Electoral Competition with Rationally Inattentive Voters[*]

Filip Matějka[†]and Guido Tabellini[‡]

First version: September 2015

Abstract

This paper studies how voters optimally allocate costly attention in a model of probabilistic voting. The equilibrium solves a modified social planning problem that reflects voters' choice of attention. Voters are more attentive when their stakes are higher, when their cost of information is lower and prior uncertainty is higher. We explore the implications of this in a variety of applications. In equilibrium, extremist voters are more influential and public goods are under-provided. The analysis also yields predictions about the equilibrium pattern of information, and about policy divergence by two opportunistic candidates. Endogenous attention can lead to multiple equilibria, explaining how poor voters in developing countries can be politically empowered by welfare programs

1

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) and emphasized by classic works like that of 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 issues that are 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, as documented in the exhaustive empirical research on US opinion polls by Page and Shapiro (1992), voters' opinions are remarkably stable and consistent. Public opinion does not fluctuate wildly in unexplained ways. On the contrary, it tends to change gradually, in reaction to new information and relevant events. Often voters miss the details, but they can nevertheless make subtle distinctions between different policies in a coherent pattern. As Page and Shapiro (1992), p.13 state, "There is evidence that most Americans do grasp the essentials of major issues (..), even if they do not know a lot of specific details". In other words, although voters are not well informed, there are regularities in what they know and don't know, and this is reflected coherently in their views about public policy.

How does this selective ignorance by 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 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 to different issues, and in which 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. 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, 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 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. 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 (more transparent) candidates have a higher probability of winning the election.

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

3

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 also 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, 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 exactly what Palfrey and Poole (1987) find in their analysis of 1980 survey data of U.S. presidential elections : voters who are highly informed about the candidate location tend to be significantly more polarized in their ideological views compared to uninformed voters.

Finally, the endogenous choice of attention can be a source of indeterminacy. The allocation of attention by voters reflects their prior beliefs about what candidates are likely to do. Candidates' behavior, in turn, reflect voters' attention. In some circumstances, the joint determination of attention and policy can give rise to multiple equilibria. We illustrate this in an example where voters' attention reflects the opportunity cost of time. This is related to welfare programs in developing countries which also empower the poor in the sense of enabling them to devote attention to politics, rather than to justsurvival activities. This finding emerges, for instance, in the empirical assessment of the PANES program in

Uruguay, where Manacorda et al. (2009) find that beneficiaries of the poor relief program are not only more likely to support the incumbent government, but also report greater attention paid to politics in general. Thus a complementarity is at work: pro-poor policies enable the poor to be more attentive and hence more influential in the political process, which in turn induces politicians to enact more pro-poor policies. This can give rise to multiple equilibria, consistently with 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), Maækowiak and Mirko Wiederholt (2009), Van Nieuwerburgh and Veldkamp (2009), or Matějka and McKay (2015). Rational inattention 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.[1] In our paper, rational inattention allows us to model how much attention voters choose to pay to each of the candidates' policies, depending mostly on the importance of such policies for the particular voters.

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.[2] Ponzetto (2011) studies a model of trade policyin 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 to their consumer choices. This asymmetry in information leads to a political bias against

---

[1]See Bartoš et al. (2014) for empirical evidence of how agents choose their levels of attention depending on the choice problem they face.

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

free trade. Although information is endogenous, 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.

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.[3]

Finally, a large body of 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 (eg.Martinelli (2006)) or focuses on small committees (eg. Persico (2003)), however, and does not ask how voters' endogenous information shapes equilibrium policies. The literature on endogenous participation has studied 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.

---

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

2   The general framework

This section presents a general model of electoral competition with rationally inattentive voters.

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. There are also two political candidates $C \in \{A, B\}$, each running on a proposed 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. Thus, throughout, superscripts refer to voters and subscripts refer to candidates. Candidates are opportunistic and only care about winning elections.

As in standard probabilistic voting models (Persson and Tabellini 2000), voters have preferences over both policies and candidates. Their preferences have two additive components. The first component $U^J(q_C)$ is a function of the policy vector and is common to all voters belonging to the same group J. The second component is a preference shock x^v in favor of candidate B. Thus, the utility function of a voter of type $\{v, J\}$ if candidate A or B wins is respectively:

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

$U^J(\cdot)$ is concave and differentiable.

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 towards candidate B, while $\tilde{x}$ is a shock favoring candidate B that is 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. Alternatively, it is possible to interpret x^v as being due to additional exogenous signals that agents receive, which are independent of the attention strategy and of the policy platforms announced by the candidates. The purpose of having the shock x^v is mainly technical, namely to smooth the candidates' objective functions. As shown below, this guarantees that the candidates' best response functions are continuous, so that an equilibrium of the electoral competition game exists.

Voters are uninformed about the candidates' policies, and they must devote costly attention to the electoral campaign. The sequence of events is as follows.

i) Voters form prior beliefs about the policy platforms of each candidate.

ii) Candidates choose their platforms, and voters choose their attention strategies.

iii) Voters observe the signals of the policy platforms and update their beliefs.

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

In Subsection 2.2 we define the equilibrium, which is a pair of 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 choices. We now describe voter's behavior in greater detail, and in Subsubsection 2.2.1 we discuss some of the modeling choices.

