The pronouncements of paranoid politicians

Guido, Cataife

Department of Economics, Washington University in St. Louis.

10 August 2007

Abstract

This paper models the strategic encounter of two office-motivated candidates who may or may not announce policy. In the case of no announcement, the voters rank the candidates according to prior beliefs. In the case of announcement, the candidates cannot avoid a degree of noise in the voters' interpretation of their announcements. We show that this simple deviation from the standard Downsian setting suffices to overcome previous impossibility results which suggest that not announcing policy can never occur in equilibrium.

Also, we extend the model to study the equilibrium when candidates are ambiguity averse. An ambiguity averse candidate is interpreted as being concerned about an ongoing negative campaign against him. This negative campaign would consist in inducing the voters to adopt some interpretation of the candidate's announcement unfavorable to his electoral performance. We show that under ambiguity aversion the candidates opt not to announce position under less stringent conditions than expected utility.

Finally, we use data on U.S. Senate elections to test an empirical implication of the model. We find that the relevant coefficient has the sign predicted by the theory and is statistically significant.

The Pronouncements of Paranoid Politicians*

PRELIMINARY

Guido Cataife†

August 08, 2007

1 Introduction

Not announcing a position on sensitive issues is a common practice of candidates across elections. In the presidential campaign of 2004, for instance, the phenomenon was so evident that the *New York Times* devoted a series of editorials to bring up "the toughest challenges facing the country" that were either "being glossed over with one-liners" or "not being mentioned" by candidates.[1] The two most striking issues were the federal budget deficit and agricultural subsidies.

In this paper, we address the following questions: Why are some important issues never discussed in an election? Why are some other less sensitive issues often discussed? In democracies, elections are *the* means by which citizens' preferences are translated into policy outcomes. The disregard of sensitive issues by candidates during political contests then undermines the efficacy of democracy in achieving its main goal: being the government by the people. Furthermore, if the disregard of the parties for certain issues is persistent over time, polities may enter into a "political trap" in which highly unpopular policies continue to be executed by politicians just because the voters lack the chance to vote them out. For example, consider the failure of the candidates to discuss the budget deficit in the presidential campaign of 2004. Given this fact, to what extent does the current budget deficit reflect a popular choice, no matter how distorted? How large are the welfare losses from having a budget deficit other than the socially desired one? How will society deal with the issue if the parties opt not to discuss it again in the next election? These rhetorical questions point out potentially deleterious effects of political non-salience. In this paper, by salience we mean the phenomenon that both candidates opt to announce a position, and by non-salience we mean the phenomenon that at least one of the candidates in an election opts not to announce a position on some issue.

*This research has been benefited from comments and discussion from Enriqueta Aragonès, Marcus Berliant, Michele Boldrin, Randall Calvert, Hideo Konishi, Amanda Friedenberg, Edward Greenberg, Gary Miller, Ignacio Monzon, Wolfgang Pesendorfer, Paul Rothstein, Norman Schofield and Robert W. Walker. I especially thank my advisors, Marcus Berliant and Norman Schofield, for their constant support. All errors remain my own.

†Department of Economics, Washington University in St. Louis, Campus Box 1208, One Brookings Drive, St. Louis, MO 63130-4899 USA. Email: gcataife@artsci.wustl.edu.

[1] http://www.nytimes.com/ref/opinion/campaign-issues.html?pagewanted=all

Several prominent scholars, including Downs (1957), Shepsle (1972), and Page (1976), pointed out the relevance of this problem a long time ago. Yet, the incentives of the candidates not to discuss relevant issues are not well understood. In fact, rational choice theory has provided strong negative results for this phenomenon. Recently, Berliant and Konishi (2005) proved that, under very mild assumptions, non-salience cannot happen in a robust equilibrium within the standard expected utility framework. In their model, two office-motivated candidates may announce or not announce policy. The candidates are uncertain regarding the most preferred policy of the voters. In case a candidate does not announce, the voters rank the candidate according to their prior beliefs regarding the policy to be implemented by the candidate if he wins the election. In this setting, risk aversion on the part of the voters suffices to guarantee that, in equilibrium, the candidates will always prefer to announce policy. The authors further suggest that robust non-salience may be obtained if one abandoned the expected utility framework in favor of the ambiguity aversion framework.

The aim of this paper is to shed some light on the conditions under which office-motivated candidates may have incentives not to announce positions. This paper shows that only minor modifications to the standard Downsian model are required to generate non-salience in equilibrium. We consider this a contribution to the literature on political ambiguity, since, to the best of our knowledge, there is no paper to date that generates ambiguity within a simple framework or without reverting to strong assumptions. In the next paragraphs we discuss the assumptions made throughout.

The first premise is that we assume a one-dimensional polity, where two office-motivated candidates know the distribution of voters' ideal points with certainty. Since our main objective is to show that, contrary to what previous impossibility results have suggested, non-salience can indeed be obtained in equilibrium, a one-dimensional proof makes our main point. Besides, for our results to be strong, we need to show that non-salience can be obtained even in the presence of Condorcet (or "unbeaten") points. Because multi-dimensional models usually present Condorcet cycles, even if we provided a multi-dimensional model, the most exigent test for the model would be to hold in the particular case of one dimension. Examples of one-dimensional models are abundant in the literature on spatial voting. See Osborne (2000), Morelli (2004), and, in the specific area of ambiguity in elections, see Aragones and Postlewaite (2000).

The second and most important premise of our research is that the candidates cannot avoid a certain degree of noise in the voters' interpretation of their announcements. In the real world, we identify two main sources of noise in the interpretation of announcements. The first one is language, which is an ambiguous means of communication. The seminal work in this area has been written by the semiotician Ferdinand Saussure (1986), who investigated the miscommunication problems generated by language mainly by elaborating on the dichotomy *signified* (*signifié*) / *signifier* (*signifiant*).[2] The second source of noise is the intermediation of third parties between politicians and voters. An obvious example would be the mass media (see Ansolabehere *et al.*, 1993). We refer to the voters' noisy signal of the announcement by a candidate as the *interpretation* (of the announcement).

If a candidate does not announce policy, the voters use their beliefs about the policy that he/she will implement. On the other hand, if a candidate does announce policy, the voters expect the candidate to implement their interpretation of his announcement. We consider this assumption the equivalent, within our model, to the standard assumption of the literature that the candidates commit

[2] We find it surprising that his work remains unexplored by economists, since it has impacted the ideas of leading scholars in several other disciplines, including philosophy, psychoanalysis and anthropology.

to implement their announcements. The usual justification for this assumption is that politicians that renege their promises are likely to be punished in future elections. Importantly, we understand that the assumption that the voters expect the candidates to implement the interpretation of their announcements is the only compatible with the premise of political accountability. If, for example, we assumed that the voters expect the winner to implement a policy randomly drawn among all the possible interpretations of his announcement, then the candidates could not be made accountable for the implementation of any specific policy. (The severity of this problem increases with the number of possible interpretations of the announcement.) Clearly, the lack of accountability undermines the very nature of a representative government, since the votes themselves no longer reflect support to any specific platform. For this reason, we suggest that the voters' expectation that the candidates will implement the interpretations of their announcements may be considered as a "focal" point, i.e., as an institutional arrangement made between politicians and voters in order to make democracy viable in a context of noisy announcements. Our last remark regarding this assumption is that, were we to assume instead, voters expect the candidates to randomize among the possible interpretations, the results would be as follows. Within expected utility, non-salience would only hold if we made the stronger assumptions that voters are risk averse and the variance of the lottery from which the policy implemented by a winner that did not announce is drawn is higher than the variance of the lottery from which the interpretation of the announcements is drawn. However, within the ambiguity aversion framework, similar results would apply.

The third premise is that, rather than maximizing probability of winning, the candidates maximize expected vote share. This assumption is very frequent in the literature; for example, see Adams (1999), Lin et al. (1999), and Schofield (2006). (For a discussion on the equivalence of the assumptions of maximizing probability of winning and maximizing vote share with probabilistic voters, see Patty, 2007). However, some scholars consider that maximization of probability of winning is a more appropriate assumption than maximization of vote share. In any case, we remark that non-salience can also be generated in this model under the assumption that the candidates maximize expected probability of winning.[3] We do not adopt the latter assumption throughout for two reasons. First, the minimum number of possible interpretations of the candidates' announcements required to obtain an equilibrium in which both candidates do not announce position increases from two (under the assumption of maximization of expected vote share) to three (under the assumption of maximization of probability of winning), complicating the model substantively. Second, the assumption that the candidates maximize expected vote share simplifies the extension of the model that assumes ambiguity averse candidates.

The main result of our paper is that non-salience may occur in equilibrium, under certain values of the parameters, for any distribution of voters' ideal points. As a general principle, in order to generate non-salience, it is required that we either fix the mean of the voters' beliefs and increase the noise that enters in the voters' interpretation of the announcements or fix the magnitude of this noise and make the mean of the voters' beliefs close enough to the electoral center. The intuition for this result is that if the mean of the beliefs is close enough to the median ideal point, then not announcing policy gives the candidates a vote share close to the optimum in the absence of noise and, at the same time, provides to them a means to circumvent the noise in their announcements.

The assumption that there are several possible interpretations of the announcements made by the

[3] This point is discussed in Section V.

candidates allows us to pursue the following natural extension of the model. A plausible concern of the candidates may be that some external agent, perhaps his opponent or the mass media, attempts to induce, among the several possible interpretations, one that is unfavorable to him. The reason that the candidates may be concerned about the possibility that *their opponents* attempt to manipulate the interpretations in an unfavorable manner is that elections are zero-sum games. The reason that they may be concerned about the possibility that *the mass media* attempt to manipulate the interpretations in an unfavorable manner is that the media may have interests different than reporting facts objectively (see Ansolabehere *et al.*, 1993).

However, the mere fact that the media can *potentially* bias the candidates' message to the voters does not necessarily imply that there actually is such a media bias. Whether or not such a bias exists is an empirical question, and there is no consensus in the literature on this issue. On one hand, mass media experts have encountered empirical evidence that although people are convinced of the existence of media bias, there is no such bias (D'Alessio and Allen 2000, Shah *et al.* 1999, Robinson and Sheenan 1983). Importantly, this literature has encountered what it has termed the *hostile media effect*: although there is no media bias, individuals firmly believe there is such a bias and, furthermore, that its direction is against their viewpoints (Dalton *et al.*, 1998, Vallone *et al.*, 1985). On the other hand, a recent study by two economists (Groseclose and Milyo, 2005) finds significant evidence of media bias. Note that both bodies of literature support the idea that people in general may be concerned about the possibility that the media alter their messages in an unfavorable manner. The following quotation by former Democratic presidential candidate Michael Dukakis, in which he explains his defeat in the 1988 election, provides evidence that the hostile media effect is present, in particular, among political candidates, and plays a role in shaping their political strategies.

> "I said in my acceptance speech at Atlanta that the 1988 election was not about ideology but about competence. I was wrong. It was about phraseology. It was about 10-second sound bites. And made-for-TV backdrops. And going negative. I made a lot of mistakes in the '88 campaign, but none was as damaging as my failure to understand this phenomenon, and the need to respond immediately and effectively to distortions of one's record and one's positions"[4]

We pursue an extension of the model that accounts for the phenomenon of hostile media effect. This is done through abandoning the expected utility framework in favor of ambiguity aversion. Recall that in the specialized literature, it is customary to see ambiguity averse agents as acting *as if* "out there" there were a malevolent third party playing a zero-sum game against them (Maccheroni *et al.*, 2006a). Therefore, once we assume the candidates are ambiguity averse, it follows that each of them acts *as if* there were a malevolent influence, presumably his opponent or the mass media, playing against him, by inducing an unfavorable interpretation of their announcements.

Two groups of papers may be distinguished for the purpose of describing where our model fits in the literature. The first group addresses the problem of ambiguity, rather than salience properly speaking, because candidates are allowed to announce lotteries. Hence, this body of literature explains why candidates may prefer to announce *many* (instead of *any*) positions on certain issues. Shepsle (1972), the seminal paper within this group, finds that in order to obtain ambiguity in equilibrium,

[4] Extracted from Johnson-Cartee and Copeland (1997). Original quotation in Butterfield (1990).

voters need to be risk lovers. Since the assumption of risk-loving voters is controversial, this result has motivated further research by social choice scholars. Since Alesina and Cukierman (1990) also include a noise term in their model, we wish to make clear that their model is very different from ours. In particular, in their model, a necessary condition to obtain ambiguity in equilibrium is that candidates have policy preferences. In our model, candidates are strictly office-motivated. Another necessary condition in their model is that candidates aim to run for re-election, and hence the dynamic nature of their model. Our model is static, and hence re-electoral concerns are implausible. Also, in their model, candidates can choose the magnitude of noise, and, in equilibrium, this magnitude is set at levels higher than zero. In our model, if candidates could choose the magnitude of noise in the interpretation of their announcements, they would set it at zero (assuming risk averse voters). In Aragones and Neeman (2000), the candidates are assumed to (i) have preferences over policy (ii) not commit to implement their promises and (iii) be concerned about their probability of re-election. They show that for sufficiently large degrees of uncertainty on the part of the candidates regarding the distribution of voters' ideal points, ambiguity occurs in equilibrium. As in Alesina and Cukierman (1990), in Aragones and Neeman (2000), ambiguity decreases the candidates' probability of winning office, and it is driven by the assumption that candidates care about policy. Aragones and Postlewaite (2002) assume that the candidates only care about office; in their setting, candidates cannot guarantee voters that their promises will be honored.

The second group of papers addresses the problem of salience properly speaking. Instead of allowing the candidates to announce any possible lottery, this group of papers gives them the candidates the option of not announcing any position. Therefore, candidates can announce any position in the policy space or not announce a position at all. Following the Downsian setting, candidates commit to implement their promises if any was made; otherwise, voters' use their beliefs regarding the policy that will be implemented by a winner that did not announce policy. Berliant and Konishi (2005) add uncertainty on the part of the candidates regarding the distribution of voters' most preferred policies, and prove that non-salience can never be a robust equilibrium. Since in their model it is just assumed that candidates are expected utility maximizers and voters are risk neutral or risk averse, their impossibility results constitute a substantial challenge to the literature. We wish to make clear the point that, although they assume that candidates maximize probability of winning, their results also hold when candidates maximize vote share. The reason is that the key to their results is that elections are zero-sum game (which is true under both premises). A formal proof follows readily from both Proposition 8 and Proposition 12 in the main body of this paper.

Berliant and Konishi (2005) seemingly contradict Glazer (1999), who offers an informal model in which candidates prefer not to announce policy. However, this seeming contradiction is resolved when one makes explicit the implicit assumption in Glazer (1999) that the candidates do not know the mean of the voters' beliefs regarding the policy that will be implemented by a winner that did not announce. The main problem with this assumption is that it violates the standard game theoretic postulate that the structure of the game is *common knowledge* (Aumann, 1976). Berliant and Konishi (2005) suggest that ambiguity aversion may be necessary to generate an equilibrium where at least some candidate does not announce a position. Our model distances itself from theirs in that, by introducing a slight deviation from the standard Downsian setting, it does succeed in generating such an equilibrium, even within expected utility. Aragones *et al.* (2007) study a repeated election model in which strictly policy-motivated candidates have the option of not announcing policy. However, their focus is not on generating non-salience, but in analyzing equilibria in which the voters threat to punish candidates who renege on their campaign promises and in which all campaign promises are believed by voters and honored by candidates.

The impossibility results of Berliant and Konishi (2005) may explain why the papers on ambiguity have significantly outnumbered the papers on salience proper, especially since the latter are more parsimonious. However, our paper, which models salience proper, finds that the impossibility results may be overcome by just introducing a simple and not unreasonable assumption. Previous results in the literature have generated ambiguity by assuming: risk lover voters (Shepsle, 1972), absence of common knowledge (Glazer, 1990), inability of the candidates to commit to implement their promises (Aragones and Postlewaite, 2002). Our results are encouraging, since our model has been able to generate both partial non-salience and non-salience by slightly deviating from the standard Downsian setting. It suffices to assume that the interpretation of announcements depends on the state of the world.

Finally, two papers which do not address the problem of salience or ambiguity but are, nonetheless, related to our paper are Ghirardato and Katz (2002) and, especially, Bade (2003). These two papers, as well as Berliant and Konishi (2005), share the common feature with our paper of introducing ambiguity aversion in the study of elections. Ghirardato and Katz (2002) apply ambiguity aversion in order to explain why voters may abstain, even when voting is assumed to be costless. Bade (2003) introduces ambiguity aversion *à la* Gilboa-Schmeidler to generate equilibrium in multi-dimensional political contests.

A direct implication of the propositions obtained in this paper is that non-salience (at least one candidate opts not to announce a position) can be obtained in equilibrium if the mean of the voters' beliefs is close enough to the Condorcet winner in the standard Downsian game. The intuition for this result is as follows. Assume that a candidate, say 1, does not announce a position. Candidate 2's opportunity cost of responding with a no-announcement increases with the distance between the voters' beliefs (about the policy that will be implemented by a candidate that won without announcing policy) and the Condorcet winner. To see this, note that given the spatial nature of the game, the less centrist the voters believe candidate 1's implemented policy will be, the larger the vote share candidate 2 could achieve by implementing a response which is epsilon away from candidate 1's implicit policy position. Being epsilon closer to the Condorcet winner, candidate 2's announcement could restrict candidate 1's appeal to the disaffected voters in an extreme of the voters' distribution, irrespective of the state of the world. Therefore, there must be a point such that if the distance between the voters' beliefs and the Condorcet winner is smaller than the distance between that point and the Condorcet winner, the cost of alienating some voters is smaller than the "cost" of announcing a "noisy" interpretation.[5]

Since it plays an important role in generating the results, we give some intuition on why the noise that enters in the interpretation of the announcements decreases the candidates' expected utility of announcing policy. Consider the following example. Suppose that: (i) the policy space is $[0, 1]$, (ii) the voters distribute uniformly over the policy space, (iii) the noise term entering additively in the

[5]The fact that we model non-salience (and not ambiguity) implies that, if the candidates decide not to announce, they are assigned a position (or, more precisely, a lottery) by default. Moreover, the candidates have no say over this lottery. Hence, there is no simple way to extend the model in order to obtain that, in equilibrium, candidates prefer not announce position for extreme values of η but not for central values of η. We believe that this is nice feature of the model, since it offers a testable implication which is robust to any small perturbation of the model.

voters' interpretation of the announcement made by candidate 2 is ± 0.1, (iv) there is no noise entering in the voter's interpretation of the announcement made by candidate 1, (v) the two possible states of the world occur with probability $\frac{1}{2}$ each. Also, fix candidate 1's announcement: $d_1 = 0.2$. In the standard Downsian model (where interpretations and announcements are identical), assuming that the payoffs of the candidates are given by their expected vote share, the best response of candidate 2 is to announce $d_2 = 0.2 + \epsilon$,[6] which gives candidate 2 a payoff of 0.8. However, in the model in which candidate 2's interpretation is noisy, $d_2 = 0.2 + \epsilon$, gives candidate 2 a payoff of $\frac{0.2+(0.2+\epsilon-0.1)}{2} \to 0.15$ with probability $\frac{1}{2}$,[7] and a payoff of $1 - \frac{0.2+(0.2+\epsilon+0.1)}{2} \to 0.75$ with probability $\frac{1}{2}$, which implies that candidate 2's expected payoff is $\frac{1}{2}(0.15) + \frac{1}{2}(0.75) = 0.45$. Note that candidate 2's expected utility is lower in the noisy model for a fixed response. But, of course, $d_2 = 0.2 + \epsilon$ need not be candidate 2's best response to $d_1 = 0.2$. Indeed, in Appendix A we show that the best response of candidate 2 in this case is $d_2 = 0.2 + 0.1 + \epsilon$. In this case, candidate 2's gets a payoff of $1 - \frac{0.2+(0.2+0.1+\epsilon-0.1)}{2} \to 0.8$ with probability $\frac{1}{2}$ and a payoff of $1 - \frac{0.2+(0.2+0.1+\epsilon+0.1)}{2} \to 0.7$ with probability $\frac{1}{2}$, which implies an expected payoff of $\frac{1}{2}(0.8) + \frac{1}{2}(0.7) = 0.75$. This example illustrates that, conditional on playing his best response, a candidate whose interpretation is subject to some noise pays a (utility) premium of $\frac{1}{2}$ times the magnitude of the noise entering in the voters' interpretation of the announcement –namely: $\frac{1}{2}(0.1) = 0.05$. The greater the magnitude of this noise, the higher the premium paid in expectation by the candidate is. Now, assume that, instead of announcing $d_1 = 0.2$, candidate 1 declares $\emptyset$, which stands for no-announcement. Also, assume that the mean of the voters' beliefs regarding the policy that will be implemented by a winner that did not announce policy is $\eta = 0.2$. If, for expositional clarity, we assume that the voters are risk neutral, this example and the previous one, in which $d_1 = 0.2$, are obviously equivalent. Note that, if candidate 2 wants to avoid paying the expected utility premium associated to announcing a position, which we said sums up to $\frac{1}{2}(0.1)$, there is only one strategy he can play. In this model (unlike in the standard Downsian setting), announcing the same strategy as the opponents' one does not guarantee an expected vote share of $\frac{1}{2}$, since the interpretation of the announcements may be different even when the strategies are identical. However, if both candidates declare $\emptyset$, there is no announcement, there is no noise and, henceforth, the candidates have identical appeal to each and every voter, implying that the expected vote share is $\frac{1}{2}$ for each of them. To sum up, candidate 2 may avoid paying the utility premium generated by the noise in the interpretation of his announcement only if he declares $\emptyset$. From candidate 2's perspective, this choice has the drawback that it deprives him of the opportunity of taking advantage of the fact that his opponent, candidate 1, has adopted an implicit position which is extreme (namely: 0.2), and therefore unappealing to the bulk of the voters.[8] Because this second example is equivalent to the first one, it follows that the best candidate 2 can do announcing policy is to announce $d_2 = 0.2 + 0.1 + \epsilon$, which gives him an expected utility of 0.75, which is higher than the expected utility of responding with a no-announcement, namely $\frac{1}{2}$. If, instead, the mean of the voters' beliefs were, say, 0.47, instead of 0.2, then the best candidate 2 can do conditional on announcing policy is to announce

[6] As we already remarked, strictly speaking, in the standard Downsian model there is no best response to any position other than the median. However, this is not really a problem, since we can always use the concept of ε-equilibrium. The latter only requires that no player can gain more than ε in expected utility by deviating from his strategy, given the strategy of his opponent.

