WARWICK

ECONOMICS

QAPEC

Quantitative and Analytical Political Economy Research Centre

Discussion Paper Series

Trust in Political Institutions and Electoral Accountability: A Political Agency Approach

Ben Lockwood and Minh Tung Le

This paper also appears as Department working paper 1620

Date: July 26

No: 30

QAPEC
Quantitative and Analytical Political Economy Research Centre

Department of Economics
University of Warwick, Coventry,
CV4 7AL, United Kingdom

https://warwick.ac.uk/fac/soc/economics/research/centres/qapec/discussionpapers

Trust in Political Institutions and Electoral Accountability: A Political Agency Approach*

Ben Lockwood[†] Minh Tung Le[‡]

July 15, 2026

Abstract

We analyze the implications of three stylized facts about trust in political institutions for the role of elections in improving discipline, constraining pandering, and improving selection of politicians. First, trust in political institutions (politicians, political parties, and the legislature) has declined in recent years. Second, within countries, there is heterogeneity across citizens in their level of trust in institutions, and third, at the individual level, trust is malleable, especially when young. We add these features to an otherwise standard model of electoral agency. Unlike the standard model, where the quality of the candidate pool is assumed known, citizens are born with imprecise and heterogeneous priors over the quality of the pool, so young citizens learn not only about the incumbent, but also about the pool quality. An increase in precision increases both the discipline and selection effects of elections, whereas an increase in heterogeneity of prior beliefs has the opposite effect. The effect of changes in average trust in the population is generally ambiguous. When candidate entry is modelled, these findings extend to endogenous pool quality. The implications for citizen welfare and civics education are discussed.

JEL Classification: H11, C72

Keywords: elections, discipline, pandering, selection, trust in political institutions, civic education.

*We would like to thanks Helios Herrera and Federico Trombetta for helpful comments.

[†]b.lockwood@warwick.ac.uk, University of Warwick and CEPR.

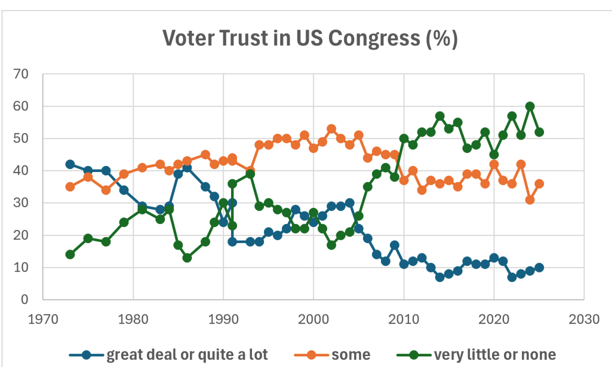
[‡]minh.t.le@warwick.ac.uk, University of Warwick.

1 Introduction

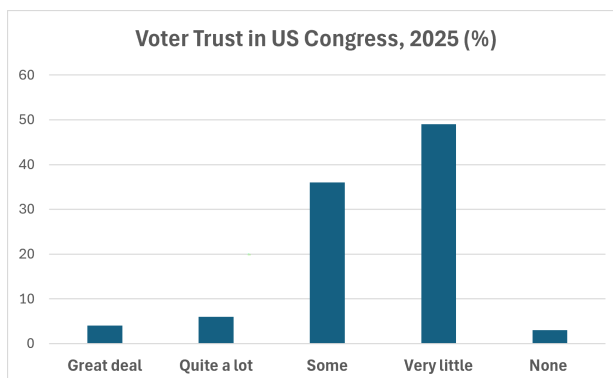
Trust in political institutions, commonly measured as voter confidence in government, parliament or Congress, political parties, or politicians, is widely regarded as a core indicator of democratic legitimacy. The starting points for this paper are three stylized facts about trust in politicians and political institutions. The first and most familiar, is that in many countries, trust is now low and has been declining in recent years (Valgarðsson et al., 2025). One possibility is that this is due to decreases in the "quality" of the political class, as measured by e.g. competence or honesty. Interestingly, these sharp declines in trust have occurred despite no hard evidence that the quality of politicians has declined, indicating a developing mismatch between voter beliefs and reality.[1] A good example is in the US, where voter trust in Congress has declined sharply over the last fifty years (Figure 1a), at the same time as the percentage of members of Congress educated at college level has increased, and prosecutions and expulsions of members for corruption remain very rare.[2]

The second is that at any point in time, in any country, there is considerable variation across the electorate in their level of trust in politicians, political parties, and the legislature, as shown for the case of the US Congress in Figure 1b. Factors such as income, education, or ethnicity often explain differences in trust *between* individuals, while changes in trust *within* individuals are more strongly linked to political and economic experiences e.g. unemployment, economic growth, changes in government, and even epidemics (Besley, Dann and Dray, 2026; Eichengreen, Saka and Aksoy, 2024; Haugsgjerd and Kumlin, 2020). However, panel data studies suggest that changes in trust within individuals in response to these factors are relatively transitory (Devine et al., 2024).

Figure 1: Voter Trust in US Congress



(a) Trust in US Congress over time



(b) Trust in US Congress, 2025

Source: Gallup

The third, which is perhaps less appreciated, is that beliefs about government are more malleable

[1] Additional evidence for this is Dal Bó et al. (2017), who find that politicians are on average significantly smarter and better leaders than the population they represent.

[2] On college education, see https://www.pewresearch.org/?p=8145, and on corruption, see https://www.bbc.co.uk/news/world-us-canada-67458673

during adolescence and early adulthood, where changes in trust are large.[3] For example, Devine et al. (2024) find, using the US Youth Panel, that average trust in government differs significantly between adolescence and adulthood.[4] Similarly, Eichengreen, Saka and Aksoy (2024) show that the negative impact of epidemics on political trust is much larger for the 18-25 age group.

In this paper, we analyse an otherwise standard model of political agency that captures these three stylised facts, namely (i) declining trust, even when the quality of political institutions may not have changed; (ii) heterogeneity in levels of trust across the electorate, and (iii) learning by young citizens about the quality of political institutions. Of course, there is a large political agency literature in economics that allows citizens to learn about the ability of *particular incumbent politicians* (Besley, 2006; Canes-Wrone, Herron and Shotts, 2001; Maskin and Tirole, 2004). However, it is very important to note that the empirical literature surveyed above relates to survey questions about trust in politicians or institutions *in general*, not about an incumbent politician.

In order to stay as close as possible to the existing political agency literature, we model trust in political institutions as beliefs about the "pool" of possible politicians. In a standard political agency model, the quality of the candidate pool, i.e. the probability that a candidate randomly drawn from this pool is "good" is common knowledge. By contrast, we allow citizens to have *incorrect, heterogeneous and imprecise beliefs* about this probability. This is the key contribution of our paper.

Specifically, we work with an infinite-horizon model of political competition, where overlapping generations of citizens live for two periods, there is an election at the beginning of every period. Young citizens do not vote but they observe the action of the incumbent, and then update their priors on the quality of the incumbent (if not term-limited) and the pool of challenger politicians.[5] They then vote in the election at the beginning of the second period of life. Thus, citizens learn how much to trust politicians early in life, consistently with the empirical evidence.

