Confirmation bias and signaling in Downsian elections

Antony Millner*[1], Hélène Ollivier†[2], and Leo Simon‡[3]

[1]UC Santa Barbara
[2]Paris School of Economics – CNRS
[3]UC Berkeley

January 8, 2020

Abstract

How do voters' behavioural biases affect political outcomes? We study this question in a model of Downsian electoral competition in which candidates have private information about the benefits of policies, and voters may infer candidates' information from their electoral platforms. If voters are Bayesian, candidates 'anti-pander' – they choose platforms that are more extreme than is justified by their private beliefs. However, anti-pandering is ameliorated if voters' inferences are subject to confirmation bias. Voter confirmation bias causes elections to aggregate candidates' information better, and all observers, whether biased or Bayesian, would like the voters in our model to exhibit *more* confirmation bias than they do themselves.

Keywords: Confirmation bias, electoral competition, pandering, signaling.
JEL codes: D72, D91

1 Introduction

Politicians often surround themselves with experts and advisers who provide them with information about the consequences of policy choices. While voters seldom have direct access to this information, they can in principle infer it from politicians' policy platforms. Since more information is presumably better, intuition might suggest that competition between multiple privately informed candidates would lead to more informed voters and better outcomes from elections. However, electoral competition can lead to undesirable outcomes when office-motivated candidates choose their platforms strategically, accounting for the inferences

*Millner: `millner@ucsb.edu`
†Ollivier: `helene.ollivier@psemail.eu`
‡Simon: `leosimon@berkeley.edu`.

rational voters will make from them. This paper studies signaling and electoral competition when voters are *not* fully rational; we assume that they exhibit a form of confirmation bias that distorts their inferences from candidates' platforms. Can elections in this context aggregate candidates' information more effectively, and give rise to better outcomes for voters?

Of all the well-known behavioural biases, confirmation bias is surely one of the most consequential for political decision-making. As Barack Obama stated in his farewell speech on January 11, 2017, we have a tendency to 'accept only information, whether true or not, that fits our opinions, instead of basing our opinions on the evidence that's out there'. Psychologists have identified several specific behaviours that the term 'confirmation bias' subsumes (Kahneman & Tversky, 1973; Nickerson, 1998). These include a tendency to overweight evidence that confirms our prior beliefs, limit the range of sources we consult to those that are consistent with our priors, and more easily recall evidence that favours our priors. Taber & Lodge (2006) show experimentally that citizens exhibit these behaviours when forming opinions on politically contentious issues such as gun control and affirmative action. In this paper we focus on analysing how confirmation bias that distorts voters' interpretation of information might affect the outcome of competitive elections.

The model we study is a version of the classical Downsian model of electoral competition. Two political candidates possess private information about policy consequences, and choose their platforms with a view to maximizing their probability of winning the election. Voters infer candidates' information from their electoral platforms, update their beliefs about which policies are optimal, and vote for the candidate whose platform is preferred given their updated beliefs. Although each candidate chooses her platform so as to appeal to voters, electoral competition between them can lead to two different kinds of distortions. Candidates who *pander* to the median voter[1] distort their platforms towards the voter's prior beliefs, and away from their private, more informed, beliefs (Heidhues & Lagerlof, 2003).[2] Similarly, candidates who *anti-pander* choose platforms that are more extreme than is justified by their private beliefs, i.e., they push their platforms away from the voter's prior, and from what they privately believe to be optimal. Kartik et al. (2015) (henceforth KST) have shown that anti-pandering equilibria arise in electoral signaling games in which candidates can choose platforms from a continuous interval of the real line.[3] Candidates' platforms are so extreme in their model that voters are worse off than they would have been if they only had access to one candidate's information. In effect, electoral competition fails to aggregate candidates' private information.

In this paper we marry a version of the KST model with a model of confirmation bias due to Epstein (2006), allowing us to illustrate how voter confirmation bias affects candi-

[1]Since we work in the Downsian framework only the median voter's preferences are relevant for determining equilibrium behaviour. We thus refer to 'the voter' from now on where there is no possibility of confusion.

[2]Other studies of pandering in a Downsian framework include Laslier & Van der Straeten (2004) and Gratton (2014).

[3]See Levy (2004, 2007) for other examples of anti-pandering.

date behaviour and information aggregation in elections. We show that when voters exhibit confirmation bias candidates' incentives to anti-pander are ameliorated – confirmation bias and strategic electoral incentives partially offset one another. If the degree of confirmation bias is not too extreme, voter confirmation bias is welfare improving from the perspective of a rational Bayesian observer. Moreover, from the perspective of *any* observer, whether biased or Bayesian, the optimal level of voters' confirmation bias is *higher* than the observer's own level of bias. In addition, if the voter's confirmation bias is above a threshold value, welfare is strictly higher in competitive elections between two privately informed candidates than it would be if the voter only had access to one candidate's information. Perhaps surprisingly, negative findings about the ability of elections to aggregate candidates' information and promote voter welfare may thus be reversed if voters suffer from confirmation bias.

Our paper contributes to two strands of literature in political economics; one studies information aggregation in elections, the other studies the welfare effects of behavioural biases in political contexts.

The literature on information aggregation in elections is vast (for a review see e.g. Austen-Smith & Feddersen, 2009). While much work in this area focuses on the ability of elections to extract voters' private information, the converse problem has also been of interest, i.e., to what extent may voters benefit from candidates' private information? Most work on this topic concludes that electoral competition does not lead to policies that aggregate candidates' information and promote voters' welfare.[4] These negative results are a consequence of pandering (e.g. Heidhues & Lagerlof, 2003), or of anti-pandering as in KST. In the model of KST, upon which this paper builds, voters are never better off than they would have been if they had observed only *one* candidate's private signal – there is an 'as if' information loss in their model.[5] By contrast, in our behavioural model there is no simple relationship between the information content of candidates' platforms and voters' realised welfare. Even though the equilibria we study exhibit similar properties to those studied in KST, voters may achieve high welfare levels, even obtaining the (Bayesian) first best for some parameter values. This demonstrates that there is in general a complex interaction between voter behaviour, candidates' strategies and their informativeness, and realised welfare. 'As if' information losses do not necessarily translate into welfare losses in behavioural models.

Our paper also contributes to an emerging literature on the potential positive impacts of behavioural biases on political equilibria.[6] In an early contribution, Levy & Razin (2015)

[4] An exception is Gul & Pesendorfer (2012), where parties balance the cost of information provision against the probability of convincing the median voter. In this war of information the voter benefits most when parties are equally matched, providing a rationale for regulating campaign expenditures.

[5] The information loss is only 'as if' since both candidates' information is correctly inferred by the voter in a fully-revealing equilibrium. However, equilibrium platforms in the KST model are equivalent to the first-best platforms that would occur if candidates had received identical copies of the same signal. Thus, even though there are two candidates with independent signals, the voter is no better off than she would have been if both candidates received identical signals. See p.12 for further discussion.

[6] Of course, behavioural biases do not always lead to welfare improvements in political contexts. Biases that

found that correlation neglect (i.e. erroneously treating correlated signals as if they were independent) can improve outcomes for voters.[7] In their setting rational voters' ideological policy preferences lead them to favour policies that are welfare sub-optimal, thus distorting the information aggregation process. However, voters who suffer from correlation neglect overweight their information, counteracting the distortion that arises due to their ideologies. Our model differs from this approach in a number of ways. Candidates (not voters) are privately informed in our model, and voters' beliefs are determined in equilibrium, rather than exogenously specified. In addition, in our baseline specification candidates and voters have no ideological policy biases. The distortions in our model are entirely due to the strategic interaction between privately informed candidates, which arise even if candidates' objectives are perfectly aligned with voters'.

On this dimension our paper is perhaps closest in spirit to Lockwood (2017), who introduces voter confirmation bias into the political agency model of Maskin & Tirole (2004). Although a limited amount of confirmation bias raises welfare in both our models, the comparative statics of pandering with respect to the degree of confirmation bias has opposite signs in Lockwood's model and ours. This difference highlights the different roles that pandering and confirmation bias play in models involving agency (Lockwood) versus electoral competition (us). In agency models candidates pander to influence the voter's inference about their *type*, i.e. whether their policy preferences are aligned with the voter's or not.[8] Confirmation bias reduces the benefits of pandering in this framework, since the voter may misinterpret signals, weakening the case for 'lying' in order to try to pass for a type who shares the voter's preferences. In our electoral competition model however, candidates anti-pander as a strategic response to the voter's inferences about their *information* about the consequences of policy. Confirmation bias weakens the response of the voter's beliefs to any information she may infer from candidates' platforms, thus reducing the strategic incentive to anti-pander. The mechanism, and direction of the effects, in these two models is thus quite different. Thus, in contrast to both Levy & Razin (2015) and Lockwood (2017), whose results depend on a tension between beliefs and ideologies (of voters or candidates), we identify a purely informational channel through which confirmation bias may work to improve welfare.

---

lead to negative welfare consequences include context-dependent voting (Callander & Wilson, 2008), and time inconsistency (Bisin et al., 2015; Lizzeri & Yariv, 2017).

[7] Levy & Razin (2015) also study a version of their model in which voters exhibit confirmation bias, but find no welfare change relative to rational voters.

[8] Candidates have two possible types in the Maskin & Tirole (2004) model: Consonant or Dissonant. Consonants share the voters' policy preferences, while Dissonants' preferences are diametrically opposed to the voter's.

2 The model

We study a model of Downsian electoral competition (Downs, 1957) in which candidates can signal their private information about the benefits of policy choices through their electoral platforms. Voters are assumed to have preferences over policies $y \in [-1, 1]$ that depend on the value of an unknown state of the world, $\theta \in \Theta \equiv \{-1, 0, 1\}$. Since we work in the Downsian framework, voters can be represented by a single median voter. In our baseline specification we assume that the cost of a non-optimal choice (for the median voter) is quadratic in the distance between the chosen policy and the optimal *ex-post* choice. We relax these assumptions in Section 3 below, where we consider a continuum of voters with heterogeneous ideological preferences. For now, the representative voter's preferences are taken to be

$$u(y) = -E(y - \theta)^2, \tag{1}$$

where E denotes the expectation operator. θ may be thought of as the optimal policy choice given perfect knowledge of the state of the world. It is clear from (1) that in general the voter's optimal policy choice is

$$\hat{y} = E\theta. \tag{2}$$

Since preferences are single peaked, the voter will always prefer platforms that are closer to $\hat{y}$ to those farther away. The probability distribution from which θ is drawn is denoted by the 3-vector

$$\mathbf{p} = (\Pr(\theta = -1), \Pr(\theta = 0), \Pr(\theta = 1)) \equiv (p_{-1}, p_0, p_1).$$

This probability distribution is itself uncertain, and beliefs about $\mathbf{p}$ are subject to revision if new information becomes available. The vector $\mathbf{p}$, and not θ itself, is the primitive uncertain quantity in the model, as we explain below.