[7] Recall that voters distribute uniformly over $[0, 1]$.

[8] To see this, note that because candidate 1 declared $\emptyset$, by risk neutrality the voters associated him with the mean of their beliefs, namely 0.2, which is far away from the median, namely 0.5.

$d_2 = 0.47 + 0.1 + \epsilon$. In this case, candidate 2's gets a payoff of $1 - \frac{0.47+(0.47+0.1+\epsilon-0.1)}{2} \to 0.53$ with probability $\frac{1}{2}$ and a payoff of $1 - \frac{0.47+(0.47+0.1+\epsilon+0.1)}{2} \to 0.43$ with probability $\frac{1}{2}$, which implies an expected payoff of $\frac{1}{2}(0.53) + \frac{1}{2}(0.43) = 0.48$, which is lower than the expected utility of responding with a no-announcement, namely $\frac{1}{2}$. In this case not announcing policy is candidate 1's best response.

This discussion makes evident that, in order to obtain non-salience, we can either fix the mean of the voters' beliefs and increase the noise that enters in the voters' interpretation of the announcements or fix the magnitude of this noise and make the mean of the voters' beliefs closer and closer to the electoral center. In the main body of the paper, we show that there is a certain threshold for the mean of the voters' beliefs for which candidate 2's incentive to deviate from $\emptyset$ in order to capture the votes of the electoral center is high enough to overcome the cost generated by announcing a position.

The example above gives some intuition as to why candidates that behave like simple expected utility maximizers may opt not to announce a position in a model that slightly deviates from the standard Downsian setting. The key to our results is that this slight modification induces a cost of announcing position, in expected utility terms, which is proportional to the magnitude of the noise entering in the interpretation of the announcements.

When we introduce ambiguity aversion, the results are stronger, in the sense that the conditions under which candidates prefer not to announce policy are less stringent. To see why, note that within the ambiguity aversion framework, the candidates are endowed with multiple beliefs, which in this case translates into multiple beliefs regarding the probability distributions of the states of the world. For example, consider the extreme case in which the candidates believe that any conceivable probability distribution of the states of the world is possible. Since this includes the degenerate probabilities (in which one state occurs with probability 1 and the other with probability 0), it follows readily that an ambiguity averse candidate will act as if, with probability 1, Nature will pick the most unfavorable scenario for him.[9] Hence, in terms of the second example above, the (maximin) expected payoff of candidate 2 becomes $0(0.53) + 1(0.43) = 0.43$ instead of 0.48. Then, not announcing position becomes candidate 2's best response for a larger set of values of the mean of the voters' beliefs. We wish to make clear that this example is simply to anticipate some intuition, and that the actual ambiguity aversion framework adopted in this paper, variational preferences (Maccheroni *et al.*, 2006a), is much richer, realistic and better accommodates the interpretation of the model.

Finally, to the best of our knowledge, there is no paper in the literature that offers an empirical study of the phenomenon of ambiguity or salience in the election context. A nice, distinguishing, feature of our model is that it generates a testable implication. We carry out an empirical analysis of this testable implication, centered on U.S. Senate elections. First, we suggest that, under reasonable premises (discussed in Section VI), a sensible measure of the cost of reducing the noise in the voters' interpretation of the announcements by the candidates is given by the cost of airing political advertising on TV in the particular state the candidate is running for Senator. Second, we suggest that the passed/failed result of the *National Political Awareness Test*, conducted on each U.S. Senate candidate by *Project Vote Smart* is a good proxy of whether or not the candidates announced positions. Then, we run a probit model to test the relationship between the cost of reducing the noise in the voters' interpretation of the announcements and the probability of announcing position. Our empirical analysis shows that, *ceteris paribus*, the higher the cost of TV advertising in a political

[9] We are implicitly assuming that every probability distribution can be induced by Nature at the same cost.

district, the lower the probability that the candidate will announce position. This relationship is found to be statistically significant at the 5% level, and robust to several minor modifications of the empirical model.

The remainder of the paper is organized as follows. In Section II, we outline the model (within the expected utility framework). Section III discusses the problem of the voters. Section IV discusses the problem of the candidates and states the main results. Section V discusses two extensions of the model within expected utility: two-candidate competition assuming maximization of expected probability of winning (instead of maximization of expected vote share) and multi-candidate competition. Section VI studies the problem under a behavioral assumption different from expected utility, namely ambiguity aversion, in order to account for the phenomenon of negative campaigning. Section VII tests an empirical implication of our model. Section VIII concludes.

2 The Model

Consider a one-issue, simultaneous-move political contest with two candidates, who have the objective of maximizing expected vote share. The candidates are assumed to know the distribution of voters' most preferred policies with certainty. The declaration of a platform by a candidate can either (i) contain the announcement of a position in the policy space, or (ii) be empty, representing no-announcement. Note that we are using the term "declare a platform" to denote either an empty announcement or a non-empty announcement, and keeping the term "announcement" to refer to a *non-empty* announcement.

The distinguishing feature of this model is that the voters receive a noisy signal of the announcements, which we call the *interpretation* (of the announcement). This signal is assumed to be the same for all the voters. Different realizations of the noise represent different *states of the world*. (Intuitively, the magnitude of the noise entering in the interpretation depends on the technology of communication available to the candidates.) Candidates are committed to execute the interpretation of their announced policy. In case the platform is empty, the voters assume that the candidate will implement a policy that will be randomly drawn from a known distribution. For simplicity, this distribution is assumed to be the same for both candidates.

To simplify the analysis, we make the following assumptions. First, the noise can only take two different values. One value represents a leftist bias, and the other a rightist bias.[10] Second, the technology of communication is neutral, in the sense that the rightist and the leftist biases have equal probability. Third, the bias in the voters' interpretation of candidate 1's announcement is drawn independently to the bias in the voter's interpretation of candidate 2's announcement.

At this point, some minimal notation needs to be introduced. We let c index candidates 1 and 2. We use $D \cup \{\emptyset\}$ to denote the policy space of the candidates, where D is a compact interval and $\emptyset$ denotes a no-announcement. There is a set of *states of the world*, $S \equiv S_1 \times S_2$, with elements denoted by $s \equiv (s_1, s_2)$. For $c \in \{1, 2\}$, s_c, represents the bias in the interpretation of candidate c's announcement. There are two possible biases: leftist and rightist. Then, $s \in S$ represents any possible combination of biases for the candidates.[11] We let the scalar $\delta_c(s)$ denote the noise that

[10] For example: a candidate announces policy a and the voters' interpretation is either $a + \delta$ or $a - \delta$, where δ is any (small and positive) scalar.

[11] We refer to any subset of S as an *event*. We let Σ denote an algebra of events and $\Delta(\Sigma)$ the set of all finitely

enters in the voters' interpretation of candidate c's announcement conditional on the state of the world s.

For the sake of exposition, we proceed to define four models that will be used throughout.

Definition 1 *Model $M(D\cup\{\emptyset\},\delta_1(s),\delta_2(s))$. The strategy space is $D\cup\{\emptyset\}$, candidates 1 and 2 face noises $\delta_1(s)$ and $\delta_2(s)$, respectively, conditional to the state of the world s.*

$M(D\cup\{\emptyset\},\delta_1(s),\delta_2(s))$ is the general model under consideration. By restricting the policy space to D (instead of $D\cup\{\emptyset\}$) and setting $\delta_1(s)=\delta_2(s)=0$ for all s, we obtain the standard Downsian setting. The definitions below follow from specializing $M(D\cup\{\emptyset\},\delta_1(s),\delta_2(s))$ in other ways.

Definition 2 *Model $M(D,\delta_1(s),\delta_2(s))$. This model corresponds to the case in which the strategy space of the candidates is restricted to D, so that $\emptyset$ is not feasible. Hence, we also refer to it as the* restricted model in which both candidates face noise *in the interpretation of their announcements.*

Although the strategy space of the model $M(D,\delta_1(s),\delta_2(s))$ is identical to the strategy space of the standard Downsian model, the two models differ in that, while in $M(D,\delta_1(s),\delta_2(s))$ the candidates face noise in the interpretation of their announcements, in the standard Downsian model they do not.

Definition 3 *Model $M(D\cup\{\emptyset\},0,\delta_2(s))$. In this model, there is no noise entering in the voters' interpretation of candidate 1's announcement.*

Then, we use $M(D\cup\{\emptyset\},0,\delta_2(s))$ to refer to the specialization of $M(D\cup\{\emptyset\},\delta_1(s),\delta_2(s))$ in which $\delta_1(s)=0$ for every state of the world s. In other words, $M(D\cup\{\emptyset\},0,\delta_2(s))$ assumes that, while candidate 2 faces an imperfect technology (so that announcement and interpretation may differ), candidate 1 has access to a perfect technology of communication (so that announcement and interpretation are identical). It is important to note that, even if the interpretation of candidate 2's announcement is identical to the announcement itself, it is still the case that the vote shares of *both* candidates are non-deterministic. This is because the vote share of each candidate depends on both his announcement and his opponent's announcement. Hence, as long as at least one candidate faces noise in the interpretation of his announcement, there will be a random component in the expected vote shares of both candidates.

Definition 4 *Model $M(D,0,\delta_2(s))$. This model corresponds to the case in which the strategy space of the candidates is restricted to D (so that $\emptyset$ is not feasible) and there is no noise entering in the voters' interpretation of candidate 1's announcement. Hence, we also refer to it as the* restricted model in which only candidate 2 faces noise *in the interpretation of his announcements.*

In other words, the model $M(D,0,\delta_2(s))$ not only restricts the space of strategies of the candidates but also assumes that only candidate 2 faces noise in the interpretation of his announcement.

For any of these models, the timing of the game is as follows. First, both candidates declare a platform. Second, Nature draws a state of the world, which is used by the voters to form their interpretation of the announcements. Third, the voters use this interpretation of the announcements to cast their votes.

additive probabilities on Σ.

3 The Voters

We index candidates by $c = 1, 2$ and assume a continuum of voters, indexed by i, whose most preferred policies ("ideal points") are distributed over the unit interval according to Ψ. We use $d_c \in D \cup \{\emptyset\}$ to denote *the declaration of platform made by candidate* c in a one-dimensional polity. Here, $d_c \in D \equiv [0,1]$ denotes an *announcement* and $d_c = \emptyset$ denotes a *no-announcement.* In case $d_c = \emptyset$ for some c, in any state of the world, the voters believe that, in case c wins the election, the policy to be implemented will be drawn from a known distribution, common to both candidates, which we leave unspecified. In case of announcement, all voters receive the same noisy signal of d_c. We use δ_c, with $c \in \{1, 2\}$, to denote the noise in the signal of candidate c's announcement; this noise depends on the coordinate of the state of the world that corresponds to candidate c. Formally, $\delta_c : S_c \to \{-\delta, \delta\}$, where δ is a (small and) positive scalar. The simplest case we will consider corresponds to $\delta_c(l) = -\delta$ and $\delta_c(r) = \delta$. However, other cases will be also studied.

The *interpretation of candidate c's announcement*, denoted $i_c : [0,1] \to [-\delta_c(l), 1 + \delta_c(r)]$, $c \in \{1, 2\}$ has the following functional form

$$i_c(d_c, s_c) = d_c + \delta_c(s_c) \quad \forall d_c \neq \emptyset \tag{1}$$

At this time, the following technical point must be addressed. Combined, the assumptions that (i) $d_c \in [0,1]$ and (ii) Ψ is a distribution function over $[0,1]$, lead to the following issue with boundaries. If a candidate declared an extreme platform, say 0, voters could never have a leftist interpretation of this platform, since such an interpretation would imply that the voters believe the candidate announced a platform to the left of 0 and, therefore, outside of the policy space. There are several alternatives to formally deal with this problem. One of them is to assume that (i) $d_c \in \mathbb{R}$ and (ii) Ψ is a distribution function over $\mathbb{R}$. However, this would force us to abandon the closed-form results that are obtain when Ψ is a distribution function over $[0,1]$. Other alternative is to assume that: (i) $d_c \in \mathbb{R}$ and (ii) Ψ is a distribution function over $[0,1]$. The latter alternative would imply that the voters' ideal points are distributed over a *subset* of the policy space. Because rational candidates will never announce policy positions whose possible interpretations would lie outside the support of the distribution of voters' ideal points, it follows that, in equilibrium, the policies announced by the candidates will necessary belong to a compact subset of $\mathbb{R}$. Hence, as a practical matter, there would be no substantive difference with the case in which we assume that (i) $d_c \in [0,1]$ and (ii) Ψ is a distribution function over $[0,1]$. We avoid this unnecessary complication sticking to the assumptions that (i) $d_c \in [0,1]$ and (ii) Ψ is a distribution function over $[0,1]$.

Conditional on state $s \in S$, the utility that voter i derives from the declaration of platform d_c by candidate c, $c \in \{1, 2\}$, is given by $u_i : [0,1] \times S_c \times [0,1] \to \mathbb{R}$; we write $u_i(d_c, s_c, x_i)$. The usual spatial-preferences given by

$$u_i(d_c, s_c, x_i) = \begin{cases} -\left|d_c + \delta_c(s_c) - x_i\right| & \text{for} \quad d_c \neq \emptyset \\ \int -\left|n - x_i\right| dF(n) & \text{for} \quad d_c = \emptyset \end{cases} \tag{2}$$

are assumed. Here, $|\cdot|$ denotes absolute value, x_i is voter i's ideal point and $F(n)$ represents the voters' beliefs regarding the policy $n \in D - \{\emptyset\}$ to be implemented if the winner does not announce policy. It is assumed that n is the realization of a random variable, N, from which the policy implemented by a candidate that wins without announcing is drawn.[12] Therefore, the distribution

[12] The symbols n and N stand for no-announcement.

$F(n)$ gives the probability that the realized policy for a winner with no-announcement is less than or equal to n. We assume that F is *common knowledge*[13] and represents the voters beliefs for both candidates; we use η to denote its mean.

Consider $M(D \cup \{\emptyset\}, 0, \delta_2(s))$. This implies that, while $i_2 = d_2 + \delta(s_2)$, $i_1 = d_1$. Then, the probability that voter i votes for candidate 2 conditional on state $s \in S$ is given by

$$\rho_{i,2}(d_2, d_1|s) = \begin{cases} \Pr\left[-\left|d_2 + \delta_2(s_2) - x_i\right| > -\left|d_1 - x_i\right|\right] & \text{for} \quad d_2 \neq \emptyset, d_1 \neq \emptyset \\ \Pr\left[\int -\left|n - x_i\right| dF(n) > -\left|d_1 - x_i\right|\right] & \text{for} \quad d_2 = \emptyset, d_1 \neq \emptyset \\ \Pr\left[\int -\left|d_2 + \delta_2(s_2) - x_i\right| > \int -\left|n - x_i\right| dF(n)\right] & \text{for} \quad d_2 \neq \emptyset, d_1 = \emptyset \\ \frac{1}{2} & \text{for} \quad d_2 = \emptyset, d_1 = \emptyset \end{cases} \tag{3}$$

Note that $\rho_{i,2}(d_2, d_1|s)$ only takes the values 0, 1 and $\frac{1}{2}$, representing support for candidate 1, support for candidate 2 and tie, respectively. Note that, for $d_2 = \emptyset$ there is no noise in the utility u of the voters –if there is no signal, there is no noise–, but there is uncertainty, represented by F. Finally, within expected utility, it must be that $\rho_{i,1}(d_1, d_2|s) = 1 - \rho_{i,2}(d_2, d_1|s)$.

Analogous procedures can be applied to derive $\rho_{i,c}(d_1, d_2|s)$ for the remainder models. We omit this step to save space.

4 The Problem of the Candidates

The *vote share of candidate c conditional to state $s \in S$* is

$$v_c(d_c, d_{-c}|s) = \int \rho_{i,c}(d_2, d_1|s) d\Psi.$$

Given that the preferences of the voters are perfectly spatial, when both candidates announce policy, the vote share of candidate c conditional on state s is the area below the density function of Ψ in between one of the extremes of the distribution and the mean of the interpretations of the announcements of the candidates.

Unless we explicitly indicate otherwise, throughout we consider $M(D \cup \{\emptyset\}, 0, \delta_2(s))$, so that the space of states of the world has only two elements. Let us use $V_c(d_c, d_{-c})$ to denote the expected vote share of candidate $c \in \{1, 2\}$, with $-c \in \{1, 2\}$ and $-c \neq c$, given the profile of strategies (d_c, d_{-c}). Within the expected utility framework it must be that

$$V_c(d_c, d_{-c}) = v_{c,l} q_l + v_{c,r} q_r,$$

where $v_{c,l}$ and $v_{c,r}$ are shortcuts for $v_c(d_c, d_{-c}|l)$ and $v_c(d_c, d_{-c}|r)$, respectively, and q_l and q_r denote the probability of a leftist and rightist interpretation, respectively. Whenever the context avoids confusion, we drop the subindex c in $v_{c,l}$ and $v_{c,r}$. Throughout, we make the neutral assumption that $q_l = q_r = \frac{1}{2}$, which implies that the leftist and the rightist interpretations of candidate 2's announcement are equally likely. The problem of candidate $c \in \{1, 2\}$ is to maximize $V_c(\cdot, d_{-c})$.

Note that to check whether a profile of strategies (d_c^*, d_{-c}^*) that constitutes an equilibrium in the restricted model $M(D, 0, \delta_2(s))$ also constitutes an equilibrium in the unrestricted model $M(D \cup \{\emptyset\}, 0, \delta_2(s))$, we just need to check that $V_c(d_c^*, d_{-c}^*) \geq V_c(\emptyset, d_{-c}^*)$ for $c, -c \in \{1, 2\}$ and $-c \neq c$.

[13] The assumption of common knowledge is fundamental. If it is violated (as in Glazer, 1990), results change drastically. However, the violation of this assumption would cast doubt on the internal consistency of the game in itself.

Throughout, we use $\widetilde{b}_c(d_{-c})$ to denote candidate c's best response to candidate's $-c$ strategy d_{-c} *in a restricted model* (i.e, in either $M(D, 0, \delta_2(s))$ or $M(D, \delta_1(s), \delta_2(s))$). Hence, $\widetilde{b}_c : D-\{\emptyset\} \to D-\{\emptyset\}$. Accordingly, we refer to $\widetilde{b}_c(d_{-c})$ as candidate c's *restricted best response* to d_{-c}. In Appendix A, we derive, for $M(D, 0, \delta_2(s))$, $\widetilde{b}_c(\cdot)$. Two considerations need to be taken into account. First, because $M(D, 0, \delta_2(s))$ assumes an asymmetry in terms of the noise structure that the candidates face, the restricted best response correspondences of the candidates need not be symmetric. Second, at least to some extent, the restricted best response correspondences depend on the particular assumption of Ψ (in this case, a uniform distribution). In this latter respect, we make clear the point that any symmetric and unimodal yields qualitatively similar results.

4.1 Equilibria within Expected Utility

In this subsection, we investigate under which conditions non-salience occurs in equilibrium. Proposition 5, below, deals with $M(D \cup \{\emptyset\}, 0, \delta_2(s))$. Instead, propositions 8 and 7, below, deal with $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$. The reason that propositions 5 deals with the case in which there is no noise in the interpretation of one candidate is that the analysis of equilibrium is greatly simplified under this assumption. The fact that Proposition 8 proves that there is a robust non-salient equilibrium under the assumption that both candidates face identical structures makes clear that our main result is not driven by any asymmetry in the candidates' problem. The proofs are contained in Appendix D, and rely heavily in the best response correspondences for $M(D \cup \{\emptyset\}, 0, \delta_2(s))$ derived in Appendix A.