2.1 Voters' behavior

Voters have 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. Each element $q_{C,i}$ is 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. All voters have identical prior beliefs and thus these variables are not indexed by type.[4]

As described above, the voters' decision process has two stages: information acquisition and voting.

2.1.1 The choice of attention

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 $s^{v,J}_C$ of independent signals on all the elements $\{1, ..., M\}$ of both candidates, A and B,

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

where the noise $\epsilon^{v,J}_{C,i}$ is drawn from a normal distribution $N(0, \gamma^J_{C,i})$, which is the same within each group, and the realization of noise is independent across voters. The level of noise is subject to a voter's choice, and defines his or her attention strategy. 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; for this reason we have omitted the v superscript from γ.

---

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

It is convenient to define the following vector $\xi^J \in [0,1]^{2M}$, which is the decision variable for attention in our model:

$$\xi^J = \left\{ [\xi^J_{A;1}..., \xi^J_{A,M}], [\xi^J_{B,1}..., \xi^J_{B,M}] \right\},$$

where

$$\xi^J_{C,i} = \frac{\sigma^2_{C,i}}{\sigma^2_{C,i} + \gamma^J_{C,i}}.$$

The more attention is paid by the voter, the closer $\gamma^J_{C,i}$ is to 0, and the closer $\xi^J_{C,i}$ is to 1. The choice of attention also determines the variance of posterior beliefs about $q_{C,i}$. This posterior variance is denoted as $\rho^J_{C,i}$, and exploiting its definition we also have:[5]

$$\xi^J_{C,i} = 1 - \frac{\rho^J_{C,i}}{\sigma^2_{C,i}}. \tag{2}$$

Thus, the choice variable $\xi^J_{C,i}$ measures the reduction of uncertainty about $q_{C,i}$. The more attention is paid, the closer $\xi^J_{C,i}$ is to 1, and the closer the posterior variance $\rho^J_{C,i}$ Is to 0. 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. To use a cost function of a particular type, we employ the standard one 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 about q 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:

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

where $\lambda^J_{C,i} \in \mathbb{R}_+$ scales the unit cost of information of voter J about $q_{C,i}$.

2.1.2 Voting

The second stage is a standard voting decision under uncertainty. After the voters receive additional information of the selected form, and knowing the realization of

---

[5]The posterior variance is $\rho^J_{C,i} = \gamma^J_{C,i} \sigma^2_{C,i} / (\sigma^2_{C,i} + \gamma^J_{C,i})$

the candidate bias x^v, they choose which candidate to vote for. Specifically, after a voter receives signals $s_C^{v,J}$, he forms posterior beliefs about utilities from policies that will be implemented by each candidate, and he votes for A if and only if:

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

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

2.1.3 Voter's objective

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

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

By (1), the realization of the candidate bias x^v is incorporated in the functions $U_C^{v,J}(q_C)$. 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.

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 policy vectors of the candidates, and candidates select a best response to the attention strategies of voters and to each other's policies. The voters' prior beliefs are not degenerate at the equilibrium policy vectors, however, they are dispersed around them with noise normally distributed with variance vectors σ_C^2 - see the discussion in Subsubsection 2.2.1 below. Specifically:

Definition 1 *Given the level of noise σ_C^2 in voters' beliefs, the equilibrium (denoted with a $*$) is a set of policy vectors chosen by each candidate, q_A^*, 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 $\bar{q}_C = q_C^*$ and noise σ_C^2.

b) The policy vector 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

This framework illustrates the main implications of voters' limited attention. Some assumptions are a reasonable reduced form of a larger model, which is left unspecified. The qualitative implications of this framework, however, are likely to be robust to such modeling choices. Our main qualitative results in the following sections are based on monotonicity arguments only. The arguments are that higher stakes imply higher attention, which in turn implies stronger voter response to a policy deviation from the expected equilibrium. Thus candidates have stronger incentives to appeal to such high-stake needs than under perfect information or with exogenously given attention.

Noise in prior beliefs. Voters' prior beliefs are exogenously dispersed about the expected equilibrium policies. How do we motivate such an assumption?

We take some level of uncertainty as given and study its implications. We simply assume that there is some room for uncertainty. Otherwise, limited attention would play no role. Nevertheless, voters have some prior beliefs about what each candidate will do, and in equilibrium the prior means are consistent with the candidates optimizing behavior. In other words, there is no systematic bias in one direction or the other.

There are several possible ways to endogenize the uncertainty in prior beliefs from more primitive assumptions. The candidates can be erratic; they can be targeting a unique equilibrium policy and can be subject to an additional random error in the implementation of the policy. Or the environment can be random, e.g., the composition of the population of voters. If the true state of the environment is random, and unknown by the voters, then the voters are uncertain about what policies are targeted by the candidates. Alternatively, voters could be uncertain about whether candidates have partisan or ideological preferences favoring some groups or certain policy instruments. And obviously, the uncertainty can also be a behavioral assumption. Instead of choosing one of these drivers of uncertainty, and endogenizing the beliefs, we take the resulting uncertainty as given but impose consistency of prior means with the candidates' actual behavior. In other words, we explore the implications of voter uncertainty and costly attention, without, however, introducing any systematic distortion in the location of prior beliefs.

