* - t^J|$, where $\hat{q}^*$ denotes the equilibrium policy target. The reason is that the utility stakes increase in this distance, due to concavity of U^J. If the cost of collecting information $\hat{\lambda}^J$ is the same for all groups of voters, then more extreme groups pay more attention to q_C. As a result, the extremists receive a higher weight in the modified planner's problem and are more influential, compared to the utilitarian optimum. Groups with a lower cost $\hat{\lambda}^J$ also receive a greater weight, for the same reason.

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

The specific implications for how the equilibrium differs from that with full information depend on the shape of the distribution of bliss-points t^J. If the distribution is asymmetric, then voters in the longer tail pay relatively more attention, and thus equilibrium under rational inattention is closer to them relative to the perfect information equilibrium. For instance, suppose that q refers to the size of government, or to a proportional income tax. Since income distribution is skewed to the right, and the rich prefer lower taxes, the distribution of bliss points t^J is then skewed to the left. In this case, the equilibrium policy under rational inattention moves to the left compared to the socially optimal policy. That is, the rich exert a disproportionate influence over the equilibrium, and the size of government is smaller than optimal. This effect is reinforced if, as is plausible, the rich also have a lower cost of gathering information (i.e. a lower $\hat{\lambda}^J$).

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

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

The policy divergence emerges because candidate A assigns a greater weight to the more extreme voters compared to candidate B, since these voters are more attentive to his policies given their higher stakes. Thus, in the size of government interpretation, the less established candidate (A) would announce a policy more favorable to the rich, compared to candidate B for which information is more easily available. More generally, this suggests that more established candidates tend to cater to the average voter, while candidates receiving less media coverage go after extremist voters. With policy divergence and different attention weights, the probability of victory differs from 1/2, and the less transparent candidate A (who receives less attention by all voters and by the centrist voters in particular) is less likely to win (since $\xi_B^J > \xi_A^J$ for all J, the value of the objective function $\sum_J m^J \xi_C^J U^J(q_C)$ at the optimum will be larger for B than for A).

To illustrate these findings, let there be three types of voters of equal masses such that $t^1 = t^2 = \frac{1}{2}$ and $t^3 = -1$. Let us also assume $U^J(q) = -(q - t^J)^2$ - thus the two candidates are identical and announce the same policies. Under perfect information, $\hat{\lambda} = 0$, the equilibrium policy coincides with the social optimum, $q = 0$. It is the average of the bliss-points in the population. However, when the cost of information increases, the equilibrium q decreases.

Figure 1 presents the equilibrium q as a function of $\hat{\lambda}$. The solid curve represents the exact solution using (38) in the Appendix, and the dashed curve is based on the first order approximation, (10). The left panel shows results for $\sigma_C^2 = 0.05$. There, when $\hat{\lambda} = 0.01$, then $q \doteq -0.02$, when $\hat{\lambda} = 0.05$, then $q \doteq -0.13$, and when $\hat{\lambda} = 0.1$, then $q \doteq -0.23$. For positive costs of information, the extreme voters $J = 3$ pay relatively more attention than $J = 1$ and $J = 2$ when q is in the neighborhood of zero, and thus the equilibrium policy moves in their direction.[17]

[17]When the cost of information increases beyond a certain level, then attention becomes uni-

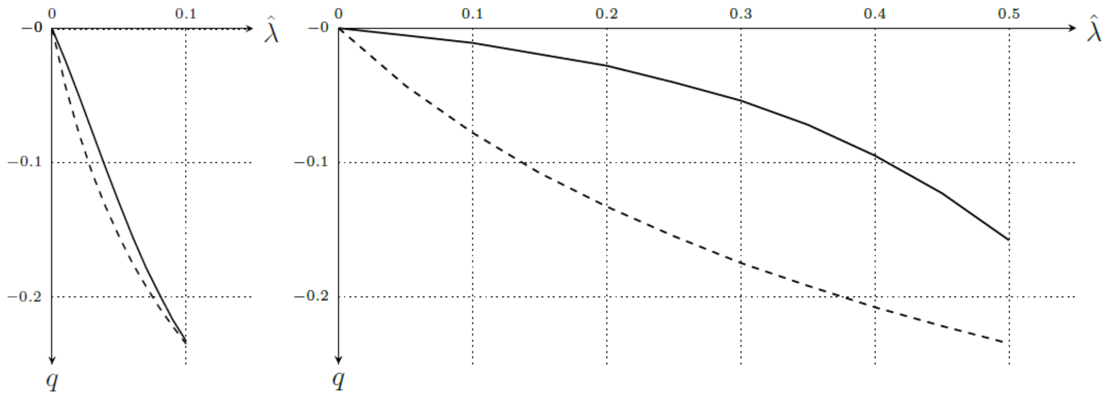

Figure 1: Effect of the cost of information, left: $\sigma_C^2 = 0.05$, right: $\sigma_C^2 = 0.25$, solid: exact solution, dashed: first-order approximation.

Note that here the variance of prior uncertainty about policies is of moderate size: it is one tenth of the total variance of bliss points in the population. We can see that the first order approximation works quite well here.

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

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

If the cost of attention is heterogeneous across voters, then the equilibrium policy reflects that, too. Preferences of voters with a lower marginal cost weigh

form again since all voters are at the lower bound for attention, ξ_0. Once this lower bound is reached, policy is again at the social optimum since all voters are weighted equally.

more in equilibrium. For instance for $\sigma_C^2 = 0.05$, if $\hat{\lambda}^3 = 0.01$ and $\hat{\lambda}^1 = \hat{\lambda}^2 = 0.1$, then in equilibrium $q = -0.34$, policy is closer to the more attentive voters $J = 3$.