Throughout, Ψ represents the distribution of ideal points of the voters, m represents the ideal point of the median voter and η represents the mean of qthe voters' beliefs regarding the policy that will be implemented by a winner that did not announce position. An alternative, more rigorous statement of Proposition 5, below, can be obtained by introducing the concept of ε-equilibrium. The latter would circumvent the fact that, strictly speaking, in the standard Downsian setting the best response of the candidates is empty for any position adopted by the opponent other than m. However, more notation would need to be introduced without gaining any further insight.

Proposition 5 *Consider $M(D \cup \{\emptyset\}, 0, \delta_2(s))$ and set $\delta_2(l) = -\delta_2(r)$. Assume $\eta \leq m$ (the case $\eta > m$ is analogous). A necessary condition for an equilibrium in which at least one candidate strictly prefers not to announce position is $2\Psi(\eta) + \Psi\left(\frac{\eta+d_2+\delta}{2}\right) - \Psi\left(\frac{\eta+d_2-\delta}{2}\right) > 1$.*

A direct corollary of Proposition 5 is that, in $M(D \cup \{\emptyset\}, 0, \delta_2(s))$, if $\delta_2(s) = 0$ for all s, so that there is no noise in the interpretation of any candidate, then, in equilibrium, the candidates do announce position. (To see this note that, if $\delta_2(s) = 0$ for all s, the condition in Proposition 5 reduces $\Psi(\eta) > \frac{1}{2}$, which can never be satisfied under the premise that $\eta \leq m$.) This corollary reinforces the idea that the noise in the interpretation of candidates' announcements is necessary to obtain non-salience. Also, Proposition 5 anticipates some of the results, in that it makes evident that necessary condition for non-salience is that one candidate η and m be close enough. The following case illustrates this point (its proof is contained in the Appendix).

Case 6 *If Ψ is the uniform distribution over the unit interval, then, for $\eta < m$, $\eta > m - \frac{1}{4}\delta$ is a necessary and sufficient condition for an equilibrium in which one candidate announces and the other does not. The equilibrium is $(d_1, d_2) = (\eta + \epsilon, \emptyset)$, where $\epsilon > 0$ is an infinitesimal scalar.*

Because the optimal strategy of the candidate that faces the smallest noise (in this case: no noise) is to make an epsilon deviation from the mean of the voters' beliefs about his opponent, it follows from this example that in equilibrium there is "convergence" in the sense that the implicit or explicit position of the candidates differs only by an epsilon.

Proposition 7 below shows that in the model in which both candidates face noise in the interpretation of their announcements, convergence to the ideal point of the median voter can be equilibrium. The theorem presents sufficient conditions for (m, m) to be an equilibrium. Of course, the necessary conditions may be less stringent. However, sufficient conditions are enough proof that our model is also compatible with salient equilibria.[14]

Proposition 7 *Consider $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$ and set $\delta_1(l) = -\delta_2(r)$ and $\delta_1(r) = -\delta_2(l)$, with $\delta_c(s) \neq 0$ for all c, s. A sufficient condition for the profile (m, m) to be a Nash equilibrium is that Ψ be symmetric and that $\eta < m + \delta_1(l) < m + \delta_1(r)$ or $\eta > m + \delta_1(r) > m + \delta_1(l)$.*

Note that Proposition 7 includes, but does not restrict to, the case $\delta_1(l) = \delta_2(l) = -\delta_1(r) = -\delta_2(r)$. Several other configurations of the parameters are compatible with its hypothesis. For example, $\delta_1(l) = -0.1$, $\delta_1(r) = 0.2$, $\delta_2(l) = -0.2$ and $\delta_2(r) = 0.1$.

Now, consider $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$ and assume that the magnitude of the noise terms is identical across candidates and states of the world. Can it be that in equilibrium both candidates prefer not to announce a policy position? This question is addressed in the next proposition, which summarizes our main result.

Proposition 8 *Consider $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$ and assume that: $\delta_1(l) = \delta_2(l) = -\delta_1(r) = -\delta_2(r)$. In this setting, the profile $(\emptyset, \emptyset)$ is a (strict) Nash equilibrium if and only if $\Psi(\eta + \delta) > 1 - \Psi(\eta) > \Psi(\eta - \delta)$.*

Proposition 8 says that there is an open set of values of the parameters in which both candidates not annnouncing position is a strict equilibrium. The condition $\Psi(\eta + \delta) > 1 - \Psi(\eta) > \Psi(\eta - \delta)$ has two possible readings. First, it may read that, for a fixed $\delta \geq 0$ the parameters η and m have to be close enough. Second, it may read that, for a fixed η and m, δ has to be large enough. The following example illustrates this point.

Case 9 *Assume Ψ is the uniform distribution over the unit interval. It follows readily from Proposition 5 that the profile $(\emptyset, \emptyset)$ is a (strict) Nash equilibrium if and only if $m + \frac{1}{2}\delta > \eta > m + \frac{1}{2}\delta$.*

Using a minimal set of assumptions, Propositions 5 and 8 overcome previous impossibility results in the literature. Indeed, Berliant and Konishi (2005) have proved that, within expected utility, in a robust equilibrium both candidates announce positions. The reason that we obtain different results is that our models are formally different. In their setting, there is no noise in the voters' interpretation of the candidates' announcement. Hence, in their model, the utility that candidate c gets from not announcing policy, assuming risk neutral voters, is identical to the utility he gets from announcing the mean of the voters' beliefs. In our model, the utility that, say, candidate 2

[14] If both candidates were assumed to expect no noise in the interpretation of his opponent, (m, m) would be a (strict) equilibrium, and would never obtain for the same configuration of parameters that a non-salient equilibrium does. Note, however, that the latter assumption would imply a violation of the premise of common knowledge.

gets from not announcing policy, assuming risk neutral voters, is identical to the utility he gets when *the voters' interpretation* coincides with the mean of their beliefs. The fact that there is a degree of imprecision in his announcement precludes the candidate from having a perfect control over how the voters will interpret his announcement. This degree of imprecision in candidate 2's technology of communication increases the payoff of not announcing relative to the payoff of announcing the best response in the restricted game, which leads the candidate to prefer not to announce for certain values of the parameters.[15]

[15] Also, the fact that there is no noise in absence of announcement in turn breaks down the translation invariance property of the preferences of the voters, key to the three proofs of the theorems in Berliant and Konishi (2005).

5 Extensions Within Expected Utility

5.1 Two-candidate competition with maximization of expected probability of winning

The insight behind Proposition 8 also applies when candidates are assumed to maximize expected probability of winning (instead of expected vote share). To see this, let us now assume that there are three (instead of two) possible interpretations of an announcement, and each of them occurs with probability $\frac{1}{3}$. We refer to these interpretations as the leftist, the rightist and the correct interpretation, and denote them by i_C, i_R and i_L, respectively. i_C and i_R are as before; i_C is the interpretation that coincides with the announcement of the candidate itself –hence, i_C may be seen as the "correct" interpretation. We refer the reader to Figure 1.

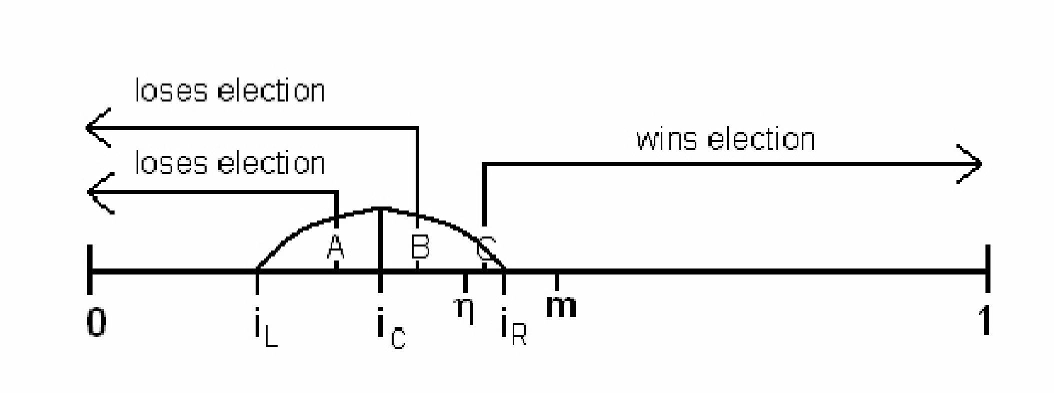

Figure 1: Any unilateral deviation from $(\emptyset, \emptyset)$ leaves the candidate worse off. Candidate 2 deviates to announce $d_2 = i_C$. The voters may interpret i_L, i_C or i_R, which bring about cut points A, B and C, respectively. While cut points A and B gives the victory to candidate 1, cut point C gives the victory to candidate 2. Since each interpretation has probability $\frac{1}{3}$, candidate 2's expected probability of victory, $\frac{1}{3}$, is lower than the probability of victory when no deviation occurs, $\frac{1}{2}$.

The figure makes evident that the profile $(\emptyset, \emptyset)$ is a Nash equilibrium. As Proposition 8 suggests, the key to obtain non-salience is that η be close enough to m. When the latter condition is satisfied, candidate 2 has basically two options. On one hand, he may announce the same implicit position than candidate 1, that is, η. In this case, if (i) $i_2 = i_C$, candidate 2 wins with probability $\frac{1}{2}$, (ii) $i_2 = i_L$, candidate 2 loses with certainty, (iii) $i_2 = i_R$, candidate 2 loses with certainty (note that this presumes that η and m are close enough or, alternatively, that δ is large enough). Since each interpretation has probability $\frac{1}{3}$, this implies that candidate 2's vote share is $\frac{1}{6}$. On the other hand, he may announce a position to the left of candidate 1's implicit position, such that the rightist interpretation of this announcement is associated to a cut point like point C in the figure, which gives him the victory. In this case, if (i) $i_2 = i_C$, candidate 2 loses with certainty, (ii) $i_2 = i_L$, candidate 2 loses with certainty, (iii) $i_2 = i_R$, candidate 2 wins with certainty. Since each interpretation has probability $\frac{1}{3}$, this implies that candidate 2's vote share is $\frac{1}{3}$. This strategy dominates the previous one, but is still strictly lower than the expected probability of winning at the profile $(\emptyset, \emptyset)$. Finally, note that to obtain non-salience it is required that, in addition to the leftist and rightist interpretation, there be also a third possible interpretation of the announcements, i_C. Otherwise, candidate 2 can generically obtain a vote share of $\frac{1}{2}$ using a strategy similar to the second one described in this paragraph.

5.2 Multi-candidate competition

The rationale behind Proposition 8 also applies in electoral competitions with more than two candidates. However, the more candidates participate in the competition, the more stringent are the conditions under which all candidates will prefer not to announce position. In fact, under conditions equivalent to those of Proposition 8, in a competition with N candidates, all candidates not announcing position is an equilibrium if and only if $\Psi(\eta+\delta) > \frac{2(N-1)}{N} - \Psi(\eta) > \Psi(\eta-\delta)$ (the proof is analogous to that of Proposition 8).

Example 10 *Assume that* Ψ *is the uniform distribution on* $[0,1]$. *Set* $N = 3$ *and* $\delta = \frac{4}{10}$. *The necessary and sufficient condition to obtain an equilibrium in which the three candidates opt for no-announcement reduces to* $\frac{8}{15} > \eta > \frac{7}{15}$.

The reason that, as N increases, the conditions under which non-salience occurs in equilibrium are more and more stringent is as follows. Since the voters' beliefs are assumed to be identical across candidates, the expected vote share for a candidate at the profile in which all candidates do not announce position is $\frac{1}{N}$. If a candidate deviates from the latter profile, the votes split as follows. A share $0 \leq \alpha \leq 1$ of the vote goes to the candidate that deviates, and the rest, $1-\alpha$, is divided in identical shares between the remainder $N-1$ candidates, so that each receives, in expectation, $\frac{\alpha}{N-1}$. When a new candidate enters the competition, the vote share that goes to the unique candidate that deviates from $\emptyset$ remains intact (namely, is equal to α), and the remaining $(N+1)-1$ candidates get now a smaller share, given by $\frac{\alpha}{N}$. Therefore, the payoff of deviating relative to declaring $\emptyset$ decreases monotonically as N increases.

The proposition that, as N increases, the conditions under which non-salience occurs become more stringent constitutes an empirical implication of our theoretical model. Moreover, if, in addition, we consider Duverger's law, our theory predicts that polities under plurality rule should be more prone to generate non-salience than polities under proportional representation.

6 Ambiguity Aversion

6.1 Candidates' Problem

The candidates, indexed by $c = 1, 2$, seek to maximize expected vote share in a one-dimensional political contest with simultaneous moves. Letting Ψ represent distribution of voters' ideal points, the *vote share of candidate c conditional on state* $s \in S$ is

$$v_c(d_1, d_2|s) = \int \rho_{i,c}(d_1, d_2|s) d\Psi, \tag{4}$$

where the function $\rho_{i,c}$ is as defined in Equation (3). Note that, because it is conditioned on a particular state of the world, there is no stochastic element entering in Equation (4)

For simplicity, throughout this section we assume that Ψ is the uniform distribution over the unit interval. It is straightforward to obtain the same qualitative results with any other symmetric and unimodal distribution. Moreover, we conjecture that unimodality alone may suffice to obtain similar qualitative results.

An *act* is a mapping from the cross product of strategy profiles and states, $D \cup \{\emptyset\} \times D \cup \{\emptyset\} \times S$, to the range of the function v_c, namely $[0, 1]$.[16] We write $v_c(d_c, d_{-c})$ to denote the act that candidate c faces when the strategy profile is (d_c, d_{-c}). A *conditional act for candidate c* is a mapping from the cross product of c's strategy set and state set, $D \cup \{\emptyset\} \times S$, to the range of the function v_c, namely $[0, 1]$; we write $v_c(\cdot, d_{-c})$. When candidate c chooses strategy d_c, he is basically choosing the conditional act $v_c(d_c, d_{-c})$ from which the electoral result will be drawn, after Nature draws the state of the world. The next step consists of endowing the candidates with preferences that allow them to rank conditional acts.

In the standard expected utility framework discussed in the previous section, it was implicitly assumed that the candidates choose among acts by computing the expected value, $E_{s \in S}[v_c(d_1, d_2)]$, using a unique and common prior belief. Now, we depart from this framework by assuming that the candidates have multiple ("ambiguous") beliefs; we use $\Delta(\Sigma)$ to denote this set of beliefs. The candidates do not know which, among the many beliefs $\Delta(\Sigma)$ is the actual probability distribution of states of the world.

In order to deal with the many beliefs, the candidates need to be endowed with a certain rule. Since in this extension of the model we aim to account for the fears that candidates may have regarding the possibility that either his opponent or, in general, the mass media, engage in "negative campaigning" against the candidate, maximin expected utility with multiple priors strikes as an appropriate rule. Namely, in a context in which interpretations and announcements may differ –such is the distinguishing premise of our baseline model–, it seems sensible to assume that each candidate $c \in \{1, 2\}$ is concerned with the possibility that his opponent, $-c$, attempts to convince the voters that the "correct" interpretation of c's announcement is whichever is most convenient for $-c$, and, therefore, less convenient for $-c$. The quotation in which Michael Dukakis explains the role played by "phraseology" and negative campaigning in the 1988 election (above in the Introduction) provides anecdotal evidence that this behavior may be present among political candidates.

[16] An intuitive (although inexact) notion of an act is its identification with the concept of *lottery.* The difference is that a lottery is an specification of consequences and their probabilities, while an act is an specification of consequences and the states of the world in which they obtain (but not their probabilities).

To model this behavior on the part of the candidates, we use the variational representation of preferences (Maccheroni *et al.*, 2006a), which generalizes maximin expected utility with multiple priors (Gilboa and Schmeidler, 1989). It has been proved that the behavioral assumptions made by these two approaches are almost identical, since the axioms on which they rely are (Maccheroni *et al.*, 2006a). Aside from their high degree of generality, some specializations of variational preferences are very tractable and, in particular, differentiable. The main advantage over maximin preferences is that calculus can be used.

A preference is called *variational* if it satisfies the following axioms: weak order, weak certainty independence, continuity, monotonicity, uncertainty aversion and nondegeneracy (see Maccheroni *et al.*, 2006a). The axiom of uncertainty aversion is usually interpreted as an ambiguity aversion axiom (among others, see Gilboa and Schmeidler 1989, and Epstein 1999). For an extended discussion of ambiguity aversion, we refer the reader to Ellsberg (1961).

Individuals endowed with ambiguity averse preferences in general (and variational preferences in particular) are customarily interpreted as *believing* that they are playing a zero-sum game against a malevolent agent. Such "paranoid" behavior has been encountered in experiments where the individuals have limited information (see, for example, Keren and Gerritsen, 1999). Variational preferences allow for different degrees of "paranoia". If extreme paranoia is assumed, then the individuals behave as a maximinimizers of expected utility with multiple priors. In such a case, the interpretation is straightforward: the individuals believe that they will confront the worst possible scenario. However, variational preferences also allow for moderate degrees of paranoia. Finally, among variational preferences, we choose *Gini preferences* to represent the preferences of the candidates, because they are very tractable and differentiable. The utility functional is introduced below, as Functional (5).

6.2 The Game

Consider model $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$. The timing of the game is as follows. In period $t = 0$, the candidates declare platforms simultaneously. A declaration of a platform by candidate $c \in \{1, 2\}$ is either (i) an announcement of policy $d_c \in D$, or (ii) a no-announcement, namely $d_c = \emptyset$. Gini preferences imply that candidate c ranks act $v_c(d_1, d_2)$ according to the following utility functional,

$$V(v_c(d_c, d_{-c})) = \min_{p_c \in \Delta(q)} \left(\int v_c(d_c, d_{-c}) dp_c + \theta \chi (p_c || q) \right) \tag{5}$$

where $\theta > 0$ is the *ambiguity coefficient*, which we assume to be the same for both candidates. The closer θ comes to zero, the more ambiguity averse the candidate is. $\chi (p||q)$ is the *relative Gini index* between distributions p and q.[17] An individual endowed with ambiguity averse preferences has multiple beliefs, $\Delta(q)$, and acts *as if* there were a malevolent influence playing a zero-sum game against him. Functional (5) makes evident that the hostile influence is allegedly exerted through the choice of a prior adverse to candidate c, denoted $p_c \in \Delta(q)$.

At the beginning of period $t = 1$, Nature draws the noise in the interpretation of candidate 1's announcement (in case there was announcement), and also the noise in the interpretation of candidate 2's announcement (in case there was announcement). Some of the propositions below consider $M(D \cup \{\emptyset\}, 0, \delta_2(s))$, in which candidate 1 faces no noise in the interpretation of his announcement, instead of $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$. This reduces the space of states of the world entering into both

[17] $\chi (p||q)$ is formally introduced below, as Equation (6).

$V(v_1(d_1, d_2))$ and $V(v_2(d_2, d_1))$ to $S = S_2 = \{l(eft), r(ight)\}$, simplifying the analysis. Because the candidates have multiple priors, they do not expect Nature to draw the noise terms from the reference distribution q; each of them expects Nature to use a distribution from $\Delta(q)$ that gives him a lower expected utility than q. It is important to note that, because each candidate believes that there is a malevolent influence acting against him, two prior beliefs regarding the state of world are relevant: (i) the prior that candidate 1 thinks the hostile influence chooses to weaken him, (ii) the prior that candidate 2 thinks the hostile influence chooses to weaken him. This is irrespective of whether both candidates or only one candidate face noise in his interpretation. (In other words, it is true for both $M(D \cup \{\emptyset\}, 0, \delta_2(s))$ and $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$) The only difference assuming noise in one or both interpretations makes is how many elements are in the set of states of the world, S.

At the end of period $t = 1$, the voters use the noises drawn by Nature, $s \in S$, to form their interpretation of the announcements of each candidate, i_c. For simplicity, we assume that the interpretation of announcements is identical across voters. If a candidate did not announce, then the voters assume that in case he wins a policy drawn from the distribution $F(n)$ will be implemented. We call $F(n)$ the *beliefs* of the voters. We simplify by assuming that these beliefs are identical across voters and are the same for both candidates. Finally, voters use these interpretations to cast their votes. Voters are assumed to have standard spatial preferences exhibiting risk neutrality. The game is depicted in Figure 2 (at the end of this paper). To avoid confusion, note that we are deviating from the game theoretic convention of Nature moving first –that is, before any player. This is only for the sake of exposition. We could assume that Nature moves first and the candidates cannot see Nature's move. We did not do this just to be consistent with our interpretation that candidates act as if there were a malevolent agent that picks the distribution from which Nature will draw.