Another feature of the prior beliefs that is worth discussing is the assumed independence of 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 needed 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).

Finally, the model would also become more complicated if we assumed that voters form their prior beliefs by conditioning on being pivotal. In this case, the form of noise would no longer be Gaussian, which would complicate all analysis tremendously. As we discuss above, we take the form of prior beliefs as given exogenously.

Voter's objective. Implicit in the formulation of the voters' objective (4) is the assumption that voters are motivated by "sincere attention". The voter chooses how much and what form of information to acquire, which helps him to choose the candidate that provides him with higher utility. Thus, when choosing attention the voter behaves *as if* he is pivotal in his subsequent voting decision.

Although this appears to be a restrictive assumption, it is actually rather innocuous, or at least not more restrictive than in most of the literature on electoral competition. Specifically, suppose that the voter expects to be pivotal with an exogenously given probability $\delta > 0$. Then the first term in (4), the expected utility from the selected policy, would be premultiplied by δ. Such a modification would be equivalent to rescaling the cost of information $\lambda_{C,i}^J$ by the factor $1/\delta$ for each voter, with no substantive change in any result.[6]

If the probability of being pivotal was endogenous and part of the equilibrium, the model would become more complicated, as we would need additional assumptions about what motivates voting. Nevertheless, most qualitative implications discussed below would remain unchanged, because it would still remain true that in equilibrium this would be equivalent to rescaling the cost of information. The first order condition (14) below would still hold, but the main difference would be that then the quantity $\lambda_{C,i}^J$ would be endogenous and determined as part of the equilibrium.

---

[6]Naturally, with a continuum of voters, the probability of being pivotal is zero, and voters would not be willing to pay any positive cost of information. This is the same issue faced by many papers in the field of political economy, which implies zero participation if the cost of voting is positive. The focus of this paper is orthogonal to this issue, and we do not aspire to solve it. Alternatively, we could either study the case of a sufficiently low number of voters N, or the limit of $N \to \infty$ with $(\lambda_{C,i}^J N)$ being fixed at a positive constant.

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 idiosincratic 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 show that the equilibrium policy solves a specific modified social welfare function which can be compared with that of standard probabilistic voting models. We then show that, if 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 -see Persson, Tabellini (2000).

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:

$$\tilde{x}^v = E[U^J(q_A)|s_A^{v,J}] - E[U^J(q_B)|s_B^{v,J}] - \tilde{x} \equiv \tilde{x}_T^{v,J} \tag{5}$$

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 , 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:

$$\pi_A = \sum_J m^J E_\epsilon^J[\Pr(\tilde{x}^v \leq \tilde{x}_T^{v,J})] = \frac{1}{2} + \phi \sum_J m^J E_\epsilon^J[\tilde{x}_T^{v,J}] \tag{6}$$

Note that (6) holds when the noise in the ideological preference shocks $\tilde{x}^v$ is sufficiently large to affect the vote with positive probability.[7]

By (5)-(6), the vote share π_A is a linear function of the popularity shock $\tilde{x}$. Since the latter is also uniformly distributed, the probability of winning for candidate A is then:

$$p_A = \frac{1}{2} + \psi \left(\sum_J m^J E_\epsilon^J \left[E[U^J(q_A)|s_A^{v,J}] - E[U^J(q_B)|s_B^{v,J}] \right] \right) \qquad (7)$$

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 (7), which in a symmetric equilibrium will always be true.

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:

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

$$\max_{q_C \in \mathbb{R}^M} \quad \sum_J m^J E_\epsilon^J \left[E[U^J(q_C)|s_C^{v,J}] \right] \qquad (8)$$

In equilibrium, candidate C maximizes the "perceived social welfare" provided by his policies. It is the weighted average (weighted by the mass of voters) of utilities from policy q_C expected by voters in each group. 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. Proposition 1 thus reveals the main difference between this framework and standard probabilistic voting models. If perceived welfare systematically differs from actual welfare, rational inattention can lead politicians to select distorted policies, even if all groups are equally influential in the sense of having the same distribution of ideological preference shocks $\tilde{x}^v$.

Finally, note that the candidates' objective (8) 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 realized policy q_C and posterior variances

---

[7]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.

are independent of it.[8]   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 (8), since they are necessary and sufficient for an optimum.