We assume that there are two purely office-motivated candidates, A and B. At the beginning of the game the voter and the candidates have common prior beliefs about the possible values of $\mathbf{p}$. We assume that prior beliefs over $\mathbf{p}$ are given by a Dirichlet distribution with probability density

$$g(\mathbf{p}; \boldsymbol{\alpha}) = \frac{\Gamma(\sum_{k\in\Theta} \alpha_k)}{\prod_{k\in\Theta} \Gamma(\alpha_k)} \prod_{k\in\Theta} p_k^{\alpha_k - 1}, \tag{3}$$

where

$$\boldsymbol{\alpha} = (\alpha_{-1}, \alpha_0, \alpha_1)$$

is a vector of hyper-parameters, and $\Gamma(\cdot)$ is the Gamma function. The Dirichlet distribution is a multidimensional generalization of the Beta distribution, and is a natural prior on the set of discrete probability distributions. The support of the prior distribution in (3) is the

2-dimensional simplex of three element probability distributions. The parameter vector $\boldsymbol{\alpha}$ controls the shape and concentration of prior beliefs on this simplex.

Given these assumptions we can write the voters' ex-ante bliss point as

$$\hat{y}_0 = E\theta = \int \left(\sum_{\theta \in \Theta} p_\theta \theta \right) g(\mathbf{p}; \boldsymbol{\alpha}) d\mathbf{p}. \tag{4}$$

Some simple calculations show that

$$\int p_\theta g(\mathbf{p}; \boldsymbol{\alpha}) d\mathbf{p} = \frac{\alpha_\theta}{\sum_{k \in \Theta} \alpha_k} \tag{5}$$

and thus we have

$$\hat{y}_0 = \frac{\alpha_{-1} - \alpha_1}{\sum_{k \in \Theta} \alpha_k}.$$

We assume a symmetric model, in which prior beliefs place equal weight on all pairs of distributions of the form $\mathbf{p} = (a, b, c)$ and $\mathbf{p}' = (c, b, a)$. Thus $\theta = -1$ and $\theta = 1$ are treated as equally probable ex-ante. This requires $\alpha_{-1} = \alpha_1$, implying that $\hat{y}_0 = 0$. In addition, we assume that prior beliefs are concentrated on probability distributions $\mathbf{p}$ in which $\theta = 0$ is the most likely realized value of θ. This requires $\alpha_0 > \alpha_1$. $\theta = -1, 1$ are thus seen as 'outliers' *a priori*, relative to the modal value of $\theta = 0$.

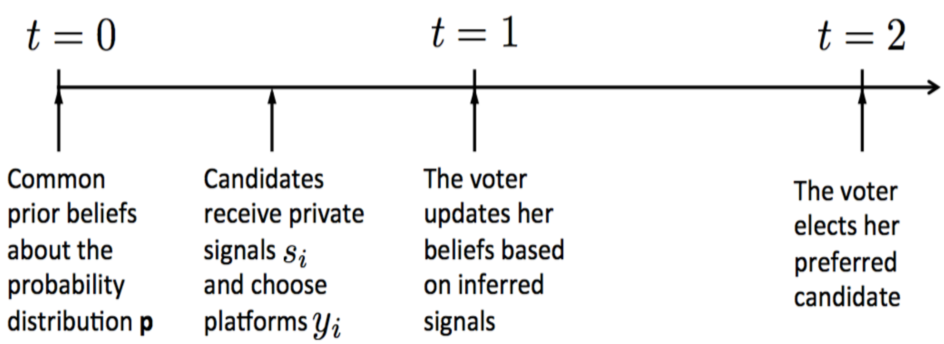
The timing of events is as follows. In period 0 each candidate $i \in \{A, B\}$ receives a private signal $s_i \in \Theta$ about the uncertain state θ. After observing their signals, candidates simultaneously choose their platforms, denoted y_A and y_B, in order to maximize their probabilities of winning the election. In period 1 the voter observes candidates' platforms, perhaps infers something about the signals the candidates received, and updates her beliefs about the probability distribution $\mathbf{p}$ accordingly. Finally in period 2 an election takes place, and the candidate whose platform is closest to the median voter's updated bliss point (2) is elected. If the voter is indifferent between two platforms she randomizes with equal probability between the two candidates. Figure 1 summarizes the sequence of events in the electoral game.

The model we have described above is superficially different from the baseline specification in KST, which focusses on a model with a continuum of signals and normally distributed prior beliefs. However, although the statistical formalism in our discrete model may be less familiar to some readers, the two models are analogous to one another. We relegate a detailed explanation of this fact, and our reasons for focussing on a discrete model, to Appendix A.

2.1 Updating with confirmation bias

The source of novelty in our approach is the behavioural model we adopt for how the voter's beliefs change when she infers candidates' information. We assume that the voter exhibits confirmation bias, i.e., she underweights information that conflicts with her prior beliefs.

Figure 1: Timing of events in the electoral game



Since we will focus on fully revealing equilibria (more on this below) we need only consider the case where the voter infers a pair of signals, one from each candidate. Let

$$S_{ij} = (s_i, s_j)$$

be an arbitrary signal pair. In addition, let $P_B(\mathbf{p}|S_{ij})$ denote the Bayesian posterior belief about $\mathbf{p}$ after receiving the signal pair S_{ij}. Our model of confirmation biased updating is:

$$P_q^*(\mathbf{p}|S_{ij}) = \begin{cases} P_B(\mathbf{p}|S_{ij}) & \text{if } S_{ij} = (0,0) \\ (1-q)P_B(\mathbf{p}|S_{ij}) + qg(\mathbf{p};\boldsymbol{\alpha}) & \text{otherwise.} \end{cases} \tag{6}$$

where $q \in [0, 1)$ is a parameter. Thus, if the voter infers that the candidates' signals are $(0, 0)$ her updating rule is unbiased (0 is the modal signal, and value of θ, *a priori*). In all other cases, her updating rule is a convex combination of the Bayesian rule and her prior beliefs, causing her to under-react to information, and biasing her beliefs towards her prior. The closer q is to 0, the closer the voter's updating rule is to Bayesian rationality; q is thus a measure of the magnitude of the voter's confirmation bias.

Appendix B derives (6) from a primitive model of a non-Bayesian agent's preferences, due to Epstein (2006). Epstein's model provides a micro-foundation for the analysis that allows us to derive confirmation-biased agents' inferences, actions, and welfare in a coherent choice-theoretic framework. However, since the main thrust of the paper can be appreciated without a full understanding of this apparatus we relegate a detailed discussion to the appendix.[9]

[9]The analysis of equilibrium behaviour in our model only requires the reduced form in (6). It is however essential to ground this reduced form in a formal model of voter preferences, as our main results will concern agents' *welfare*. Non-paternalistic welfare computations are undefined in the absence of a fully specified preference structure, which the Epstein formalism delivers.

Since the voter's bliss point (2) is a sufficient statistic for her preferences between candidate's platforms, we will be interested in computing her posterior expected value of θ conditional on receiving a signal S_{ij}, denoted by $E_q[\theta|S_{ij}]$. Some calculations with the Dirichlet distribution (3) and the updating rule (6) show that:

$$E_q[\theta|S_{ij}] = \int \left(\sum_{\theta \in \Theta} p_\theta \theta \right) P_q^*(\mathbf{p}|S_{ij}) d\mathbf{p} \Rightarrow \begin{cases} E_q[\theta|(1,0)] = \frac{1-q}{2\alpha_1+\alpha_0+2} = -E_q[\theta|(-1,0)] \\ E_q[\theta|(0,0)] = E_q[\theta|(-1,1)] = E_q[\theta|(1,-1)] = 0 \\ E_q[\theta|(1,1)] = \frac{2(1-q)}{2\alpha_1+\alpha_0+2} = -E_q[\theta|(-1,-1)]. \end{cases} \tag{7}$$

2.2 Equilibria

A pure strategy for candidate i is a function $y_i(.) : \Theta \to [-1,1]$, where $y_i(s_i)$ is the platform chosen by i when his signal is s_i. We focus throughout on *symmetric, fully revealing* strategy profiles, in which $y_i(\cdot){=}y(\cdot)$, for $i{=}A,B$, with the property that if $s_i{\neq}s_i'$ then $y(s_i){\neq}y(s_i')$.[10] Given such a profile, and a vector of platforms $\boldsymbol{\zeta} = (\zeta_i, \zeta_j) \in y(\Theta) \times y(\Theta)$, denote by $S_{ij}(\boldsymbol{\zeta})$ the uniquely identified signal pair $\left(y^{-1}(\zeta_i), y^{-1}(\zeta_j)\right)$ that the voter can infer from observing $\boldsymbol{\zeta}$. Next, given a bias level q, let $\omega_i^y(\boldsymbol{\zeta}; q)$ denote the probability that the voter selects i if the candidates' symmetric strategy is y and the announced platform vector is $\boldsymbol{\zeta}$, with the restriction:

$$\text{if } \boldsymbol{\zeta} \in y(\Theta) \times y(\Theta) \text{ then } \omega_i^y(\boldsymbol{\zeta}; q) = \begin{cases} 1 & \text{if } |\zeta_i - E[\theta|S_{ij}(\boldsymbol{\zeta})]| < |\zeta_j - E[\theta|S_{ij}(\boldsymbol{\zeta})]| \\ 0 & \text{if } |\zeta_i - E[\theta|S_{ij}(\boldsymbol{\zeta})]| > |\zeta_j - E[\theta|S_{ij}(\boldsymbol{\zeta})]| \\ 0.5 & \text{otherwise} \end{cases}. \tag{8}$$

Restriction (8) is the analog of the restriction in sender-receiver games that if a sender makes an "on the equilibrium path" deviation from his declared strategy, then the receiver must use Bayes' Law to update her beliefs about the sender's private information. When $\boldsymbol{\zeta} \notin y(\Theta) \times y(\Theta)$, we impose no *a priori* restriction on $\omega_i^y(\boldsymbol{\zeta}; q)$; this is analogous to the assumption underlying Perfect Bayesian equilibria that after observing an "off the equilibrium path" deviation, the receiver's beliefs are unrestricted. (When we support an equilibrium with specific off-equilibrium path beliefs, we will further restrict ω_i^y.)

We study the *Symmetric Perfect Confirmation-Biased Equilibria (SPCBE)* of the game between candidates and the voter. A strategy profile $y(\cdot)$ is a SPCBE if: A) the voter updates her beliefs in accordance with the non-rational updating rule (6) for some q, then chooses between the candidates so as to maximize her expected utility (1), conditional on these beliefs; and B) for each signal s_i received by candidate i, the platform $y(s_i)$ maximizes his posterior probability of being elected given the voter's behavior rule A). In symbols, let

[10] We explain our focus on this class of equilibria, and discuss the existence of other equilibria, in Footnote 12.

$\rho(s_j|s_i) = \frac{\alpha_j + I_{j=i}}{\sum_\theta \alpha_\theta + 1}$ denote candidate i's posterior predictive probability that candidate j's signal is s_j, given that i's signal is s_i. If i observes signal s_i and chooses platform ζ against j's strategy $y(\cdot)$, i's expected probability of being elected is $\sum_\Theta \rho(s_j|s_i)\omega_i(\zeta, y(s_j); q)$. Then $y(\cdot)$ is a SPCBE if for $i = A, B$,

$$\forall s_i \in \Theta, \forall \zeta \in [-1,1], \quad \sum_\Theta \rho(s_j|s_i)\Big(\omega_i^y(\zeta, y(s_j); q) \; - \; \omega_i^y(y(s_i), y(s_j); q).\Big) \quad \leq \quad 0 \tag{9}$$

Our first result shows that there is a unique fully revealing, symmetric equilibrium (all proofs are contained in the appendix):

Proposition 1. *For each $q \in [0,1)$, there is a unique fully revealing, SPCBE in pure strategies $y_q(\cdot)$, given by:*

$$y_q(-1) = -\bar{y}_q;\ y_q(0) = 0;\ y_q(1) = \bar{y}_q, \ \textit{where}\ \bar{y}_q = \frac{2(1-q)}{2\alpha_1 + \alpha_0 + 2}. \tag{10}$$

Each candidate is elected with probability 1/2 regardless of his signal realization.