Otherwise, the model is standard. There are binary states 0 and 1, and actions 0 and 1, where citizens prefer the action to match the state in each period. The bad politician prefers action 0 in both states, whereas the good politician has the same preferences as the citizens. An interpretation is that the bad politician is ideologically committed to a particular action, which may or may not be optimal.[6] However, both types also have a non-policy payoff from being in office (office rent),

[3] This fact is consistent with the trend that over recent years, many education systems across the OECD have increasingly prioritised civic education at secondary school level to promote democratic citizenship in response to falling political participation rates (Borhan, 2025).

[4] The USYP is unique in that its first wave was conducted when all but three respondents were 18 or 19 years old, following up with them 17 and 32 years later. The mean change is much larger for those who are young in their first wave (i.e., under 16, 20 or 25) than for those who are over those ages, and this is especially true for those under 16 (Devine et al., 2024).

[5] Politicians are limited to a maximum of two terms of office, so some elections will be between two "new" politicians drawn from the pool.

[6] An alternative leading model of pandering, Maskin and Tirole (2004), models the bad politician as having "non-congruent" preferences i.e. prefers the opposite action to the optimal one in both states. Following much of the more recent literature e.g. Trombetta (2020), we work with the Besley model, as in many

which in equilibrium, creates an incentive for the bad incumbent to be *disciplined* in state 1 by choosing action 1, and simultaneously, an incentive for both types of politician to *pander* in state 0 by choosing action 1.

We model trust as follows. Specifically, the true quality of the candidate pool is denoted by π, the probability that a randomly drawn politician from the pool is a good type. We suppose that there are n groups of citizens in the population, and that a young citizen of group $i = 1, \ldots, n$ has a prior belief about π that follows a continuous distribution with mean τ_i and strictly positive variance.[7] The mean τ_i can be lower than π, capturing low trust. The variation of τ_i across groups captures the heterogeneity of beliefs (at a point in time) observed in survey data. The fact that the prior belief has a positive variance means that the young citizens will update their belief about π, the quality of the pool, as well as the incumbent, when they observe the action of the incumbent.[8]

Our key findings are as follows. First, the precision of beliefs is important for the structure of the equilibria. When precision of beliefs is high, there is a unique equilibrium which can involve discipline/pandering with any probability, depending on parameters. When precision of beliefs is low, there are only corner equilibria, where discipline/pandering occurs with either probability zero or one. For some parameters, both these equilibria occur at once; so, low precision gives rise to multiple equilibria.

Next, we show that looking across both types of equilibria, both precision and heterogeneity of beliefs have clear effects on discipline/pandering. First, an increase in precision of beliefs always increases discipline/pandering. The intuition is the following. In our setting, the action chosen by the incumbent is informative to the citizens not only about the incumbent himself, but also about the quality of the pool from which the challenger is drawn. This second effect, in equilibrium, offsets the first, but is smaller, the more precise are citizens' beliefs about the quality of the pool.

Second, heterogeneity of beliefs, modelled as a mean-preserving spread in the τ_i, always reduces discipline/pandering. This is due to the fact that Bayesian updating of beliefs implies that the electoral reward from choosing action 1 over action 0 is a concave function of the prior mean beliefs τ_i, so a mean-preserving spread in these beliefs mechanically reduces this electoral reward. Third, in contrast to these clear results, an increase in the mean level of trust in the population (the mean of the τ_i) unambiguously increases discipline only in the special case where there is no incentive to pander for the incumbent.

We also consider the effects of these changes on the role of elections as selection devices. Our measure of selection is a standard one; the probability that the incumbent is good, given that he wins re-election. Here, we show that when there is a fully separating equilibrium, the effectiveness

---

settings, it seems more realistic.

[7] We choose a Beta distribution because it is the conjugate prior for the Bernoulli distribution, and signals of candidate quality are Bernoulli.

[8] For simplicity, we model the payoff to both good and bad politicians from taking their short-run optimal action as random draws from the same distribution, which implies that in any equilibrium, the probability that the bad politician is disciplined is the same as the probability that the good incumbent panders - these behaviors are two sides of the same coin.

of selection is strictly increasing in precision, and decreasing in heterogeneity.

We then extend the model to allow for candidate entry, using the contest approach pioneered by Klumpp and Polborn (2006), which endogenizes π, the quality of the candidate pool. Here, again the same pattern emerges: the quality of the pool is increasing in precision, and decreasing in heterogeneity. Finally, we consider citizen welfare. Simulation results show a similar picture, whether or not candidate entry is modelled; if the discipline effect of elections dominates the pandering effect, an increase in precision or a decrease in heterogeneity increases welfare, whereas the reverse is true if the pandering effect dominates. We conclude by interpreting the effect of civic education as both increasing mean trust and the precision of beliefs, and discuss the implications for welfare.

2 Related Literature

To our knowledge, there are just two related papers that model trust in politicians in a similar way to this paper. The first is Bracco et al. (2021), which considers a simplified version of our model where there is a single voter (no heterogeneity) who holds a precise belief about the quality of the pool of candidates (no learning about the pool). However, this belief can be biased i.e. $\tau_i \neq \pi$. Proposition 1 in that paper shows that the higher is τ_i, the more disciplined is the bad politician; Proposition 6 in our paper extends and qualifies this result, by showing that it generally does not hold in equilibria with pandering.[9]

The second is Herrera and Trombetta (2024), which develops a dynamic political agency model in which voter trust evolves endogenously and shapes the adoption of populist worldviews. However, there are a number of key differences between our paper and theirs. First, in theirs, there is a second dimension to the politicians' type; as well as being good or bad, they may be populist or traditional.[10] Second, trust in politicians changes not by learning within a generation, but by transmission of beliefs from each one-period lived generation to the next. Third, in line with these two features, the main focus of the paper is on the dynamics of trust and the alternation of power between traditional and populist politicians.[11] The only result where there is overlap is their Proposition 1, which identifies conditions under which the bad-type politician is undisciplined as a function of the initial level of trust of the voter.[12] This relates to our results on the effect of average trust on discipline and pandering in Propositions 5 and 6.[13] By construction, their model does not allow for heterogeneous trust between groups or imprecision of prior beliefs, which are two key features of our model. Finally, their set-up does not generate pandering in equilibrium, as the good

[9] This paper now supersedes the theoretical section in Bracco et al. (2021): a new revised version of the latter with the empirical results only is Bracco et al. (2026).

[10] In their setting, a traditional politician knows the true probability of the state but the populist mistakenly believes that the "world is simple" i.e. only one state is possible.

[11] See e.g. Proposition 3 of Herrera and Trombetta (2024) on trust traps.

[12] In Section 6.2, Herrera and Trombetta (2024) study entry by a populist party, but this does not affect the probability in their model that the traditional politician is good, which remains fixed at γ in their model.

[13] Specifically, their finding that discipline can be non-monotonic in trust is similar to our Proposition 5.

politicians are assumed to act non-strategically. So, overall, we view our paper as complementary to theirs.

Other related literatures are as follows. One related literature concerns the effects of voter information on electoral accountability (Besley, 2006; Ashworth and Bueno de Mesquita, 2014). A common theme in the literature is that improved voter information may not raise voter welfare, as it impacts the discipline or selection effects of elections in different ways. For example, both Besley (2006) and Ashworth and Bueno de Mesquita (2014), in somewhat different settings, show that better information about the incumbent's type either by observing it directly (Besley, 2006), or observing the action, not just the payoff (Ashworth and Bueno de Mesquita, 2014) increases electoral discipline, but reduces the selection effect, and can lower voter welfare overall. However, all of this literature assumes that the quality of the candidate pool (i.e. the ex ante probability that a politician is good or bad) is fully known by the voters.