3.2   Small noise approximation

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

How should costly attention be allocated to alternative components of the policy vector? The following lemma provides a clear perspective on the tradeoffs that arise when answering this question. Let us denote by

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

the marginal utility for a voter in group J of a change in the ith component of the policy vector, evaluated at the equilibrium policies. Thus, $u_{C,i}^J$ measures intensity of preferences about $q_{C,i}$ for voters in group J in a neighborhood of the equilibrium. Suppose that the noise in prior beliefs σ_C^2 is small, and that the difference in expected utilities from the two candidates is small relative to the support of the preference shock x^ν. Then Appendix proves:

---

[8]Variance of posterior belief can be expressed in terms of prior variance and the attention vector ξ^J:

$$\rho_{J,i} = (1 - \xi_i^J)\sigma_i^2, \tag{9}$$

and upon acquisition of a signal $s_{C,i}^{v,J}$, the posterior mean $\hat{q}_{C,i}$ is:

$$\hat{q}_{C,i} = \xi_{C,i}^J s_{C,i}^{v,J} + (1 - \xi_{C,i}^J)\bar{q}_{C,i}, \tag{10}$$

where $s_{C,i}^{v,J} = q_{C,i} + \epsilon_{C,i}^{v,J}$, which makes the distribution of posterior means about $\hat{q}_{C,i}$, conditional on the prior, drawn from

$$N\left(\xi_{C,i}^J q_{C,i} + (1 - \xi_{C,i}^J)\bar{q}_{C,i}, \xi_{C,i}^J \sigma_i^2 \right).$$

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

$$\sum_{C \in \{A,B\}, i=1}^{M} \left(\xi_{C,i}^J (u_{C,i}^J)^2 \sigma_{C,i}^2 + \hat{\lambda}_{C,i}^J \log(1 - \xi_{C,i}^J) \right), \tag{11}$$

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

$$\xi_{C,i}^J = \max \left(\xi_0, 1 - \frac{\hat{\lambda}_{C,i}^J}{(u_{C,i}^J)^2 \sigma_{C,i}^2} \right). \tag{12}$$

As shown in the appendix, the term

$$\sum_{C \in \{A,B\}, i=1}^{M} \xi_{C,i}^J (u_{C,i}^J)^2 \sigma_{C,i}^2 \tag{13}$$

in (11) corresponds to the variance of the difference in expected utilities under each of the two candidates, conditional on posterior beliefs. Thus, the voter maximizes the variance of expected utilities of posteriors, less the cost of information (up to a constant of proportionality that reflects the distribution of the preference shocks for candidates, $x^v = \tilde{x} + \tilde{x}^v$).[9] The variance of expected utilities enters because if both candidates provide the same utility, then there is no gain from information. The larger the discovered difference in utilities is, the larger the gain is, since then the voter can choose the candidate that provides the higher utility.

Note that $\xi_{C,i}^J \sigma_{C,i}^2 = (\sigma_{C,i}^2 - \rho_{C,i})$ also corresponds to the difference between prior $(\sigma_{C,i}^2)$ and posterior $(\rho_{C,i})$ variance of beliefs, a measure of reduced uncertainty. Thus, the net of the cost of attention, the voter maximizes a weighted average of the reduction in uncertainty, where the weights in (13) correspond to the (squared) marginal utilities from deviations in $q_{C,i}$. That is, the voter aims to achieve a greater reduction in uncertainty when the policy stakes are higher. Quite intuitively, the solution (12) implies that, for a given cost of information, the voter pays more attention to those elements $q_{C,i}$ for which the cost of information $\lambda_{C,i}^J$ is lower, prior uncertainty $\sigma_{C,i}^2$ is higher, and which have higher utility-stakes $|u_{C,i}^J|$ from deviations in $q_{C,i}$.

Exploiting Proposition 1 and the steps in the proof of the previous lemma, the Appendix also demonstrates proof of:

---

[9]Given the distributional assumptions made in section 2, the p.d.f of preference shock $x^v = \tilde{x} + \tilde{x}^v$, evaluated at point 0, is $Min(\psi, \phi)$.

Proposition 2 *The equilibrium policies satisfy the following first order conditions, where the attention weights $\xi_{C,i}^J$ are given by (12):*

$$\sum_{J=1}^{N} m^J \xi_{C,i}^J u_{C,i}^J = 0, \qquad \forall i. \tag{14}$$

This proposition emphasizes the main forces in electoral competition with rationally inattentive voters. If all voters paid the same attention to a given policy instrument, so that $\xi_{C,i}^J = \xi$ for all J, C, i, then both candidates would enact the same policies as a benevolent social planner, since (14) would imply the same optimality conditions as for a 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 deviation.

Also note that if $\lambda^J \to 0$ for all J, then $\xi_{C,i}^J = 1$ for all groups J and all policy instruments i. In this case (14) takes the form of

$$\sum_{J=1}^{N} m^J u_{C,i}^J = 0, \qquad \forall i, \tag{15}$$

and the equilibrium coincides with the utilitarian optimum. Intuitively, if voters are fully informed, then the equilibrium is equivalent to that of a probabilistic voting model when all groups are equally influential.

Finally, the attention weights $\xi_{C,i}^J$ also depend on the identity of the candidate, because prior uncertainty $\sigma_{C,i}^2$ or the cost of information $\lambda_{C,i}$ 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 the two 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 only if candidates have policy preferences themselves (see Persson and Tabellini 2000). Subsection 4.1 below illustrates this result with an example.

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 policy instruments, and finally show that rational inattention can lead to multiple equilibria.

4.1 Heterogeneous bliss-points and costs of information

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 2 the equilibrium with rational inattention can be computed as the solution to a modified social planning problem, where each candidate C maximizes

$$\sum_J m^J \xi_C^J U^J(q_C) \tag{16}$$

and where the weights ξ_C^J are taken as given and in equilibrium are given by

$$\xi_C^J = max\left(\xi_0, 1 - \frac{\hat{\lambda}^J}{(u_C^J)^2 \sigma_C^2}\right) \tag{17}$$

with the marginal utilities u_C^J evaluated at the equilibrium policies.