To understand the important features of the equilibrium in Proposition 1, begin by considering the case of a Bayesian voter, i.e. $q = 0$. In this case candidates' equilibrium platforms exhibit anti-pandering: their platforms are *more extreme* than their private beliefs about the voter's posterior bliss point. To see this, note that

$$E[\theta|1] = \frac{1}{2\alpha_1 + \alpha_0 + 1} < \bar{y}_0 \ \text{ and } \ E[\theta|-1] = \frac{-1}{2\alpha_1 + \alpha_0 + 1} > -\bar{y}_0, \tag{11}$$

where expectation operators without subscripts are Bayesian.[11] Thus, for example, a candidate who receives a signal 1 believes that the voter's most preferred platform will be at $E[\theta|1]$, however in equilibrium this candidate plays $\bar{y}_0 > E[\theta|1]$. Candidates thus overreact to their private information, choosing equilibrium platforms that are more extreme than they believe they should be.

To understand why candidates anti-pander it is helpful to see why a strategy profile in which each candidate plays his true posterior expected value of θ is *not* an equilibrium. Suppose that candidates play their true posterior expectations, i.e. on receiving signals $(-1, 0, 1)$ they play $(-y^*, 0, y^*)$ where $y^* = E[\theta|1]$. To make the discussion concrete, suppose that

[11] This calculation follows from the properties of the Dirichlet distribution (3):

$$\text{Prob}(\mathbf{p}|1) \propto \text{Prob}(1|\mathbf{p})g(\mathbf{p};\boldsymbol{\alpha}) = p_1 g(\mathbf{p};\boldsymbol{\alpha}) \Rightarrow \text{Prob}(\mathbf{p}|1) = g(\mathbf{p};(\alpha_1, \alpha_0, \alpha_1 + 1))$$

and hence

$$E[\theta|1] = \int \left(\sum_{\theta \in \Theta} p_\theta \theta\right) \text{Prob}(\mathbf{p}|1)d\mathbf{p} = \frac{1}{2\alpha_1 + \alpha_0 + 1}(-\alpha_1 + 0 + (\alpha_1 + 1)) = \frac{1}{2\alpha_1 + \alpha_0 + 1}.$$

$\alpha_0 = 1 + \delta$ and $\alpha_1 = 1 - \frac{\delta}{2}$ for some $\delta > 0$. In this case we have $y^* = E[\theta|1] = \frac{1}{4}$. Now consider what happens if candidate A plays 0, while B plays y^*. In this case the voter would infer the signal pair $(0, 1)$, and from (7), her posterior expectation of θ would be $E[\theta|(0, 1)] = \frac{1}{5}$. Since $\frac{1}{5}$ is closer to $y_B = \frac{1}{4}$ than to $y_A = 0$, candidate A will lose the election in this case. Thus, A's strategy admits a profitable deviation: if A plays y^* rather than 0 when he observes a signal 0, he will tie, rather than lose, the election when B plays $\pm y^*$ and win rather than tie when B plays 0. The proposed 'truthful revelation' strategy thus cannot be an equilibrium. To eliminate the possibility of a profitable deviation for A, we must increase the value of $y(1)$ from y^* to the point where $y(1)$ and 0 are equidistant from the voter's posterior, conditional on her inferring the signal pair $(0, 1)$. This point is given by $\bar{y}_0$ in (10). This equilibrium exhibits *anti-pandering*: each candidate's platform after observing signals ± 1 is greater in absolute value than their private expectations.

The ultimate reason for anti-pandering when voters act as Bayesians is that the voter's posterior beliefs after inferring both candidates' signals are more extreme than the average of each candidate's private expectations, which are based on the observation of just one signal. Indeed, some calculations using the expressions for $E[\theta|(s_i, s_j)]$ in (7), and $E[\theta|s_i]$ in (11), show that

$$E[\theta|(s_i, s_j)] = \left[2 \times \frac{2\alpha_1 + \alpha_0 + 1}{2\alpha_1 + \alpha_0 + 2}\right] \left[\frac{1}{2}\left(E[\theta|s_i] + E[\theta|s_j]\right)\right] \tag{12}$$

$$\Rightarrow |E[\theta|(s_i, s_j)]| \geq \left|\frac{1}{2}\left(E[\theta|s_i] + E[\theta|s_j]\right)\right|, \tag{13}$$

where the inequality follows since $2 \times \frac{2\alpha_1+\alpha_0+1}{2\alpha_1+\alpha_0+2} \in (1, 2)$, and is strict whenever $|s_i| \neq |s_j|$. Thus, if the candidates truthfully reported their private expectations, then whenever the voter infers signals that are non-identical in absolute value, her posterior beliefs will be closer to the private expectations of the candidate with the more extreme signal. This gives rise to incentives for candidates to overreact to their private information, choosing platforms that are more extreme than they believe to be optimal. Indeed, using equations (10), (11), and (12) we see that

$$y_0(1) = E[\theta|(1, 1)] = -y_0(-1) = -E[\theta|(-1, -1)]. \tag{14}$$

Thus, each candidate chooses his platform at what would be the voter's bliss point, had she inferred that both candidates received the *same* signal. Since the candidates' platforms are ex post ties, the voter is no better off than she would have been having observed only one of the candidates' signals – there is an 'as if' information loss in equilibrium. These results are similar to those in Kartik et al. (2015), who show that such 'ex post' anti-pandering equilibria arise in a large class of electoral signaling games in which the set of possible platforms is continuous.

Now consider how confirmation bias modifies the equilibrium of the electoral game. When the voter is subject to confirmation bias, her posterior beliefs are *less* responsive to her inferences about candidates' signals than those of a Bayesian voter. Equation (7) shows that her posterior expected value of θ after inferring signal pair S_{ij} is now

$$E_q[\theta|S_{ij}] = (1-q)E[\theta|S_{ij}] \quad \Rightarrow |E_q[\theta|S_{ij}]| \leq |E[\theta|S_{ij}]|$$

where the inequality is strict as long as $S_{ij} \neq (0,0)$. Confirmation bias thus draws the voter's posterior expected value of θ towards her prior (note that $E_q[\theta] = 0$ *a priori*). Rational candidates will adapt their strategies to the biased voter's beliefs. Since confirmation bias makes the voter's posterior beliefs *less* extreme, this counteracts the strategic incentive to overreact to information we observed in the Bayesian case. Indeed, we see that platforms in the equilibrium with confirmation bias satisfy

$$y_q(1) = E_q[\theta|(1,1)] < E[\theta|(1,1)] = y_0(1).$$

2.3 Welfare

Because confirmation bias counteracts the overreaction effects we observed when the voter is a Bayesian, a Bayesian observer may view voter confirmation bias as welfare improving. To see the intuition for this consider the effect of a very small amount of confirmation bias on candidates' platforms. Equation (14) shows that when $q = 0$ candidates' equilibrium platforms are ex-post optimal from the perspective of a Bayesian observer whenever candidates' signal pairs are $(0,0), (1,1), (-1,-1)$, or $(-1,1)$. Thus, increasing q to a small positive value $\epsilon \ll 1$ has only a second order negative effect on ex-post Bayesian welfare if candidate's receive these signal pairs. However when the inferred signal pair is $(0,1)$ or $(0,-1)$ candidates' platforms are *not* ex-post optimal according to the Bayesian observer. When $q = \epsilon$ the Bayesian observer would calculate ex-post welfare if signals are $(0,1)$ as

$$\frac{E[(y_\epsilon(0)-\theta)^2|(0,1)]}{2} + \frac{E[(y_\epsilon(1)-\theta)^2|(0,1)]}{2} = \frac{(y_\epsilon(1) - E[\theta|(0,1)])^2}{2} + \text{terms independent of } \epsilon$$

where we have used the fact that $E(y-\theta)^2 = (y - E\theta)^2 + \text{Var}(\theta)$, and $y_\epsilon(0) = 0$. Since

$$E[\theta|(0,1)] < y_\epsilon(1) < y_0(1) = E[\theta|(1,1)]$$

if ϵ is small, increasing confirmation bias from 0 to ϵ has a first order positive effect on ex-post welfare in this case. Similar reasoning holds when the inferred signal pair is $(0,-1)$. Since the positive effects of a small increase in confirmation bias are first order, but the negative effects are second order, a small increase in a Bayesian voter's confirmation bias *increases* ex-ante Bayesian welfare.[12]

[12]Our model has other equilibria, some of which *reduce* welfare relative to the benchmark of a Bayesian voter. Since we assume that candidates are identical *a priori*, we focus on equilibria that are symmetric with respect to candidates' strategies. Of the equilibria that are symmetric with respect to candidates, there

We can make this more precise by using equations (1) and (10) to define ex-ante Bayesian welfare as follows:

$$W = -\sum_{s_i \in \Theta} \rho(s_i) \sum_{s_j \in \Theta} \rho(s_j|s_i) \left(\frac{1}{2} \sum_{t \in \{s_i, s_j\}} E\left[(y_q(t) - \theta)^2 | (s_i, s_j) \right] \right), \tag{15}$$

where

$$\rho(s_i) = \int \text{Prob}(s_i|\mathbf{p}) g(\mathbf{p}; \boldsymbol{\alpha}) d\mathbf{p} = \frac{\alpha_i}{\sum_{k \in \Theta} \alpha_k}$$

is the prior probability of receiving signal s_i, and we have used the fact that the voter randomizes between candidates' platforms in equilibrium. Straightforward, if lengthy, computations[13] show that

$$W \propto -[(\alpha_0 + 2\alpha_1 + 1)q^2 - (\alpha_0 + 2\alpha_1)q] + \text{terms independent of } q.$$

W is thus a concave function of q, which achieves its maximum at some $q^* > 0$. Calculations with the full analytic expression for W yield the following result:

Proposition 2. *A Bayesian observer believes that confirmation bias improves equilibrium welfare (relative to the case of a Bayesian voter) if*

$$q < q_C \equiv \frac{\alpha_0 + 2\alpha_1}{\alpha_0 + 2\alpha_1 + 1}.$$

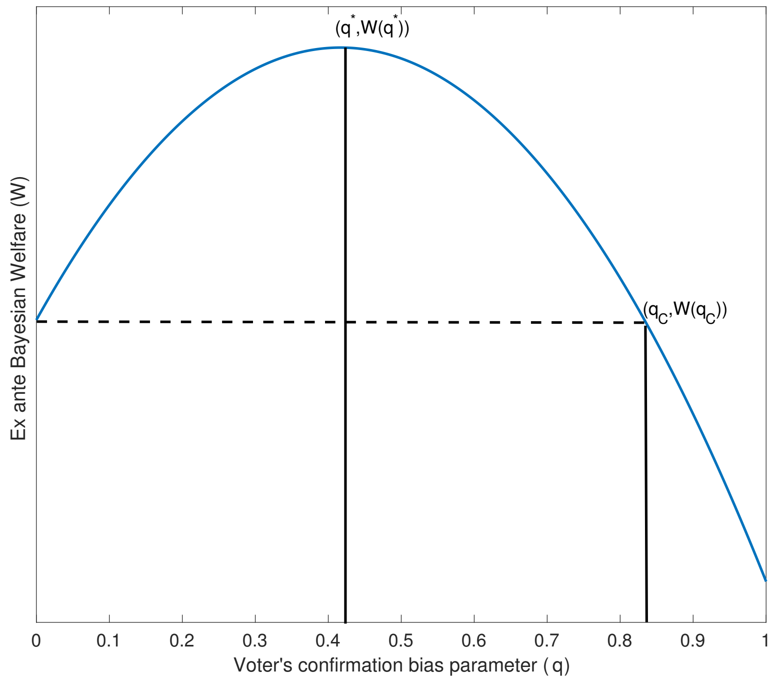
The value of q that maximizes ex-ante Bayesian welfare is $q^* = q_C/2$.