6.3 Gini Preferences

The problem of the candidates is to maximize Functional (5). Hence, from a formal point of view, the model can be closed by establishing (i) a functional form for $\chi(p_c||q)$ and, (ii) an equilibrium concept. However, it may be useful to first discuss the interpretation of Functional (5).

We focus on one particular aspect of Functional (5), namely

$$\arg \min_{p_c \in \Delta(q)} \left(\int v_c(d_c, d_{-c}) dp_c + \theta \chi(p_c||q) \right).$$

In our interpretation of the model, this problem can be stated as: What distribution of noise candidate c *believes* Nature will draw the noises from? We refer to this as *the hostile influence's problem.*[18]

Under the Gini preferences assumption, each candidate believes that the hostile influence will choose a prior, p_c, from the set $\Delta(q)$, that weakens his expected vote share. However, this paranoia needs not be absolute. Indeed, unlike the maximin multiple priors framework, variational preferences do not imply that the candidates expect the malevolent agent to choose the prior that *minimizes* their expected utility. This is because the choice of distribution is constrained by a structure, as we explain next.

[18]It should be clear by now that because the hostile influence is only a mental construction of the players, the hostile influence's problem is, more precisely, the problem that each candidate *believes* the *alleged* hostile influence faces.

On one hand, the hostile influence bears a certain cost of implementing her choice of a prior distribution. This cost is given by the term $\theta\chi(p_c||q)$ in Functional (5) above. The larger the "distance" between the choice of distribution and an (exogenously given) *reference distribution*, the costlier it becomes for the hostile influence the implementation of that choice.[19] Hence, the hostile influence must pay a cost in order to force Nature to sample from an alternative distribution. (And the candidates understand that the larger the deviation is, the costlier its implementation becomes.) On the other hand, the hostile influence gets utility from weakening the candidate's position; or, equivalently put, it gets disutility from every unitary increase in the candidate's expected vote share. This disutility is given by the term $\int v_c(d_c, d_{-c})dp_c$ in Functional (5) above.

Note that each candidate thinks of the media as a rational actor in the sense that, given the cost of implementing a choice of prior and the disutility of the candidate's expected vote share, it implements the *optimal* prior distribution. Namely, each candidate acts as if the hostile influence were endowed with a utility function and seeks to maximize it. This utility function would be equal to the opposite of the sum of the candidate's expected utility and the cost of inducing such expected utility conditional on the chosen profile of strategies (by implementing a particular distribution of noise). In order to aggregate these two heterogenous bads, a ratio that represents the marginal rate of substitution between them (for the hostile influence) is required. We refer to this ratio as the *coefficient of ambiguity aversion*. For simplicity, we assume that both candidates believe that the hostile influence has the same coefficient of ambiguity aversion. The coefficient of ambiguity aversion is given by the constant θ in Expression (5). Note, finally, that from our perspective, the reference distribution is the real distribution of noise.

It may be helpful to discuss how Functional (5) works when the coefficient of ambiguity aversion, θ, tends to its extreme values: 0 and ∞. Consider first a coefficient close to 0. This means that the hostile influence is willing to provide any amount of effort in exchange for the reduction of one vote in the candidate's expected vote share. (Equivalently, the disutility of spending an extra unit of effort in reducing the candidates' expected vote share is zero). Not surprisingly, a coefficient close to 0 is equivalent to assuming that the candidates behave as maximinimizers of expected utility with multiple priors (Gilboa and Schmeidler, 1989). Consider now a coefficient approaching to ∞. This means that the hostile influence is willing to give up any amount of candidates' vote share in exchange for a reduction of its effort in one unit. (Equivalently, the marginal utility of a malevolent act is zero). Not surprisingly, a coefficient approaching ∞ is equivalent to assuming that the candidates behave as standard maximizers of expected utility. For further details on variational preferences, and for the proof of the insights offered in this paragraph, we refer the reader to Maccheroni *et al.* (2006).

We use $p_c = (p_{c,l}, p_{c,r})$ to denote any distribution that candidate $c \in \{1, 2\}$ may expect to be implemented against him. And we use $p_c^* = \left(p_{c,l}^*, p_{c,r}^*\right)$ to denote the distribution that maximizes the hostile influence's utility function. The cost to induce $p_c = (p_{c,l}, p_{c,r})$ is given by the *Gini relative index*[20] between p_c and the reference distribution q. The *Gini relative index* between p_c and q is given by

[19] To measure the distance between distributions, we use relative entropy.

[20] The Gini relative index is a divergence.

$$\chi\left(p_c||q\right) \equiv \frac{p_{c,l}^2}{q_l} + \frac{p_{c,r}^2}{q_r} - 1. \tag{6}$$

The reference distribution q gives the distribution of noise *in the absence of any malevolent influence* for any ordered pair of announcements by candidates 1 and 2, namely (d_1, d_2). We assume

$$q_l = q_r = \frac{1}{2} \tag{7}$$

where q_l represents the probability of event s_l, and q_r the probability of event s_r. To rule out any asymmetry, the reference distribution is the same for both candidates.

6.4 Equilibrium Concept

A *strategy* for candidate c is a choice of action $d_c \in D \cup \{\emptyset\}$. An equilibrium for this game is a *strategy profile* $(d_1, d_2) \in D \cup \{\emptyset\} \times D \cup \{\emptyset\}$ satisfying certain conditions. Note that for any given strategy profile, candidate 1 and candidate 2's beliefs about the "optimal" choice of distribution by the media, p_1^* and p_2^*, respectively, are endogenously given by Functional (5).

Definition 11 *A Nash equilibrium for this game is a profile of strategies $(d_1^*, d_2^*) \in D \cup \{\emptyset\} \times D \cup \{\emptyset\}$ such that, for $c \in \{1, 2\}$, $V(v_c(d_c^*, d_{-c}^*)) \geq V(v_c(d_c, d_{-c}^*))$ for all $d_c \in D \cup \{\emptyset\}$.*

We refer to

$$p_c^* = \arg \min_{p_c \in \Delta(q)} \left(\int v_c(d_c^*, d_{-c}^*) dp_c + \theta \chi\left(p_c||q\right) \right) \tag{8}$$

as candidate c's belief, *for $c \in \{1, 2\}$.*

6.5 Restricted Best Response Correspondences Summarized

Throughout, m represents the most-preferred policy of the median voter and η represents the mean of the voters' beliefs regarding the policy that will be implemented by a winner that did not announce position.

In Appendix B, we characterize the restricted best response correspondences of candidate 1 and candidate 2. There are three possible restricted best responses for candidate 2 –namely for, say, $\eta \leq m$, $d_2 + \delta + \epsilon$, $d_2 + \delta$ and $1 - d_1$– and two possible restricted best responses for candidate 1 –namely for, say, $\eta \leq m$, $d_1 + \delta + \epsilon$, $d_1 + \delta$ and $1 - d_2$.

Recall that throughout this (ambiguity aversion) section, we simplify the analysis by assuming that Ψ is the uniform distribution over the unit interval. Consider the best response of candidate 1 to $d_2 \in [0, 1]$. For, say, $\eta \leq m$, it must be that $\widetilde{b}_1(d_2) = \arg\max_{d_1 \in \{d_2+\delta+\epsilon, d_2+\delta, 1-d_2\}} V(v_1(d_1, d_2))$. Of course, which one of the three alternatives is the actual best response depends on the value that d_2 takes. In Appendix B we show that $V_1\left(1 - d_2, d_2\right) = \frac{1}{2} + \frac{1}{2}\delta$, and that $V_1\left(d_2 + \delta + \epsilon, d_2\right) = 1 - d_2 - \frac{1}{2}\delta - \frac{1}{16\theta}\delta^2$. Hence, it follows that $V_1\left(1 - d_2, d_2\right) \geq V_1\left(d_2 + \delta + \epsilon, , d_2\right)$ if and only if

$$\frac{1}{2} + \frac{1}{2}\delta > 1 - d_2 - \frac{1}{2}\delta - \frac{1}{16\theta}\delta^2,$$

which after algebraic manipulation reduces to

$$d_2 > m - \delta - \frac{1}{16\theta}\delta^2.$$

Also, $V_1(d_2+\delta, d_2) = \frac{3}{4} - \frac{1}{2}d_2 + \frac{1}{16\theta}d_2 - \frac{1}{16\theta}d_2^2 - \frac{1}{64\theta}$. Hence, it follows that $V_1(1-d_2, d_2) \geq V_1(d_2+\delta, d_2)$ if and only if

$$\frac{1}{2} + \frac{1}{2}\delta > \frac{3}{4} - \frac{1}{2}d_2 + \frac{1}{16\theta}d_2 - \frac{1}{16\theta}d_2^2 - \frac{1}{64\theta},$$

which after algebraic manipulation reduces to

$$d_2 \geq m - \delta.$$

Proceeding in an analogous fashion, it is possible to obtain the following specification.

$$\begin{array}{rcccc}
V_1(1-d_2, d_2) & > & \left\{\begin{array}{c} V_1(d_2+\delta+\epsilon, d_2) \\ V_1(d_2+\delta, d_2) \end{array}\right. & \text{if} & \begin{array}{c} 0 < m - d_2 < \delta + \frac{1}{16\theta}\delta^2 \\ 0 < m - d_2 < \delta \end{array} \\
V_1(1-d_2, d_2) & > & \left\{\begin{array}{c} V_1(d_2-\delta-\epsilon, d_2) \\ V_1(d_2-\delta, d_2) \end{array}\right. & \text{if} & \begin{array}{c} 0 < d_2 - m < \delta + \frac{1}{16\theta}\delta^2 \\ 0 < d_2 - m < \delta \end{array} \\
V_2(1-d_1, d_1) & > & V_2(d_1+\delta+\epsilon, d_1) & \text{if} & 0 < m - d_1 < \frac{1}{16\theta}\delta^2 \\
V_2(1-d_1, d_1) & > & V_2(d_1-\delta-\epsilon, d_1) & \text{if} & 0 < d_1 - m < \frac{1}{16\theta}\delta^2
\end{array}$$

Note that $\widetilde{b}_c(d_{-c}) = 1 - d_{-c}$ for "central" values of d_{-c}, that is, for d_{-c} close enough to m. In particular, it can be readily seen that $(d_c, d_{-c}) = (m, m)$ will always be an equilibrium in $M(D, 0, \delta_2(s))$, since for any values of δ and θ, $\widetilde{b}_c(m) = 1 - m = m$. Also, as δ becomes larger, and θ becomes lower, more equilibria will emerge in this restricted model. To see why, fix any d_{-c}, and note that as θ decreases and δ increases, $\widetilde{b}_c(d_{-c}) = 1 - d_{-c}$ will eventually become the restricted best response to d_{-c}. Because the median is $m = \frac{1}{2}$, it follows that the profile of strategies $(1 - d_{-c}, 1 - d_c)$ is symmetric around m. For example, consider $\theta \downarrow 0$. In this case, the set of equilibria surrounding m will increase in size until becoming $\{(m - \alpha\frac{\delta}{2}, m + \alpha\frac{\delta}{2})\}_{\alpha\in[0,1]}$. It must be that, in equilibrium, $\alpha \in [0, 1]$, because otherwise the distance between the announcements of the candidates will be larger than δ.

It is straightforward to extend all these results to any unimodal and symmetric distribution Ψ. Qualitatively, the results would be identical. Quantitatively, however, there would be some difference, since $V(v_c(d_{-c} + \delta + \epsilon, d_{-c}))$ would yield a different value. We conjecture that similar qualitative results may be obtain for Ψ asymmetric but unimodal.

6.6 The equilibria in the extended game

The best response correspondences summarized in the previous section allows us to obtain the equilibria in the *restricted* game $M(D, 0, \delta_2(s))$. Now, we build on these results to shed some light on the equilibria of $M(D \cup \{\emptyset\}, 0, \delta_2(s))$ and $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$. The rationale for our procedure is as follows.

The strategy $d_c = \emptyset$ is a best response to $d_{-c} \in D$ in $M(D \cup \{\emptyset\}, 0, \delta_2(s))$ if and only if $V_c(\emptyset, d_{-c}) > V_c(\widetilde{b}_c(d_{-c}), d_{-c})$. Also, since voters are risk neutral, $d_2 = \emptyset$ is a best response to $d_1 = \emptyset$ in $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$ if and only if $V_2(\emptyset, \emptyset) \geq V_2(\widetilde{b}_2(\eta), \eta)$ in $M(D \cup \{\emptyset\}, 0, \delta_2(s))$. (Recall that η is the mean of the voters beliefs regarding the policy that a candidate that did not announce policy will implement in case he wins.)

Proposition 12 states the necessary and sufficient conditions for non-salience to be an equilibrium in $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$.

Proposition 12 *Consider $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$ and assume that Ψ is the uniform distribution over the unit interval. Assume that both candidates face identical parameters: $\delta_1(l) = \delta_2(l) = -\delta_1(r) = -\delta_2(r)$. For any $\theta \geq \frac{\delta}{4}$, the profile $(\emptyset, \emptyset)$ is a Nash equilibrium if and only if*

$$m + \frac{1}{2}\delta + \frac{1}{16}\frac{\delta^2}{\theta} > \eta > m - \frac{1}{2}\delta - \frac{1}{16\theta}\delta^2.$$

It follows readily from Propositions 12 and Case 9 that ambiguity aversion amplifies the likelihood of candidates opting for no-announcement. That is, as the candidates become more and more ambiguity averse, the values of the parameters for which non-salience occurs in equilibrium is larger and larger. It also follows that non-salience is only possible for non-zero values of δ. Otherwise, we go back to standard Downsian problem and the impossibility results of Berliant and Konishi (2005) obtain. The larger δ, the more likely the candidates will opt for no-announcement in equilibrium. Also, if $\eta = m$, it is always the case that the candidates prefer not to announce.

The need of the threshold $\theta \geq \frac{\delta}{4}$ stems from the fact that the probabilities p^* depend on θ (see Appendix C). After this threshold, the probabilities become degenerate and lower levels of θ no longer bring about a change in $V(v_c(d_c, d_{-c}))$. This implies that the values of η for which there is non-salience can only be enlarged, using ambiguity aversion, up to some point.[21]

In order to investigate the salient equilibria, it is analytically helpful to restrict our attention to $M(D \cup \{\emptyset\}, 0, \delta_2(s))$.

Proposition 13 *Consider $M(D \cup \{\emptyset\}, 0, \delta_2(s))$ and assume that Ψ is the uniform distribution over the unit interval. For any $\theta \geq \frac{\delta}{4}$, the profile (m, m) is an equilibrium if and only if either $\eta \leq m - \delta$ or $\eta \geq m + \delta$.*

Proposition 13 says that $\delta \leq |m - \eta|$ is a sufficient condition for convergence to the median to be an equilibrium of the game $M(D \cup \{\emptyset\}, 0, \delta_2(s))$. Note that this condition is independent of θ. Then, even for θ close to zero (high levels of ambiguity aversion) it is possible to obtain robust salience, as long as η is sufficiently distant from m. As η becomes closer to m (or, alternatively, δ increases), convergence to the median is no longer an equilibrium.

Altogether, propositions 12 and 13 suggest that our model is able to generate both salient and non-salient equilibria within the ambiguity aversion framework. Similar considerations follow from the propositions in the previous section, within the expected utility framework. In this respect, the model is *prima facie* congruent with actual politics, where both salient and non-salient outcomes are observed. However, a nice feature of our model is that it allows a much more challenging empirical test. In the next section, we describe and pursue this empirical test using data on actual U.S. Senate elections.

[21] The largest the set of values of η for which non-salience occurs is $[m - \frac{3}{4}\delta, m + \frac{3}{4}\delta]$, which corresponds to $\theta = \frac{\delta}{4}$.

7 Testable Implications

No matter whether the candidates are expected utility maximizers or ambiguity averse, a robust implication of our theoretical model is that the larger the magnitude of noise entering in the voters' interpretation of the announcements, the more likely the candidates will prefer not to announce policy (see Propositions 5, 8 and 12 above). The aim of this section is to test this empirical implication.

In our theoretical model, the candidates have only one possible means of delivering their messages to the population. In the real world, the literature distinguishes two types (or, in our terms, "technologies") of communication within TV campaigns: paid ads (PA) and news coverage (NC). Ansolabehere *et al.* (1993) report that senatorial campaigns rely primarily on PA and only secondarily on NC.

Two differences between NC and PA are relevant to our analysis. First, while NC can alter, bias or misrepresent the events of a campaign (and, in particular, a candidate's speech), PA circumvents any intervention by the media in the delivery of a message to the voters, and therefore prevents any source of distortion. Moreover, since advertising is protected by the constitutional right of free speech, the candidates can deliver any message they want to the voters, no matter how untruthful it may be.[22] In brief, the messages that, through NC technology, would be subjected to censorship, distortion or critique by journalists or anchormen, can be delivered straightforward to the voters using PA technology. In terms of the parameters of our theoretical model, it seems reasonable to say that the delivery of a message through NC is associated with a large magnitude of noise in the voters' interpretation of the announcements of the candidates, and that the delivery of a message through PA is associated with a small magnitude of noise. The second difference is that, while NC has no cost at all for the candidates, PA is a rather expensive technology for delivering messages to the voters. (This specific issue will be further discussed in the next subsection.)

Our empirical analysis relies on the following premises:

(i) when considering strategic choices, candidates are concerned about the possibility that voters misinterpret their announcements,

(ii) any eventual misinterpretation by the voters can be (at least partially) corrected by airing PA that clarify the issue,

(iii) candidates may engage in costly fund-raising activity in order to collect money to buy PA,

(iv) the cost structure of fund-raising is similar across states.

Premises (i)-(iv) let us informally extend the theoretical model in the main body of this paper to a more realistic setting. In this setting, the candidates can not only announce or not announce but also have the option of making further announcements in order to correct eventual voters' misinterpretations.

Premise (i) is the mainstay of the theoretical model in the main body of this paper. Premise (ii) suggests that PA is the technology that gives the candidates a level of precision enough to either minimize the likelihood of misinterpretations in their announcements or emend observed misinterpretations.[23] It follows from this assumption that the price of PA represents the cost of conveying

[22] The prevalence of "ad-watches" may partially restrain the candidates' from delivering untruthful messages. However, Ansolabehere and Reeves (1997, page) point out that, because they increase the exposure of the candidates, ad-watches usually cause an effect opposite to the desired.

[23] Because one could argue that PAs may correspond to negative, rather than positive, ads, some discussion may be needed. First, and perhaps most important, to substantiate our arguments, all we need to assume is that the candidates

a message whose interpretation will be (at least partially) free of noise. Premise (iii) provides the candidates a means to raise the money necessary to purchase this technology.[24] Premise (iv) has been already used in the literature (see Gerber, 1998). The two most compelling justifications are that Senators usually raise individual donations from out of state (Sorauf, 1992) and a substantive share of the contributions come from Political Action Committees (PACs), which contribute in exchange for the senators' commitment to vote in a certain way on particular issues.

The following simple example may help to understand the rationale behind our test of the empirical implication of the model. Consider two states, which only differ in their population size. State L has a large population and State S has a small population. It is important to keep in mind that we are not considering competing candidates. Instead, we considering two candidates running different races; one of them runs for a Senate seat to represent State L and the other runs for a Senate seat to represent State S. The candidates are compulsorily subject to news coverage (NC), but have the option to deliver messages using paid ads (PA). NC is free, but, unlike messages delivered through PA, it may lead to the misinterpretation of the announcements made by the candidates. Premise (i) says that the candidates are concerned about the chance that voters misinterpret their announcements. Therefore, in evaluating whether or not to announce position, the candidates will consider the consequences, in term of their expected payoff, of an eventual misinterpretation. By assumption (ii), the candidates know that, in case of misinterpretation, they will have to give away an amount of money proportional to the cost of airing PA that emend the misinterpretation. By premise (iii), this monetary cost can, in turn, be reduced to time, effort and promises made in exchange of contributions (these three are the main inputs of the fund-raising campaigns). At this point, we let the reader know that a well-established fact of TV media markets is that the cost of airing a TV ad is mainly determined by the size of the TV market (this is thoroughly discussed in the next subsection). This implies that the candidate running for State L will face a larger cost of airing TV ads than the candidate running for State S. By premise (iv), the cost that the candidates running for states L and S must incur to finance PA is similar.