Finally, this example can also speak to how elections aggregate dispersed information on other issues. Suppose that there is uncertainty about the benefit of addressing a specific issue, say global warning or financial instability, while the cost is well known. Voters receive different realizations of noisy signals about the unknown benefit, and this induces heterogeneous beliefs and hence heterogeneity in policy preferences. Our findings imply that policy can over-react to such issues. The reason is that voters with extreme beliefs are more attentive to the policy, because they have more at stake, and thus are more influential in the electoral competition. This is interesting because if voters are fully informed about the policy itself, then the equilibrium policy typically under-reacts to imperfect information about a new issue (since prior beliefs dampen the reaction to shocks). This can explain why a large shock that is interpreted differently by different voters, like the recent financial crisis, could lead to over-reactions (eg. excessive financial regulation).

4.2 Targeted transfers and public good provision

When the policy is multi-dimensional, rational inattention has additional implications, because voters also have to 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. This in turn affects the politicians' incentive. In this example we show that rational inattention leads to under-provision of public goods and over-reliance on distorting taxes in order to finance targeted redistribution.

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

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

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

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

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

$$U^J(q) = y - b^J - T(\tau) - S(g - \sum_K b^K - N\tau)/N + H(g). \tag{11}$$

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.

It is easy to verify that the socially optimal policy vector satisfies $s = \tau = 0$, i.e., eliminates all distorting taxes, and sets the public good so as to satisfy Samuelson optimality condition; namely $H'(\hat{g}) = 1/N$. Thus the optimal level of the public good is financed through targeted lump sum taxes. The allocation of tax burden across groups is indeterminate because of linearity in consumption.

Next consider the policy outcome under electoral competition. To express the first order conditions (8) we use: $u^J_J = -1 + S'/N$, $u^J_{-J} = S'/N$, $u^J_\tau = T' - S'$ and $u^J_g = H' - S'/N$, 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 with respect to g and τ respectively; all derivatives are evaluated at the equilibrium policy targets. The equilibrium 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:

$$
\begin{aligned}
-S'/N + H' &= 0 \qquad (12)\\
-T' + S' &= 0 \qquad (13)
\end{aligned}
$$

The reason is that all types J pay the same level of attention to g and τ, and thus ξ^J_g and ξ^J_τ do not enter these expressions.[18] What could drive equilibria away from the social optimum is heterogeneity in ξ^J_i across different voters, only, which does not arise with these uniform tax instruments.

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

$$\xi^J_J(-1 + S'/N) + (N-1)\xi^J_{-J}S'/N = 0$$

or equivalently as:

$$[1 + (N-1)\frac{\xi^J_{-J}}{\xi^J_J}]S'/N = 1 \qquad (14)$$

At the social optimum, $S' = 1$ (since $s = 0$), which in turn implies that $\xi^J_{-J} < \xi^J_J$, since $N > 2$ - cf (10). Namely, at the socially optimal policy, all groups pay more attention to their own taxes than to taxes paid by other groups. But if $\xi^J_{-J} < \xi^J_J$, then equation (14) implies $S' > 1$, a contradiction. Hence in equilibrium, it must be that $S' > 1$, and hence that $\hat{s} > 0$. Equations (12)-(13) then imply that $H\prime > 1/N$ and that $T' > 1$. Thus, in equilibrium there is under-provision of the public good

[18]This can be seen from (10) and from the fact that u^J_τ and u^J_g are common to all voters.

relative to the social optimum, and the government relies on distorting (observable and unobservable) sources of revenues, despite the availability of lump sum taxes. In fact, if the marginal tax distortions T' and S' do not rise too rapidly, it is even possible that the equilibrium entails negative values of $\hat{b}^J$. That is, both candidates collect revenue through distorting taxes from all citizens, and then give it back to each group in the form of targeted transfers (i.e. there is fiscal churning). The source of these distortions is the asymmetry in attention: voters pay more attention to the targeted instruments, because (in equilibrium) the stakes are higher, and they neglect the instruments that have the same effects on all citizens, for the same reason. Moreover, they pay more attention to their own targeted taxes (or transfers) than to the targeted instruments affecting others. This in turn induces both candidates to deviate from efficient allocation, in order to appear to please each group. The higher is the cost of information λ and the larger is N, the larger is the distortion

Finally, note that in equilibrium $u_\tau^J = T' - S' = 0$ and $u_g^J = H' - S'/N = 0$. By (10) this in turn implies that $\xi_g^J = \xi_\tau^J = \xi_0$. Namely, in equilibrium all voters pay minimal attention to public goods and to the uniform distorting tax, as if they were non-observable. The reason is that there is no disagreement amongst voters regarding these policy instruments, and hence all voters expect both candidates to set these general instruments at their optimal values (from the individual voter's selfish perspective). Given these prior beliefs and the first order approximation, voters have no incentive to devote costly attention to these items. This does not apply to targeted taxes, where there is disagreement amongst voters, and where the individual returns from attention are higher.[19]

The result that in equilibrium voters are inattentive to policies on which everyone agrees (such as g and τ in the model) while they pay attention to divisive issues (such as targeted instruments), is consistent with existing evidence on the content of Congressional debates and on the focus of US electoral campaigns. Ash et al.

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

(2015) construct indicators of divisiveness in the floor speeches of US congressmen. Exploiting within-legislator variation, they show that the speeches of US senators become more divisive during election years, consistently with the idea that voters' attention is greater on the more divisive issues. Moreover, Hillygus and Shields (2008) show that divisive issues figure prominently in US presidential campaigns, contrary to the expectation that candidates instead try to avoid divisive policy positions in order to win more widespread support.