Suppose that prior uncertainty is the same for both candidates, so that $\sigma_C^2 = \sigma^2$ for $C \in \{A, B\}$. Then, both candidates receive equal attention by all voters and announce the same policies. But groups with more extreme policy preferences receive a greater weight and hence are more influential in the political equilibrium, compared to the utilitarian optimum (since ξ^J is increasing in u_C^J). The reason is that the stakes are higher for these more extremist groups, and hence they pay more attention to policy deviations. Groups with a lower cost of collecting information (a lower λ^J) also receive a greater weight, for the same reason.

The specific implications for how the equilibrium differs from that with full information depends on the shape of the distribution of voters' preferences. A plausible assumption is that the distribution of bliss points t^J is asymmetric. 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 λ^J).

The size of this deviation from the utilitarian optimum increases with the size of the information cost. Specifically, suppose that $\lambda^J = \lambda$ for all J. The first order condition that characterizes the equilibrium with inattentive voters is

$$\sum_J m^J \xi^J u^J(q) = 0 \qquad (18)$$

The derivative of the above expression 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 λ rises. Notice that this holds for negatively skewed distributions of t^J.

The model has an additional implication. If the cost of information is the same for everyone, by (17), voters with more extreme policy preferences (the very rich or very poor in the income tax example) pay more attention and thus are more informed. Voters in the middle do not bother to collect much information (or pay no attention at all), since the stakes for them are much lower, given that they expect either candidates will choose equilibrium policies close to their bliss points.This prediction of the model is exactly in line with the evidence uncovered by Palfrey and Poole (1987). Studying the survey data of U.S. presidentialelections held in 1980, they 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.

Finally, suppose that candidates differ in terms of transparency or media attention, so 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 (17), the difference in attention given by voters between the two candidates depends on u^J, and it is higher in the center 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 (17)). This in turn affects the incentives of both candidates and leads to policy divergence. Candidate A now assigns a greater weight to the
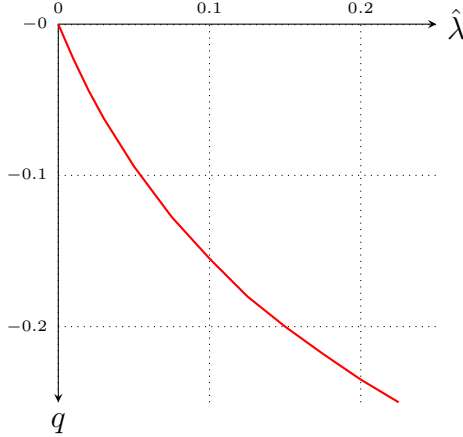
Figure 1: Effect of the cost of information.

more extreme voters compared to candidate B, since these candidates 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, compared to candidates receiving less media coverage who have a stronger incentive to go after voters with extremist policy preferences. With policy divergence and different attention weights, the probability of victory will differ 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 (18) at the optimum will be larger for B than for A).

To illustrate this example numerically, 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 $U^J(q) = -(q-t^J)^2$ and $\sigma_{C,i}^2 = 0.1$ - 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, then the equilibrium q decreases.

Figure 1 presents the equilibrium q as a function of $\hat{\lambda}$. When $\hat{\lambda} = 0.01$, then $q \doteq -0.02$, when $\hat{\lambda} = 0.1$, then $q \doteq -0.15$, and when $\hat{\lambda} = 0.2$, then $q \doteq -0.23$. For positive costs of information, the extreme voters $J = 3$ pay relatively more attention than the $J = 1$ and $J = 2$ when q is in the neighborhood of zero, and thus the equilibrium policy moves in their direction.[10] The same results hold

---

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

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, 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 more in equilibrium. For instance, in the example above, if $\hat{\lambda}^3 = 0.01$ and $\hat{\lambda}^1 = \hat{\lambda}^2 = 0.05$, then in equilibrium $q = -0.35$.

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, in 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

---

policy is again at the social optimum since all voters are weighted equally.

distorting lump sum tax targeted to each group, b^J, with negative values of b^J corresponding to targeted transfers to the group; 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:

$$g = \sum_J b^J + N\tau + s$$
$$c^J = y - b^J - T(\tau) - S(s)/N.$$

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:

$$U^J(q) = y - b^J - T(\tau) - S(g - \sum_K b^K - N\tau)/N + H(g). \tag{19}$$

The observable policy vector is

$$q = [b^1, ..., b^N, g, \tau],$$

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

It is easy to verify that the socially optimal policy vector $\tilde{q} = [\tilde{\tau}^1, ..., \tilde{\tau}^N, \tilde{g}]$ satisfies $s = \tau = 0$, i.e., eliminates all distorting taxes, and sets the public good so as to satisfy the Samuelson optimality condition; namely $H'(g) = 1/N$. 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 (18) we use: $u_J^J = -1 + S'/N$, $u_{-J}^J = S'/N$, $u_\tau^J = T' - S'$ and $u_g^J = H' - S'/N$, where the J and $-J$ subscripts refer to partial derivatives of U^J with respect to a voters' own taxes b^J, and taxes targetedat 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. The equilibrium first order condition with respect

to g and τ, as long as attention to these instruments is positive, are the same as for the social planner's problem, respectively:

$$-S'/N + H' = 0 \tag{20}$$
$$-T' + S' = 0 \tag{21}$$

The reason is that all types J pay the same level of attention to g and τ,[11] and thus ξ_g^J and ξ_τ^J do not enter the FOC. What could drive equilibria away from the social optimum is heterogeneity in ξ_i^J across different voters, only, which we assumed away.