In this context, the empirical implication of our theoretical model can be formulated as a simple comparative-static exercise. On one hand, the candidate running for State L faces a higher cost of PA than the candidate running for State S. On the other hand, the candidate running for State L has the same endowment of time and political freedom than the candidate running for State S. Clearly, time and political freedom are the main inputs of any fund-raising campaign. The assumption that candidates face similar fund-raising structures implies that both candidates can transform time and political freedom into dollars at similar rates (i.e., with similar levels of effectiveness). Since the cost of emending a misinterpretation by the voters (in terms of time and political freedom) is higher for the candidate running for State L than for the candidate running for State S, it readily follows that opportunity cost of announcing position is higher for the candidate running for State L than for the

have the means to make precise announcements *if they feel the need* to do so. That is, the concern of the candidates or campaign directors need not be to make precise announcements *ex ante*. More likely, what they may be concerned about is the lack of possibilities to emend a wrong interpretation of their positions once this interpretation has been made. Then, even an observed share of 100% of negative ads would not be, at least theoretically, a concern for our line of reasoning. Second, Ansolabehere and Iyengar (1996, p. 49) have shown, using lab experiments, that in both Presidential and Senate campaigns, "voters come away from positive and negative advertisements with about the same level of issue information".

[24] One could think of a formal model where the candidates are endowed with some scarce resource (like time or political freedom) which can be used to raise funds in order to buy PA and improve the electoral performance.

candidate running for State S. We can now state the empirical implication that corresponds to this version of our theoretical model as follows. *Ceteris paribus*, the higher the price of PA in a political district, the lower the probability that the candidate will announce position.

The example above illustrates some important points. First, differences in the population size across states suffice to generate variation in the opportunity cost of announcing position. This suggests that there are aspects of the phenomenon of non-salience deeply rooted in the fundamentals of a polity, rather than on social norms or institutions.

Second, the example shows that the money available to the candidates can be easily endogeneized in the model. As a result, the amount of money available to the candidates is ultimately determined by the fund-raising efforts of the candidates. Namely, fund-raising is costly in terms of time, money and political degrees of freedom.

Third, and most importantly, the key independent variable to test the model is the cost of PA across states, and not the number of PA aired by the candidates. Namely, what is relevant to the decision problem of the candidates is the expected cost of emending the voters' beliefs, and not the ex-post cost. The theoretical model developed in the main body of this paper is static in nature, assumes that the candidates know with certainty that the voters will misinterpret their announcements (although they do not know the direction of this misinterpretation) and, above all, assumes that candidates have no means to emend possible misinterpretations by the voters. The reason why these simplifications have been made in our theory part is that they simplify greatly the analysis. However, in a more realistic setting, when the candidates make their strategic decisions on whether to announce or not to announce position, they take into account the fact that if an announcement is made, the voters' interpretation will be observed and possible misinterpretations can be emended (at a certain cost) by airing more and more accurate messages, which will update the voters' interpretation. [25] Clearly, in this setting the key variable in the candidates' cost-benefit analysis that will decide their strategy is the expected cost of emending eventual misinterpretations by the voters. Unlike this latter expected cost, the actual number of PA or the money spent in PA could not be valid measures for the opportunity cost of announcing position. This is because our analysis requires a measure of the opportunity cost of announcing position at the moment of making the strategic decision, not after this decision has been made.[26]

The remainder of this section is devoted to empirically test the mentioned theoretical implication. We carry out a cross-sectional analysis of the Senate elections in the USA. As we will show, this analysis benefits from the fact that prices of PA vary a lot across states –much more than across time. The choice of Senate, instead of House races, is due to the fact that House campaigns do not rely on TV ads as much as Senate candidates (to the point that some candidates do not advertise on TV at all).

[25] Although presumably complex, this richer version of the theoretical game may be technically feasible even with ambiguity averse candidates. Technical frameworks appropriate for this kind of models are Gilboa and Schmeidler (1993) and Maccheroni *et al.* (2006b).

[26] Assuming that a candidate did announce position, the actual number of PA or the money spent in PA is a measure of the ex-post cost, i.e., it is a measure of the cost of having emended the distortions in the interpretation of his announcement. However, since the idea of our analysis is to model the decision itself of announcing (i.e., the probability of announcing), we can not use a measure that is valid only if the candidates are assumed to have announced. In other words, using the actual number of PA or the money spent in PA would be a highly distorted measure of this cost, since those candidates that did announce position will have aired more PA (and have spent more money) than those candidates that did not announce position.

7.1 Measures and Data description

We use the Passed NPAT / Failed NPAT status of the *National Political Awareness Test* (NPAT) as a measure of whether the candidates announced or did not announced policy. NPAT is a key component of *Project Vote Smart* (PVS), a well-known, serious non-profit organization that aims to provide citizens with abundant and accurate information about the U.S. political candidates at any level. The aim of the NPAT, as described by Project Vote Smart, is to measure each candidate's willingness to provide citizens with their issue positions. Each election, PVS organizes an effort of over 200 media organizations and political leaders who write, call, and repetitiously encourage each and every candidate to provide essential information regarding his/her position on different relevant issues. Candidates who agree to fill out a form of questions regarding their positions on many issues, sign it and submit it back, knowing that this information will be uploaded and distributed freely, are given the status "Passed NPAT". Candidates that do not accept, or, simply, do not respond to the queries of PVS and the 200 media organizations and political leaders, are given the status "Failed NPAT".[27] Citizens can access the failed/passed status, as well as the complete form filled out by the candidates specifying their positions in every issue, free of any charge, either online or by phone, at any time. Project Vote Smart reports to have reached, for the last election, 16 million database hits a day (Project Vote Smart, 2006). Although PVS has been collecting NPAT tests since its foundation in 1988, only data on current office holders is publicly available. Project Vote Smart generously agreed to provide us the NPAT of the candidates that lost the 2006 Senate election. Therefore, our data set comprises the NPAT of all the Senators in office at the current time (June 2007), who may have been elected in 2002, 2004 or 2006, plus the NPAT of all the candidates that lost the 2006 Senate election ($n = 131$). The NPAT is a dummy variable, where 0 denotes that the candidate failed the NPAT and 1 denotes that the candidate passed the NPAT. Passed NPAT means that the candidate filled out, signed and submitted Project Vote Smart's form asking the candidate to report his position on several issues, and Failed NPAT means that the candidate failed to take this series of actions.

We use cost per point of spot TV advertisement, which comes from the Media Market Guide (2004), as a basis for the construction of our measure of the cost of paid ads (PA). The *cost per point* (CPP) is an estimate of the dollars required to deliver one rating point (or one percent of the audience) of any designated population within a spot TV market (DMA) area. Importantly, TV market areas, called *designated market areas* (DMAs) need not coincide with political districts. On one hand, a particular DMA may extend across counties of more than one state. For example, the so-called New York DMA comprises, among many others, the counties of Fairfield (CT), Essex (NJ), Kings (NY), and Pike (PA). On the other hand, different counties within a particular state may be spanned by different DMAs. For example, California comprises, among many others, the DMAs of Los Angeles, Sacramento-Stockton-Modesto, San Diego and San Francisco-Oakland-San Jose. As we will explain shortly, the lack of a one-to-one correspondence between political districts and DMAs has important implications that need to be accounted for in our analysis.

[27] Textually from Project Vote Smart, "[t]he National Political Awareness Test (NPAT) asks candidates which items they will support if elected. It does not ask them to indicate which items they will oppose. Through extensive research of public polling data, we have discovered that voters are more concerned with what candidates would support when elected to office, not what they oppose. If a candidate does not select a response to any part or all of any question, it does not necessarily indicate that the candidate is opposed to that particular item. "

TV cost across DMAs varies greatly. While the CCP in 2004 reaches \$ 1,477 in the most expensive DMA, New York (NY), it is just \$ 5 in the most inexpensive DMA, North Platte (NE); the median CPP among all markets is \$ 150, which corresponds to Milwaukee (WI).[28] Being the basic unit used to construct media plans and budgeting schedules across markets, CPP is available for different demographic groups, quarterly and for different day parts. For the sake of our research, any of these variants of CPP could be used, since the relative CPP across media markets remains more or less the same for any possible combination of demographic group, quarter, year and day time. In our data, the demographic group is households, the quarter is the second quarter of 2004 and the day time is *late news*.[29] Because CPP is the cost *per one percent of the audience*, the difference in CPP between, say, New York (NY) and North Platte (NE) is primarily driven by the fact that one percent of New York's audience comprises a much larger number of households that one percent of North Platte's audience. We use this measure, rather than, say, cost per thousand households, because the impact of an ad seen by one thousand households in a large market like is obviously very different than the impact of an ad seen by one thousand households in a small market. The use of a measure that accounts for a given percentage, rather than a given number, of households, makes the cross-market measures comparable.

A more debatable methodological aspect of our study is the construction of the cost of political advertising. The question is: what is the cost of political advertising in each state? If there were a one-to-one correspondence between TV markets and political districts, the solution would be simple. But there is none. The first problem is that to advertise in say, Newark (NJ), a political candidate is forced to pay the whole price of advertising in the New York DMA. Note that from the candidate's point of view, this is a rather inefficient investment, since to deliver his message to his potential voters, the candidate is forced to also pay for the delivery of his message, among others, to all New Yorkers (who do not belong to his political district). Conversely, if a candidate of, say, Alabama, wants to deliver his message to all the potential voters in his state, he would be forced to purchase time in not one but nine DMAs, while a candidate of, say, Utah, would only have to purchase time in one DMA. This examples illustrate that the degree of fragmentation of the media markets varies enormously across states. One could argue that, to the extent that CPM reflects the cost per percentage point of the markets' audience, fragmentation is not really a problem. To see why this could be argued, note that if the nine DMAs within Alabama were merged into one big DMA, then the CPM of this big DMA would approximately equal the sum of the nine smaller DMAs, since by construction CPM accounts for the size of the population of its media market. Following this line of thinking, a possible measure of the cost of political advertising would be the sum of CPMs across all DMAs within a political district. Such approach has been taken by Ansolabehere *et al.* (2001). However, one could posit the following concerns.

Consider the case of Alabama. As we said, there are nine DMAs that extend across over this state. One of them is the expensive DMA of Atlanta, which spans over two counties of Alabama: Clebune and Randolph. Combined, these two counties only sum up to 0.8 % of Alabama's population. Would political candidates running for an Alabama's seat pay the cost of advertising in the expensive Atlanta TV market just to reach a 0.8 % of Alabama's voters? Most likely not. The candidates will find more

[28] This data corresponds to the day part known as *late news*, and comes from Media Market Guide (2004).

[29] The choice of 2004 is because this year is the median of the three electoral cycles pooled in the data (2002, 2004 and 2006). The choice of late news is because this is the day part in which most political ads are aired. Similar results can be obtained using prime time. The choice of the second quarter is arbitrary.

convenient to use the same money to purchase extra TV time in, say, the DMA of Birmingham TV market, which spans over several counties that sum up to more than 39% of Alabama's population, at a cost that equals 20% of the cost of Atlanta's cost. This example illustrates that, even when CPM adjusts for the population within its DMA market, a large part of the population within a DMA may correspond to a political district different than the one in which the candidate is running. To deal with this kind of problems, in their study of U.S. House races, Ansolabehere *et al.* (2001) considered, as additional measures of the cost of political advertising, the cost of reaching not the total of the voters, but a certain proportion, say 60 %. However, this approach creates a new dilemma. Namely, how is this 60 % obtained? One possibility is to start with the most inexpensive markets and move progressively to the most expensive ones. It seems to us that, although appealing from an economic perspective, this approach fails to take into account that in reality politicians tend to concentrate in a few media markets. These few media markets tend to be highly correlated with the most populated districts (which are usually not the most inexpensive ones). Moreover, there are good reasons to think that the most inexpensive media markets may correspond to the most "partisan" districts (i.e., the ones with small proportion of issue-oriented voters relative to party-oriented voters), and therefore, the least appealing for issue-related advertising. (Such is the class of ads relevant to our study). Hence, a plausible alternative approach is to calculate the, say, 60 % of the state's population starting with the largest markets and moving progressively to the smaller markets. This approach would capture the fact that Senate campaigns usually focus on the main city (or cities) of each state. Finally, irrespective of how we calculate this fraction of the population, some problems will still arise. To start with, the proportion of the population needs to be arbitrarily set. Why not 20, 40 or 80 % instead of 60%? Even more: why even assuming a constant percentage across states? The implications of a 60% threshold are not the same for, say New York (where the population is highly concentrated in one TV market) than for, say, Alabama, where the population is widespread across the state. For instance, a 60% threshold uniform across states would imply that political candidates in New York devote all their attention to the most educated citizens of that state, while political candidates in Alabama devote most of their resources to second, third-tier and even fourth-tier counties in terms of the level of education of its inhabitants. Since different degrees of education may be highly correlated with level of political partisanship and issue-oriented voting, this may play an important role in our analysis. For instance, it is likely that Alabama's candidates do not attempt to convince voters of low-density population with issue-oriented TV ads. Forcing the percentage of population constant across states will implicitly make the assumption that they do.

To bypass at least some of these problems, our approach consists in using the CPM of the DMA that spans over the most populated county of a state as the measure of the cost of political advertising for that particular state. Then, for example, since Kings is the most populated county within the state of New York, and because the DMA market that spans over Kings is the New York DMA, we use the CPM that corresponds to New York DMA as our measure of the cost of political advertising for the state of New York. Similarly, since the most populated county in Alabama is Jefferson, and because the DMA that spans over Jefferson is Birmingham, we use the CPM that corresponds to Birmingham as our cost measure for Alabama. Throughout, we refer to our construction of the cost of political advertising as the TV cost index.

We make clear that the TV cost index is not meant to be a valid measure in abstract, but in particular for the empirical implication of our theoretical model. First, this measure is not meant to summarize, in any way, the *total* cost of political advertising across states. Instead, it is meant to summarize one particular component of this total cost: the cost of reaching *issue-oriented* voters. As the literature suggests, political ads may or not be issue-oriented, and the population target for each of these types of ads need not be the same. There are good reasons to think that issue-oriented ads are more likely to be aired on urban rather than rural areas. If, instead, we calculated the TV cost index as the sum across counties within a particular state, we would be implicitly assuming that a candidate running for, say, Missouri, has the same level of concern that his issue position would be misinterpreted in St. Louis county (liberal district, home of world-wide renowned universities) than in, say, Macon county (which only has 15,700 inhabitants). Second, this measure of TV cost index is not meant to summarize, in any way, the average measure of the CPM of the DMAs within each state, for similar reasons. Third, this measure of TV cost index does not intend to reflect the degree of TV market fragmentation of each state. This is because (i) it is usually the case that campaigns focus in (if they do not restrict to) the most populated counties, and (ii) the degree of fragmentation of a particular state is likely to be only weakly correlated to the number of DMAs market within a state in which *issue-oriented ads* are aired. Fourth, ideally, in order to test our model, one would like to count with a measure of TV cost index that be a weighted sum of all DMAs within each state, where the weights would be given by some proxy of the degree to which: (i) the candidates may campaign in that particular market, (ii) the degree to which the voters of that market are issue-oriented instead of party-oriented. However, it is not clear that such measure would yield results significantly different than the ones given by our measure of TV cost as the CPM of the most populated county. Also, the construction of this "ideal" measure, which is a very heavy task, would rely on highly subjective appreciations.

Previous studies have used different measures of the cost of political advertising. In particular, Campbell *et al.* (1984) and Stewart and Reynolds (1990) address the question of whether higher TV costs increase the incumbency advantage. These papers use measures (such as fragmentation) of how well political districts match media markets. Since for their theoretical model issue-oriented ads and non issue-oriented oriented ads may be considered substitutes, their measure of market is appropriate. The work of Ansolabehere *et al.* (2001) address the question of whether TV advertising prices are to blame for the increase in campaign expenditure in the last three decades. Accordingly, they calculate the cost of political advertising as the sum across DMAs within each state to reach (i) 1%, (ii) 50%, (iii) 66% and (iv) 100% of the state's population. The pros and cons of this approach have been discussed above. Note that the three papers use measures different than ours in order to address questions different than ours. For the sake of comparison, in addition to using the CPM of the most populated county as a proxy of the TV cost index of political advertising in issue-oriented districts, we also calculate an alternative measure, namely the sum of CPMs across the counties that sum up to certain percentages of the state's population. We discard the measures of political advertising that rely exclusively on the match between political districts and media markets because (i) Ansolabehere *et al.* (2001) find no strong correlation between these measures and actual costs, and (ii) as we already noted, accounting for market fragmentation may not be a good idea when we want to measure the cost of political advertising in districts where votes are primarily issue-oriented.

Finally, we use the following control variables: incumbency, open seat, party of the candidate and whether the candidate won the election, median income by state and a dummy variable for the year 2006 (we did not control for the years 2004 and 2002 because of the small number of observations for each of these years). Median income by state has been obtained from US Census data; the source for the remaining controls is Wikipedia. Table 1 in Appendix E provides the summary statistics for the dependent variable and different versions of the TV cost index that will be used throughout.

7.2 The Model and Results

We regress the probit model

$$\Pr(NPAT = 1|CPP, X) = \Phi(\beta_0, \beta_1 CPP, \boldsymbol{\gamma}\mathbf{X}),$$

where $\Phi(\cdot)$ denotes the standard Normal cumulative distribution function.

In this equation, the dependent variable NPAT may take two values: 0 represents F(ailed NPAT), and 1 represents P(assed NPAT), in the last election run by the candidate. The coefficient we are interested in is β_1, which captures the effect that an increase in the TV cost index produces in the probability that $NPAT = 1$. Our theory suggests that there must be a negative correlation between the TV cost index and the probability of announcing position: the larger the cost of the PA technology (which allows the candidates to minimize the degree of ambiguity in the interpretation of their announcements), the more likely the candidates will not announce policy. As we already discussed, we measure the TV cost index for a given state as (the logarithms of) the CPP corresponding to the DMA that spans over the most populated county within that state. (See the examples of states of Alabama and New York in the previous subsection.) The vector $\boldsymbol{\gamma}$ and the covariates $\mathbf{X}$ are in bold type to emphasize that they are, respectively, a vector and a matrix. The covariates $\mathbf{X}$ are the control variables and $\boldsymbol{\gamma}$ is the associated vector of coefficients. Controls are as follows. *Democrat Party* is a dummy variable that takes value 1 if the candidate represents the Democrat party and 0 otherwise. *Incumbent* is a dummy variable that takes value 1 if the candidate is an incumbent in the election he is running for, and 0 otherwise. *Open seat* is a dummy variable that takes value 1 if the candidate is running for an open seat, and 0 otherwise . *Year 2006* is a dummy variable that takes value 1 if the candidate is running for the 2006 Senate election, and 0 otherwise.[30] *Vote share democrats* represents the vote share of the Democrats in the election run by the candidate in that particular observation. This variable aims at capturing closeness between the candidates.

[30] Note that 0 represents either the 2004 Senate election or the 2002 Senate election.

Table 2 Probit Estimates for Announcement of Platform, U.S. Senate Races	
	Pr(NPAT= *passed*)
TV cost index	-0.27 [0.12]**
Democrat party	-0.64 [0.26]**
Incumbent	-0.30 [0.31]
Open seat	-0.60 [0.42]
Year 2006	-0.02 [0.27]
Vote share democrats	0.00 [0.01]
Constant	1.38 [0.73]*
N	129
Pseudo R-squared	0.09

Notes: Standard errors in brackets. * significant at 10%; ** significant at 5%. The TV cost index equals log(CPP).

We refer the reader to Table 2. As our theory suggests, the TV cost index, which equals $log(CPP)$, exhibits a coefficient of negative sign. This coefficient is statistically significant at the 5 % level. Among the control variables, the only statistically significant coefficient is Democratic party. The regression shows, then, that the probability that a U.S. Senate candidate makes a clear stand on his platform decreases (i) with the cost of TV advertising in his district, (ii) conditional on the candidate being a Democrat. Before addressing point (i), which is of utmost relevance for us, let us make some comments regarding point (ii). Since Project Vote Smart is a non-partisan organization committed to political impartiality, it is one of its policies that "[a]ll Founding Board members have an ideological opposite to provide balance and ensure strict impartiality." This implies that the ratio of republicans to democrats is close to one. However, because our empirical analysis is restricted to the US Senate, a plausible concern is that, among PVS Founding Board members, there are actually more republican *Senators* than democrat Senators. If this were the case, we may conclude that the lower probability of announcement conditional on the candidate representing the Democrat party may be the result of PVS members having more resources (contacts) to urge the political candidates to take the NPAT. Indeed, we checked that, within PVS Founding Board members, there are five republican U.S. senators and only two democratic U.S. senators. This suggests that the statistically significance of the coefficient of the covariate Democratic party might be due to a sample bias.

Most important to us is the coefficient associated to the TV cost index. This coefficient is negative, statistically significant at 5% level and robust to minor modifications of the regression. To check the robustness, we proceed as follows. First, we regress the model omitting each of the control variables one at a time, and no substantial changes occur. Second, we regress the model omitting two variables and three variables at a time. Again, no substantial changes occur.