The result that lack of information implies fiscal churning and under-provision of public goods is similar to findings in Gavazza and Lizzeri (2009). In that paper, however, the pattern of information is exogenous and does not result from the optimal allocation of attention by voters. Moreover, the equilibrium is sustained by particular out of equilibrium beliefs. Gavazza and Lizzeri also argue that exogenous provision of information on taxes vs spending has opposite welfare effects, with more information on spending being welfare improving, while information on taxes is counter-productive. 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. Specifically, the equilibrium would become less distorted if the cost of information on instruments targeted at others (λ^J_{-J}) fell, while the cost of information on instruments targeted at themselves (λ^J_J) increased. This can be seen from (14): a higher λ^J_J and a lower λ^J_{-J} would raise the ratio $\frac{\xi^J_{-J}}{\xi^J_J}$, leading to less seignorage, more public good provision and less distorting taxation. Intuitively, voters would pay more attention to benefits targeted at other groups, raising the political costs of targeting. Of course, there is a limit to how much these costs can be exogenously changed through increased fiscal transparency, since the cost of observing instruments targeted at one-self will generally be lower than the cost of instruments targeted at others (see Ponzetto (2011) for a specific example of this point with regard to trade policy). Moreover, transparency is also a policy choice, and it is not clear that politicians would always benefit from it.

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.[20]

4.3 Empowering the poor

In the previous examples, the cost of political attention is exogenously given. In this subsection we consider what happens when policy affects the opportunity cost of time, and hence the cost of political attention. The example that follows is motivated by the observations in Mani et al. (2013) and Banerjee and Mullainathan (2008), that often poor individuals in developing countries are impaired in their cognitive functions by the stress induced by survival activities. As suggested by Mani et al. (2013), "poverty-concerns consume mental capacities, leaving less for other tasks". Poverty alleviation by the government can thus free up human resources and empower the poor, making them more effective in their social activities, including politics. Conversely, an absence of welfare programs directed towards the poor leaves them hampered not only in their material interests, but also in their ability to influence the political process.

[20]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).

In other words, a complementarity is at work: pro-poor policies make the poor more attentive to and influential in the political process, which in turn reinforces the political inclination to support the poor. Vice versa, an absence of effective welfare programs forces the poor to devote almost exclusive attention to survival activities, de facto excluding them from the political process and reinforcing the anti-poor political bias. This can explain why otherwise similar societies might end up on different political and economic trajectories. This multiplicity result is reminiscent of those emphasized by Benabou and Tirole (2006) and Alesina and Angeletos (2005), but the mechanism at work is quite different.

To illustrate this idea, suppose that there are two equally sized groups, the rich and the poor, indexed by $J = R, P$. The rich have income ω and enjoy linear utility from consumption. The income of the poor, y, depends on their effort, e. Effort can be high ($\bar{e}$) or low ($\underline{e}$). High efforts gives higher income ($\bar{y}$) but entails high disutility costs, $\bar{d}$. Low effort gives lower income ($\underline{y}$) but entails low disutility costs $\underline{d}$. The poor's utility from consumption is strictly concave, $U(.)$, with $u(.)$ denoting the marginal utility of consumption for the poor.

Policy consists of a lump sum subsidy to the poor, s, financed by a corresponding lump sum tax on the rich. Thus, the indirect utility function of the rich is: $W^R(s) = \omega - s$, and the indirect utility function of the poor is $W^P(s) = U(y+s) - d$, where y and d can be high or low, depending on the choice of effort.

The choice of effort by the poor depends on the expected subsidy. Let $\bar{s}$ denote the prior mean of the subsidy that will be enacted by both candidates. That is, as in the previous sections, voters have prior beliefs about the forthcoming subsidy, these beliefs are normally distributed, with mean $\bar{s}$ and variance σ^2, and are the same for both candidates. Let $\tilde{s}$ denote the value of the prior mean that leaves the poor indifferent between choosing high or low effort. It is easy to verify that $\tilde{s}$ is defined implicitly by:

$$\int [U(\bar{y} + s) - U(\underline{y} + s)] dN(\tilde{s}, \sigma^2) = \bar{d} - \underline{d} \qquad (15)$$

By concavity of $U(.)$, if $\bar{s} \geq \tilde{s}$ then the poor choose low effort, and if $\bar{s} < \tilde{s}$ they choose high effort.

Throughout, we assume that the income of the rich ω is sufficiently large, and that $\bar{y} - \underline{y} > \bar{d} - \underline{d} > 0$. Then the socially optimal subsidy s^* equates the marginal utility of income of rich and poor individuals, and induces high effort by the poor; it is defined by $u(\bar{y} + s^*) = 1$.[21]

Now consider the equilibrium under electoral competition with rational inattention. Suppose that the (rescaled) cost of information by the rich is $\hat{\lambda}^R = \hat{\lambda}$, while the cost of information for the poor can be high or low, depending on their choice of economic effort. If economic effort is high ($e = \bar{e}$), then the poor have little time left for political attention, and the cost of information for poor voters is also high, $\hat{\lambda}^P = \hat{\lambda}^h$. Conversely, if economic effort by the poor is low ($e = \underline{e}$), then they can afford to spend more time on political attention, and their cost of information is low, $\hat{\lambda}^P = \hat{\lambda}^l$, with $\hat{\lambda}^h > \hat{\lambda}^l$.

The timing of events is as follows. First, voters form their prior beliefs and choose their attention strategies, and the poor choose effort levels. Then candidates choose target policies and actual policies are realized. Finally, voters gather information and vote. The actual policy s is imperfectly observed, as in the previous sections. Repeating the previous steps, and considering the small noise approximation, by Proposition 1 the equilibrium policy target solves

$$Max_s[\xi^R W^R(s) + \xi^P W^P(s)],$$

taking the choice of effort by the poor and the weights ξ^J as given. The optimality condition for the equilibrium policy target can be written as.

$$u = \frac{\xi^R}{\xi^P} \tag{16}$$

where the poor's marginal utility of income, u, is computed at the equilibrium policy target, and where as before $\xi^J = Max[\xi_0, 1 - \frac{\hat{\lambda}^J}{\sigma^2 (W_s^J)^2}]$, with W_s^J denoting the derivative of $W^J(s)$ with respect to s. After some simplifications, and neglecting

[21]If instead $0 < \bar{y} - \underline{y} < \bar{d} - \underline{d}$, then the optimal subsidy would still set the marginal utility of the poor equal to 1 (when evaluated at low income $\underline{y}$), but it would induce low effort by the poor. Nothing important hinges on this, although the first case seems more plausible.