Figure 2a illustrates this result, and the dependence of ex-ante Bayesian welfare on the confirmation bias parameter q.

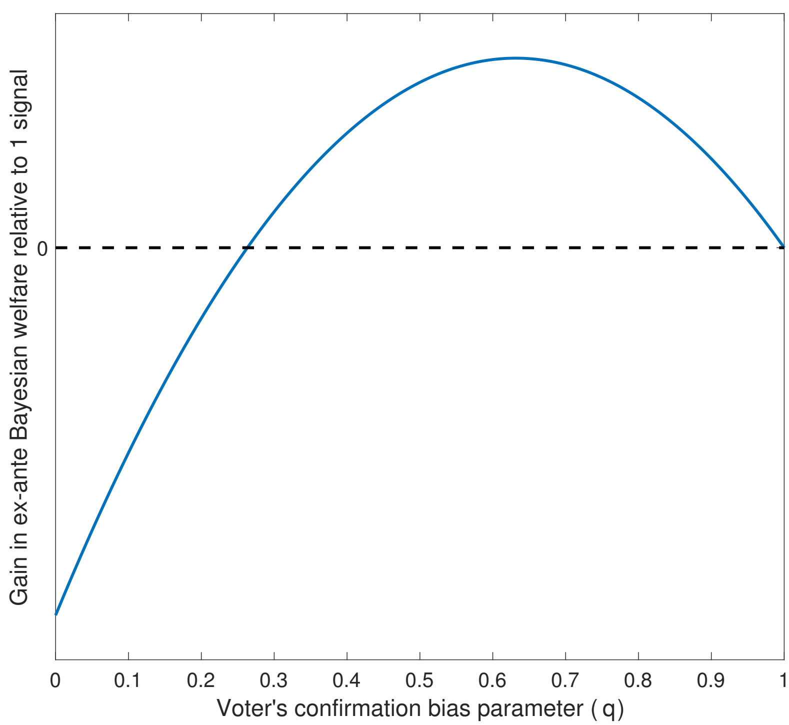
To interpret the optimal value of voter confirmation bias q^* further, notice from (10), (11),

are two kinds. The first is a 'babbling equilibrium' in which both candidates announce zero, irrespective of the signals they observe. This equilibrium has no interesting comparative statics properties, as it does not vary with the voter's degree of confirmation bias. The second is a one-dimensional family of semi-pooling equilibria. Specifically, assume without loss of generality that each candidate announces $y^- < 0$ if signal -1 is observed, and $y^+ > 0$ otherwise. The voter will be indifferent between the two candidates' platforms unless exactly one of them observes -1. It follows that this platform profile will be an equilibrium if and only if, after drawing the appropriate inferences, the bliss point for the voter with bias parameter q equals $(y^- + y^+)/2$, and so is equidistant from y^- and y^+. Now suppose this property is satisfied. Choose $\xi^+ > 0$ arbitrarily and let $\xi^- = 1 + (1 - \xi^+) y^+ / y^-$. In this case, the semi-pooling platform pair $(\xi^+ y^+,\ \xi^- y^-)$, will also be an equilibrium, since $(\xi^+ y^+ + \xi^- y^-)/2 = (y^+ + y^-)/2$. It follows that for any q, there are equilibria which exhibit either pandering or anti-pandering, for one or both platforms. However, since for each q there is no natural criterion by which we can select one value of ξ^+ over the others, the effect of a change in confirmation bias on the equilibrium is indeterminate for this family. We thus focus on the only symmetric equilibria for which changes in voter confirmation bias are both non-trivial and have an unambiguous meaning.

[13] These computations are best performed with a computer algebra package. As the proofs of the results in this subsection are simple, if lengthy, algebraic manipulations, we do not provide details here. The Matlab code used to generate the results is available at `https://tinyurl.com/ybfkc7c2`.



(a) Dependence of ex-ante Bayesian welfare on confirmation bias. $\alpha_0 = 3, \alpha_1 = 1$.



(b) Welfare gain relative to one signal, as a function of voter confirmation bias. $\alpha_0 = 3, \alpha_1 = 1$.

and the definition of q^* that

$$y_{q^*}(1) = \frac{2(1-q^*)}{2\alpha_1+\alpha_0+2} = 2\frac{1-\frac{1}{2}\frac{\alpha_0+2\alpha_1}{\alpha_0+2\alpha_1+1}}{2\alpha_1+\alpha_0+2} = \frac{1}{\alpha_0+2\alpha_1+1} = E[\theta|1].$$

Thus, at the optimal level of voter confirmation bias, candidates play their unbiased private best guess of the value of θ. They thus act as if they were in a non-strategic situation, effectively forgetting that the voter will in fact infer *two* signals in equilibrium. At first this may seem surprising, given that welfare depends on the ex-post expected value of θ conditional on *both* signals being observed. To understand this further, note that the optimal value of $y(1)$ (i.e. candidates' platforms when they receive signal 1) for an ex-ante Bayesian observer must satisfy

$$\max_{y(1)} \left\{ -\sum_{s_j\in\Theta} \rho(s_j|1)E[(y(1)-\theta)^2|(s_j,1)] \right\}.$$

Denoting the solution of this optimisation problem by $y^*(1)$, we have

$$\begin{aligned}
y^*(1) &= \sum_{s_j\in\Theta} \rho(s_j|1)E[\theta|(s_j,1)] \\
&= \sum_{s_j\in\Theta} \rho(s_j|1)\int \left(\sum_{\theta\in\Theta} p_\theta\theta\right) \mathrm{Prob}(\mathbf{p}|(s_j,1))d\mathbf{p} \\
&= \int \left(\sum_{\theta\in\Theta} p_\theta\theta\right) \left[\sum_{s_j\in\Theta} \mathrm{Prob}(\mathbf{p}|(s_j,1))\rho(s_j|1)\right] d\mathbf{p} \\
&= \int \left(\sum_{\theta\in\Theta} p_\theta\theta\right) \mathrm{Prob}(\mathbf{p}|1)d\mathbf{p} \\
&= E[\theta|1].
\end{aligned}$$

Since the voter's payoff function is quadratic, the platform that maximizes ex-ante Bayesian welfare when one of the candidates receives a signal 1 is just the expected value of θ, conditional on 1 being observed. This expectation runs over the values of the other candidate's signal s_j. However, by the Law of Iterated Expectations, the expected value of θ conditional on the signal pair $(s_j, 1)$ being observed, averaged across all values of s_j, is equivalent to the expected value of θ if just the single signal 1 were observed. Unbiased strategies are thus ex-ante optimal from the perspective of a Bayesian observer.

In Figure 2b we illustrate the extent to which candidates' signals provide an informational welfare benefit, despite the distortions in their platforms. To quantify this we compare ex-ante welfare when there are *two* privately informed candidates to ex-ante welfare when there is only *one* candidate who seeks to maximise the voter's welfare.[14] If candidate's platforms were undistorted by political competition, two candidates would always be better than one, as information has positive value. However, a key finding of KST is that electoral competition between two candidates gives rise to a welfare *loss* relative to a single informed candidate (this is a consequence of anti-pandering). Figure 2b illustrates that this result may be reversed if the voter exhibits confirmation bias. Ex-ante Bayesian welfare is larger with two candidates than with one whenever the voter's bias parameter $q > 0.26$ in this example.

The Bayesian welfare measure we have used thus far provides a natural normative benchmark for welfare judgements. However, it is also interesting to ask how a *confirmation-biased* observer would evaluate the equilibrium outcome. In Appendix B we show that the appropriate welfare measures for an observer who has confirmation bias parameter q_w is:

$$W_{q_w} = -\sum_{s_i \in \Theta} \rho(s_i) \sum_{s_j \in \Theta} \rho_{q_w}(s_j|s_i) \left(\frac{1}{2} \sum_{t \in \{s_i, s_j\}} E_{q_w} \left[(y_q(t) - \theta)^2 | (s_i, s_j) \right] \right), \tag{16}$$

where the expectation E_{q_w} is now taken with respect to the observer's biased posterior beliefs, given by (6) with $q = q_w$. Notice that two occurences of confirmation bias enter this welfare calculation: the observer's bias q_w and the voter's bias q, and only the latter influences the candidates' strategies. Aside from modifying the observer's views on the 'correct' platform conditional on candidate's signals S_{ij}, confirmation bias also affects the observer's views on the posterior predictive probability $\rho_{q_w}(s_j|s_i)$. Using an updating rule analogous to (6) but for one signal,[15] and the properties of the Dirichlet distribution, one can show that

$$\rho_{q_w}(s_j|s_i) = \begin{cases} \frac{\alpha_j + I_{s_j=0}}{\sum_k \alpha_k + 1} & s_i = 0 \\ \\ (1-q_w)\frac{\alpha_j + I_{s_i = s_j}}{\sum_k \alpha_k + 1} + q_w \frac{\alpha_j}{\sum_k \alpha_k} & s_i \neq 0. \end{cases} \tag{17}$$

Appendix B explains that the prior probability of receiving a signal s_i, $\rho(s_i)$, is unaffected by confirmation bias in the Epstein (2006) framework that underpins the model.

Calculations similar to those used in Proposition 2 now yield the following:

Proposition 3. *For all $q_w \in [0, 1)$, the biased ex-ante welfare measure W_{q_w} is a concave*

[14] Formally, the one candidate benchmark welfare level is

$$W^1(q) = -\sum_{s_i \in \Theta} \rho(s_i) E\left[(y_q^1(s_i) - \theta)^2 | s_i \right]$$

where $y_q^1(s_i) = E_q[\theta|s_i]$.

[15] Note that, unlike Bayesian updating, updating with confirmation bias will in general be sensitive to the order in which signals are received (see also Rabin & Schrag (1999)). Thus, posterior beliefs after observing two signals in sequence will differ from posterior beliefs after observing the same two signals simultaneously. Since signals are always inferred simultaneously in our model this poses no difficulties for us, and we use the same updating rule for individual signals s_i and simultaneous signal pairs S_{ij}.

function of the voter's confirmation bias parameter q. It attains its maximum at $q = q^*(q_w) \in (0,1)$, where the function $q^*(q_w)$ is increasing, and satisfies $q^*(q_w) > q_w$ for all $q_w \in [0,1)$.