In order to challenge our results, one could think of two alternative theories to explain why candidates may opt not to announce positions. A first plausible theory is that the candidates become more and more willing to announce position as the race becomes closer and closer. This theory finds no support in our estimation, since the coefficient associated to *vote share democrats* is zero.[31] A second plausible theory is that the candidates become more and more willing to announce position as they run for seats in states with higher and higher levels of education. Although we did not include education level as a covariate in our analysis, we did check for *median income* by states, which should be highly correlated with education level. The result is that median income is negative correlated to the probability of announcing position, what indicates that neither this theory finds support in our analysis. After observing the results in Table 2, the negative coefficient of median income should be no surprise, since median income and TV cost index are expected to be highly correlated (and they are: correlation is 0.64). We did not include median income in the regression reported in Table 2 since its negative sign provides strong evidence that median income (and therefore the educational level) is not an omitted variable driving the results. (Of course, due to its collinearity with TV cost index, the standard deviation of the coefficient associated to the latter variable is affected when the two variables are included in the regression.)

In Table 2, it can be observed that $N = 129$; this is because we drop the two observations corresponding to Alaska, since the DMAs are not assigned for some counties within this state. The fact that we do not aim to maximize explanatory power but to test a specific relation between two variables may explain the relatively low R^2. A feature of probit models is that the omission of relevant variables tends to inflate the standard deviations of the coefficients even when the omitted variables are not correlated with the covariates (Cramer, 2003). This suggests that significance levels in Table 2 may be too conservative. This, together with the low R^2, implies that in a fully specified model, β_1 may be significant even at levels lower than 5 %. Similar considerations follow from the small size of the sample.

To give an idea of the marginal effect of the TV cost index on the probability of passing the NPAT test, consider two Senators running for a seat in the U.S. Senate: one for Alabama (TV cost index: 110) and the other for Massachusetts (TV cost index: 569). Assume that both represent the Republican party and are running for an open seat in the 2006 election. Also, set the vote share of the democratic party at the sample mean, 48%. From our regression, it follows that the probability of passing the NPAT is 14 percentage points higher for the candidate from Alabama than for the candidate from Massachusetts.

7.2.1 Alternative measures of the cost of political advertising

As we already discussed, the choice of a measure of the cost of political advertising may be debatable. In this section we consider a different measure, closer, but not identical, to the one used in Ansolabehere *et al.* (2001). This measure defines the TV cost index as the sum, within each state, of as many DMAs as needed in order to reach a certain percentage of the population of that state. From now on, we refer to this percentage of the population as the *population threshold*. Our new measure of cost of political advertising equals the sum of the CPP of the DMAs that span over the

[31] Similar results are obtained when the square of *vote share democrats* is used.

most populated counties of the state, until the population threshold is reached. For example, suppose the threshold is 30 %. In the case of Alabama, the cost of reaching 30% of the population is equal to the cost of advertising in the DMAs of Birmingham, Mobile-Pensacola and Huntsville-Decatur, since, altogether, these three DMAs span over the 30% of Alabama's population, and they cover the most populated counties.

We consider several population thresholds in the range 1%-100%. In every case, the sign of the coefficient of the TV cost index was negative, which is consistent with the empirical implication of our theoretical model. However, in general, as the population threshold increases, the *p-value* associated to this coefficient increases. For small values, approximately up to 10%, we obtained similar results than using the CPP of the DMAs that covers the most populated county within the state (that is, than the results reported in Table 2). For a threshold close to 15%, the level of significance drops to 10%. For some value of the population threshold between 10% and 20%, the level of significance drops to more than 10%, implying that we can no longer reject the null that β_1 is equal to zero. As we move beyond the 20% population threshold, the *p-value* keeps increasing. Following we provide our interpretation of this pattern.

As we already discussed, the empirical implication of our theoretical model is that candidates will be more likely to announce positions whenever they believe that *issue-oriented* voters are less likely to misinterpret their issue announcements. TV market fragmentation increases the cost of advertising because, in order to reach small towns or rural districts, the candidate needs to advertise in several DMAs in addition to the DMAs that cover the main cities of the state. Such extra cost will be more and more reflected in the TV cost index as we set the population threshold higher and higher. Importantly, although such extra cost reflects the fact that the candidates reach more voters, the latter come from increasingly smaller, more rural areas, whose population is presumably less educated. In contrast, when the TV index is constructed so that only the DMA that covers the most populated county is considered, the effect of market fragmentation, and therefore of the voters that are more likely to cast their vote based on partisan considerations (*vis à vis* issue-position considerations), is minimized. This suggests that, when making strategic decisions like whether or not to announce positions, the candidates are sophisticated in the sense that their cost-benefit analysis weights the districts within the state in order to account for their share of issue-oriented voters.

8 Conclusion

Previous impossibility results found that, in a robust equilibrium, rational candidates will always prefer to announce position. We developed a model with the novel feature that the voters may misinterpret the announcements made by the candidates. Within this model, we investigated several different specifications of the model, namely: expected utility candidates, ambiguity averse candidates, two-candidate elections, multi-candidate elections, candidates that maximize expected vote share and candidates that maximize expected probability of vote.

We showed that, in this model, the impossibility theorems are overcome. Moreover, this result is robust to any of the mentioned changes in the specification of the model. The probability that the candidates do not announce in equilibrium depends on the distance between the mean of the voters' beliefs regarding the policy that a candidate that wins without announcement will implement and the ideal point of the median voter. As this distance decreases, the likelihood of robust non-salience is increased.

Finally, we estimated a probit model that supports a testable implication of our model. Ansolabehere *et al.* (2001) infer that the cost structure faced by the candidates has a substantive impact on the mediatic strategy of political campaigns. Our empirical work shows that the impact of the cost structure goes even further, since it affects key electoral strategies such as whether or not to make clear stands on important issues.

8.1 Appendix A: Restricted Best Responses with Expected Utility in $M(D, 0, \delta_2(s))$

Throughout, $v_{c,l}$ and $v_{c,r}$ are shortcuts for the expected vote share of candidate c conditional on the leftist and rightist realization of noise, respectively. We restrict our attention to $M(D, 0, \delta_2(s))$ –that is, there is only noise in the interpretation of candidate 2's announcement. Hence, while, say, $v_{2,l}$ represents the vote share of candidate 2 when the state of the world that determines the voters' interpretation of *his own announcement* is $l(eft)$, $v_{1,l}$ represents the vote of candidate 1 when the state of the world that determines the voters' interpretation of *his opponent's announcement* is $l(eft)$. We drop the subindex c in $v_{c,r}$ whenever the context avoids confusion. We use m to denote the ideal point of the median voter.

Also, we use $\widetilde{b}_c(d_{-c})$ to denote candidate c's best response to candidate's $-c$ strategy d_{-c} in the restricted model $M(D, 0, \delta_2(s))$. (We save $b_c(d_{-c})$ to denote the best response in the unrestricted models.) Hence, $\widetilde{b}_c : D - \{\emptyset\} \to D - \{\emptyset\}$. Accordingly, we refer to $\widetilde{b}_c(d_{-c})$ as candidate c's *restricted best response* to d_{-c}.

Consider the following mutually exclusive and collectively exhaustive cases: (i) $\widetilde{b}_c(d_{-c}) \geq d_{-c} + \delta$, (ii) $\widetilde{b}_c(d_{-c}) \leq d_{-c} - \delta$, and (iii) $d_{-c} + \delta > \widetilde{b}_c(d_{-c}) > d_{-c} - \delta$. Following, for $M(D, 0, \delta_2(s))$, we characterize the restricted best response $\widetilde{b}_c(d_{-c})$ for each of these three cases.

8.1.1 Candidate 1

Case (i): $\widetilde{b}_1(d_2) \geq d_2 + \delta$

(i.a) Consider any d_2 such that $d_2 + \delta < m$. A simple separation argument shows that $\widetilde{b}_1(d_2) = d_2 + \delta + \epsilon$, and that $v_l = 1 - \Psi(d_2)$ and $v_r = 1 - \Psi(d_2 + \delta)$. Combined with $q_l = q_r = \frac{1}{2}$, this implies that

$$V_1(\widetilde{b}_1(d_2), d_2) = 1 - \frac{1}{2}\Psi(d_2) - \frac{1}{2}\Psi(d_2 + \delta)$$

And, in the case $\Psi\tilde{\ }U[0, 1]$,

$$V_1(\widetilde{b}_1(d_2), d_2) = v_l q_l + v_r q_r = 1 - d_2 - \frac{1}{2}\delta$$

(i.b) Consider any d_2 such that $d_2 + \delta \geq m$. A simple separation argument shows that $\widetilde{b}_1(d_2) = d_2 + \delta$, and that $v_l = 1 - \Psi(d_2)$ and $v_r = \frac{1}{2}$, so that

$$V(\widetilde{b}_1(d_2), d_2) = \frac{3}{4} - \frac{1}{2}\Psi(d_2)$$

And, in the case $\Psi\tilde{\ }U[0, 1]$,

$$V_1(\widetilde{b}_1(d_2), d_2) = \frac{3}{4} - \frac{1}{2}d_2$$

Irrespective of the specific form that Ψ takes, it follows that candidate 1's best response is $d_1 = d_2 + \delta + \epsilon$ if $\frac{1}{2} > \Psi(d_2 + \delta)$ and $d_1 = d_2 + \delta$ otherwise.

Case (ii): $\widetilde{b}_1(d_2) \leq d_2 - \delta$

(ii.a) Consider any d_2 such that $d_2 - \delta > m$. A simple separation argument shows that $\widetilde{b}_1(d_2) = d_2 - (\delta + \epsilon)$, that $v_l = \Psi(d_2 - \delta)$ and $v_r = \Psi(d_2)$. This implies that

$$V_1(\widetilde{b}_1(d_2), d_2) = \frac{1}{2}\Psi(d_2 - \delta) + \frac{1}{2}\Psi(d_2)$$

And, in the case $\Psi \tilde{} U[0,1]$,

$$V_1(\widetilde{b}_1(d_2), d_2) = d_2 - \frac{1}{2}\delta.$$

(ii.b) Consider any d_2 such that $d_2 - \delta \leq m$. A simple separation argument shows that $\widetilde{b}_1(d_2) = d_2 - \delta$, and that $v_l = \frac{1}{2}$ and $v_r = \Psi(d_2)$, so that

$$V_1(\widetilde{b}_1(d_2), d_2) = \frac{1}{4} + \frac{1}{2}\Psi(d_2)$$

And, in the case $\Psi \tilde{} U[0,1]$,

$$V_1(\widetilde{b}_1(d_2), d_2) = \frac{1}{2}d_2 + \frac{1}{4}$$

Irrespective of the specific form that Ψ takes, it follows that candidate 1's best response is $d_1 = d_2 - (\delta + \epsilon)$ if $\Psi(d_2 - \delta) \geq \frac{1}{2}$, and $d_1 = d_2 - \delta$ otherwise.

Case (iii): $d_2 + \delta > \widetilde{b}_1(d_2) > d_2 - \delta$

In this case, $v_l = 1 - \Psi\left(\frac{d_1+d_2-\delta}{2}\right)$ and $v_r = \Psi\left(\frac{d_1+d_2+\delta}{2}\right)$. Hence,

$$V_1(d_1, d_2) = \frac{1}{2} - \frac{1}{2}\left[\Psi\left(\frac{d_1 + d_2 - \delta}{2}\right) - \Psi\left(\frac{d_1 + d_2 + \delta}{2}\right)\right]$$

And, in the case $\Psi \tilde{} U[0,1]$,

$$V_1(d_1, d_2) = \frac{1}{2} + \frac{1}{2}\delta$$

Note that, for $\Psi \tilde{} U[0,1]$, $V_1(d_1, d_2)$ is independent of d_1 and d_2. Any $d_1 \in [d_2 - \delta, d_2 + \delta]$ gives the exact same expected utility, namely $V_1(d_1, d_2) = \frac{1}{2} + \frac{\delta}{2}$.

8.1.2 Candidate 2

(i) Case $d_1 + \delta \leq \widetilde{b}_2(d_1)$

(i.a) Consider any d_1 such that $d_1 < m$. A simple separation argument shows that $\widetilde{b}_2(d_2) = d_1 + \delta + \epsilon$, and that $v_l = 1 - \Psi(d_1)$, $v_r = 1 - \Psi(d_1 + \delta)$.

This implies that

$$V(v_2(\widetilde{b}_2(d_1), d_1)) = 1 - \frac{1}{2}\left[\Psi(d_1) + \Psi(d_1 + \delta)\right].$$

And, in the case $\Psi \tilde{} U[0,1]$,

$$V(v_2(\widetilde{b}_2(d_1), d_1)) = 1 - d_1 - \frac{1}{2}\delta. \tag{9}$$

(i.b) Consider any d_1 such that $d_1 \geq m$. A simple separation argument shows that $\widetilde{b}_2(d_1) = d_1 + \delta$, and that $v_l = \frac{1}{2}$ and $v_r = 1 - \Psi(d_1 + \delta)$.

This implies that

$$V(v_2(\widetilde{b}_2(d_1), d_1)) = \frac{3}{4} - \frac{1}{2}\Psi(d_1 + \delta)$$

And, in the case $\Psi \tilde{} U[0,1]$,

$$V(v_2(\widetilde{b}_2(d_1), d_1)) = \frac{3}{4} - \frac{1}{2}d_1 - \frac{1}{2}\delta \tag{10}$$

Irrespective of the specific form that Ψ takes, it follows that $\widetilde{b}_2(d_2) = d_1 + \delta + \epsilon$ dominates $\widetilde{b}_2(d_1) = d_1 + \delta$ for any $d_1 < m$. If $d_1 = m$, candidate 2 is indifferent between them and, for any $d_1 > m$, $\widetilde{b}_2(d_1) = d_1 + \delta$ dominates $\widetilde{b}_2(d_2) = d_1 + \delta + \epsilon$.

Case (ii) $d_1 - \delta \geq \widetilde{b}_2(d_1)$

(ii.a) Consider any d_1 such that $d_1 > m$. A simple separation argument shows that $\widetilde{b}_2(d_1) = d_1 - (\delta + \epsilon)$, and that $v_l = \Psi(d_1 - \delta)$ and $v_r = \Psi(d_1)$.

This implies that

$$V(v_2(\widetilde{b}_2(d_1), d_1)) = \frac{1}{2}\Psi(d_1 - \delta) + \frac{1}{2}\Psi(d_1)$$

And, in the case $\Psi\tilde{}U[0,1]$,

$$V(v_2(\widetilde{b}_2(d_1), d_1)) = d_1 - \frac{1}{2}\delta. \tag{11}$$

(ii.b) Consider any d_1 such that $d_1 \leq m$. A simple separation argument shows that $\widetilde{b}_2(d_1) = d_1 - \delta$, and that $v_l = \Psi(d_1 - \delta)$ and $v_r = \frac{1}{2}$.

This implies that

$$V(v_2(\widetilde{b}_2(d_1), d_1)) = \frac{1}{2}\Psi(d_1 - \delta) + \frac{1}{4}$$

And, in the case $\Psi\tilde{}U[0,1]$,

$$V(v_2(\widetilde{b}_2(d_1), d_1)) = \frac{1}{2}d_1 - \frac{1}{2}\delta + \frac{1}{4} \tag{12}$$

Irrespective of the specific form that Ψ takes, it follows that $\widetilde{b}_2(d_1) = d_1 - (\delta + \epsilon)$ dominates $\widetilde{b}_2(d_1) = d_1 - \delta$ for any $d_1 > m$. If $d_1 = m$, candidate 2 is indifferent between them and, for any $d_1 < m$, $\widetilde{b}_2(d_1) = d_1 - \delta$ dominates $\widetilde{b}_2(d_1) = d_1 - (\delta + \epsilon)$.

Case (iii): $d_1 + \delta > \widetilde{b}_2(d_1) > d_1 - \delta$

In this case, $v_l = \Psi\left(\frac{d_1+d_2-\delta}{2}\right)$ and $v_r = 1 - \Psi\left(\frac{d_1+d_2+\delta}{2}\right)$. This implies that

$$V_2(v_2(d_2, d_1)) = \frac{1}{2} - \frac{1}{2}\left[\Psi\left(\frac{d_1 + d_2 + \delta}{2}\right) - \Psi\left(\frac{d_1 + d_2 - \delta}{2}\right)\right]$$

And, in the case $\Psi\tilde{}U[0,1]$,

$$V_2(v_2(d_2, d_1)) = \frac{1}{2} - \frac{1}{2}\delta$$

In this case, when $\Psi\tilde{}U[0,1]$, $V_2(v_2(d_2, d_1))$ is independent of d_1 and d_2, implying that any announcement d_2 such that $|d_1 - d_2| < \delta$ gives candidate 2 the same utility.

8.2 Appendix B: Best Response Correspondences with Variational Preferences

We assume throughout that there are only two states: left (l) and right (r), and that the reference distribution for these states is $q_l = q_r = \frac{1}{2}$.

We focus on the optimal behavior of a generic candidate, and adopt the following notational shortcuts: $v_s \equiv v_c(d_1, d_2|s)$ and $p_{c,s} \equiv p_s$ for any $s \in S$. From Appendix C, the media's optimal probabilities become

$$p_r = \frac{1}{8\theta}(4\theta + v_l - v_r)$$

$$p_l = \frac{1}{8\theta}(v_r - v_l - 4\theta) + 1$$

if $\theta > \left|\frac{v_r - v_l}{4}\right|$, $p_r^* = 1$ if $\theta < \frac{v_l - v_r}{4}$, and $p_r^* = 0$ if $\theta < \frac{v_r - v_l}{4}$. To avoid dealing with the degenerate cases, we restrict our attention to $\theta > \left|\frac{v_r - v_l}{4}\right|$.

8.2.1 The restricted best response correspondence of Candidate 1

In this subsection, we characterize the restricted best response correspondence of the candidates, $\widetilde{b}_c(d_{-c})$, in the model $M(D, 0, \delta_2(s))$.

(i) Case $\widetilde{b}_1(d_2) \geq d_2 + \delta$

(i.a) Consider any d_2 such that $d_2 + \delta < m$.

Since $\widetilde{b}_1(d_2) \geq d_2 + \delta$, it must follow that $v_l = 1 - \frac{d_2 - \delta + d_1}{2}$ and $v_r = 1 - \frac{d_2 + \delta + d_1}{2}$. Let $p^* \equiv \arg\min_{p_c \in \Delta(q)} \left(\int v_c(d_c, d_{-c}) dp_c + \theta \chi \left(p_c || q\right)\right)$, with $p^* = (p_l^*, p_r^*)$. From Appendix C, $p_r^* = \frac{1}{8\theta}\left(4\theta + v_l - v_r\right)$ and $p_l^* = \frac{1}{8\theta}\left(v_r - v_l - 4\theta\right) + 1$. Since $\chi^2\left(p||q\right) \equiv \int\limits_S \frac{p(s)^2}{q(s)} - 1 ds$ is strictly convex, it follows that the variational preference functional is everywhere differentiable (see Maccheroni *et al.*, 2005). For all d_1 such that $d_1 > d_2 + \delta$, $\frac{\partial V}{\partial d_1} = -\frac{1}{2}$, implying that $V_1(v_1(d_1, d_2))$ decreases monotonically in d_1 in the interval $(d_2 + \delta, 1]$. Since $d_2 + \delta < m$, $V_1(v_1(d_2 + \delta, d_2)) < V_1(v_1(d_2 + \delta + \epsilon, d_2))$, where ϵ is an infinitesimal scalar. It follows that $\widetilde{b}_1(d_2) = d_2 + \delta + \epsilon$. Replacing $\widetilde{b}_1(d_2)$ in v_l and v_r above, we get that $v_l = 1 - d_2$ and $v_r = 1 - d_2 - \delta$. Combined with $q_l = q_r = \frac{1}{2}$, this implies that

$$V(v_1(\widetilde{b}_1(d_2), d_2)) = v_l p_l + v_r p_r + \theta\left(\frac{p_l^2}{q_l} + \frac{p_r^2}{q_r} - 1\right)$$

or

$$V(v_1(\widetilde{b}_1(d_2), d_2)) = 1 - d_2 - \frac{1}{2}\delta - \frac{1}{16\theta}\delta^2 \tag{13}$$

(i.b) Consider any d_2 such that $d_2 + \delta \geq m$.

For all d_1 such that $d_1 > d_2 + \delta$, $\frac{\partial V}{\partial d_1} = -\frac{1}{2}$, implying that $V_1(v_1(d_1, d_2))$ decreases monotonically in d_1 in the interval $(d_2 + \delta, 1]$. Since $d_2 + \delta \geq m$, $V_1(v_1(d_2 + \delta, d_2)) > V_1(v_1(d_2 + \delta + \epsilon, d_2))$, where ϵ is an infinitesimal scalar. It follows that $\widetilde{b}_1(d_2) = d_2 + \delta$, and that $v_l = 1 - d_2$ and $v_r = \frac{1}{2}$, so that $V(\widetilde{b}_1(d_2), d_2) = (1 - d_2)\, p_l + \frac{1}{2} p_r + \theta\left(\frac{p_l^2}{\frac{1}{2}} + \frac{p_r^2}{\frac{1}{2}} - 1\right)$, which gives

$$V(v_1(\widetilde{b}_1(d_2), d_2)) = \frac{3}{4} - \frac{1}{2}d_2 + \frac{1}{16\theta}d_2 - \frac{1}{16\theta}d_2^2 - \frac{1}{64\theta} \tag{14}$$

A comparison of Expressions 13 and 14 indicates that $V(d_2 + \delta + \epsilon, d_2) \geq V(d_2 + \delta, d_2)$ for any $d_2 \leq m - \delta$.