the lower bound in ξ, (16) can be rewritten as:

$$\sigma^2 u^2 + (\hat{\lambda} - \sigma^2)u - \hat{\lambda}^P = 0 \tag{17}$$

where $\hat{\lambda}$ is the cost of information for the rich. Equation (17) can be solved for u, selecting the positive root to avoid negative marginal utility, and this yields:

$$u = F(\hat{\lambda}^P) \equiv \frac{\sigma^2 - \hat{\lambda} + \sqrt{(\sigma^2 - \hat{\lambda})^2 + 4\sigma^2\hat{\lambda}^P}}{2\sigma^2} \tag{18}$$

Equation (18) thus pins down the marginal utility of the poor in equilibrium. Note that the function $F(\hat{\lambda}^P)$ is increasing in $\hat{\lambda}^P$ and at the point $\hat{\lambda}^P = \hat{\lambda}$ we have $F(\hat{\lambda}^P) = 1$. Thus, if the marginal cost of information of rich and poor is the same (i.e. if $\hat{\lambda}^P = \hat{\lambda}$), then (18) implies $u = 1$, as in the social optimum. If, on the other hand, $\hat{\lambda}^P > \hat{\lambda}$, then in equilibrium $u > 1$; namely the rich are more influential because they pay more attention, and the equilibrium policy stops short of equalizing the marginal utility of rich and poor individuals. More generally, the higher the information costs of the poor $\hat{\lambda}^P$, the higher is their marginal utility u in equilibrium, and hence the smaller are equilibrium subsidies. Thus, equilibrium subsidies are a decreasing function of $\hat{\lambda}^P$, the information costs of the poor. This can be seen formally. Inverting u we obtain the equilibrium subsidy targeted by both candidates as a function of $\hat{\lambda}^P$, namely

$$\hat{s} = u^{-1}[F(\hat{\lambda}^P)] - y \equiv S(\hat{\lambda}^P) - y \tag{19}$$

Since $F(.)$ is increasing and u^{-1} is decreasing, the function $S(.)$ is decreasing in $\hat{\lambda}^P$.

An important implication of (19) is that there may be multiple equilibria. Suppose that the poor expect that in equilibrium both candidates will announce low subsidies, so that their prior mean is in the range $\bar{s} < \tilde{s}$. Then they devote high economic effort, their cost of information is high ($\hat{\lambda}^P = \hat{\lambda}^h$), and their income is also high $y = \bar{y}$. By (15) and (19) this is indeed an equilibrium, call it $\hat{s}^h$, if $\hat{s}^h = S(\hat{\lambda}^h) - \bar{y}$ and if $\hat{s}^h = \bar{s} < \tilde{s}$. The other equilibrium is obtained under the

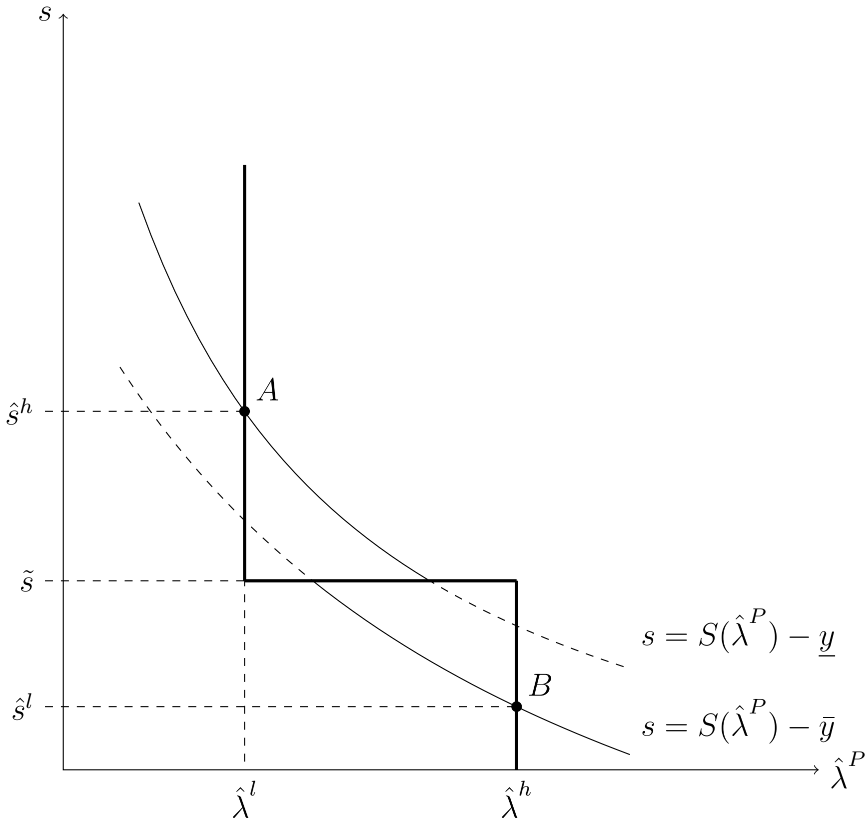

Figure 2: Two equilibrium levels of subsidy.

assumption that the poor expect both candidates to announce high subsidies, so that the prior mean is in the range $\bar{s} > \tilde{s}$. In this case, the poor exert low effort, their cost of information is low ($\lambda^P = \hat{\lambda}^l$), and their income is low as well, $y = \underline{y}$. In this second equilibrium, call it $\hat{s}^l$, equilibrium subsidies are $\hat{s}^l = S(\hat{\lambda}^l) - \underline{y}$ and $\hat{s}^l = \bar{s} > \tilde{s}$. Since $S(.)$ is increasing in $\hat{\lambda}^P$, and since $\hat{\lambda}^h > \hat{\lambda}^l$ and $\bar{y} > \underline{y}$, we must have $\hat{s}^l > \hat{s}^h$. Existence of multiple equilibria thus requires that the prior mean that leaves the poor indifferent between exerting high or low effort, $\tilde{s}$, lies in between these two values, namely $\hat{s}^l > \tilde{s} > \hat{s}^h$.