The first order condition (18) with respect to b^J can be written as:

$$\xi_J^J(-1 + S'/N) + (N-1)\xi_{-J}^J S' = 0$$

or equivalently as:

$$[1 + (N-1)\frac{\xi_{-J}^J}{\xi_J^J}]S'/N = 1 \tag{22}$$

At the social optimum, $S' = 1$ (since $s = 0$), which in turn implies that $\xi_{-J}^J < \xi_J^J$, since $N > 2$ - cf (12). 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 (22) implies $S' > 1$, a contradiction. Hence in equilibrium, it must be that $S' > 1$, and hence that $s > 0$. Equations (20)-(21) then imply that $H\prime > 1/N$ and that $T' > 1$. Thus, in equilibrium there is under-provision of the public good relative to the social optimum, and the government relies on distorting (observable and unobservable) sources of revenues, despite the availability of lump sum taxes. In fact, if the marginal tax distortions T' and S' do not rise too rapidly, it is even possible that the equilibrium entails negative values of 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 away from efficient allocation, in order to appear to please each group. The higher the cost of information λ is and the larger is N, the larger the distortion

Finally, note that in equilibrium $u_\tau^J = T' - S' = 0$ and $u_g^J = H' - S'/N = 0$. By (12) this in turn implies that $\xi_g^J = \xi_\tau^J = \xi_0$. Namely, in equilibrium all voters

---

[11]This can be seen from (12) and from the fact that $u_{C,N+1}^J = H'$ is common to 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. 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.[12]

The finding 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 (22): 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 exogenosuly changed through increased fiscal transparency, since the cost of observing instruments targetedat 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 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

---

[12]This is because, when attention to g and τ is zero, then the first order conditions (14) 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.

of taxation is higher, shifting the tax burden on less visible sources of revenue. This prediction would follow almost immediately from a modifed 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 witholding 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.[13]

4.3   Empowering the poor

Endogenous choice of attention can also be a source of multiple equilibria. This occurs because the allocation of attention reflects prior beliefs, which in turn are formed before policies are actually announced. Thus, voters' prior beliefs and their resulting attention strategies can impact the outcome of electoral competition, inducing political candidates to fulfill those prior beliefs. In this section we illustrate how this could happen.

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. 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 similar to those emphasized in other papers such as Benabou and Tirole (2006) and Alesina and Angeletos (2005), but the mechanism at work is quite different.

---

[13]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).

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 s^e 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 s^e and variance σ^2, $s \sim N(s^e, \sigma^2)$, 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:

$$\int [U(\bar{y} + s) - U(\underline{y} + s)] dN(\tilde{s}, \sigma^2) = \bar{d} - \underline{d} \tag{23}$$

By concavity of $U(.)$, if $s^e \geq \tilde{s}$ then the poor choose low effort, and if $s^e < \tilde{s}$ 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 $u(\bar{y} + s^*) = 1$.[14]

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 forpoor 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 policies. Finally, voters gather information and vote. The policy s is

---

[14]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.

imperfectly observed, as in the previous sections. Repeating the previous steps, and considering the small noise approximation, by Proposition 2 the equilibrium policy 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 can be written as.

$$u = \frac{\xi^R}{\xi^P} \tag{24}$$

where the poor's marginal utility of income, u, is computed at the equilibrium policy, 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 the lower bound in ξ, (24) can be rewritten as:

$$\sigma^2 u^2 + (\hat{\lambda} - \sigma^2)u - \hat{\lambda}^P = 0 \tag{25}$$

where $\hat{\lambda}$ is the cost of information for the rich. Equation (25) can be solved for u, selecting the positive root to avoid negative marginal utility, and this yields:

$$u = F(\hat{\lambda}^P) \equiv \frac{\sigma^2 - \hat{\lambda} + \sqrt{(\sigma^2 - \hat{\lambda})^2 + 4\sigma^2 \hat{\lambda}^P}}{2\sigma^2} \tag{26}$$