The upshot of this proposition is that all observers, whether biased or Bayesian, would prefer the voter to be more biased than themselves. Figure 3a illustrates this result. The intuition is very similar to that in the Bayesian case. Consider the case where the voter's confirmation bias parameter q is just above the observer's, i.e. $q = q_w + \epsilon$. A voter with $q = q_w$ induces candidates to choose platforms that are ex-post optimal from the perspective of an observer with bias q_w when the inferred signals are $(-1,1), (0,0), (1,1)$ or $(-1,-1)$. A small increase in q gives rise to second order losses in these cases. However, platforms are non-optimal when inferred signals are $(0,1)$ or $(0,-1)$, and a small increase in q gives rise to first order gains in these cases since

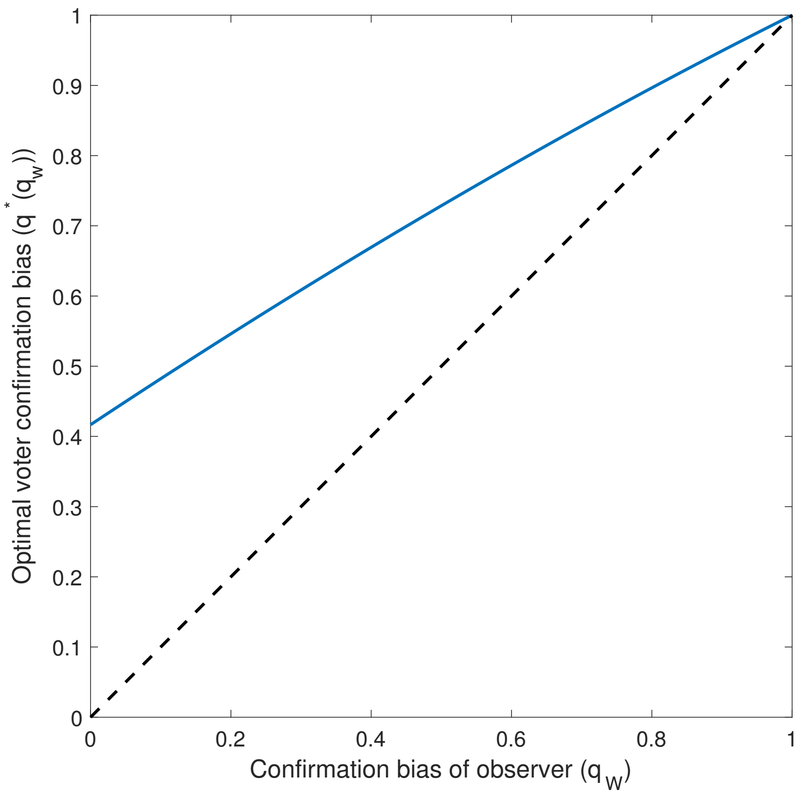
$$E_{q_w}[\theta|(0,1)] < y_{q_w+\epsilon}(1) < y_{q_w}(1) = E_{q_w}[\theta|(1,1)].$$

Thus, increasing q above q_w is always welfare improving for the biased observer. Similar logic shows that a small decrease in q below q_w reduces welfare, since in addition to the familiar second order losses, this also gives rise to first order losses when inferred signals are $(0,1)$ or $(0,-1)$. Since welfare is concave and locally strictly increasing in a neighborhood of q_w, it cannot be higher than at q_w when $q < q_w$. Thus all observers, whether biased or Bayesian, would like the voter to be more biased than themselves.

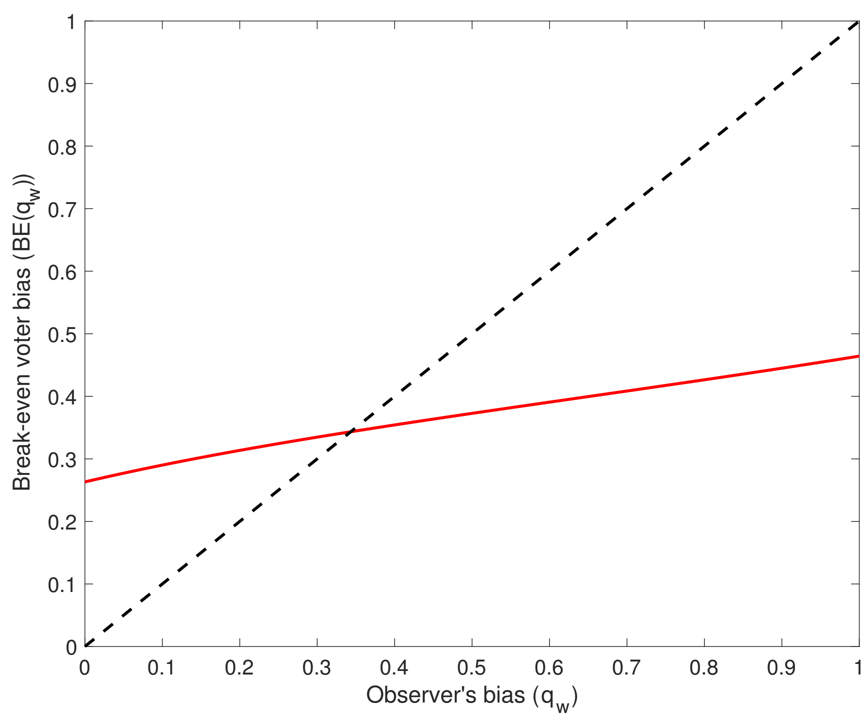
Recall that in Figure 2b we showed that Bayesian welfare with two privately informed competitive candidates is higher than with only one candidate provided the voter is sufficiently biased ($q > 0.26$ for that example). Figure 3b shows that this finding generalises to confirmation-biased welfare measures. The figure plots the 'break-even' value of the voter's bias at which confirmation-biased welfare with two candidates is equal to confirmation-biased welfare with one candidate, denoted by $BE(q_w)$ (solid red line). Whenever $q > BE(q_w)$, welfare in the two candidate election is higher than welfare when there is only one candidate, from the perspective of an observer with bias q_w.

3 Discussion

Our results thus far have all focussed on the simplest version of the model in which there is a single representative voter, and candidates are purely office motivated and commit to implementing the platforms they announce. We now examine how (and if) our results might change if these assumptions are relaxed.



(a) Optimal level of voter confirmation bias $q^*(q_w)$ according to a biased observer with confirmation bias q_w. $\alpha_0 = 3, \alpha_1 = 1$.



(b) Minimum voter bias required for welfare with *two* informed candidates to be higher than welfare with *one* candidate, as a function of the observer's bias.

3.1 Heterogeneous voter ideologies

Our main analysis used the fiction of a single representative voter to simplify the presentation. We now show that all our welfare results are preserved in a model with a continuum of voters with idiosyncratic policy preferences. Suppose that voters are indexed by their *a priori* ideological bliss points b, which affect their payoffs from a platform choice y thus:

$$u(y; b) = -E(y - (\theta + b))^2.$$

Voter bliss points are distributed on $[-\bar{b}, \bar{b}]$ according to the cumulative distribution function $G(b)$. Let b_m be the bliss point of the median voter, i.e., b_m satisfies $G(b_m) = 0.5$. Since voter preferences are still single-peaked, and all voters update their beliefs in a common manner, it is clear that the unique fully revealing SPCBE of the electoral game will be for candidates to choose platforms $\tilde{y}_q(s_i)$ given by

$$\tilde{y}_q(s_i) = b_m + y_q(s_i)$$

where $y_q(s_i)$ is the equilibrium platform from Proposition 1. The ex-ante welfare of a voter with bliss point b in this equilibrium, according to an observer with confirmation bias parameter q_w, is:

$$\tilde{W}_{q_w}(b) = -\sum_{s_i \in \Theta} \rho(s_i) \sum_{s_j \in \Theta} \rho_{q_w}(s_j|s_i) \left(\frac{1}{2} \sum_{t \in \{s_i, s_j\}} E_{q_w} \left[(\tilde{y}_q(t) - (\theta + b))^2 | (s_i, s_j) \right] \right).$$

Total ex-ante voter welfare according to this observer is

$$V_{q_w} = \int_{-\bar{b}}^{\bar{b}} \tilde{W}_{q_w}(b) G'(b) db.$$

Proposition 4. *Propositions 2 and 3 are unchanged by the introduction of heterogeneous voter bliss points.*

This result follows from the fact (proved in the appendix) that:

$$V_{q_w} = W_{q_w} + \int_{-\bar{b}}^{\bar{b}} (b_m - b)^2 G'(b) db, \tag{18}$$

where W_{q_w} is defined in (16). Since the confirmation bias parameters q and q_w only enter total voter welfare V_{q_w} through the first term W_{q_w} in (18), the results described in Section 2.3 apply for arbitrary distributions of voter bliss points.

3.2 Mixed motives

While candidates in our baseline model are purely office-motivated, our result is robust to the introduction of a small degree of policy motivation. Following KST we consider the following augmentation of candidate i's preferences to include a policy motivation:

$$u(y, b_i, \rho_i) = -\rho_i E(y - (\theta + b_i))^2 + (1 - \rho_i) I_{W=i} \tag{19}$$

where $I_{W=i}$ is 1 if i wins the election and 0 otherwise, b_i is candidate i's *a priori* bliss point, and $\rho_i \in [0, 1]$ measures the strength of i's policy motivation. Obviously, when $\rho_i = 0$, we recover our baseline specification.

Assume now that player j plays the equilibrium strategy specified in Proposition 1. Assume also that the voter's beliefs are as specified in equation (27) of Appendix C. If candidate i plays the same strategy as j, he wins the election with probability 0.5. From (19), his expected payoff is

$$-\rho_i E[(y_q(s_i) - \theta - b_i)^2 | s_i] \;+\; (1 - \rho_i)/2$$

As we show in Appendix C, if he deviates from this strategy and plays $z(s_i)$, he wins the election with probability strictly less than 1/3, so that his payoff from winning is bounded above by $(1 - \rho_i)/3$. Thus his net gain to deviating is strictly bounded above by

$$\rho_i \left[E[(y_q(s_i) - \theta - b_i)^2 - (z(s_i) - \theta - b_i)^2 | s_i] \right] \;-\; (1 - \rho_i)/6 \tag{20}$$

For given b_i, (20) will be negative for $\rho_i \in \big(0, \bar{\rho}_i]$, where $\bar{\rho}_i = (6E[(y_q(s_i) - \theta - b_i)^2 - (z(s_i) - \theta - b_i)^2 | s_i] + 1)\big)^{-1}$. That is, if the candidate's weight ρ_i on the policy-related argument of his utility function is sufficiently small (at most $\bar{\rho}_i$), there is no profitable deviation from the strategy specified in Proposition 1. Thus, our results are robust to a small amount of mixed motives for the candidates.

This result may also be applied to another extension of the model. Thus far we have assumed, in alignment with most of the literature on signalling games, that all agents have a common prior. This is a natural modelling assumption, as all the effects in the model then flow through candidates' private information and their electoral incentives; there are no other sources of heterogeneity between voters and candidates. However, it is also of interest to ask how the results might change if candidates and voters have different priors. This could occur, for example, if all agents form their priors by observing a sequence of public signals prior to the election, but candidates and voters have different degrees of confirmation bias. If candidates are purely office-motivated their priors are irrelevant for determining the equilibrium – they simply choose platforms based on voters' inferences. However, if candidates are partially policy-motivated, their priors could matter. The analysis above shows that provided ρ_i is not too large the results are robust to these concerns. There is nothing in the analysis above that requires the expectation in (19) to be taken with respect to the same prior as the voter, and so the results can be applied unchanged to this case.[16]

3.3 No commitment

We have so far assumed that candidates can commit to their announced platforms. This assumption is reasonable if candidates, concerned about re-election, wish to build a reputation for competence and reliability (Alesina, 1988). It is also empirically grounded: a meta-analysis of research that reports quantitative measures of election promise fulfilment in Northern America and Europe found that candidates fulfil 67 percent of their promises on average (Petry & Collette, 2009). This reassuringly indicates that some degree of electoral commitment occurs in the real world. However, it is still of interest to assess how our results might differ if we relaxed this assumption.