The equilibria are illustrated in Figure 2. The stepwise boldface function depicts how the poor's information cost λ^P varies with subsidies. By (15), at $\hat{s} = \tilde{s}$ the poor are just indifferent between high and low effort. For $\hat{s} > \tilde{s}$, they exert low effort into economic activities, freeing up attention for politics, thus their cost of attention is low ($\lambda^P = \hat{\lambda}^l$). And viceversa, if $\hat{s} < \tilde{s}$ then the poor find it optimal to devote more time to survival activities and their cost of political attention is high ($\hat{\lambda}^P = \hat{\lambda}^h$). The downward sloping lines depict the subsidies targeted in political equilibrium, corresponding to (19). There are two lines, because the poor's income can be high or low, depending on expected subsidies. If $\hat{s} < \tilde{s}$ then economic effort is high and so is income, $y = \bar{y}$. Vice versa, if $s > \tilde{s}$, then economic effort is low and $y = \underline{y}$. The two equilibria in pure strategies are at points A and B in Figure 2, where the political equilibrium curve intersects the stepwise function of the information costs.

At point B, the poor expect both candidates to enact low subsidies. Hence they are forced to allocate their attention away from politics and into survival activities. Their cost of gathering political information is high, which makes them less influential. Both candidates then find it optimal to enact policies that please the rich, and thus make the expectations of the poor self-fulfilling. Vice versa, at point A, the poor expect the political process to lead to more favorable policies and high subsidies, and this is indeed delivered by the political process.[22]

Of course the model is highly stylized, and its main purpose is to illustrate some implications of endogenous attention. Nevertheless, the evidence on the political effects of welfare programs in Latin America is consistent with this simple example. A large literature finds that federal support programs for the poor in Latin America, such as the Progresa program in Mexico or similar programs in other countries, are associated with increased participation by the poor in national elections, and increased interest in politics by the poor - see for instance De la O (2013) on Mexico, Manacorda et al. (2009) on Uruguay, Baez et al. (2012) on Colombia. More importantly, Idoux (2015) finds that in Mexico, municipalities that were included in the federal Progresa program allocate a greater fraction of local spending towards projects benefiting the poor. That is, where the federal

[22]This simple model could yield multiple equilibria even under a benevolent government. This is because the assumed timing (effort is chosen before the government commits to a subsidy) implies that government policy lacks credibility. This can be seen also in Figure 2, where in a neighborhood of $\hat{s} = \tilde{s}$ one or the other downward sloping equilibrium curve could be the relevant one depending on the expectations of the poor. The political mechanism stressed in this example, however, is quite different from the traditional time inconsistency argument.

government alleviates poverty, the poor participate more in politics and local governments also adopt pro-poor policies. An interpretation of these findings by Idoux (2015) is precisely that these federal welfare programs induced poor voters to pay more attention to politics, because they changed their prior beliefs about what the political process could deliver, and perhaps because it freed up some of their scarce time. This made the poor voters more influential, and as a result local politicians also started to enact policies more in line with their demands.

5 Concluding remarks

Voters tend to be poorly informed about policy issues raised during an electoral campaign, and about the political process in general. This fact is well known and undisputed. Nevertheless, not much is known about the specific patterns of voters' lack of information, and how it interacts with the behavior of politicians. This paper seeks to fill this gap, studying how voters allocate costly attention in a simple model of electoral competition. The approach of this paper could be extended to study several other aspects of the political process.

Perhaps the single most important future extension is competition for voters' attention. Here politicians react to the attention strategies of voters, but they don't take any action to grab attention. If they could, they would like to attract more attention, so as to better explain their policy platforms. This can be seen, for instance, from the candidates' objective function in Subsection 4.1, that increases in the attention weights. Studying how active competition for voters' attention changes politicians' behavior in the course of electoral campaigns or in primaries, and how this depends on voters' behavior, is an important open question.

Addressing this question could also shed light on the role of parties, as ideological labels that save voters' attention.[23] By consistently taking positions in defense of specific economic interests, or according to specific ideological views,

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

political parties can save voters the cost of collecting information on different issues or over time. This role of parties as labels can be illustrated by a simple extension of the one-dimensional policy application discussed in Subsection 4.1. Suppose that there is one national electoral district and two regional districts. A one dimensional policy has to be chosen at each level of government, and voters care about both the national and regional policies. The three elections are run simultaneously. Each voter participates in two elections, in his region and in the nation. There are two political parties, each running in all three elections. But now suppose that, before voters choose attention, each party chooses whether to coordinate policy across elections, or to let the policy be set independently at the regional vs national level. Coordination amounts to a commitment to run on the same electoral platform at the national and regional level.

The important piece here is that voters know whether polices are set nationally, or independently across regions. The presence of a party organization allows for such labeling across electoral districts. The advantage of a coordinated policy is that, by increasing the voters' stakes, it increases their attention. If the policy is coordinated, then attention devoted to this policy is useful in two elections (regional and national) rather than in one only. If voters draw the same utility from the national and the regional policy, coordination has the same effect as a four-fold reduction in λ (see (10), where stakes enter squared). As a result, the equilibrium policy gets closer to the social optimum and this increases the party's probability of winning both elections (see the example in Subsection 4.1). This benefit of a single coordinated policy is offset by the cost of a worse local fit; the cost is higher the more voters' policy preferences differ across districts. Under perfect information, both parties would always prefer full decentralization, rather than a single coordinated policy. But if heterogeneity is not too large and the cost of attention is high, then it can be shown that both parties may prefer to coordinate national and regional policies, so as to grab more attention. Similar forces may be at work in a dynamic setting, where electoral platforms could be coordinated over time and across policy issues. Exploring more in detail this role of political parties as ideological labels when voters are inattentive is a promising direction for future research.