A more recent variant of this literature endogenises the sources of voter information by explicitly modelling the media (Wolton, 2019; Li, Raiha and Shotts, 2022). Again, the focus is on the welfare effects of biased media which can be counter-intuitive. However, these papers also rule out both heterogeneity in beliefs and learning by voters about the candidate pool; for example, Wolton (2019) assumes a single voter who knows the quality of the pool of three different types.

Another relevant literature concerns the effects of increased polarization in voter ideology scale on electoral accountability in political agency settings (Schultz, 1996; Testa, 2012). Our results are somewhat related; in particular, an increase in heterogeneity in trust, modelled as a mean-preserving spread, is of course an increase in polarization, but of *trust*, not ideology. Our finding that this can sometimes increase welfare is related to the findings of Testa (2012).

Finally, our modelling of candidate entry relates to a literature on the effect of politician wages on entry in a citizen-candidate framework (Besley, 2004; Messner and Polborn, 2004; Caselli and Morelli, 2004) and also a literature on US-style primary contests (Klumpp and Polborn, 2006; Barbieri and Serena, 2025). We combine these two literatures by modeling the candidate selection procedure as a Tullock contest.[14]

3 The Model

The basic structure of the model is the following. There are an infinite number of periods $t = 1, 2, \dots$ and overlapping generations of citizens, who learn in the first period of life and vote in the second. At the beginning of each period t, there is an election, then a state $s_t \in \{0, 1\}$ is realised, and then an incumbent politician chooses an action $a_t \in \{0, 1\}$. We assume that the state of the world

[14] We extend these papers in the following ways. First, Klumpp and Polborn (2006) and Barbieri and Serena (2025) only study primary, not general elections. Messner and Polborn (2004) and Caselli and Morelli (2004) do not study the agency problem between politicians and voters once elected. Besley (2004) is the closest to our contribution but he assumes random selection of the candidate from the pool of candidates willing to serve, with this pool also being infinitely large, so no candidate has a strict incentive to stand.

is independent and identically distributed over time with $P(s_t = 1) = p$. Further details are as follows.

3.1 Citizens

Each generation of citizens is composed of a large number of individuals, formally a continuum $\mathcal{J}$. Without loss of generality, citizens only have payoffs from government actions when they are old, in which case we refer to them as voters. All voters receive a per-period policy-related payoff of 1 if the action matches the state $a_t = s_t$, and zero otherwise. We will also assume probabilistic voting as in Persson and Tabellini (2002). Any voter $j \in \mathcal{J}$ at t has an individual non-policy related preference for the incumbent, composed of a voter-specific plus a common term, $\varepsilon_{j,t} + z_t$. The term $\varepsilon_{j,t}$ is a random draw from a uniform distribution on $\left[-\frac{1}{2\rho}, \frac{1}{2\rho}\right]$, and z_t is independently and identically uniformly distributed on $\left[-\frac{1}{2\theta}, \frac{1}{2\theta}\right]$. When valence terms $\varepsilon_{j,t}, z_t$ are realized, $\varepsilon_{j,t}$ is observed by voter j only, and the common shock z_t is observed by all voters in $\mathcal{J}$. A sufficient condition for the incumbent's vote share to lie strictly between 0 and 1 is $\frac{1}{2} > \theta > \frac{\rho}{1-2\rho}$ and we maintain this assumption in what follows.

3.2 Politicians

There are two types of politicians, $\nu \in \{G, B\}$. The per period payoffs of politicians are the following. Both types have a payoff of $E > 0$ from holding office i.e. office rent. We normalize the politicians' payoffs when out of office to zero. The good type also shares the voters' objectives when in office, that is, he gets policy payoff if $s_t = a_t$ and zero otherwise. The bad politician does not care about the payoff of the voters, but always prefers action 0. To rule out mixed strategy equilibria (for reasons discussed below) we introduce a small amount of noise into politicians' payoffs as follows. The good type is assumed to get $u_{G,t}$ if the action matches the state, and the bad type gets $u_{B,t}$ if $a_t = 0$, where $u_{G,t}, u_{B,t}$ are random draws from a uniform distribution on $[1-\delta, 1+\delta]$, where $\delta < 1$. This can be thought of as a perturbation à la Harsanyi of the underlying game where politicians have payoffs of 1 from their short-run preferred actions. So, taking the limit as $\delta \to 0$ will identify a particular mixed strategy in the underlying game that is consistent with equality of the payoffs for both politician types (Fudenberg and Tirole, 1991).

We also assume that the politicians are *term-limited*, that is, any elected politician is limited to serve at most two terms in office. At the end of his second term in office, the incumbent cannot run for re-election and the election at the beginning of the next period is between two randomly drawn politicians from an infinite pool of candidates.

3.3 Citizen Beliefs

The *objective* probability that a politician randomly drawn from the pool is of G type is fixed at π. However, citizens do not know π with certainty. There are $N = \{1, \ldots, n\}$ groups of citizens with different levels of trust born in every period t. There are a continuum of citizens in each group, and $\lambda_i \in [0,1]$ is the size of group $i \in N$ with $\sum_{i \in N} \lambda_i = 1$.[15] A citizen of group i is born at time t with a prior belief about the distribution of π, denoted π_i^e, where π_i^e follows a Beta distribution with mean τ_i and concentration parameter κ. The larger the concentration parameter, the more concentrated the distribution is around its mean, i.e. the lower the variance. That is,

$$E[\pi_i^e] = \tau_i, \quad \mathrm{Var}[\pi_i^e] = \frac{\tau_i(1-\tau_i)}{\kappa+1}. \tag{1}$$

We call τ_i the *trust* parameter. When $\tau_i \neq \pi$, voters in group i have *biased beliefs* about the politicians' quality. Also, note two important limit cases. First, as $\kappa \to \infty$, the Beta distribution converges to a point mass at τ_i, that is $\Pr(\pi_i^e = \tau_i) = 1$. The maximum possible variance is in the case $\kappa \to 0$, in which case from (1), $\mathrm{Var}[\pi_i^e] \to \tau_i(1-\tau_i)$, and π_i^e follows a Bernoulli distribution with $\mathbb{P}(\pi_i^e = 1) = \tau_i, \mathbb{P}(\pi_i^e = 0) = 1 - \tau_i$. In what follows, we let $\overline{\tau} = \sum_{i \in N} \lambda_i \tau_i$ the mean trust of the electorate and $(\boldsymbol{\lambda}, \boldsymbol{\tau}) = (\lambda_i, \tau_i)_{i \in N}$ the distribution of trust among young citizens.