(ii) Case $\widetilde{b}_1(d_2) \leq d_2 - \delta$

(ii.a) Consider any d_2 such that $d_2 - \delta > m$. This case is symmetric to case (i.a). Then, $\widetilde{b}_1(d_2) = d_2 - (\delta + \epsilon)$, $v_l = d_2 - \delta$, $v_r = d_2$, and

$$V(v_1(\widetilde{b}_1(d_2), d_2)) = d_2 - \frac{1}{2}\delta - \frac{1}{16\theta}\delta^2 \tag{15}$$

(ii.b) Consider any d_2 such that $d_2 - \delta \leq m$. This case is symmetric to case (i.b). Then, $\widetilde{b}_1(d_2) = d_2 - \delta$, and that $v_l = \frac{1}{2}$, $v_r = d_2$, so that

$$V(v_1(\widetilde{b}_1(d_2), d_2)) = \frac{1}{4} + \frac{1}{2}d_2 + \frac{1}{16\theta}d_2 - \frac{1}{16\theta}d_2^2 - \frac{1}{64\theta}. \tag{16}$$

A simple comparison of Expressions 15 and 16 indicates that candidate 1's best response is $\widetilde{b}_1(d_2) = d_2 - (\delta + \epsilon)$ when $d_2 > m + \delta$.

(iii) Case $\left|\widetilde{b}_1(d_2) - d_2\right| < \delta$

In this case, $v_l = 1 - \frac{d_1+d_2-\delta}{2}$ and $v_r = \frac{d_1+d_2+\delta}{2}$. Hence,

$$V(v_1(d_1,d_2)) = \frac{1}{2} + \frac{1}{2}\delta + \frac{1}{8\theta}\left(d_1 + d_2 - d_1d_2 - \frac{1}{2}d_1^2 - \frac{1}{2}d_2^2 - \frac{1}{2}\right) \tag{17}$$

The FOC is

$$\frac{1}{8\theta}\left(1 - d_2 - d_1\right) = 0$$

and the SOC is

$$-\frac{1}{8\theta} \leq 0.$$

Note that the FOC implies that

$$d_1^* = 1 - d_2, \tag{18}$$

and the SOC implies that for any finite θ, d_1^* is a global maximum. For $\theta \uparrow \infty$, the strict concavity of $V(v_1(d_1,d_2))$ is lost, and any $d_1 \in [d_2-\delta, d_2+\delta]$ gives the exact same V, namely $V(v_1(d_1,d_2)) = \frac{1}{2}+\frac{\delta}{2}$.

Plugging in Expression 18 in Equation 17 yields

$$V(v_1(d_1^*(d_2),d_2)) = \frac{1}{2} + \frac{1}{2}\delta$$

8.2.2 The restricted best response correspondence of Candidate 2

(i) Case $\widetilde{b}_2(d_1) - (d_1+\delta) > 0$

(i.a) Consider any $\widetilde{b}_2(d_1)$ such that $\widetilde{b}_2(d_1)+\delta < m$. Then, $v_l = 1 - \frac{d_1+d_2-\delta}{2}$, $v_r = 1 - \frac{d_1+d_2+\delta}{2}$. The problem is symmetric to that of candidate 1, case (i.a). Then, $\widetilde{b}_2(d_1) = d_1 + (\delta+\epsilon)$ and

$$V(v_2(\widetilde{b}_2(d_1),d_1)) = 1 - d_1 - \frac{1}{2}\delta - \frac{1}{16\theta}\delta^2 \tag{19}$$

(i.b) Consider the case in which $\widetilde{b}_2(d_1)+\delta \geq m$. A simple separation argument shows that $\widetilde{b}_2(d_1) = d_1+\delta$. This implies that $v_l = \frac{1}{2}$, $v_r = 1 - d_1 - \delta$ so that

$$V(v_2(\widetilde{b}_2(d_1),d_1)) = \frac{3}{4} - \frac{1}{2}d_1 - \frac{1}{2}\delta + \frac{1}{16\theta}\delta - \frac{1}{64\theta} + \frac{1}{16\theta}d_1 - \frac{1}{8\theta}\delta d_1 - \frac{1}{16\theta}\delta^2 - \frac{1}{16\theta}d_1^2 \tag{20}$$

A simple comparison of Expressions 19 and 20 indicates that $\widetilde{b}_2(d_1) = d_1 + (\delta+\epsilon)$ is the best response of candidate 2 when $d_1 \leq \frac{1}{2}$.

(ii) **Case** $(d_1-\delta) - \widetilde{b}_2(d_1) > 0$

(ii.a) Consider any $\widetilde{b}_2(d_1)$ such that $\widetilde{b}_2(d_1) - \delta > m$. The problem is symmetric to that of candidate 1 case (ii.a). Then, $v_l = \frac{d_1+d_2-\delta}{2}$, $v_r = \frac{d_1+d_2+\delta}{2}$,

$$V(v_2(\widetilde{b}_2(d_1),d_1)) = d_1 - \frac{1}{2}\delta - \frac{1}{16\theta}\delta^2 \tag{21}$$

and $\widetilde{b}_2(d_1) = d_1 - (\delta+\epsilon)$,

Now, consider the case in which $\widetilde{b}_2(d_1) - \delta \leq m$. A simple separation argument shows that $\widetilde{b}_2(d_1) = d_1 - \delta$, and that $v_l = d_1 - \delta$, $v_r = \frac{1}{2}$ so that

$$V(v_2(\widetilde{b}_2(d_1),d_1)) = \frac{1}{2}d_1 - \frac{1}{2}\delta + \frac{1}{4} - \frac{1}{64\theta} - \frac{1}{16\theta}\delta + \frac{1}{16\theta}d_1 + \frac{1}{8\theta}\delta d_1 - \frac{1}{16\theta}d_1^2 \tag{22}$$

A simple comparison of Expressions 21 and 22 indicates that the best response of candidate 2 is $\widetilde{b}_2(d_1) = d_1 - (\delta + \epsilon)$ when $d_1 > \frac{1}{2}$; otherwise the best response is $\widetilde{b}_2(d_1) = d_1 - \delta$.

(iii) Case $\left|\widetilde{b}_2(d_1) - d_1\right| < \delta$

A simple separation argument shows that $v_l = \frac{d_1+d_2-\delta}{2}$ and $v_r = 1 - \frac{d_1+d_2+\delta}{2}$. It follows that

$$V(v_2(d_2,d_1)) = \frac{1}{2} - \frac{1}{2}\delta + \frac{1}{8\theta}\left(d_1 + d_2 - d_1 d_2 - \frac{1}{2}d_1^2 - \frac{1}{2}d_2^2 - \frac{1}{2}\right) \tag{23}$$

The FOC implies

$$d_2^* = 1 - d_1 \tag{24}$$

and the SOC implies that, for any finite θ, d_2^* is a global maximum. For $\theta \uparrow \infty$, the strict concavity of $V(d_1, d_2)$ is lost, and any d_2 such that $d_1 \in [d_2 - \delta, d_2 + \delta]$ gives $V(v_2(d_1, d_2)) = \frac{1}{2} - \frac{\delta}{2}$.

Plugging in Expression 24 in Equation 23 yields $V(v_2(d_2^*(d_1), d_1)) = \frac{1}{2} - \frac{1}{2}\delta$.

8.3 Appendix C: Deriving Probability p^*

First note that $\phi(t) \equiv \frac{1}{2}(t-1)^2$ is strictly convex, so that the divergence $\chi^2(p||q) \equiv \int_S \frac{p(s)^2}{q(s)} - 1ds$ is strictly convex and then, the variational preference functional is everywhere differentiable (see Maccheroni *et al.*, 2005). We now look for $p_c^* \in \arg\min_{p_c \in \Delta(q)} \left(\int v_c(s|d_c, d_{-c})dp_c + \theta\chi^2(p_c||q)\right)$. In our case, $S = \{l, r\}$, denoting leftist bias, and rightist bias, respectively. So,

$$\chi^2(p||q) \equiv \frac{p_l^2}{q_l} + \frac{p_r^2}{q_r} - 1. \tag{25}$$

The first order condition for $\min_{p_c \in \Delta(q)} \left(\int v_c(s|d_c, d_{-c})dp_c + \theta\chi^2(p_c||q)\right)$ are $v_l + \theta\frac{2p_l}{q_l} = k$, $v_r + \theta\frac{2p_r}{q_r} = k$ and $p_l + p_r = 1$. It then follows that

$$v_l - v_r = 2\theta\left(\frac{p_r}{q_r} - \frac{p_l}{q_l}\right),$$

Using $p_l + p_r = 1$, we obtain

$$v_l - v_r = \frac{2\theta}{q_l q_r}\left(q_l p_r - q_r + p_r q_r\right),$$

Solving this system of equations yields

$$p_l^* = \frac{1}{2\theta q_l + 2\theta q_r}\left(q_l q_r v_r - q_l v_l q_r - 2\theta q_r\right) + 1$$

$$p_r^* = \frac{1}{2\theta q_l + 2\theta q_r}\left(2\theta q_r + q_l v_l q_r - q_l q_r v_r\right).$$

Specialized to $q_l = \frac{1}{2}$ $q_r = \frac{1}{2}$, the system reduces to

$$(p_l^*, p_r^*) = \begin{cases} \left(\frac{1}{8\theta}(v_r - v_l - 4\theta) + 1, \frac{1}{8\theta}(4\theta + v_l - v_r)\right) & \text{if } \theta \geq \frac{|v_r - v_l|}{4} \\ (0,1) & \text{if } \theta < \frac{v_l - v_r}{4} \\ (1,0) & \text{if } \theta < \frac{v_r - v_l}{4} \end{cases}$$

8.4 Appendix D: Proofs of propositions

Proof of Proposition 7. Since candidates are identical, it suffices to check for candidate, say 2. First, assume that $\delta_1(l) = \delta_2(l)$. Since $\delta_1(l) = -\delta_2(r)$ and $\delta_1(r) = -\delta_2(l)$, it follows that $\delta_1(r) = \delta_2(r)$. Since candidates are identical and adopt identical positions, it must follow that $V_2(m.m) = \frac{1}{2}$. Second, assume that $\delta_1(l) \neq \delta_2(l)$. Since $\delta_1(l) = -\delta_2(r)$ and $\delta_1(r) = -\delta_2(l)$, it follows that $\delta_1(r) \neq \delta_2(r)$. W.l.o.g. assume that $\delta_2(l) < \delta_1(l)$, so that $\delta_2(r) < \delta_1(r)$ (since both $\delta_1(l)$ and $\delta_2(l)$ are negative). In this case,

$$\begin{aligned} V(m,m) &= \frac{1}{4}\Psi\left(\frac{d_2+\delta_2(l)+d_1+\delta_1(l)}{2}\right)+\frac{1}{4}\left[1-\Psi\left(\frac{d_2+\delta_2(r)+d_1+\delta_1(l)}{2}\right)\right] \\ &+\frac{1}{4}\Psi\left(\frac{d_2+\delta_2(l)+d_1+\delta_1(r)}{2}\right)+\frac{1}{4}\Psi\left(\frac{d_2+\delta_2(r)+d_1+\delta_1(r)}{2}\right). \end{aligned}$$

Using $\delta_1(l) = -\delta_2(r)$ and $\delta_1(r) = -\delta_2(l)$, the latter expression reduces to $V(m,m) = \frac{1}{4} + \frac{1}{4}\left[\Psi\left(\frac{m+\delta_1(l)+\delta_2(l)}{2}\right)+\Psi\left(\frac{m+\delta_1(r)+\delta_2(r)}{2}\right)\right]$. Symmetry of Ψ suffices to reduce the latter expression to $V_2(m,m) = \frac{1}{2}$. We proceed now to check that no deviation is profitable.

The following cases are mutually exclusive and collectively exhaustive: (a) d_2 such that $i_2(d_2,l) < i_2(d_2,r) < i_1(m,l)$, (b) d_2 such that $i_2(d_2,l) < i_2(d_2,r) = i_1(m,l)$, (c) d_2 such that $i_2(d_2,l) < i_1(m,l) < i_2(d_2,r) < i_1(m,r)$, (d) d_2 such that $i_2(d_2,l) = i_1(m,l) < i_2(d_2,r) = i_1(m,r)$, (e) d_2 such that $i_1(m,l) < i_2(d_2,l) < i_1(m,r) < i_2(d_2,r)$, (f) d_2 such that $i_1(m,l) < i_2(d_2,l) = i_1(m,r) < i_2(d_2,r)$, (g) $i_1(m,l) < i_1(m,r) < i_2(d_2,l) < i_2(d_2,r)$. A simple separation argument shows that the only possibly profitable deviations from (m,m) are cases (c), (d) and (e). Since cases (c) and (e) are symmetric, it suffices to check for cases (c) and (d). For any d_2 that satisfies case (c), it follows that

$$\begin{aligned} V_2(d_2^{case\_c},m) &= \frac{1}{4}\Psi\left(\frac{d_2+m+\delta_1(l)+\delta_2(l)}{2}\right)+\frac{1}{4}\left[1-\Psi\left(\frac{m+d_2+\delta_2(r)+\delta_1(l)}{2}\right)\right] \\ &+\frac{1}{4}\Psi\left(\frac{m+d_2+\delta_2(l)+\delta_1(r)}{2}\right)+\frac{1}{4}\Psi\left(\frac{m+d_2+\delta_1(r)+\delta_2(r)}{2}\right) \end{aligned}$$

The latter expression reduces to $V_2(d_2^{case\_c},m) = \frac{1}{4}+\frac{1}{4}\left[\Psi\left(\frac{m+d_2+\delta_1(l)+\delta_2(l)}{2}\right)+\Psi\left(\frac{m+d_2+\delta_1(r)+\delta_2(r)}{2}\right)\right]$. Symmetry of Ψ implies that $V_2(d_2^{case\_c},m) < \frac{1}{2}$ for any $d_2 < m$. For any d_2 that satisfies case (d), it follows that

$$\begin{aligned} V_2(d_2^{case\_d},m) &= \frac{1}{4}\frac{1}{2}+\frac{1}{4}\left[1-\Psi\left(\frac{d_2+m+\delta_1(l)+\delta_2(r)}{2}\right)\right] \\ &+\frac{1}{4}\Psi\left(\frac{m+d_2+\delta_1(r)+\delta_2(l)}{2}\right)+\frac{1}{4}\Psi\left(\frac{m+d_2+\delta_1(r)+\delta_2(r)}{2}\right) \end{aligned}$$

The latter expression reduces to $V_2(d_2^{case\_d},m) = \frac{1}{2}$. Hence, $V_2(d_2,m) \leq V_2(m,m)$ for any $d_2 \in D$ and $V_2(d_2,m) = V_2(m,m)$ if and only if d_2 is such that $i_2(d_2,l) = i_1(m,l) < i_2(d_2,r) = i_1(m,r)$. It remains to check that $V_2(\emptyset,m) < V_2(m,m)$ for η such that $\eta < m+\delta_1(l) < m+\delta_1(r)$ or $\eta > m+\delta_1(r) > m+\delta_1(l)$. Consider $\eta < m+\delta_1(l) < m+\delta_1(r)$. In this case, $V_2(\emptyset,m) = \frac{1}{2}\Psi\left(\frac{m+\eta+\delta_1(l)}{2}\right)+\frac{1}{2}\Psi\left(\frac{m+\eta+\delta_1(r)}{2}\right)$. It follows that $V_2(\emptyset,m) < \frac{1}{2}$ iff $\Psi\left(\frac{m+\eta+\delta_1(l)}{2}\right)+\Psi\left(\frac{m+\eta+\delta_1(r)}{2}\right) > 1$. Clearly, this condition is satisfied for any $\eta+\delta_1(r) < m$, for any $\delta_1(r) \neq 0$. Similar reasoning applies for $\eta > m+\delta_1(r) > m+\delta_1(l)$. ■

Proof of Proposition 8. Assume that $\eta < m$ (the case $\eta \geq m$ is analogous) Since the model is $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$ and the parameters are assumed to be identical across candidates, it suffices to check the value of the parameters for which no deviation from $(\emptyset, \emptyset)$ is profitable for candidate, say, 2. By construction, $V_2(\emptyset, \emptyset) = \frac{1}{2}$. By construction, for any $d_2 \in D$, $V_2(d_2, \emptyset)$ in $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$ is equal to $V_2(d_2, \eta)$ in $M(D \cup \{\emptyset\}, 0, \delta_2(s))$. Then, we can use the restricted best response correspondence of candidate 2 for $M(D \cup \{\emptyset\}, 0, \delta_2(s))$ (in Appendix A) to check if there is any $d_2 \in D$ such that $V_2(d_2, \eta) > V_2(\emptyset, \eta)$. From Appendix A, only three deviations may be profitable: (i) $d_2^{case\_i}(\eta) = \eta + \delta + \epsilon$, which gives $V_2(d_2^{case\_i}(\eta), \eta) = 1 - \frac{1}{2}\left[\Psi(\eta) + \Psi(\eta + \delta)\right]$, (ii) $d_2^{case\_ii}(\eta) = \eta - (\delta + \epsilon)$, which gives $V_2(d_2^{case\_ii}(\eta), \eta) = \frac{1}{2}\Psi(\eta - \delta) + \frac{1}{4}$, (iii) $d_2^{case\_iii}(\eta) \in (\eta + \delta, \eta - \delta)$ which gives $V_2(d_2^{case\_iii}(\eta), \eta) = \frac{1}{2} - \frac{1}{2}\left[\Psi\left(\frac{\eta + d_2 + \delta}{2}\right) - \Psi\left(\frac{\eta + d_2 - \delta}{2}\right)\right]$. The profile $(\emptyset, \emptyset)$ is an equilibrium for the values of the parameters such that $\frac{1}{2}$ is strictly greater than each of these values. We proceed case by case. Case (i): $\frac{1}{2} > 1 - \frac{1}{2}\left[\Psi(\eta) + \Psi(\eta + \delta)\right]$ reduces to $\Psi(\eta + \delta) > 1 - \Psi(\eta)$. Since $\eta \leq m$ and $\delta \geq 0$, this condition is satisfied: (i) for a fixed $\delta > 0$, for values of η close enough to m, (ii) for a fixed η, for values of δ large enough. Case (ii): $\frac{1}{2} > \frac{1}{2}\Psi(\eta - \delta) + \frac{1}{4}$ reduces to $\frac{1}{2} > \Psi(\eta - \delta)$. Since $\eta \leq m$ and $\delta > 0$, this condition is trivially satisfied. Case (iii): $\frac{1}{2} > \frac{1}{2} - \frac{1}{2}\left[\Psi\left(\frac{\eta + d_2 + \delta}{2}\right) - \Psi\left(\frac{\eta + d_2 - \delta}{2}\right)\right]$ reduces to $\Psi\left(\frac{\eta + d_2 + \delta}{2}\right) > \Psi\left(\frac{\eta + d_2 - \delta}{2}\right)$. Since $\delta > 0$, this condition is trivially satisfied. Hence, the necessary and sufficient condition for $\eta < m$ can be written as $\Psi(\eta + \delta) > 1 - \Psi(\eta)$. Applying the same line of reasoning, for $\eta \geq m$, $1 - \Psi(\eta) > \Psi(\eta - \delta)$. It follows that the necessary and sufficient condition can be summarized as $\Psi(\eta + \delta) > 1 - \Psi(\eta) > \Psi(\eta - \delta)$. ■

Proof of Proposition 5. For candidate 1, $V_1(\emptyset, d_2) = V_1(\eta, d_2)$ for any $d_2 \in D \cup \{\emptyset\}$. Then, in an equilibrium in which at least one candidate strictly prefers not to announce position, $d_2 = \emptyset$. Since the model is $M(D \cup \{\emptyset\}, 0, \delta_2(s))$ and $d_2 = \emptyset$, the vote shares of the candidates are deterministic and perfectly spatial. Hence, a simple separation argument shows that $d_1(\emptyset) = \eta + \epsilon$ is candidate 1's best response. A necessary condition for candidate 2 to strictly prefer $\emptyset$ is that $V_2(d_2^{case\_i}(\eta + \epsilon), \eta + \epsilon) < V_2(\emptyset, \eta + \epsilon)$ (see Appendix A). $V_2(d_2^{case\_i}(\eta + \epsilon), \eta + \epsilon) = 1 - \frac{1}{2}\left[\Psi(\eta) + \Psi(\eta + \delta)\right]$, and $V_2(\emptyset, \eta + \epsilon) = \Psi(\eta)$. After some algebra, $V_2(d_2^{case\_i}(\eta + \epsilon), \eta + \epsilon) < V_2(\emptyset, \eta + \epsilon)$ reduces to $2\Psi(\eta) + \Psi\left(\frac{\eta + d_2 + \delta}{2}\right) - \Psi\left(\frac{\eta + d_2 - \delta}{2}\right) > 1$. ■

Proof of Case 6. Assume $m > \eta$ (the case $m < \eta$ is analogous). Step 1: no deviation from $(\eta + \epsilon, \emptyset)$ is profitable for candidate 1. By construction, $\delta_1(s) = 0$ for all s, implying that $i_1(\eta + \epsilon, s_1) = \eta + \epsilon$ for all s_1. Since $d_2 = \emptyset$, and voters are risk neutral, $i_2(d_2, s_2) = \eta$ for all s_2.