A second set of issues that could be fruitfully studied in this framework is the endogenous supply of information, by the media or by political actors. In this paper we have focused on what induces voters to collect and process information, when it is costly. A natural theoretical extension is to imbed this in a more general framework, where available information is not random, but originates from the equilibrium behavior of others, such as media or interest groups. This would entail abandoning the simplifying assumption that the signals received by voters are independent. It would also entail studying the incentives of whoever provides this information, and how this interacts with rational inattention. The literature on lobbying has studied the role of organized groups in providing information to voters, but much of this literature makes very demanding assumptions on the voters' ability to process information (eg. Coate 2004, Prat 2006). Studying how individuals choose to pay attention to information provided by others (media or lobbies), and how this interacts with electoral competition, is a difficult but important area for future research.

Finally, 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.[24]

[24] Prato and Wolton (2015) study a signalling model where voters' attention can endogenously be high or low. Diermeier and Li (2015) study electoral control by behavioral and non-strategic voters.

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

Alesina, Alberto, and George-Marios Angeletos (2005), "Fairness and Redistribution: Us Vs. Europe," *American Economic Review*, 95, 913-35.

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

Ansolabhere , Stephen, Marc Meredith and Eric Snowberg (2014), "Mecro-Economic Voting: Local Information and Micro_perceptions of the macro_Economy", *Economics and Politics*, Vol. 6 (3): 380-410

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

Baez, Javier E., Adriana Camacho, Emily Conover, and Roman A. Zarate (2012), "Conditional Cash Transfers, Political Participation and Voting Behavior," World Bank Working Paper Series 6215.

Banerjee, Abhijit V., and Sendhil Mullainathan (2008), "Limited attention and income distribution," *American Economic Review*, 98(2), 489-493.

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: (2014), *"Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition,"* *IZA Discussion Paper*, 3, 8058.

Benabou, Roland and Jean Tirole (2006), *"Belief in a just world and redistributive politics,"* *The Quarterly Journal of Economics*, 121(2), 699-746.

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

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

Bordalo, P., N. Gennaioli and A. Shleifer (2015), "Competition for Attention", *Review of Economic Studies*, forthcoming

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.

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

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.

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

De la O, Ana L. (2013), "Do Conditional Cash Transfers Affect Electoral Behavior? Evidence from a Randomized Experiment in Mexico," *American Journal of Political Science*, 57(1), 1-14.

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", Harper and Row.

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.

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.

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.

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

Idoux, Clémence (2015), "Local policy feedback to wide national programs: Evidence from Mexico", Mimeo, Università Bocconi.

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

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.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao (2013), "Poverty Impedes Cognitive Function", *Science*, 341(976), 976-980.

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.

Mill, John Stuart (1861), "*Considerations on Representative Government,*" Parker, Son, & Bourn.

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.

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

Persson, Thorsten, 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.

Poole, Keith, and Howard Rosenthal (1997), *"Congress: A Political-Economic History of Roll-Call Voting,"* Oxford University Press.

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 Stephane Wolton (2015), "Rational Ignorance, Elections and Reform", Georgetown University, mimeo

Schumpeter, Joseph A. (1943), "*Capitalism, Socialism and Democracy,*" Unwin University Books.

Selten, Reinhard. (1975), "Reexamination of the perfectness concept for equilibrium points in extensive games," *International journal of game theory*, 4(1), 25–55.

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

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

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

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

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:

$$\tilde{x}^v = E[U^J(q_A)|s_A^{v,J}] - E[U^J(q_B)|s_B^{v,J}] - \tilde{x} \equiv \tilde{x}_T^{v,J} \tag{20}$$

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:

$$\pi_A = \sum_J m^J E_\epsilon^J[\Pr(\tilde{x}^v \leq \tilde{x}_T^{v,J})] = \frac{1}{2} + \phi \sum_J m^J E_\epsilon^J[\tilde{x}_T^{v,J}] \tag{21}$$

Note that (21) holds when the noise in the ideological preference shocks $\tilde{x}^v$ is sufficiently large to affect the vote with positive probability.[25]

By (20)-(21), 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

[25]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.

candidate A is then:

$$p_A = \frac{1}{2} + \psi \left(\sum_J m^J E^J_{\epsilon,q_A,q_B} \left[E[U^J(q_A)|s^{v,J}_A] - E[U^J(q_B)|s^{v,J}_B] \right] \right) \qquad (22)$$

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.

6.2 Small noise approximations or quadratic utility

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

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

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

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 is:

$$\begin{aligned} E[U^J(q_C)|s^{v,J}_C] &\simeq E[\tilde{U}^J(q_C)|s^{v,J}_C] = \\ &= U^J(\bar{q}_C) + \sum_{i=1}^{M} u^J_{C,i}(\check{q}_{C,i} - \bar{q}_{C,i}) \\ &\quad + \frac{1}{2}\sum_{i,j=1}^{M,M} u^J_{C,i,j} E\left[(q_{C,i} - \bar{q}_{C,i})(q_{C,j} - \bar{q}_{C,j})|s^{v,J}_C\right], \end{aligned} \qquad (23)$$

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

$$
\begin{aligned}
& \frac{1}{2}\sum_{i,j=1}^{M,M} u^J_{C,i,j} E\Big[\Big((q_{C,i}-\check{q}_{C,i})-(\bar{q}_{C,i}-\check{q}_{C,i})\Big)\Big((q_{C,j}-\check{q}_{C,j})-(\bar{q}_{C,j}-\check{q}_{C,j})\Big)|s^{v,J}_C\Big] \\
= \; & \frac{1}{2}\sum_{i,j=1}^{M,M} u^J_{C,i,j}(\check{q}_{C,i}-\bar{q}_{C,i})(\check{q}_{C,j}-\bar{q}_{C,j}) + \frac{1}{2}\sum_{i=1}^{M} u^J_{C,i,i}(1-\xi_{C,i})\sigma^2_{C,i}. \qquad (24)
\end{aligned}
$$