3.4 Timing

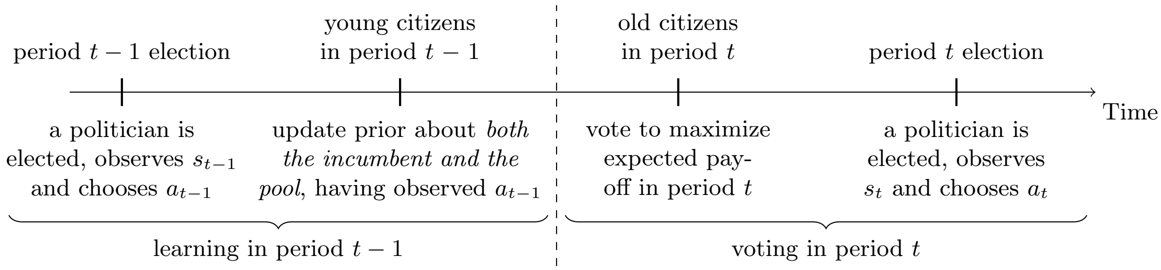
In each period t, the order of events is as follows:

1. Young citizens are born and the valence term $\varepsilon_{j,t} + z_t$ is determined for old citizen $j \in \mathcal{J}$.
2. Period t election takes place. If an incumbent in $t-1$ has served one term, an election takes place between the incumbent and a challenger, who is randomly drawn from the politician pool. If the incumbent in $t-1$ has served two terms, two politicians are randomly drawn from the pool to contest the election. Only old citizens ("voters") can vote in the election.
3. The state of the world in period t, $s_t \in \{0,1\}$, is determined.
4. The elected politician chooses an action $a_t \in \{0,1\}$ after observing s_t, his own type $\nu \in \{G,B\}$ and his type's policy payoff $u_{\nu,t}$. All citizens, young and old, observe a_t.
5. Young citizens update beliefs about *the type of the incumbent* and learn about *the quality of the politician pool.*
6. Old citizens receive their policy payoffs.

The entire life-cycle of the citizen is shown in Figure 2 below. Note a key point: by the time of the election at t, the voters (old citizens) at t have *already observed* a_{t-1} when young, and have updated both their beliefs about the probability that the incumbent is good, and their beliefs about the distribution of π_i^e. So, the voters update not only the probability that the incumbent is good, but also their beliefs about the *quality of the pool of candidates.* This update of beliefs about the pool is a key feature of the analysis.

[15] In what follows, i always denotes a citizen group, and j an individual citizen.

Figure 2: Citizen's Life Cycle



3.5 Equilibrium Concept

We solve for all stationary perfect Bayes-Nash equilibria in this model, and we call these electoral equilibria. A stationary perfect Bayes-Nash equilibrium is one where: (i) the incumbent's strategies depend only on the term of the incumbent (first or second), his type $\nu \in \{G, B\}$ and his type-specific policy payoff $u_{\nu,t}$, as well as voters' priors at the beginning of the period; (ii) decisions of voters depend only on the observed political action and the valence shock $\varepsilon_{j,t} + z_t$; (iii) beliefs of voters must be consistent with their priors and the incumbent's strategies via Bayesian updating on the equilibrium path. Given that both voters' and politicians' decisions are conditional on continuously distributed random variables ($\varepsilon_{j,t} + z_t$ and $u_{\nu,t}$ respectively), we can assume w.l.o.g, that voters and politicians choose pure strategies.

3.6 Discussion

At this point, it is helpful to explain some modelling choices, which at first sight, may seem like complications. Two-period models are often used to analyse political agency problems; however, this implies that at the end of the first period, the voter choice is between a term-limited incumbent and challenger that is also effectively term-limited, because of the two-period horizon. This understates the advantage of replacing the incumbent. The assumption of an infinite horizon is made to avoid this bias.

Second, the assumption of probabilistic voting is made to create a smooth trade-off between the policy and valence payoffs of the voter. Without this assumption, the voters' belief about the exact quality of the candidates in the election would not matter for voter choice; rather, the voter would vote for the incumbent if the expected probability that the incumbent is good is greater than the probability for the challenger, and vice-versa. Even with a single voter, equilibrium might involve a mixed strategy for the voter, making the analysis more difficult without changing the basic mechanisms at work. This would lead to even more complications with heterogeneous voters, making the analysis intractable.

4 Preliminary Results

4.1 Term-Limited Incumbents

The case of the incumbent in his second term is straightforward. As there is no re-election incentive, each type of politician chooses his most preferred action. That is, a good incumbent will take an action that matches the state $a_t = s_t$ and a bad incumbent will simply choose $a_t = 0$. In the election at the beginning of the next period, a voter j of group i believes that both (new) candidates are equally likely to be good, and thus votes only according to her valence preference $\varepsilon_{j,t} + z_t$. As $\varepsilon_{j,t} + z_t$ is symmetrically distributed around zero, the probability that either candidate wins is therefore 0.5.

So, in what follows, we focus on the election at the beginning of the first term of the incumbent, and drop all time subscripts. Let α_G^s, α_B^s be the probabilities that a first-term incumbent of type G and B respectively chooses $a = 1$ when the state is $s \in \{0, 1\}$.

4.2 Voter Learning About the Pool

Let $P_i(a)$ be the probability that the voters of group $i \in N$ assess the incumbent to be a good type, having observed $a \in \{0, 1\}$ in the previous period. We take $P_i(a)$ as fixed for the moment; it will be determined in equilibrium as described below. Then using the fact that the Beta distribution is the conjugate prior for a Bernoulli distribution, the posterior expected value of π_i^e, conditional on the type of the incumbent, is

$$E[\pi_i^e \,|\, G] = \frac{\tau_i \kappa + 1}{\kappa + 1}, \quad E[\pi_i^e \,|\, B] = \frac{\tau_i \kappa}{\kappa + 1}. \tag{2}$$

However, the voters do not observe $\nu \in \{G, B\}$ directly, but only the action $a \in \{0, 1\}$ chosen by the incumbent in his first term, which is a noisy signal of the incumbent type, with the probability that the incumbent is good conditional on a being $P_i(a)$. So conditional on observed action a, the voters update as follows:

$$E[\pi_i^e \,|\, a] = E[\pi_i^e \,|\, G] P_i(a) + E[\pi_i^e \,|\, B][1 - P_i(a)], \tag{3}$$

where the equality follows from the fact that a provides no additional information about π_i^e conditional on type G, B. That is, the posterior mean is a weighted average of $E[\pi_i^e \,|G], E[\pi_i^e \,|B]$, weighted by the posterior probabilities of the type being G, B conditional on $a \in \{0, 1\}$. So, having observed a, the voters assess the average quality of the pool from which the challenger is drawn to be $E[\pi_i^e \,|a]$.

4.3 Voter Choices and Re-election Probabilities

A key variable in the analysis will be $r(a)$, the probability that the incumbent is re-elected, given that he chooses action $a \in \{0, 1\}$. In particular, what will be of interest is $r(1) - r(0)$, which can be interpreted as the electoral reward for choosing $a = 1$ rather than $a = 0$. The calculation of this object from voter choices is lengthy but straightforward, and is given in the Appendix. It can be written as follows:

$$r(1) - r(0) = \underbrace{\theta \sum_{i \in N} \lambda_i p \left[P_i(1) - P_i(0)\right]}_{\text{performance incentive}} - \underbrace{\theta \frac{v_G - v_B}{\kappa + 1} \sum_{i \in N} \lambda_i \left[P_i(1) - P_i(0)\right]}_{\text{learning about the pool}} \tag{4}$$

Here, as defined above, $P_i(a)$ is the probability that the incumbent is good conditional on a being observed, and v_G, v_B are per period policy payoffs to all voters from a type G, B incumbent respectively serving a first term;

$$v_G = p\,\alpha_G^1 + (1-p)\,(1-\alpha_G^0),\;\; v_B = p\,\alpha_B^1 + (1-p)\,(1-\alpha_B^0). \tag{5}$$

To interpret this expression, note that $P_i(1) - P_i(0)$ is the change in the posterior belief of voters in group i that the *incumbent* is good from observing action 1 rather than 0, and $\frac{1}{\kappa+1}(P_i(1) - P_i(0))$ is the change in the posterior belief of voters in group i that the *quality of the pool* is good from observing action 1 rather than 0. So, we can see that there are two elements to $r(1) - r(0)$:

1. The *performance incentive*, which is proportional to the average change in the posterior belief that the *incumbent* is good if he chooses $a = 1$ rather than $a = 0$;

2. The *learning about the pool*, which is proportional to the average change in the posterior belief about the quality of the *pool of challengers* after observing $a = 1$ rather than $a = 0$, weighted by the difference in first-term performance between good and bad politicians.