To explore this possibility, we follow KST and consider the case in which candidates cannot commit to implement their announced platforms, i.e. platforms are "cheap talk". In this scenario candidates are free to implement any policy they please once elected. If candidates are purely office motivated the resulting pure communication game has many equilibria, both informative and uninformative. KST argue that, unlike in the version of the model with commitment, all informative equilibria of this game are implausible, as they do not survive a basic equilibrium refinement for pure communication games (neologism-proofness). Their reasoning applies without modification to our setup. This leaves the uninformative equilibria, of which there are a great many. In uninformative equilibria the voter chooses each candidate with probability 0.5, and candidates can choose *any* policy ex-post. There is thus a sense in which 'anything goes' when candidates cannot commit. In the generic plausible no-commitment equilibrium, candidates' ex-post policies are independent of the voter's beliefs, and confirmation bias thus plays no role in determining realized voter welfare.

It is also of interest to consider what might happen if candidates have mixed motives as in (19) *and* cannot commit to platforms. If platform announcements are non-binding, the voter anticipates that the elected candidate will choose the policy that is optimal given his bliss point. Therefore, the voter will, with probability 1, elect the candidate whose bliss point is closest to her own, or will randomize between candidates if they have symmetric bliss points. Once more we see that the voter's beliefs, and hence her degree of confirmation bias, are irrelevant for determining realized policy outcomes, and hence voter welfare.

Thus we see that some degree of policy commitment is essential for our analysis. Put differently, commitment is required for the electoral model we study to be of interest.

[16] We are grateful to an anonymous referee for suggesting this variation on the model to us.

4 Concluding remarks

We have presented a stylized model of the effect of voter confirmation bias on electoral competition, demonstrating that confirmation bias can counteract strategic distortions to candidates' electoral platforms, and hence increase equilibrium welfare. Indeed, we have shown that all observers, whether biased or Bayesian, would prefer the voters in our model to exhibit more confirmation bias than they do themselves. In addition, confirmation bias can reverse existing results on the ability of elections to aggregate candidates' private information. While electoral competition between two candidates gives rise to lower ex-ante welfare than if there were only a single candidate when voters are Bayesian, this need not hold when voters exhibit confirmation bias. These results are robust to arbitrary distributions of voters' ideological preferences, and also to a small amount of mixed motives for candidates.

Unlike previous work, our model does not rely on a tension between voters' (or candidates') ideologies and beliefs. The results hold even if voters have no ideological biases, and voters' and candidates' objectives are perfectly aligned. In addition, voters in our model are not prone to 'cognitive mistakes' in their interpretation of information; they simply update in a non-Bayesian manner. We also show that although information may be lost in equilibrium (voters in our model are no better off than they would have been if they had observed only one candidate's signal), this does not imply that equilibrium welfare is low. Indeed, confirmation biased voters may realise the Bayesian first best, despite the loss of information.

Our work contributes to an emerging group of papers suggesting that while deviations from rationality may be unambiguously undesirable for individuals acting in isolation, the implications of behavioural biases for society at large are more complex. These observations illustrate the theory of the second-best (Lipsey & Lancaster, 1956) in political contexts. Since politics invariably leads to non-optimal policy choices, decreasing the 'rationality' of political actors does not necessarily worsen outcomes, and may sometimes be welfare enhancing.

References

A. Alesina (1988). 'Credibility and policy convergence in a two-party system with rational voters'. *American Economic Review* **78**(4):796–805.

D. Austen-Smith & T. J. Feddersen (2009). 'Information aggregation and communication in committees'. *Philosophical Transactions of the Royal Society B* **364**:763–769.

A. Bisin, et al. (2015). 'Government policy with time inconsistent voters'. *The American Economic Review* **105**(6):1711–1737.

S. Callander & C. H. Wilson (2008). 'Context-dependent voting and political ambiguity'. *Journal of Public Economics* **92**(3):565–581.

A. Downs (1957). 'An economic theory of political action in a democracy'. *Journal of Political Economy* **65**(2):135–150.

L. G. Epstein (2006). 'An axiomatic model of non-Bayesian updating'. *The Review of Economic Studies* **73**(2):413–436.

G. Gratton (2014). 'Pandering and electoral competition'. *Games and Economic Behavior* **84**:163–179.

F. Gul & W. Pesendorfer (2001). 'Temptation and self-control'. *Econometrica* **69**(6):1403–1435.

F. Gul & W. Pesendorfer (2012). 'The War of Information'. *Review of Economic Studies* **79**(2):707–734.

P. Heidhues & J. Lagerlof (2003). 'Hiding Information in Electoral Competition'. *Games and Economic Behavior* **42**(1):48–74.

D. Kahneman & A. Tversky (1973). 'On the psychology of prediction'. *Psychological Review* **80**(237–251).

N. Kartik, et al. (2015). 'Information Revelation and Pandering in Elections'. *Working Paper* .

J.-F. Laslier & K. Van der Straeten (2004). 'Electoral competition under imperfect information'. *Economic Theory* **24**(2):419–446.

G. Levy (2004). 'Anti-herding and strategic consultation'. *European Economic Review* **48**:503–525.

G. Levy (2007). 'Decision making in committees: Transparency, reputation, and voting rules'. *American Economic Review* **97**(1):150–168.

G. Levy & R. Razin (2015). 'Correlation neglect, voting behavior, and information aggregation'. *The American Economic Review* **105**(4):1634–1645.

R. G. Lipsey & K. Lancaster (1956). 'The General Theory of Second Best'. *The Review of Economic Studies* **24**(1):11–32.

A. Lizzeri & L. Yariv (2017). 'Collective Self-Control'. *American Economic Journal: Microeconomics.* **9**(3):213–244.

B. Lockwood (2017). 'Confirmation Bias and Electoral Accountability'. *Quarterly Journal of Political Science* **11**(4):471–501.

E. Maskin & J. Tirole (2004). 'The Politician and the Judge: Accountability in Government'. *American Economic Review* **94**(4):1034–1054.

R. S. Nickerson (1998). 'Confirmation Bias: a ubiquitous phenomenon in many guises'. *Review of General Psychology* **2**(2):175–220.

F. Petry & B. Collette (2009). 'Measuring how political parties keep their promises: A positive perspective from political science'. In L. Imbeau (ed.), *Do They Walk Like They Talk?*, vol. 15 of *Studies in Public Choice*. Springer, New York, NY.

M. Rabin & J. L. Schrag (1999). 'First impressions matter: A model of confirmatory bias'. *The Quarterly Journal of Economics* **114**(1):37–82.

C. S. Taber & M. Lodge (2006). 'Motivated Skepticism in the Evaluation of Political Beliefs'. *American Journal of Political Science* **50**(3):755–769.

Appendix

A Relationship between 3 signal model and continuous model in KST

Although the statistical formalism that arises in our discrete model may be less familiar to some readers, the symmetric three signal model we use is a direct analogue of the continuous signal model in KST's baseline specification. In their model agents' beliefs about θ are given by a normal distribution with unknown mean μ, and prior beliefs about μ are also given by a normal distribution. The voter infers normally distributed signals drawn from the true distribution of θ, which allows her to update the mean and variance of her prior beliefs about μ. Each value of μ corresponds to a different probability distribution over θ, and the voter attempts to learn the value of μ (and hence the correct distribution over θ) from candidates' signals. Since μ is symmetrically distributed around the prior mean, the set of possible probability distributions for θ is also symmetric around the prior mean.

Exactly the same structure applies to our model. Each probability vector $\mathbf{p}$ is analogous to a probability distribution for θ *given* a value of μ in the continuous normal model of KST. In exact analogy, the agent doesn't know $\mathbf{p}$ *a priori*, and attempts to learn it from observed signals. Signals are drawn from the true distribution of θ (which is now discrete), allowing the agent to update her beliefs about $\mathbf{p}$. Again in analogy, the probability distributions $\mathbf{p}$ are symmetrically distributed around the peak of the prior, which is normalized to occur at zero in our model.

We have simplified to a 3 signal model as this facilitates the equilibrium analysis in a model with confirmation biased voters. The crucial feature of the KST model that ensures the existence of a pure strategy fully revealing equilibrium is the 'ex post' property – each candidate's equilibrium strategy would be a best response even if the other candidate's signal were known. This property is in turn a consequence of the fact that candidates' signals s_i, s_j enter the voter's posterior beliefs through the sum $s_i + s_j$. With confirmation bias and a continuum of signals however, the weights on signals in the voters' posterior will vary with their distance from the peak of the voters' prior, and so will be unequal in general. This destroys the ex post property, which in turn prohibits the existence of a pure strategy informative equilibrium. This difficulty does not arise with only three symmetric signals, allowing us to compute a pure strategy equilibrium and perform clean comparative statics on the degree of confirmation bias.

B Micro-foundation of confirmation-biased updating and welfare

This appendix provides a micro-foundation for the updating rule in (6), and the welfare measures in (16) via the model of non-Bayesian updating in Epstein (2006). In Epstein's model agents act as if they are subject to self-control problems: they are tempted to deviate from rational behaviour after information has been received, and sustain a cost for retaining a measure of self control. The tension between pursuing 'rational' objectives and not paying too high a self-control cost gives rise to a model that allows us to parameterise the degree of confirmation bias the voter exhibits. Another virtue of this framework for our purposes is that it allows us to define a measure of a confirmation biased agent's welfare that is analogous to standard Bayesian welfare measures.[17]

To get an intuitive understanding of Epstein's model it is best to work directly with his representation result.[18] Let $F(S)$ be a menu of acts (maps between states and outcomes), conditional on receiving a signal S about the state of the world σ. Denote a generic act by $f = f(\sigma) \in F(S)$. Let $P_B(\sigma|S)$ be the rational Bayesian update of the agent's prior beliefs about σ conditional on receiving signal S. We can think of the Bayesian rule as the updating procedure the agent considers 'correct' before she has received the signal S. In addition, let $P_T(\sigma|S)$ be an alternative 'non-rational' updating rule which the agent is *tempted* to use once S has been received. Epstein shows that agents who obey the behavioural axioms he stipulates will choose the act f that solves:

$$\max_{f\in F(S)} \left\{ \int U(f(\sigma))P_B(\sigma|S)d\sigma + \gamma(S) \left[\int U(f(\sigma))P_T(\sigma|S)d\sigma - \max_{f'\in F(S)} \int U(f'(\sigma))P_T(\sigma|S)d\sigma \right] \right\} \tag{21}$$

The interpretation of the components of this expression is as follows. The first term represents the agent's rational Bayesian expected utility if she were able to commit herself to actions ex-ante, i.e. before receiving the signal S. However the agent suffers from self-control problems: once S has been received, she is tempted to update her beliefs using the non-rational rule $P_T(\sigma|S)$. If the agent manages to partially avoid the temptation to use the non-rational rule, she pays a self-control cost, represented by the terms in square brackets. The cost of self-

[17]Epstein's model is in turn based on Gul & Pesendorfer's (2001) work on temptation and self control in intertemporal choice. A perhaps more widely known model of confirmation bias appeared in Rabin & Schrag (1999) (RS). We use Epstein's framework rather than RS for several reasons. First, an attractive feature of the former is that it introduces confirmation bias in a way that is consistent with the classical axioms of choice theory: Epstein simply adds an extra term to the classical utility function. By contrast, confirmation bias in RS involves a 'cognitive mistake': with some probability, the agent misinterprets one signal for another and then, next period, forgets that she has done so. Also, RS is less well suited to our electoral signaling game, as agents' beliefs update probabilistically in their setup. This complicates standard equilibrium concepts for signaling games, which normally rely on posterior beliefs being deterministic functions of observations (see Lockwood, 2017, for further discussion).