It follows that the vote share of both candidates are deterministic and perfectly spatial, so that the vote shares can be computed as the area below Ψ in between one of the extremes of the distribution and $\frac{\eta + \epsilon + \eta}{2}$. Then, a simple separation argument shows that, for any $\eta < m$, $\widetilde{b}_1(\eta) = \eta + \epsilon$. (See Appendix A for a definition of $\widetilde{b}_1(\cdot)$). Since $V_1(\eta + \epsilon, \emptyset) = 1 - \eta > \frac{1}{2}$, and, by construction, $V_1(\emptyset, \emptyset) = \frac{1}{2}$, it follows that $V_1(\widetilde{b}_1(\eta), \emptyset) > V_1(\emptyset, \emptyset)$. Step 2: no deviation from $(\eta - \epsilon, \emptyset)$ is profitable for candidate 2. For any $d_2 \neq \emptyset$, $i_2(d_2, s_2)$ depends on s_2. From Appendix A, it follows that $\widetilde{b}_2(\eta + \epsilon)$ takes any of the following three forms: (i) $d_2^{case\_i}(\eta + \epsilon) \equiv \eta + (\delta + \epsilon)$, (ii) $d_2^{case\_ii}(\eta + \epsilon) \equiv \eta - \delta$, and (iii) any $d_2^{case\_iii}(\eta + \epsilon) \in (\eta - \delta, \eta + \delta)$, since $V(d_2, d_1)$ is constant for any $d_2 \in (\eta - \delta, \eta + \delta)$. From Appendix A, $V_2(d_2^{case\_i}(\eta + \epsilon), \eta + \epsilon) = 1 - (\eta + \epsilon) - \frac{1}{2}\delta$, which as $\epsilon \downarrow 0$, converges to $1 - \eta - \frac{1}{2}\delta$, $V_2(d_2^{case\_ii}(\eta + \epsilon), \eta + \epsilon) = \frac{1}{2}\eta - \frac{1}{2}\delta + \frac{1}{4}$, and $V_2(d_2^{case\_iii}(\eta + \epsilon), \eta + \epsilon) = \frac{1}{2} - \frac{1}{2}\delta$. It then follows that $V_2(d_2^{casei\_i}(\eta + \epsilon), \eta + \epsilon) \geq V_2(d_2^{case\_iii}(\eta + \epsilon), \eta + \epsilon)$ for any value of η. And $V_2(d_2^{casei\_i}(\eta + \epsilon), \eta + \epsilon) \geq$

$V_2(d_2^{case\_ii}(\eta+\epsilon), \eta+\epsilon)$ if and only if $m > \eta$. Because $m > \eta$, it must be that $\widetilde{b}_2(\eta+\epsilon) = d_2^{casei\_i}(\eta+\epsilon)$. It remains to check the value of the parameters such that $V_2(\emptyset, \eta+\epsilon) > V_2(\widetilde{b}_2(\eta+\epsilon), \eta+\epsilon)$. Because as $\epsilon \downarrow 0$, $V_2(\eta+\epsilon, \emptyset)$ converges to η, candidate 2 will not deviate from $\emptyset$ to $\widetilde{b}_2(\eta+\epsilon)$ if and only if $\eta > 1 - \eta - \frac{1}{2}\delta$, which is equivalent to $\eta > m - \frac{1}{4}\delta$. ■

Proof of Proposition 12. Assume $\eta \leq m$ (an analogous argument holds for $\eta > m$). Since candidates are identical, it suffices to check that for, say, candidate 2, $V_2(\emptyset, \emptyset) \geq V_2(b_2, \emptyset)$ for all $b_2 \in D$, in $M(D \cup \{\emptyset\}, \delta_1(s), \delta_2(s))$. Since $d_1 = \emptyset$, and voters are risk neutral, this is equivalent to check that, for candidate 2, $V_2(\emptyset, \emptyset) \geq V_2(\widetilde{b}_2(\eta), \eta)$, in $M(D \cup \{\emptyset\}, 0, \delta_2(s))$. Using the characterization of $\widetilde{b}_2(d_1)$ (in the main body of the paper), it follows that either $\widetilde{b}_2(d_1) = d_1 + \delta + \epsilon$ or $\widetilde{b}_2(d_1) = 1 - d_1$. We check for what values of the parameters it is true that (i) $V(v_2(\emptyset, \emptyset)) > V(v_2(\eta+\delta+\epsilon, \emptyset))$ and (ii) $V(v_2(\emptyset, \emptyset)) > V(v_2(1-\eta, \emptyset))$. By construction, $V(v_2(\emptyset, \emptyset)) = \frac{1}{2}$. Also,

$$V(v_2(\eta+\delta+\epsilon, \emptyset)) = (1-\eta)\, p_l + (1-\eta-\delta)\, p_r + \theta\left(\frac{p_l^2}{q_l} + \frac{p_r^2}{q_r} - 1\right). \tag{26}$$

In $M(D \cup \{\emptyset\}, 0, \delta_2(s))$, for the profile of strategies $(\emptyset, \emptyset)$, $v_{2,l} = 1 - \eta$ and $v_{2,r} = 1 - \eta - \delta$. Hence, the condition $\theta \geq \frac{|v_{2,r} - v2_{,l}|}{4}$ (in Appendix C) reduces $\theta \geq \frac{\delta}{4}$. Plugging in the values of p_l^* and p_r^* that correspond to the case $\theta \geq \frac{|v_r - v_l|}{4}$ (from Appendix C), expression (26) reduces to $V(v_c(\eta+\delta+\epsilon, \emptyset)) = 1 - \eta - \frac{1}{16\theta}\delta^2 - \frac{1}{2}\delta$. Then, $V(v_2(\emptyset, \emptyset)) > V(v_c(\eta+\delta+\epsilon, \emptyset))$ if and only if $\frac{1}{2} > 1 - \eta - \frac{1}{16\theta}\delta^2 - \frac{1}{2}\delta$, which is equivalent to $\eta > m - \frac{1}{2}\delta - \frac{1}{16\theta}\delta^2$. Since $V(v_2(1-\eta, \emptyset)) = \frac{1}{2} - \frac{1}{2}\delta$, the condition $V(v_2(\emptyset, \emptyset)) > V(v_c(\eta+\delta+\epsilon, \emptyset))$ is trivially satisfied. ■

Proof of Proposition 13. Assume $\eta \leq m$ (the case $\eta > m$ is symmetric). Step 1: check for deviations for candidate 1. $V(v_1(m, m)) = \frac{1}{2} + \frac{1}{2}\delta$. It follows readily from the characterization of $\widetilde{b}_1(\cdot)$ in Appendix B that $V(v_1(\widetilde{b}_1(m), m)) \leq \frac{1}{2} + \frac{1}{2}\delta$. Then, the only potentially profitable deviation is to $d_1 = \emptyset$. Two possible cases: $|m - \eta| > \delta$ and $|m - \eta| \leq \delta$. If $|m - \eta| > \delta$, then $V(v_1(\emptyset, m)) = \frac{\eta+m-\delta}{2} p_l + \frac{\eta+m+\delta}{2} p_r + \left(\frac{p_l^2}{\frac{1}{2}} + \frac{p_r^2}{\frac{1}{2}} - 1\right)$. Since $v_1(\emptyset, m|l) = 1 - \frac{\eta}{2} - \frac{m}{2} + \frac{\delta}{2}$, and $v_1(\emptyset, m|r) = 1 - \frac{\eta}{2} - \frac{m}{2} - \frac{\delta}{2}$, the condition $\theta \geq \frac{|v_r - v_l|}{4}$ (in Appendix C) reduces $\theta \geq \frac{\delta}{4}$. Plugging in the values of p_l^* and p_r^* that correspond to the case $\theta \geq \frac{|v_r - v_l|}{4}$ (from Appendix C), $V(v_1(\emptyset, m)) = \frac{1}{4} + \frac{1}{2}\eta - \frac{1}{8\theta}\delta^2 + \frac{1}{16\theta^2}\delta^2$. Then, $V(v_1(m, m)) \geq V(v_c(\emptyset, m))$ if and only if $\frac{1}{2} + \frac{1}{2}\delta > \frac{1}{4} + \frac{1}{2}\eta - \frac{1}{8\theta}\delta^2 + \frac{1}{16\theta^2}\delta^2$. This condition is always satisfied, since the *l.h.s* is greater than $\frac{1}{2}$ and the *r.h.s* is smaller than $\frac{1}{2}$. If $|m - \eta| \leq \delta$, then $V(v_1(\emptyset, m)) = \left(1 - \frac{\eta+m-\delta}{2}\right) p_l + \frac{\eta+m+\delta}{2} p_r + \left(\frac{p_l^2}{\frac{1}{2}} + \frac{p_r^2}{\frac{1}{2}} - 1\right)$. Substituting the value of p_l^* and p_r^* that correspond to the case $\theta \geq \frac{|v_r - v_l|}{4}$, we get

$$V(v_1(\emptyset, m)) = \frac{1}{2} + \frac{1}{2}\delta - \frac{1}{32\theta} + \frac{1}{64\theta^2} + \frac{1}{8\theta}\eta - \frac{1}{8\theta}\eta^2 - \frac{1}{16\theta^2}\eta + \frac{1}{16\theta^2}\eta^2.$$

Algebraic manipulation shows that $V(v_1(m, m)) \geq V(v_1(\emptyset, m))$ if and only if $\eta \leq 0.5$. Since the latter holds trivially, candidate 1 never finds profitable to deviate from (m, m).

Step 2: check for deviations for candidate 2. $V(v_1(m, m)) = \frac{1}{2} - \frac{1}{2}\delta$. It follows readily from the characterization of $\widetilde{b}_2(\cdot)$ in Appendix B that $V(v_2(\widetilde{b}_1(m), m)) \leq \frac{1}{2} - \frac{1}{2}\delta$. Then, the only potentially profitable deviation is to $d_2 = \emptyset$. Because $V(v_2(\emptyset, m)) = \frac{\eta+m}{2}$, it follows that $V(v_2(m, m)) \geq V(v_2(\emptyset, m))$ if and only if $\frac{1}{2} - \frac{1}{2}\delta \geq \frac{\eta+m}{2}$, which reduces to $\eta \leq m - \delta$. ■

8.5 Appendix E: Summary Statistics

Table 1 Summary Statistics			
	Mean	Median	Std. Dev.
NPAT test (dummy)	0.31	-	0.46
TV cost index (U.S. $)	321.9	150	376.1
TV cost index II (10%) (U.S. $)	326.3	156	374.1
TV cost index III (50 %) (U.S. $)	526	406	439.1
Note: TV cost indexes for 2004 constructed based on CPP values, Media Market Guide (2004).			
The TV cost indexes correspond to 30 second ads reaching 1% of the markets' households.			

References

Adams, James. (1999). "Multiparty Spatial Competition with probabilistic voting". *Public Choice* 99, 259-274.

Alesina, Alberto and Alex Cukierman. "The Politics of Ambiguity". *Quarterly Journal of Economics*, 105.

Ansolabehere, Stephen, Roy Behr and Shanto Iyengar. (1993). *The Media Game: American Politics in the Television Age.* McMillan.

Ansolabehere, Stephen, Shanto Iyengar. (1996). "Going negative: how political advertisements shring and polarize the electorate". The Free Press, NY.

Ansolabehere, Stephen, Alan Gerber and James Snyder. (2001). "Does TV advertising Explain the Rise of Campaign Spending? A study of Campaign Spending and Broadcast Advertising Prices in US House Elections in the 1990s and the 1970s". mimeo.

Aragones, Enriqueta, Thomas Palfrey, and Andrew Postlewaite. (2007). "Political Reputations and Campaign Promises". *Journal of the European Economic Association*, 5(4), 846-884.

Aragones, Enriqueta and Andrew Postlewaite. (2002). "Ambiguity in Election Games". *Review of Economic Design 7*, p.233-255.

Aragones, Enriqueta and Zvika Neeman. (2000). "Strategic Ambiguity in Electoral Competition". *Journal of Theoretical Politics* 12, p. 183-204.

Aumann, Robert. (1976). "Agreeing to Disagree", *Annals of Statistics 4*, 1236-9.

Bade, Sophie. (2003). "Electoral Competition with Uncertainty Averse Parties". mimeo, NYU.

Berliant, M., H. Konishi. (2005). "Salience: Agenda Choices by Competing Candidates". *Public Choice*, 125:129-149.

Butterfield, F. (1990). Dukakis says race was harmed by TV. *New York Times*, Section 1, p. 2, April 22.

Campbell, James E., John R. Alford and Keith Henry. (1984). "Television Markets and Congressional Elections". *Legislative Studies Quarterly* 9, 665-678.

Cramer, Jan Salomon. (2003). *Logit Models from Economics and Other Fields.* Cambridge University Press: NY.

D'Alessio, Dave and Mike Allen. (2000). "Media Bias in Presidential Elections: A Meta Analysis". *Journal of Communication*, 50, p. 133-156.

Dalton, Russell and Paul Beck. (1998). "Partisan Cues and the Media: Information Flows in the 1992 Presidential Election". The *American Political Science Review* 92, p. 111-126.

Downs, Anthony. (1957). *An economic Theory of Democracy.* New York: Harper Collins.

Ellsberg, D. (1961). "Risk, Ambiguity, and the Savage Axioms". *Quarterly Journal of Economics* 75:643-669.

Enelow James, Melvin J. Hinich. (1981). "A New Approach to Voter Uncertainty in the Downsian Spatial Model". *American Journal of Political Science 25*: 483-493.

Epstein, L. G, M. Schneider. (2003) "Recursive Multiple Priors". *Journal of Economic Theory*, 113:1-31.

Gentzkow, Matthew and Jesse M. Shapiro. (2005) "Media Bias and Reputation", unpublished.

Gerber, Alan. (1998). "Estimating the Effect of Campaign Spending on Senate Election Outcomes Using Instrumental Variables." *American Political Science Review*, 92, 401-411.

Ghirardato, Paolo and John Katz. (2002). "Indecision Theory: Quality of Information and Voting Behavior". Discussion Paper 1106, CALTECH.

Gilboa, I, D. Schmeidler. (1989). "Maxmin expected utility with a non-unique prior". *Journal of Mathematical Economics*, 18, 141-153.

Gilboa, I, D. Schmeidler. (1993). "Updating Ambiguous Beliefs". *Journal of Economic Theory* 59, 33-49.

Glazer, A. (1990). "The strategy of candidate ambiguity".*The American Political Science Review*, 84, 237-241

Groseclose, Tim and Jeff Milyo. (2005). "A Measure of Media Bias", *The Quarterly Journal of Economics* 120, p. 1191-1237.

Iyengar, Shanto, Richard Reeves, eds. (1997). *Do the Media Govern?: Politicians, Voters and Reporters in America.* SAGE.

Johnson-Cartee, Karen, and Gary A. Copeland. (1997). *Manipulation of the American Voter.* Praeger: Westport, Connecticut.

Keren, G., and L.E. Gerritsen. (1999). "On the Robustness and Possible Accounts of Ambiguity Aversion". *Acta Pshychologica*, 103, 149-172.

Kuhberger, A., and J. Perner. (2003). "The role of Competition and Knowledge in the Ellsberg Task". *Journal of Behavioral Decision Making*, 16, 181-191.

Lin, Min-Tse, James Enelow, Han Dorussen. (1999). "Equilibrium in multicandidate probabilistic spatial voting". *Public Choice* 98, 59-82,

Maccheroni, F, M. Marinacci and A. Rustichini. (2006a). "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences", *Econometrica,* 74, 1447-1498.

Maccheroni, F, M. Marinacci and A. Rustichini. (2006b). "Dynamic Variational Preferences". *Journal of Economic Theory*, 128.

McKelvey, Richard, and John Patty. (2006). "A theory of voting in large elections". Games and Economic Behavior 57, 155-180.

Morelli, Massimo (2004). "Party Formation and Policy Outcome Under Different Electoral Rules". *Review of Economic Studies* 71: 829-853.

Osborne, Martin. (2000). "Entry-deterring policy differentiation by electoral candidates". *Mathematical Social Sciences* 40, 41-62.

Page, Benjamin. (1976). "The role of Political Ambiguity". *American Political Science Review* 70(3):742-752

Patty, John. W. (2007). "Generic Difference of Expected Vote Share and Probability of Victory Maximization in Simple Plurality Elections with Probabilistic Voters." *Social Choice and Welfare*, forthcoming.

Project Vote Smart. (2006). "2006 Voter's self-defense manual".

Robinson, Michael and Margaret Sheehan. (1983). *Over the Wire and on TV.* New York: Russell Sage Foundation.

Saussure, F. (1986) [1916]. *Course in General Linguistics*. Open Court Publishing Company.

Schofield, Norman. (2006). "The Mean Voter Theorem: Necessary and Sufficient Conditions for Convergent Equilibrium". *Review of Economic Studies* (forthcoming)

Shah, D, M. Watts D. Domke, D. Fan, M. Fibison. (1999). "News coverage, economic cues, and the public's presidential preferences: 1984-1996". *Journal of Politics*, 61, p. 914-943.

Shepsle, Kenneth. (1972). "The strategy of Ambiguity: Uncertainty and Electoral Competition". *American Political Science Review 66*, p. 555-568.

Sorauf, Frank. (1992). *Inside Campaign Finance, Myths and Realities.* New Have, CT: Yale University Press.

Stewart, Charles and Mark Reynolds. (1990). "Television Markets and U.S. Senate Elections". *Legislative Studies Quarterly*, 15, 495-523

Vallone, R, Ross L and M Lepper. (1985). "The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre". *Journal of personality and Social Psychology*, 49, 577-585.

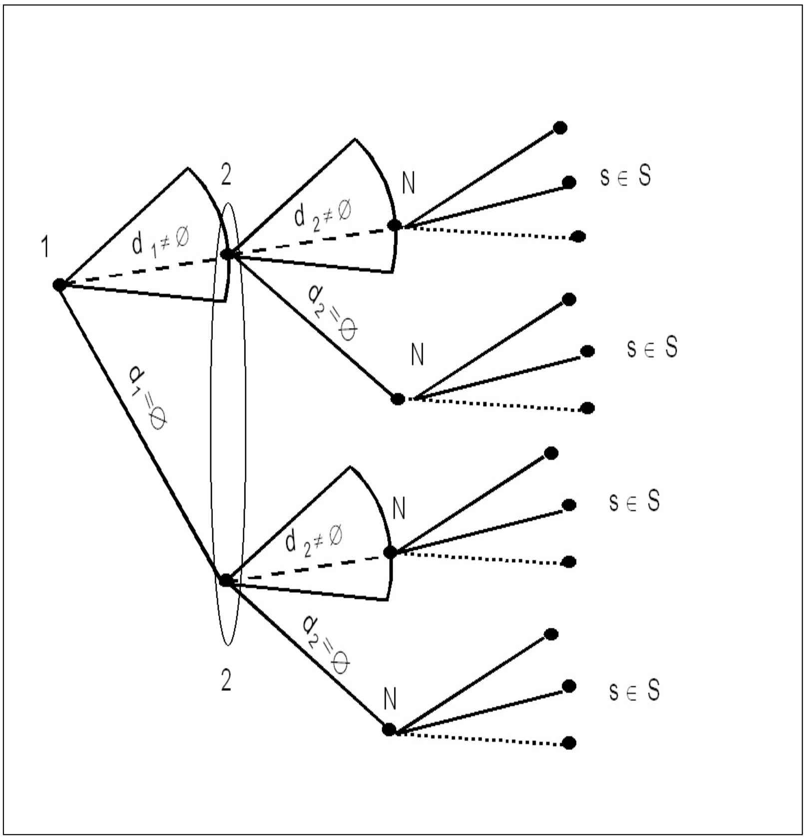

Figure 2: The game as a tree. In the picture, 1 represents candidate 1, 2 represents candidate 2, and N represents a (fair) Nature. Contrary to the convention, Nature moves last to be consistent with our proposed interpretation. We have only depicted three branches for Nature's move to save space, but there actually as many branches as elements in $S_1 \times S_2$, namely 4.