Equation (26) 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 (26) 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 their marginal utility u in equilibrium is, 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 as a function of $\hat{\lambda}^P$, namely

$$s^* = u^{-1}[F(\hat{\lambda}^P)] - y \equiv S(\hat{\lambda}^P) - y \tag{27}$$

Since $F(.)$ is increasing and u^{-1} is decreasing, the function $S(.)$ is decreasing in $\hat{\lambda}^P$.
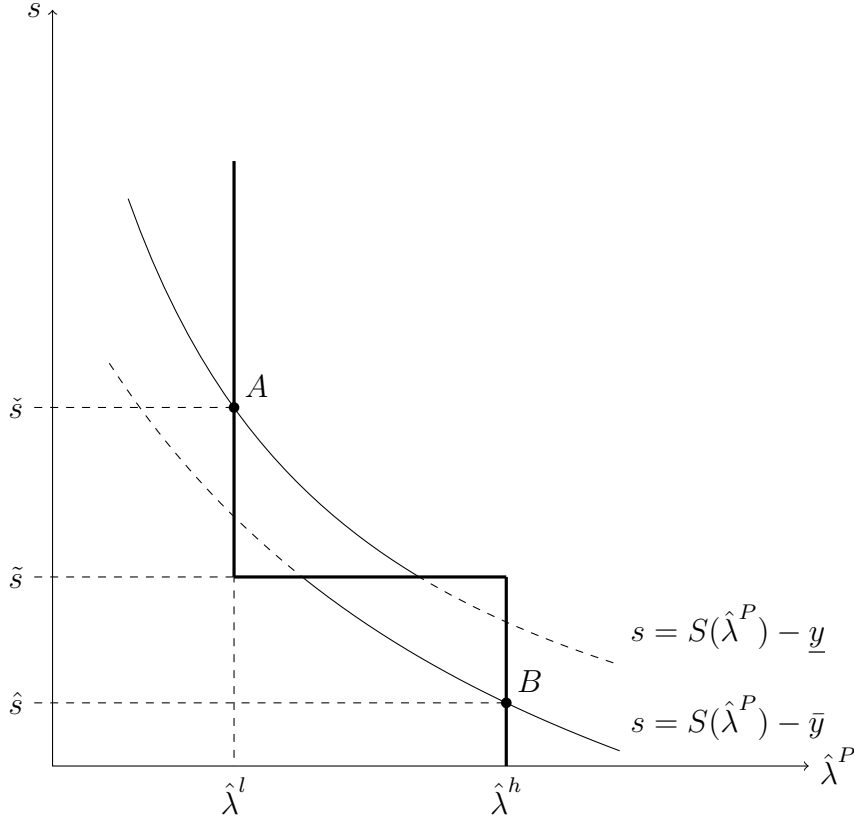
Figure 2: Two equilibrium levels of subsidy.

An important implication of (27) 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 $s^e < \tilde{s}$. 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 (23) and (27) This is indeed an equilibrium, call it $\hat{s}$, if $\hat{s} = S(\hat{\lambda}^h) - \bar{y}$ and if $\hat{s} = s^e < \tilde{s}$. The other equilibrium is obtained under the assumption that the poor expect both candidates to announce high subsidies, so that the prior mean is in the range $s^e > \tilde{s}$. 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 $\check{s}$, equilibrium subsidies are $\check{s} = S(\hat{\lambda}^l)] - \underline{y}$ and $\check{s} = s^e > \tilde{s}$. Since $S(.)$ is increasing in $\hat{\lambda}^P$, and since $\hat{\lambda}^h > \hat{\lambda}^l$ and $\bar{y} > \underline{y}$, we must have $\check{s} > \hat{s}$. 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 $\check{s} > \tilde{s} > \hat{s}$.

The equilibria are illustrated in Figure 2. The stepwise boldface function depicts how the poor's information cost λ^P varies with subsidies. By (23), at $s = \tilde{s}$ the poor are just indifferent between high and low effort. For $s > \tilde{s}$, 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 $s < \tilde{s}$ 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 announced in political equilibrium, corresponding to (27). There are two lines, because the poor's income can be high or low, depending on expected subsidies. If $s < \tilde{s}$ 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 equilibrium 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.[15]

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 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

---

[15]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 $s = \tilde{s}$ 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.

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. Neverthless, 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.

In this paper we have focused on what induces voters to collect and process information, when this 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 other actors, 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, seems a difficult but fruitful area for future research.

A second theoretical aspect that is missing from our simple framework is competition for voters' attention. Here politicians react to the attention strategies of voters, but they are not allowed to take any action prior to the voters' choice of attention. If they could, they would like to attract more attention, because this would allow them to better explain to the voters their true policy platforms. This can be seen, for instance, from the objective function (16) in subsection 4.1, that is increasing 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.

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 assumes that voters' information, although incomplete, is exogenous. Endogenizing what voters pay attention to in a framework of retrospective voting, and how this in turn affects the behavior of incumbents is likely to yield other novel insights.

References

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

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.

Vojtech Bartos, Michal Bauer, Julie Chytilova, and Filip Matejka: (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.

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.

Delli Carpini ,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.

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.

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.

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

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.

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.

Maækowiak, 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.

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.

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.