[18]Our presentation of this result is simplified for expositional purposes. In particular, we do not emphasize Epstein's axioms on state-contingent menus of acts, but rather work directly with his representation result for choices conditional on a given menu. We refer the reader to the original paper for further details.

control increases with the distance between her actual choice and the choice she would have made if she had given in to temptation. The agent's choices are thus a compromise between pursuing her ex-ante rational objectives, and avoiding large self-control costs ex-post. The cost of exhibiting self-control may be signal dependent, and is captured by the weight function $\gamma(S) \geq 0$.

To apply this preference representation in our model we need to specialise it further. In a fully revealing equilibrium the voter will be able to infer the private signals of both candidates from their platform choices. Thus, the signals the voter infers are of the form $S_{ij} = (s_i, s_j)$ where $s_i, s_j \in \Theta$. Each observed signal pair S_{ij} allows the voter to update her beliefs about the probability vector $\mathbf{p}$, which determines the likely values of the state of the world θ. To model confirmation bias we make a specific choice for the 'temptation' part of the voter's preferences (i.e., the second term in (21)). We assume that after inferring a signal S_{ij}, the voter is tempted to deploy an updating rule of the following form:

$$P_T(\mathbf{p}|S_{ij}) = (1 - \lambda(S_{ij}))P_B(\mathbf{p}|S_{ij}) + \lambda(S_{ij})g(\mathbf{p}; \boldsymbol{\alpha}) \tag{22}$$

where $g(\mathbf{p}; \boldsymbol{\alpha})$ is the voter's prior beliefs about $\mathbf{p}$ given in (3), and $\lambda(S_{ij})$ is a signal-dependent weight defined below.[19] By assumption, the agent's prior beliefs place more weight on the central value $\theta = 0$ than the extreme values $\theta = \pm 1$ (i.e. $\alpha_0 > \alpha_1$). Thus, if she infers a signal $(0, 0)$, this confirms her prior expectations. Any other signal, however, diverges from her prior beliefs. A voter who exhibits confirmation bias will treat signals that confirm her prior beliefs differently from those that do not. We adopt the simplest possible choice of the weight function $\lambda(S_{ij})$ that achieves this:

$$\lambda(S_{ij}) = \begin{cases} 0 & \text{if } S_{ij} = (0,0) \\ \bar{\lambda} & \text{otherwise} \end{cases} \tag{23}$$

where $\bar{\lambda} \in (0, 1)$ is a parameter. Substituting these choices into the representation (21), and assuming that the cost of self control function $\gamma(S)$ is a constant $\bar{\gamma}$, it is easily seen that the voter will act as if she updates her beliefs according to the compromise rule:

$$P^*(\mathbf{p}|S_{ij}) = \left(1 - \frac{\bar{\gamma}\lambda(S_{ij})}{1 + \bar{\gamma}}\right) P_B(\mathbf{p}|S_{ij}) + \frac{\bar{\gamma}\lambda(S_{ij})}{1 + \bar{\gamma}} g(\mathbf{p}; \boldsymbol{\alpha}). \tag{24}$$

Defining a new parameter $q \equiv \frac{\bar{\gamma}\bar{\lambda}}{1+\bar{\gamma}} \in [0, 1)$, this updating rule can be written simply as:

$$P_q^*(\mathbf{p}|S_{ij}) = \begin{cases} P_B(\mathbf{p}|S_{ij}) & \text{if } S_{ij} = (0,0) \\ (1-q)P_B(\mathbf{p}|S_{ij}) + qg(\mathbf{p}; \boldsymbol{\alpha}) & \text{otherwise.} \end{cases} \tag{25}$$

[19]An axiomatic justification for this choice for $P_T(\mathbf{p}|S_{ij})$ is provided by Epstein (2006).

This is the updating rule in (6).

To define a confirmation-biased welfare measure that is consistent with this framework, recall that agents in the Epstein model are only tempted to deviate from rational Bayesian updating *after* they receive a signal that conflicts with their priors. Their *a priori* beliefs about which signals they are likely to receive are thus unaffected by confirmation bias, but once a signal has been received it affects their conditional expectations about the state of the world, and about which additional signals might be observed. These observations, coupled with equations (1), (21), and (24), lead us to conclude that ex-ante voter welfare in the equilibrium of Proposition 1, according to an observer who has confirmation bias parameter q_w, is:

$$W_{q_w} = -\sum_{s_i \in \Theta} \rho(s_i) \sum_{s_j \in \Theta} \rho_{q_w}(s_j|s_i) \left(\frac{1}{2} \sum_{t \in \{s_i, s_j\}} E_{q_w} \left[(y_q(t) - \theta)^2 | (s_i, s_j) \right] \right). \tag{26}$$

Here the expectation E_{q_w} is now taken with respect to the observer's biased posterior beliefs, given by (6) with $q = q_w$, $\rho_{q_w}(s_j|s_i)$ is the confirmation-biased posterior predictive distribution given in (17), and $y_q(\cdot)$ is candidates' equilibrium strategy when the voter has bias q.

C Proof of Proposition 1 (existence)

To prove the proposition, we need to construct a belief system for the median voter. For any platform $y \in [-1, 1]$ announced by candidate i, a belief system specifies a probability distribution over the set $\{-1, 0, 1\}$ of signals that i may have privately observed. The equilibrium requirement is as follows: suppose that candidate i announces a platform y; assume that the voter updates her beliefs having observed y, and then votes for her preferred candidate; then the payoff that i obtains cannot strictly exceed the payoff he would have obtained if he had played the strategy prescribed in Proposition 1. The only restriction on beliefs required by the Perfect Confirmation Biased Equilibrium criterion is that when the voter observes an "on-the-equilibrium-path" platform—i.e., observes $y \in [-\bar{y}, 0, \bar{y}]$—her beliefs are determined by applying the updating rule in (6). We impose an additional requirement which, though not required, is extremely natural: voter's beliefs must be continuous in the platform announced by player i.

For $\bar{y}$ as specified in equation (10) of Proposition 1, for $\delta \in [0, \bar{y})$, and $\eta(\delta) = (\delta + \bar{y})/2 \in (\delta, 1)$, the following belief system supports (10) as a SPCBE: for $i \in \{A, B\}$

$$\text{if } i \text{ plays } \begin{cases} y' \leq -\bar{y} \\ -\bar{y}+\delta, \\ \bar{y}-\delta, \\ y' > \bar{y} \end{cases} \text{ voter assign beliefs } \begin{cases} (1,0,0) \\ (1-\frac{\eta(\delta)}{\bar{y}}, \frac{\eta(\delta)}{\bar{y}}, 0) \\ (0, \frac{\eta(\delta)}{\bar{y}}, 1-\frac{\eta(\delta)}{\bar{y}}) \\ (0,0,1) \end{cases} \text{ to } i\text{'s signal vector } (-1,0,1). \quad (27)$$

Using these beliefs we can extend (in the obvious way) the specification (8) of $\omega_i^y(\boldsymbol{\zeta}; q)$, which is the probability that the voter chooses candidate i having observed the platform pair $\boldsymbol{\zeta}$. We can then define inferred probabilities over signals following any unilateral deviation from the candidate equilibrium strategies $\bar{y}(\cdot)$. To verify that the strategy profile (10) is a best response to itself it is straightforward to check that for any signal pair the two platforms defined by $\bar{y}$ are equidistant from the voter's posterior bliss point conditional on that pair. This follows from the definition of $\bar{y}$ and the expressions in (7) for the voter's posterior expectations of θ.

We now show that candidate i cannot do better by deviating to any off-equilibrium-path platform in $[-1, 1]$. We begin by explaining the structure of the argument. It is straightforward to check that if i announces a platform outside the interval $[-\bar{y}, \bar{y}]$, he loses the election with probability 1, regardless of the voter's beliefs. Now suppose that i deviates to the platform $-\bar{y} + \delta$, for $\delta \in (0, \bar{y})$. Having observed this platform, the voter believes (see (27)) that i's signal was -1 with some probability, and 0 with the remaining probability. She simultaneously observes the (on-equilibrium-path) platform selected by j. She thus has to make three choices, one for each realization of j's platform. We will show below that in each case, she would prefer to vote for i rather than j, given one of the possible realizations of i's signal, and for j rather than i given the other realization in the support of her conditional belief. The belief system (27) is specified so that when j's platform is $-\bar{y}$, on balance she prefers to vote for i, while when j's platform is either 0 or $\bar{y}$, on balance she prefers to vote for j. But from (5), the ex ante probability that j will announce $-\bar{y}$, and hence the ex ante probability that i will win the election, is $\frac{\alpha_{-1}}{2\alpha_{-1}+\alpha_0} < 1/3$. Therefore i strictly prefers to play his equilibrium strategy, and win with probability 1/2.

We now consider each of the possible cases in detail. If:

1) i plays $y' < -\bar{y}$; the voter believes that i's signal was -1 with probability 1. For each possible realization of s_j, it follows immediately from (7) that y' is further than $y_j(s_j)$ from the voter's ideal point $E[\theta|(-1, s_j)]$; thus i loses the election for any signal realization by candidate j.

2) i plays $-\bar{y} + \delta$, for $\delta \in (0, \bar{y})$; the voter assigns beliefs $\left(1 - \frac{\eta(\delta)}{\bar{y}}, \frac{\eta(\delta)}{\bar{y}}, 0\right)$ to the possibilities that $s_i = (-1, 0, 1)$. We will show that with these beliefs, i wins the election with probability $< 1/3$.