Effect 2 is specific to this model of learning. As shown below, it is always positive in equilibrium, and so the effect of learning is to reduce $r(1) - r(0)$, other things equal. So, learning about the pool effect offsets the standard incentive to perform well. The intuition is simple; if the voter infers having observed $a = 1$, that the quality of the pool is higher, she is more willing to replace the incumbent with a new challenger, *even though* she also assesses the incumbent to be good with a higher probability having observed $a = 1$. One important observation is that the more precise are initial beliefs i.e. the higher κ, the less learning about the pool matters, and so other things equal, the bad politician's incentive to behave well is stronger. This is explored more in Section 6 below.

5 Characterization of Equilibrium

5.1 Preliminary Results

Note that the bad politician gets an overall payoff $r(1)(1+E)$ if he chooses $a = 1$, and $u_B + r(0)(1+E)$ if he chooses $a = 0$, whatever the state. Also, if $s = 1$, the good politician gets $u_G + r(1)(1+E)$ if he chooses $a = 1$, and $r(0)(1+E)$ if he chooses $a = 0$. If $s = 0$, the good politician gets $u_G + r(0)(1+E)$ if he chooses $a = 0$, and $r(1)(1+E)$ if he chooses $a = 1$.

Given these payoffs, and without any further assumptions, there is an uninteresting pooling equilibrium where all politicians choose $a = 0$ with probability one.[16] To eliminate this equilibrium, we assume an off-equilibrium-path belief that voters always update in favor of the good incumbent after observing $a = 1$, formally, we assume $P_i(1) \geq E[\pi_i^e]$ for all $i \in N$. This ensures that in equilibrium $r(1) > r(0)$, a fact that is proved in the Appendix.[17]

Now we can focus on the case where $r(1) > r(0)$. Then we know from the previous discussion that the bad politician's probability of choosing $a = 1$ is

$$\alpha_B^s = \Pr\left(u_B \leq [r(1) - r(0)](1+E)\right). \tag{6}$$

Writing out this condition in full, using the fact that u_B is uniformly distributed on $[1-\delta, 1+\delta]$:

$$\alpha_B^s = \left\{ \begin{array}{c} 1 \\ \frac{1}{2\delta}\left[(r(1) - r(0))(1+E) - (1-\delta)\right] \\ 0 \end{array} \right\} \text{ if } \left\{ \begin{array}{c} \big[r(1) - r(0)\big](1+E) > 1+\delta \\ \big[r(1) - r(0)\big](1+E) \in [1-\delta, 1+\delta] \\ \big[r(1) - r(0)\big](1+E) < 1-\delta \end{array} \right\}. \tag{7}$$

Note that in any equilibrium, this probability is independent of the state i.e. $\alpha_B^s = \alpha_B$.

Turning to the good politician, we have just established that if $s = 1$, the good politician gets $u_G + r(1)(1+E)$ if he chooses $a = 1$, and $r(0)(1+E)$ if he chooses $a = 0$. As $r(1) > r(0)$, and $u_G > 0$, we have $\alpha_G^1 = 1$ in any equilibrium. If $s = 0$, however, from the above discussion, α_G^0 must also satisfy (6) and so $\alpha_G^0 = \alpha_B$.

These probabilities have the following interpretation. If $s = 1$, α_B is the probability that the bad type is disciplined i.e. chooses the optimal action for the citizens, even though it is short-term suboptimal for him. If $s = 0$, α_G^0, α_B are the probabilities that the good and bad incumbents *pander* i.e. choose $a = 1$ when $s = 0$ in order to get re-elected. Note that $\alpha_G^0 = \alpha_B$, i.e. the probability of pandering is always equal to the probability of discipline. This is a feature of the assumed symmetry of policy payoffs for the two types (i.e. both payoffs u_G, u_B are drawn from the same distribution). The positive association between discipline and pandering is a more general feature, however, as the payoff $r(1) - r(0)$ that incentivizes the one behavior also incentivizes the other. From now on, to lighten notation, we define $\alpha_G^0 = \alpha_B \equiv \alpha$.

Given $\alpha_G^0 = \alpha_B \equiv \alpha$, $\alpha_G^1 = 1$, we can now derive explicit formulae for $P_i(1), P_i(0)$ and substitute these into (4) to get an expression for $r(1) - r(0)$ just as a function of α and parameters. This is

[16] Assume that both types of incumbent choose $a = 0$ with probability 1, whatever s. Then $P_i(0) = \tau_i, P_i(1) = 0$ is consistent with Bayesian updating. Then, substituting $\alpha_B^s = \alpha_G^s = 0$, $P_i(0) = \tau_i$, $P_i(1) = 0, v_G = v_B = 1 - p$ into formula (4) gives $r(0) - r(1) = \theta p \sum_{i \in N} \lambda_i \tau_i$. If $\theta p \sum_{i \in N} \lambda_i \tau_i > \frac{1+\delta}{1+E}$, then $r(0) - r(1) > \frac{1+\delta}{1+E}$. But this last condition ensures that both types choose $a = 0$ for all s, verifying the assumption.

[17] All proofs are in the Appendix.

done in the Appendix, where we show:

$$\Delta(\alpha) \equiv r(1) - r(0) = \theta p \left(\frac{\kappa + \alpha}{\kappa + 1}\right) \sum_{i \in N} \lambda_i \frac{\tau_i(1-\tau_i)p}{[\tau_i p + \alpha(1-\tau_i p)](1-\tau_i p)} \tag{8}$$

5.2 Equilibrium Regimes

With (7) and (8) in hand, we now show that there are two equilibrium regimes, depending on the precision of beliefs, as measured by κ, the concentration parameter. When beliefs are precise i.e. κ high enough, there is a unique equilibrium where discipline/pandering may be full ($\alpha = 1$), partial ($0 < \alpha < 1$), or zero ($\alpha = 0$). When beliefs are imprecise, there are at most two equilibria, one with $\alpha = 1$, and one with $\alpha = 0$. In that case, for some parameter values, both equilibria can exist simultaneously. So, the introduction of imprecise beliefs into the standard agency model has fundamental implications for the structure of equilibria.