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

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 (24), which is

$$
\begin{aligned}
& \frac{1}{2}E_{\epsilon,e}\Big[\sum_{i,j=1}^{M,M} u^J_{C,i,j}\xi^J_{C,i}\xi^J_{C,j}(\hat{q}_{C,i}+e_i+\epsilon^J_{C,i}-\bar{q}_{C,i})(\hat{q}_{C,j}+e_j+\epsilon^J_{C,j}-\bar{q}_{C,j})\Big] \\
= \; & \frac{1}{2}\sum_{i=1}^{M} u^J_{C,i}(\xi^J_{C,i})^2(\sigma^2_{C,i}+\frac{1-\xi^J_{C,i}}{\xi^J_{C,i}}\sigma^2_{C,i}) \\
& +\frac{1}{2}\sum_{i,j=1}^{M,M} u^J_{C,i,j}\xi^J_{C,i}\xi^J_{C,j}(\hat{q}_{C,i}-\bar{q}_{C,i})(\hat{q}_{C,j}-\bar{q}_{C,j}), \qquad (25)
\end{aligned}
$$

where $\frac{1-\xi^J_{C,i}}{\xi^J_{C,i}}\sigma^2_{C,i}$ is the variance of $\epsilon^J_{C,i}$. Putting (23)-(25) together, we get

$$
\begin{aligned}
E_{\epsilon,e}\Big[E[U^J(q_C)|s^{v,J}_C]\Big|\hat{q}_C\Big] \; \simeq \; & U^J(\bar{q}_C) + \sum_{i=1}^{M}\xi^J_{C,i}u^J_{C,i}(\hat{q}_{C,i}-\bar{q}_{C,i}) + \frac{1}{2}\sum_{i=1}^{M}u^J_{C,i,i}\sigma^2_{C,i} \\
& +\frac{1}{2}\sum_{i,j=1}^{M,M} u^J_{C,i,j}\xi^J_{C,i}\xi^J_{C,j}(\hat{q}_{C,i}-\bar{q}_{C,i})(\hat{q}_{C,j}-\bar{q}_{C,j}). \qquad (26)
\end{aligned}
$$

Therefore, derivative of the RHS of (26) with respect to $\hat{q}_{C,i}$, evaluated at the equilibrium $\hat{q}_C = \bar{q}_C$, is

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

Weighting this by m^J, we get (7) ∎

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

$$
\begin{aligned}
E\left[\max_{C\in\{A,B\}} E[U_C^{v,J}(q_C)|s_C^{v,J}]\right] - \text{cost of info} \quad = \quad & \frac{1}{2}E\left[E[U_A^{v,J}(q_A)|s_A^{v,J}] + E[U_B^{v,J}(q_B)|s_B^{v,J}]\right] + \\
& + \frac{1}{2}E\left[\left|E[U_A^{v,J}(q_A)|s_A^{v,J}] - E[U_B^{v,J}(q_B)|s_B^{v,J}]\right|\right] - \\
& - \text{cost of info.} \qquad (27)
\end{aligned}
$$

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

Using similar steps in the proof of Proposition 1 and imposing $\hat{q}_C = \bar{q}_C$, the second-order approximation of the first term on the RHS of (27) yields:

$$
\begin{aligned}
& \frac{1}{2}E\Big[\sum_{C\in\{A,B\}} E[U_C^{v,J}(q_C)|s_C^{v,J}]\Big] \\
\simeq \quad & \frac{1}{2}E\Big[\sum_{C\in\{A,B\}} E[U_C^{v,J}(\bar{q}_C) + \sum_{i=1}^{M} u_{C,i}^J(q_{C,i} - \bar{q}_{C,i}) + \frac{1}{2}\sum_{i,j=1}^{M,M} u_{C,i,j}^J(q_{C,i} - \bar{q}_{C,i})(q_{C,j} - \bar{q}_{C,j})|s_C^{v,J}]\Big] \\
= \quad & \frac{1}{2}\sum_{C\in\{A,B\}}\Big(U^J(\bar{q}_C) + \frac{1}{2}\sum_{i,j=1}^{M,M} u_{C,i,j}^J E\Big[E\Big[\Big((q_{C,i} - \check{q}_{C,i}) - (\bar{q}_{C,i} - \check{q}_{C,i})\Big) \\
& \qquad \Big((q_{C,j} - \check{q}_{C,j}) - (\bar{q}_{C,j} - \check{q}_{C,j})\Big)|s_C^{v,J}\Big]\Big]\Big) \\
= \quad & \frac{1}{2}\sum_{C\in\{A,B\}}\Big(U^J(\bar{q}_C) + \frac{1}{2}\sum_{i=1}^{M}(u_{C,i,i}^J\xi_{C,i}\sigma_{C,i}^2 + u_{C,i,i}^J(1-\xi_{C,i})\sigma_{C,i}^2)\Big) \\
= \quad & \frac{1}{2}\sum_{C\in\{A,B\}}\Big(U^J(\bar{q}_C) + \frac{M}{2}u_{C,i,i}^J\sigma_{C,i}^2\Big) \qquad (28)
\end{aligned}
$$

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.

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

$$\frac{1}{2}\Delta^v = \frac{1}{2}\Big(E[U_A^{v,J}(q_A)|s_A^{v,J}] - E[U_B^{v,J}(q_B)|s_B^{v,J}]\Big) \tag{29}$$

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 $\Delta > 0$):

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)|\Delta - x|dx &= \int_{-\infty}^{\Delta} f(x)(\Delta - x)dx - \int_{\Delta}^{\infty} f(x)(\Delta - x)dx \\
&= \Delta\Big(\int_{-\infty}^{\Delta} f(x)dx - \int_{\Delta}^{\infty} f(x)dx\Big) + \\
&\quad +\Big(-\int_{-\infty}^{\Delta} f(x)xdx + \int_{\Delta}^{\infty} f(x)xdx\Big) \\
&= \Delta\int_{-\Delta}^{\Delta} f(x)dx + 2\int_{\Delta}^{\infty} f(x)xdx.
\end{aligned} \tag{30}$$

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

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

$$\begin{aligned}
\Delta\int_{-\Delta}^{\Delta} f(x)dx &\simeq 2f(0)\Delta^2, \\
2\int_{\Delta}^{\infty} f(x)xdx &= 2\int_{0}^{\infty} f(x)xdx - 2\int_{0}^{\Delta} f(x)xdx \simeq E_f[|x|] - f(0)\Delta^2.
\end{aligned} \tag{31}$$

Therefore, conditional on Δ, the expectation of $|\Delta^v|$ equals $(E_f[|x|] + f(0)\Delta^2)$. Now we just need to express the unconditional expectation of Δ^2, i.e., of the square of difference between expected utilities from the two candidates after signals are acquired, evaluated at $\hat{q}_C = \bar{q}_C$.