a) suppose $s_j = -1$. The voter has to consider two possibilities: if $s_i = -1$, then $E[\theta|(s_i, s_j)] = -\bar{y}$, so she receives utility $-\delta^2$ if she votes for i and zero if she votes for j; if $s_i = 0$, then $E[\theta|(s_i, s_j)] = -\bar{y}/2$, so she receives utility $-(\bar{y}/2 - \delta)^2$ if she votes for i and $-\bar{y}^2/4$ if she votes for j. Her expected utility difference between voting for i and j is thus

$$
\begin{aligned}
& -\left(1 - \frac{\eta(\delta)}{\bar{y}}\right)\delta^2 \quad + \quad \frac{\eta(\delta)}{\bar{y}}\Big(- (\bar{y}/2 - \delta)^2 - (-\bar{y}^2/4)\Big) \\
= \quad & -\delta^2 + \frac{\eta(\delta)\delta^2}{\bar{y}} \quad + \quad \frac{\eta(\delta)}{\bar{y}}\Big(\delta\bar{y} - \delta^2\Big) \quad = \quad \delta(\eta(\delta) - \delta) \quad > \quad 0;
\end{aligned}
$$

so that the voter votes for i

b) suppose $s_j = 0$. If $s_i = -1$, then $E[\theta|(s_i, s_j)] = -\bar{y}/2$, so the voter receives utility $-(\bar{y}/2-\delta)^2$ if she votes for i and $-\bar{y}^2/4$ if she votes for j; if $s_i = 0$, then $E[\theta|(s_i, s_j)] = 0$, so she receives utility $-(\bar{y} - \delta)^2$ if she votes for i and 0 if she votes for j. The expected utility difference between voting for i and j is thus

$$
\begin{aligned}
& \left(1 - \frac{\eta(\delta)}{\bar{y}}\right)\left(-(\bar{y}/2-\delta)^2 - \bar{(} - y^2)/4\right) \quad - \quad \frac{\eta(\delta)}{\bar{y}}\Big(\bar{y} - \delta\Big)^2 \\
= \quad & \left(y\delta/2 \ + \ \delta^2/2 \ - \ y^2/2 \ - \ \frac{\delta^3}{2y}\right) \quad + \quad \left(\frac{\delta^3}{2y} \ - \ \delta^2 \ + \ \delta y/2\right) \\
= \quad & -\Big(\delta - \bar{y}\Big)^2 \quad < \quad 0
\end{aligned}
$$

so that the voter votes for j

c) suppose $s_j = +1$. If $s_i = -1$, then $E[\theta|(s_i, s_j)] = 0$; so the voter receives utility $-(\bar{y}-\delta)^2$ if she votes for i and $-\bar{y}^2$ if she votes for j; if $s_i = 0$, then $E[\theta|(s_i, s_j)] = +\bar{y}/2$, so she receives utility $-(3\bar{y}/2 - \delta)^2$ if she votes for i and $-\bar{y}^2/4$ if she votes for j. The expected utility difference between voting for i and j is thus

$$
\begin{aligned}
& \left(1 - \frac{\eta(\delta)}{\bar{y}}\right)\Big(- (\bar{y} - \delta)^2 \ - \ (-\bar{y}^2)\Big) \quad + \quad \frac{\eta(\delta)}{\bar{y}}\Big(- (3\bar{y}/2 - \delta)^2 \ - \ (-\bar{y}^2/4)\Big) \\
= \quad & \delta\left\{\left(\frac{\delta^2}{2\bar{y}} \ - \ 3\delta/2 \ + \ \bar{y}\right) \quad + \quad \left(\bar{y}/2 \ + \ \delta \ - \ \bar{y}^2/\delta \ - \ \frac{\delta^2}{2\bar{y}}\right)\right\} \\
= \quad & -(\bar{y} - \delta)(2\bar{y} - \delta)/2 \quad < \quad 0
\end{aligned}
$$

Once again, the voter votes for j.

Thus, if i plays $-\bar{y} + \delta$, for $\delta \in (0, \bar{y})$, she wins the election with probability $\frac{\alpha_1}{2\alpha_1 + \alpha_0} < 1/3$.

3) i plays $\bar{y} - \delta$, for $\delta \in (0, \bar{y})$; the voter assigns beliefs $\left(0, \frac{\eta(\delta)}{\bar{y}}, 1 - \frac{\eta(\delta)}{\bar{y}}\right)$ to the possibilities that $s_i = (-1, 0, 1)$. The proof that i wins the election with probablity $< 1/3$ is the mirror

image of the proof step 2).

4) i plays $y' > \bar{y}$; the voter believes that with probability 1, i's signal was $+1$. The proof that i loses the election, for any signal realization by j, is the mirror image of the proof step 1).

D Proof of Proposition 1 (uniqueness)

Let $y(\cdot)$ be a fully revealing SPCBE.

1) All type-pairings result in a tied election: Clearly (A) if $s_i = s_j$, then the election result is a tie, because i and j play identical strategies. Second, observe that (B) if candidate j of type 0 is paired with a candidate of type $s_i \neq 0$ then the result must be a tie. To see this, suppose (w.l.o.g) that $s_i = -1$.

   a) *if j wins against i:* Then condition (9) fails for $s_i = -1$ and $\zeta = y_i(0)$, since:

$$\sum_{\Theta} \rho(s_j|-1)\omega_i^y(y(-1), y(s_j); q) \quad \leq \quad \frac{1}{1+\sum_\theta \alpha_\theta}\Big((\alpha_{-1}+1)/2 \quad + \quad \alpha_{+1}\Big) \quad = \quad \frac{3\alpha_{-1}+1}{2(1+\sum_\theta \alpha_\theta)}$$

   since $y(-1)$ ties against $y(-1)$, loses against $y(0)$ and, at best, wins against $y(1)$, while

$$\sum_{\Theta} \rho(s_j|-1)\omega_i^y(y(0), y(s_j); q) \quad \leq \quad \frac{1}{1+\sum_\theta \alpha_\theta}\Big((\alpha_{-1}+1) \quad + \quad \alpha_0/2\Big) \quad = \quad \frac{2\alpha_{-1}+\alpha_0+2}{2(1+\sum_\theta \alpha_\theta)}$$

   since $y(0)$ wins against $y(-1)$, ties against $y(0)$ and, at worst, loses against $y(1)$. Since $\alpha_0 > \alpha_{-1}$, this shows $y(0)$ is a profitable deviation for i from $y(-1)$ conditional on $s_i = -1$.

   b) *if j loses against i:* Then condition (9) fails for $s_j = 0$ and $\zeta = y(-1)$, since:

$$\sum_{\Theta} \rho(s_i|0)\omega_j^y(y(0), y(s_i); q) \quad \leq \quad \frac{1}{1+\sum_\theta \alpha_\theta}\Big((\alpha_0+1)/2 \quad + \quad \alpha_{+1}\Big) \quad = \quad \frac{\alpha_0+1+2\alpha_{-1}}{2(1+\sum_\theta \alpha_\theta)}$$

   since $y(0)$ loses against $y(-1)$, ties against $y(0)$ and, at best, wins against $y(1)$, while

$$\sum_{\Theta} \rho(s_i|0)\omega_j^y(y(-1), y(s_i); q) \quad \leq \quad \frac{1}{1+\sum_\theta \alpha_\theta}\Big((\alpha_{-1}+1)/2 \quad + \quad \alpha_0\Big) \quad = \quad \frac{\alpha_{-1}+1+2\alpha_0}{2(1+\sum_\theta \alpha_\theta)}$$

   since $y(-1)$ ties against $y(-1)$ wins against $y(0)$, and, at worst, loses against $y(1)$. This shows $y(-1)$ is a profitable deviation for j from $y(0)$ conditional on $s_j = 0$.

   It now follows from (B) that if candidate i of type -1 is paired with candidate j of type $+1$, then the election result must be a tie. To see this, suppose w.l.o.g. that to the contrary

c) *i loses against j:* Then condition (9) fails for $s_i = -1$ and $\zeta = y(+1)$, since:

$$\sum_{\Theta} \rho(s_j|-1)\omega_i^y(y(-1), y(s_j); q) \quad = \quad \frac{1}{1+\sum_\theta \alpha_\theta}\Big((\alpha_{-1}+1)/2 \quad + \quad \alpha_0/2\Big) \quad = \quad \frac{\alpha_{-1}+\alpha_0+1}{2(1+\sum_\theta \alpha_\theta)}$$

since $y(-1)$ ties against $y(-1)$ and $y(0)$, and loses against $y(+1)$, while

$$\sum_{\Theta} \rho(s_j|-1)\omega_i^y(y(+1), y(s_j); q) \quad = \quad \frac{1}{1+\sum_\theta \alpha_\theta}\Big(3(\alpha_{-1}+1)/2 \quad + \quad \alpha_0/2\Big) \quad = \quad \frac{3\alpha_{-1}+\alpha_0+1}{2(1+\sum_\theta \alpha_\theta)}$$

This shows $y(+1)$ is a profitable deviation for i from $y(-1)$ conditional on s_i=-1.

2) The unique vector of platforms s.t. all type pairings result in ties is (10):
Assume that $y(\cdot)$ satisfies this property. The following properties are all implications of the calculations (7).

i) $E[\theta|(-1,+1)] = 0$ implies that $y(-1) + y(+1) = 0$.
ii) property i), together with $|E[\theta|(-1,0)]| = |E[\theta|(0,+1)]|$, implies $y(0) = 0$.
iii) $y(0) = 0$ together with $E[\theta|(-1,0)] = \frac{-(1-q)}{2\alpha_1+\alpha_0+2}$ implies $y(-1) = \frac{-2(1-q)}{2\alpha_1+\alpha_0+2}$.
iv) $y(0) = 0$ together with $E[\theta|(0,+1)] = \frac{1-q}{2\alpha_1+\alpha_0+2}$ implies $y(+1) = \frac{2(1-q)}{2\alpha_1+\alpha_0+2}$.

E Proof of Proposition 4

Consider the ex-ante welfare of a voter with bliss point b in the equilibrium in which candidates play the strategy

$$\tilde{y}_q(s_i) = b_m + y_q(s_i).$$

According to an observer with bias parameter q_w, this voter's welfare is:

$$\tilde{W}_{q_w}(b) = -\sum_{s_i\in\Theta}\rho(s_i)\sum_{s_j\in\Theta}\rho_{q_w}(s_j|s_i)\left(\frac{1}{2}\sum_{t\in\{s_i,s_j\}}E_{q_w}\left[(\tilde{y}_q(t)-(\theta+b))^2|(s_i,s_j)\right]\right).$$

Consider the term

$$\begin{aligned}E_{q_w}\left[(\tilde{y}_q(t)-(\theta+b))^2|(s_i,s_j)\right] &= E_{q_w}\left[((y_q(t)-\theta)+(b_m-b))^2|(s_i,s_j)\right]\\ &= E_{q_w}\left[(y_q(t)-\theta)^2|(s_i,s_j)\right]+(b_m-b)^2+2(b_m-b)E_{q_w}\left[(y_q(t)-\theta)|(s_i,s_j)\right].\end{aligned}$$

From this we see that

$$\tilde{W}_{q_w}(b) = W_{q_w} + (b_m-b)^2 + 2(b_m-b)\sum_{s_i\in\Theta}\rho(s_i)\sum_{s_j\in\Theta}\rho_{q_w}(s_j|s_i)\left(\frac{1}{2}\sum_{t\in\{s_i,s_j\}}\left(E_{q_w}\left[((y_q(t)-\theta)|(s_i,s_j)]\right)\right)\right). \tag{28}$$

The last term in this expression is identically zero. To see this, note that

$$y_q(s_i) = E_q[\theta|(s_i, s_i)]$$

and by symmetry

$$\sum_{s_i \in \Theta} \rho(s_i) \sum_{s_j \in \Theta} \rho_{q_w}(s_j|s_i) E_q[\theta|(s_i, s_i)] = 0.$$

Similarly, by symmetry,

$$\sum_{s_i \in \Theta} \rho(s_i) \sum_{s_j \in \Theta} \rho_{q_w}(s_j|s_i) E_{q_w}[\theta|(s_i, s_j)] = 0.$$

Thus, both the expectations that make up the last term in (28) are zero. Hence, we conclude that

$$\tilde{W}_{q_w}(b) = W_{q_w} + (b_m - b)^2$$

The voters' total ex-ante welfare is

$$\int_{-\bar{b}}^{\bar{b}} \tilde{W}_{q_w}(b) G'(b) db = W_{q_w} + \int_{-\bar{b}}^{\bar{b}} (b_m - b)^2 G'(b) db.$$

Since the confirmation bias parameters q and q_w only enter this expression through the first term W_{q_w}, our welfare results are robust for arbitrary distributions of voter bliss points.