To proceed, we can use (8) to substitute out $r(1) - r(0) \equiv \Delta(\alpha)$ in (7), and re-arranging, we see that the equilibrium α^* must satisfy:

$$\alpha^* = \begin{cases} 1 & \text{if} \quad 1 + E \geq \frac{1+\delta}{\Delta(1)} \\ \frac{\Delta(\alpha^*)(1+E)-(1-\delta)}{2\delta} & \text{if} \quad \frac{1-\delta}{\Delta(0)} < 1 + E < \frac{1+\delta}{\Delta(1)} \\ 0 & \text{if} \quad 1 + E \leq \frac{1-\delta}{\Delta(0)} \end{cases} \tag{9}$$

There are then two cases. The first is where beliefs are precise i.e.

$$\max_{i \in N} \tau_i < \frac{\kappa}{p(\kappa + 1)}. \tag{10}$$

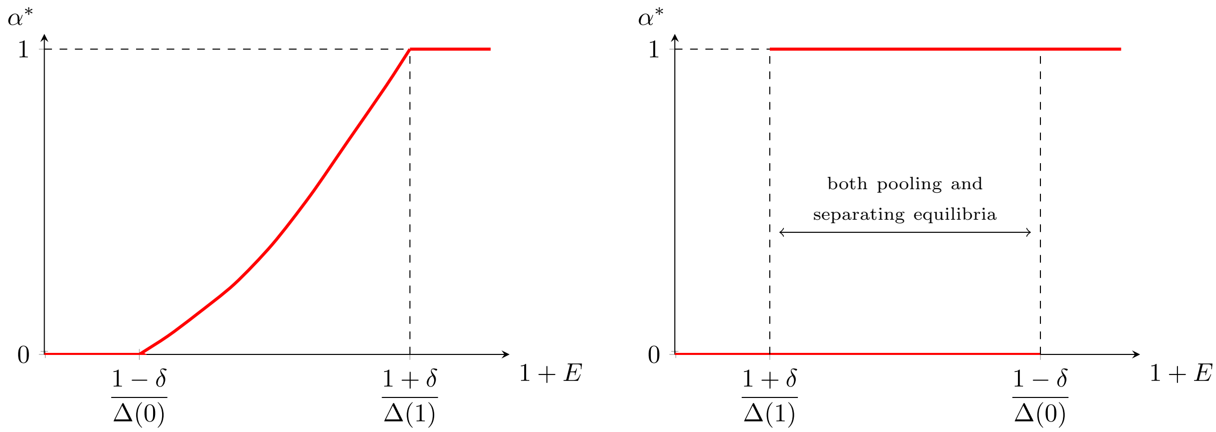
In this case, it can be shown that $\frac{1-\delta}{\Delta(0)} < 1 + E < \frac{1+\delta}{\Delta(1)}$ so from (9), there are parameter values where there might be an interior equilibrium $\alpha^* \in (0, 1)$. Moreover, we can show that this interior equilibrium, if it exists, is unique.

Proposition 1 *When beliefs are precise, i.e. (10) holds, then there is a unique electoral equilibrium where the good politician matches the state when $s = 1$, and panders with probability α^* in state $s = 0$, and the bad politician is disciplined with probability α^* in both states. In particular, if α^* is interior i.e. $\alpha^* \in (0, 1)$ then it is the unique solution to $\alpha^* = \frac{\Delta(\alpha^*)(1+E)-(1-\delta)}{2\delta}$.*

Figure 3a illustrates this equilibrium as E varies. Note that Proposition 1 always applies in the standard case with no bias in trust ($\tau_i = \pi$) and precise beliefs ($\kappa \to \infty$). Moreover, in this case, (10) reduces to $\pi < 1$, so the equilibrium is also unique.

We now turn to the second case where beliefs are imprecise i.e. κ is low. In this case, it is possible that $\Delta(0) < \Delta(1)$ so for small enough δ, the condition for interior equilibrium does not hold. In particular, we can prove:

Figure 3: Equilibrium Possibilities



(a) An interior equilibrium $\alpha^* \in (0, 1)$ exists.

(b) Only corner equilibria $\alpha^* \in \{0, 1\}$ exist.

Proposition 2 *When beliefs are imprecise i.e.* $p \min_{i \in N} \tau_i > \frac{\kappa}{\kappa+1}$*, then there exists a sufficiently small* δ *such that no interior equilibrium exists because* $\frac{1-\delta}{\Delta(0)} > \frac{1+\delta}{\Delta(1)}$*. In this case, there are three possibilities:*

(a) if $1 + E > \frac{1-\delta}{\Delta(0)}$*, there is a pure pooling electoral equilibrium* $(\alpha^* = 1)$*;*

(b) if $1 + E < \frac{1+\delta}{\Delta(1)}$ *there is a pure separating electoral equilibrium* $(\alpha^* = 0)$;

(c) if $\frac{1-\delta}{\Delta(0)} \geq 1 + E \geq \frac{1+\delta}{\Delta(1)}$*. there are both pure pooling and pure separating equilibria.*

Figure 3b illustrates this equilibrium. The intuition for multiple equilibria here is as follows. When condition (10) fails, prior beliefs are imprecise, and so policy choices have a strong effect on voters' learning. This effect induces the incumbent to adopt extreme strategies — either "go for the win" or "extract office rent." In particular, any intermediate choice $\alpha \in (0, 1)$ does not generate a sufficiently large increase in re-election probability to offset the associated loss in policy payoff.[18] For intermediate values of E and certain values of model parameters, both strategies may be sustained as viable equilibria.

Finally, consider the limit behavior of the equilibria as $\delta \to 0$. It is easily calculated that the classification of equilibria depending on the precision of beliefs does not change qualitatively. So, the two types of equilibria are not an artifact of the "noise" in the politicians' utility functions.

[18]When E is sufficiently high, "go for the win" becomes a strictly dominant strategy for all values of $u_G, u_B \in [1 - \delta, 1 + \delta]$, and all types of the incumbent chooses $a = 1$ to signal quality. Conversely, when E is sufficiently low, "extract office rent" is strictly dominant for all values of u_G and u_B, and the incumbent chooses action a to maximize policy payoff $(\alpha^* = 0)$.

6 Comparative Statics

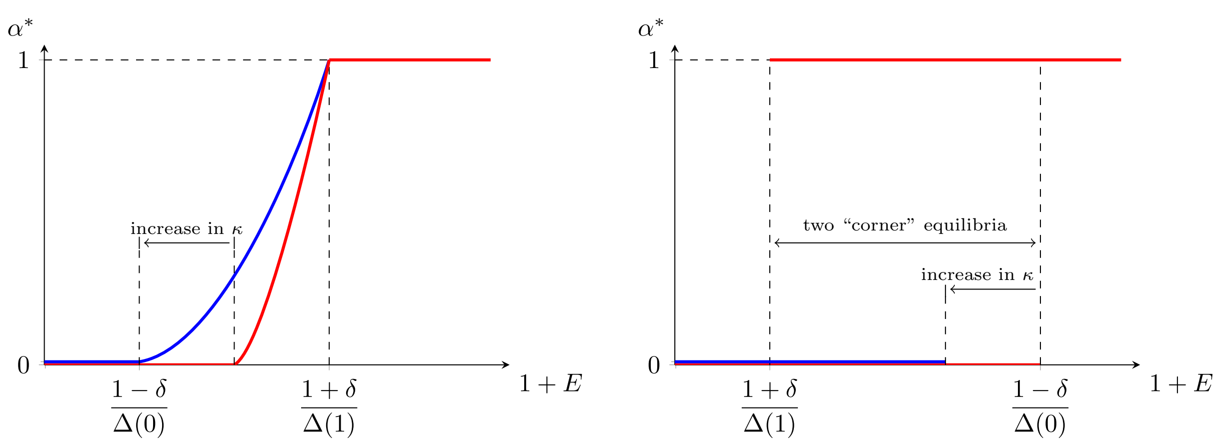
6.1 Discipline and Pandering

Recall that the introduction identified the stylized fact that trust in a given set of national political institutions varies widely across citizens of that country, and that at the individual level, there is evidence of lack of precision in beliefs, especially among younger voters. We are now in a position to investigate how changes in the precision of beliefs, κ, and heterogeneity of trust in the population affect discipline/pandering, α^*. We will also consider the effects of changes in trust within a group i, τ_i, on α^*. We define an increase in heterogeneity of trust as a *mean-preserving spread* in the distribution of trust across groups in the population.[19] This is a standard definition of increased heterogeneity that includes e.g. an increase in the variance.