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

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

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

6 Appendix

Proof of Lemma 1: 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:

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

$$
+\frac{1}{2}E\left(\Big|E[U_A^{v,J}(q_A)|s_A^{v,J}] - E[U_B^{v,J}(q_B)|s_B^{v,J}]\Big|\right) -
$$

$$
-\text{cost of info.} \tag{28}
$$

The inner expectation is over a realized posterior belief. The outer expectation is over realizations of $\epsilon_C^{\nu,J}$ and x^ν. Since we assume that the noise in beliefs is small, we can use the first order approximation to expected utility conditional on posterior belief:

$$
E[U^J(q_C)|s_C^{v,J}] \simeq U^J(\bar{q}_C) + \sum_{i=1}^M u_{C,i}^J(E[q_{C,i}|s_C^{v,J}] - \bar{q}_{C,i})
$$

$$
= U^J(\bar{q}_C) + \sum_{i=1}^M u_{C,i}^J \xi_i^J(q_{C,i} - \bar{q}_{C,i} + \epsilon_{C,i}^{v,J}), \tag{29}
$$

where we apply equation (10). Using the first-order approximations (29), the term in the first bracket on the RHS of (28) equals $\frac{1}{2}(U_A^J(\bar{q}_A) + U_B^J(\bar{q}_B))$. The outer expectation is over all realizations of q_C, noise in signals and preference shocks, which have expectation zero. This term is thus independent of the choice of attention and does not affect the optimum, which is then given by the maximization of the expectation of the absolute value of:

$$
\frac{1}{2}\Delta^v = \frac{1}{2}\Big(E[U_A^{v,J}(q_A)|s_A^{v,J}] - E[U_B^{v,J}(q_B)|s_B^{v,J}]\Big) \tag{30}
$$

less the cost of information. Let

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

which is 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 $\Delta > 0$):

$$
\begin{aligned}
\int_{-\infty}^{\infty} f(x)|x+\Delta| dx &= -\int_{-\infty}^{-\Delta} f(x)(x+\Delta) dx + \int_{-\Delta}^{\infty} f(x)(x+\Delta) dx \\
&= \Delta\left(-\int_{-\infty}^{-\Delta} f(x) dx + \int_{-\Delta}^{\infty} f(x) dx\right) + \\
&\quad + \left(-\int_{-\infty}^{-\Delta} f(x) x dx + \int_{-\Delta}^{\infty} f(x) x dx\right) \\
&= \Delta\int_{-\Delta}^{\Delta} f(x) dx + 2\int_{\Delta}^{\infty} f(x) x dx. \quad (31)
\end{aligned}
$$

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

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

$$
\begin{aligned}
\Delta\int_{-\Delta}^{\Delta} f(x) dx &\simeq 2f(0)\Delta^2, \\
2\int_{\Delta}^{\infty} f(x) x dx &= 2\int_{0}^{\infty} f(x) x dx - 2\int_{0}^{\Delta} f(x) x dx \simeq E_f[|x|] - f(0)\Delta^2. \quad (32)
\end{aligned}
$$

Therefore, conditional on Δ, the expectation of $|\Delta^v|$ equals $(E_f[|x|]+f(0)\Delta^2)$. Now we just need to express the unconditional expectation of Δ^2, i.e., of the square of difference between expected utilities from the two candidates after signals are acquired. First, approximation (29) implies that $\Delta = E[U^J(q_A)|s_A^{v,J}] - E[U^J(q_B)|s_B^{v,J}]$ has a normal distribution with mean $\left(U_A^J(\bar{q}_A) - U_B^J(\bar{q}_B)\right)$. By (29) the variance of this distribution is:

$$
\begin{aligned}
variance(\Delta) &= \sum_{C\in\{A,B\},i=1}^{M} \left(u_{C,i}^J \xi_i^J\right)^2 \left(variance(q_{C,i}) + variance(\epsilon_{C,i}^{v,J})\right) \\
&= \sum_{C\in\{A,B\},i=1}^{M} \left(u_{C,i}^J \xi_i^J\right)^2 \sigma_{C,i}^2 \left(1 + \frac{1-\xi_i^J}{\xi_i^J}\right) \\
&= \sum_{C\in\{A,B\},i=1}^{M} \xi_i^J (u_{C,i}^J)^2 \sigma_{C,i}^2. \quad (33)
\end{aligned}
$$

Second, the expectation of y^2 for $y \sim N(\bar{y}, \sigma_y^2)$ is $(\bar{y}^2 + \sigma_y^2)$. Therefore,

$$
\frac{1}{2}E[|\Delta^v|] = \frac{1}{2}\left[E_f[|x|] + f(0)\left((U_A^J(\bar{q}_A) - U_B^J(\bar{q}_B))^2\right)\right] + \frac{1}{2}\sum_{C\in\{A,B\},i=1}^{M} \xi_{C,i}^J (u_{C,i}^J)^2 \sigma_{C,i}^2
$$
$$
(34)
$$

This equation implies (11), where we omit the additive constant $E_f[|x|] + f(0) \left(U_A^J(\bar{q}_A) - U_B^J(\bar{q}_B) \right)^2$, which is independent of ξ^J, and where $f(0) = Min(\psi, \phi)$ given the distributional assumption on $x^v = \tilde{x} + \tilde{x}^\nu$.

Finally, the solution (12) is an immediate consequence of (11). The max operator is included since $\xi_{C,i}^J$ must be nonnegative. ∎

Proof of Proposition 2: First, the change in the candidate's (8) due to a change in q_C is proportional to weighted changes in expected utilities of the voters. Second, according to (29), the change in expected utility of voter (v, J) due to an infinitesimal change $\Delta q_{C,i}$ is $u_{C,i}^J \xi_i^J$. ∎