Using the second order approximation, and manipulations similar to those in (24), we get:

$$\begin{aligned}\Delta &\simeq U^J(\bar{q}_A) - U^J(\bar{q}_B) + \sum_{i=1}^{M}\Big(u^J_{A,i}(\check{q}_{A,i} - \bar{q}_{A,i}) - u^J_{B,i}(\check{q}_{B,i} - \bar{q}_{B,i})\Big) \\ &+\frac{1}{2}\sum_{i=1}^{M}\Big(u^J_{A,i,i}((\check{q}_{A,i} - \bar{q}_{A,i})^2 + (1-\xi^J_{A,i})\sigma^2_{A,i}) - u^J_{B,i,i}((\check{q}_{B,i} - \bar{q}_{B,i})^2 \quad (32) \\ &+(1-\xi^J_{B,i})\sigma^2_{B,i})\Big). \quad (33)\end{aligned}$$

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

$$\begin{aligned}E[\Delta^2] &\simeq \Big(U^J(\bar{q}_A) - U^J(\bar{q}_B)\Big)^2 + \sum_{i=1, C\in\{A,B\}}^{M} \xi^J_{C,i}(u^J_{C,i})^2\sigma^2_{C,i} \\ &+\frac{1}{4}E\Big[\Big(\sum_{i=1}^{M} u^J_{A,i,i}((\check{q}_{A,i} - \bar{q}_{A,i})^2 + (1-\xi^J_{A,i})\sigma^2_{A,i}) - u^J_{B,i,i}((\check{q}_{B,i} - \bar{q}_{B,i})^2 + (1-\xi^J_{B,i})\sigma^2_{B,i})\Big)^2\Big] \\ &+\Big(U^J(\bar{q}_A) - U^J(\bar{q}_B)\Big)\Big(\sum_{i=1}^{M} u^J_{A,i,i}\sigma^2_{A,i} - u^J_{B,i,i}\sigma^2_{B,i}\Big). \quad (34)\end{aligned}$$

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

$$
\begin{aligned}
- \quad & 2\sum_{i,j=1}^{M,M} u^J_{A,i,i}u^J_{B,j,j}\sigma^2_{A,i}\sigma^2_{B,j} + 2\sum_{i,j=1,C\in\{A,B\}}^{M,M} u^J_{C,i,i}u^J_{C,j,j}\xi^J_{C,i}(1-\xi^J_{C,j})\sigma^2_{C,i}\sigma^2_{C,j} \\
& + \sum_{i,j=1,C\in\{A,B\}}^{M,M} u^J_{C,i,i}u^J_{C,j,j}(1-\xi^J_{C,i})(1-\xi^J_{C,j})\sigma^2_{C,i}\sigma^2_{C,j} \qquad (35) \\
& + \sum_{i,j=1,C\in\{A,B\}}^{M,M} u^J_{C,i,i}u^J_{C,j,j}\xi^J_{C,i}\xi^J_{C,j}\sigma^2_{C,i}\sigma^2_{C,j} + 2\sum_{i=1,C\in\{A,B\}}^{M} (u^J_{C,i,i})^2(\xi^J_{C,i})^2(\sigma^2_{C,i})^2 \\
= \quad & -2\sum_{i,j=1}^{M,M} u^J_{A,i,i}u^J_{B,j,j}\sigma^2_{A,i}\sigma^2_{B,j} + \sum_{i,j=1,C\in\{A,B\}}^{M,M} u^J_{C,i,i}u^J_{C,j,j}\sigma^2_{C,i}\sigma^2_{C,j} \\
& +2\sum_{i=1,C\in\{A,B\}}^{M} (u^J_{C,i,i})^2(\xi^J_{C,i})^2(\sigma^2_{C,i})^2. \qquad (36)
\end{aligned}
$$

The first term on the LHS of (35) 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 forth and fifth are product of the terms including $(\check{q}_{C,i} - \bar{q}_{C,i})^2$ and $(\check{q}_{C,j} - \bar{q}_{C,j})^2$, and the last term being a correction of the forth one for $i = j$, since if $x \sim N(0, \sigma^2)$, then $E[x^4] = 3(\sigma^2)^2$.

Therefore, putting everything together and omitting constants independent of ξ^J, the objective equivalent to (27) 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

$$F(\xi^J) = \sum_{i=1,C\in\{A,B\}}^{M} \Big(\xi^J_{C,i}\sigma^2_{C,i}(u^J_{C,i})^2 + 2(\xi^J_{C,i})^2(\sigma^2_{C,i})^2(u^J_{C,i,i})^2\Big). \qquad (37)$$

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 (37); and we also denote $\hat{\lambda}^J_{C,i} = 2\lambda^J_{C,i}/Min(\psi,\phi)$.

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

$$\xi^J_{C,i} = \max\left(\xi_0, \frac{4\sigma^2_{C,i}(u^J_{C,i,i})^2 - (u^J_{C,i})^2 + \sqrt{(4\sigma^2_{C,i}(u^J_{C,i,i})^2 + (u^J_{C,i})^2)^2 - 16\hat{\lambda}^J_{C,i}(u^J_{C,i,i})^2}}{8\sigma^2_{C,i}(u^J_{C,i,i})^2}\right). \tag{38}$$

∎