We will look at the effects of changes in each of three separately, but in each case, we consider both types of equilibria i.e. the cases covered by both Propositions 1 and 2 to get a complete picture. Our first result is on the effects of an increase in the precision of beliefs, κ :

Proposition 3 *(a) An increase in κ strictly reduces the set of parameter values for which a corner equilibrium $\alpha^* = 0$ exists. (b) For an interior equilibrium at $\alpha^* \in (0,1)$, for all values of p, an increase in κ strictly increases α^*. (c) An increase in κ does not affect the set of parameter values for which a corner equilibrium $\alpha^* = 1$ exists.*

Figure 4: The effect of κ on discipline/pandering α^*.



(a) Effect of κ on $\alpha^* \in (0,1)$. (b) Effect of κ on $\alpha^* \in \{0,1\}$.

[19]Formally, we say that $(\boldsymbol{\lambda}', \boldsymbol{\tau}')$ is a mean-preserving spread of $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ if

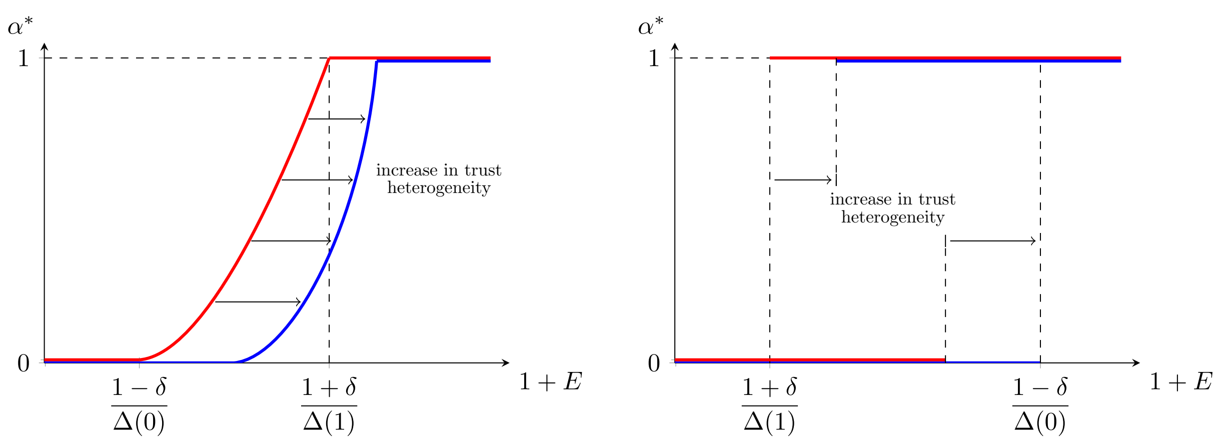
$$\sum_{i \in N} \lambda_i \tau_i = \sum_{i \in N} \lambda'_i \tau'_i \text{ and } \sum_{i \in N} \lambda_i \phi(\tau_i) \leq \sum_{i \in N} \lambda'_i \phi(\tau'_i) \text{ for every convex function } \phi.$$

Proposition 3 is illustrated in Figure 4a below. So, looking across equilibria, we can see that there is a general pattern; an increase in precision increases both discipline and pandering. The intuition for this is as follows. If beliefs of voters are already quite precise, the voters learn relatively little about the quality of the pool from the action taken by the incumbent. In particular, we see from equation A.5 that a higher κ reduces learning about the politician pool. This weakens the learning-from-the-pool term and therefore raises the electoral reward $\Delta(\alpha)$, strengthening discipline/pandering.

We now turn to the effects of an increase in heterogeneity.

Proposition 4 *(a) A mean preserving spread in* $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ *increases the set of parameter values for which a corner equilibrium* $\alpha^* = 0$ *exists. (b) For an interior equilibrium* $\alpha^* \in (0, 1)$*, a mean preserving spread in* $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ *decreases* α^**. (c) A mean preserving spread in* $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ *decreases the set of parameter values for which a corner equilibrium* $\alpha^* = 1$ *exists. The comparative statics holds strictly for* $p < 1$.

Figure 5: The effect of trust heterogeneity on discipline/pandering α^*.



(a) Effect of trust heterogeneity on $\alpha^* \in (0, 1)$. (b) Effect of trust heterogeneity on $\alpha^* \in \{0, 1\}$.

Proposition 4 is illustrated in Figure 5a below. The intuition for this result is the following. From (A.5), the reward to good behavior, $\Delta(\alpha)$, is a non-linear function of the prior beliefs of the voters, $\tau_i,\ i \in N$ via the term

$$\sum_{i \in N} \lambda_i \big[P_i(1) - P_i(0)\big] = \sum_{i \in N} \lambda_i \frac{\tau_i(1-\tau_i)p}{[\tau_i p + \alpha(1-\tau_i p)](1-\tau_i p)} \tag{11}$$

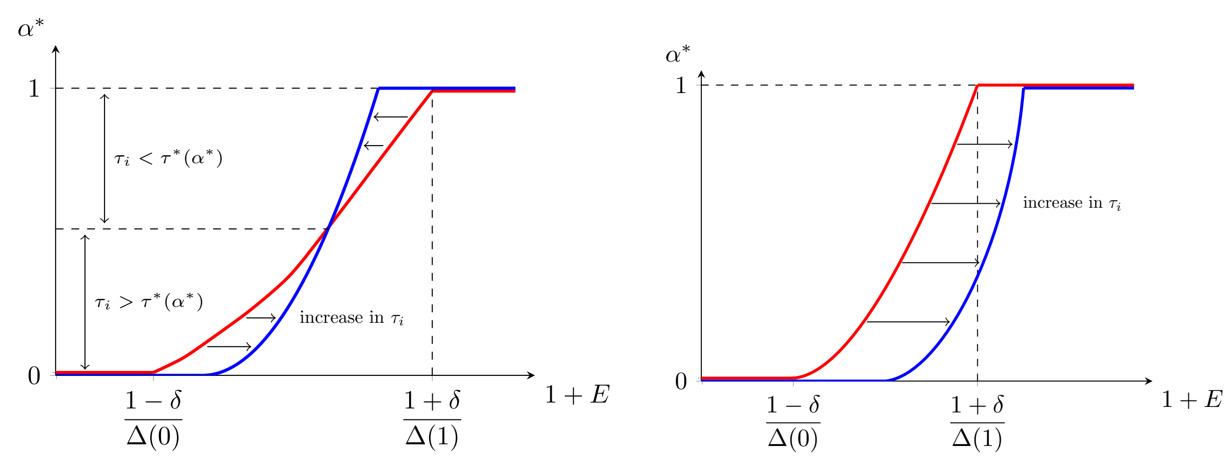
This function is derived mechanically from Bayesian updating, and measures the average sensitivity of voter posteriors to the incumbent choosing $a = 1$ rather than $a = 0$. It is easy to check that this function is concave in τ_i for all p, α and it holds strictly for $p < 1$. So, a mean-preserving spread in the distribution of τ_i *reduces the aggregate updating of the population* in response to a choice of $a = 1$ over $a = 0$, and thus reduces the incentive to behave well.

It is also natural to ask what happens if the level of trust in a particular group, τ_i, increases. The effect of a change in τ_i on the equilibrium α^* will also work through the term (11). It then turns out that we can prove the following result, which shows that the probability p of state $a = 1$ occurring plays a key role.

Proposition 5 *(a) An increase in* τ_i *increases the set of parameter values for which a corner equilibrium* $\alpha^* = 0$ *exists. (b) For an interior equilibrium* $\alpha^* \in (0,1)$*, then there exists a unique* $\tau(\alpha^*) \in (0,1)$ *such that an increase in* τ_i *increases (resp. decreases)* α^* *if* $\tau_i < \tau^*$ *(resp.* $\tau_i > \tau^*$*). (c) An increase in* τ_i *increases the set of parameters for which a corner equilibrium* $\alpha^* = 1$ *exists if* $\tau_i < \frac{1-\sqrt{1-p}}{p}$*, and decreases this set if* $\tau_i > \frac{1-\sqrt{1-p}}{p}$.

Proposition 5 is illustrated in Figure 6 below.

Figure 6: The effect of individual trust on discipline/pandering α^*.



(a) Effect of individual trust when $\tau_i < \frac{1-\sqrt{1-p}}{p}$. (b) Effect of individual trust when $\tau_i > \frac{1-\sqrt{1-p}}{p}$.

So, the picture here is more complex than in the cases of changing heterogeneity and precision of beliefs; an increase in trust by a particular group may increase or decrease discipline/pandering. However, we can obtain a stronger result if we consider the limiting case where $p = 1$; this is where the bad state never occurs so the good politician is never in a position where he has an incentive to pander, so α^* is just the probability that the bad politician is disciplined. Then we can show that across equilibria, an increase in trust in any group i at least weakly increases discipline.

Proposition 6 *Assume* $p = 1$*. (a) An increase in* τ_i *does not affect the set of parameter values for which a corner equilibrium* $\alpha^* = 0$ *exists. (b) For an interior equilibrium* $\alpha^* \in (0,1)$*, an increase in* τ_i *increases* α^**. (c) An increase in* τ_i *increases the set of parameters for which a corner equilibrium* $\alpha^* = 1$ *exists.*

The intuition for this simpler result is simply that when $p = 1$, an increase in τ_i raises $P_i(1) - P_i(0)$, the sensitivity of group $i's$ posterior belief to a choice of $a = 1$ over $a = 0$. In particular, from (11), when $p = 1$, we see that $P_i(1) - P_i(0) = \frac{\tau_i}{\tau_i + \alpha(1-\tau_i)}$, which is clearly increasing in τ_i.

We close this section with two observations. First, it is obvious that Propositions 5 and 6 give sufficient conditions for an increase in mean trust $\overline{\tau}$ across groups to increase discipline/pandering. Second, Propositions 3-6 also apply to the mixed-strategy equilibrium in the limit game where politician payoffs are deterministic.

6.2 Selection

In this section, we consider the impact of voter learning about the pool on the selection effect of elections. Let "re" be the event that the incumbent is re-elected. So, conditional on re-election, the voters' expected policy payoff from a term-limited incumbent is

$$W_T = \Pr(G\,|\text{re}) + [1 - \Pr(G\,|\text{re})](1-p) = 1 - p + p\Pr(G\,|\text{re}) \tag{12}$$

Because of the two-term limit, there are no discipline effects of elections in the second term.[20] So, $\Pr(G|\text{re})$ is a natural measure of the effectiveness of selection. Also, let $r = \Pr(\text{re})$ be the unconditional probability of this event. Then from Bayes' rule:

$$\Pr(G|\text{re}) = \frac{\Pr(\text{re}|G)\,\pi}{r}. \tag{13}$$

Note that we calculate this given the objective probability $\Pr(G) = \pi$, as we are interested in the objective probability that the re-elected incumbent is good, not its subjective perception by citizens.[21] We can then prove:

Proposition 7 $\Pr(G\,|re) \geq \pi$, *with* $\Pr(G\,|re) = \pi$ *only when* $\alpha^* = 1$. *In the separating equilibrium* $\alpha^* = 0$ *: (a) selection is strictly increasing in* κ*, (b) selection is strictly decreasing in trust heterogeneity, and (c) selection is strictly increasing in* τ_i *if* $\tau_i > \frac{1-\sqrt{1-p}}{p}$*. At* $p = 1$*, selection is strictly decreasing in* τ_i*.*

Proposition 7 first makes the straightforward point that as long as the equilibrium does not involve the bad incumbent fully imitating the good incumbent in both states ($\alpha^* = 1$), then there is a positive selection effect i.e. $\Pr(G\,|\text{re}) > \pi$. That is, unless there is full pooling, re-election is strictly better than random selection of a new incumbent from the pool.

The more interesting question is how our trust parameters affect the effectiveness of selection. This occurs through two channels. The first is a direct effect on the re-election probabilities $r(1)$

[20] The bad politician is completely undisciplined i.e. always chooses $a = 0$ and the good politician always matches the action to the state i.e. $a = s$.

[21] Here we follow an established approach in behavioral welfare economics, we calculate voter welfare given objective probabilities, not subjective ones.

and $r(0)$. The second is an indirect effect operating through the impact of the parameter change on discipline, α^*. At $\alpha^* = 0$, the latter channel vanishes, allowing us to isolate the direct effect of parameter changes on selection, as in Proposition 7.

In a separating equilibrium, $r(0)$ also represents the re-election probability of bad politicians. Moreover, an increase in the precision of beliefs, κ, strictly decreases $r(0)$ and strictly increases $r(1) - r(0)$. Hence, greater belief precision strengthens the selection effect of elections by reducing the likelihood that bad incumbents are re-elected. Conversely, an increase in trust heterogeneity decreases $r(1) - r(0)$ while increasing $r(0)$, thereby raising the probability that a bad incumbent is re-elected. As with discipline, the effect of trust on selection is more nuanced than the effects of heterogeneity and belief precision. In particular, there exists a threshold level of trust above which further increases in trust improve selection. The intuition is that, at sufficiently high levels of trust, the effect of trust on updating about the incumbent dominates its effect on learning about the pool of politicians. As a result, higher trust raises the probability of re-election for good incumbents by more than for bad incumbents, thereby strengthening selection. However, when $p = 1$, this threshold converges to 1, implying that the effect of trust is reversed: in this case, higher trust weakens selection.

For the case of an interior solution, $\alpha^* \in (0, 1)$, it is not possible to establish clear analytical results due to the fact that $\alpha^* \in (0, 1)$ now also depends on trust parameters. So, we resort to numerical analysis to get further insights. We consider the case where there are two equal-sized groups of voters $N = \{1, 2\}$, that is, $\lambda_1 = \lambda_2 = 0.5$. Also, $\tau_1 = \overline{\tau} + \sigma, \tau_2 = \overline{\tau} - \sigma$, so σ parametrizes heterogeneity in trust across groups. We then analyze the effect of σ, precision of beliefs κ and mean trust $\overline{\tau}$ on $\Pr(G \,|\, \text{re})$. Throughout we fix parameters at $\theta = 0.45$, $p = 0.85$, $\delta = 0.1$ and $\pi = 0.5$.[22]

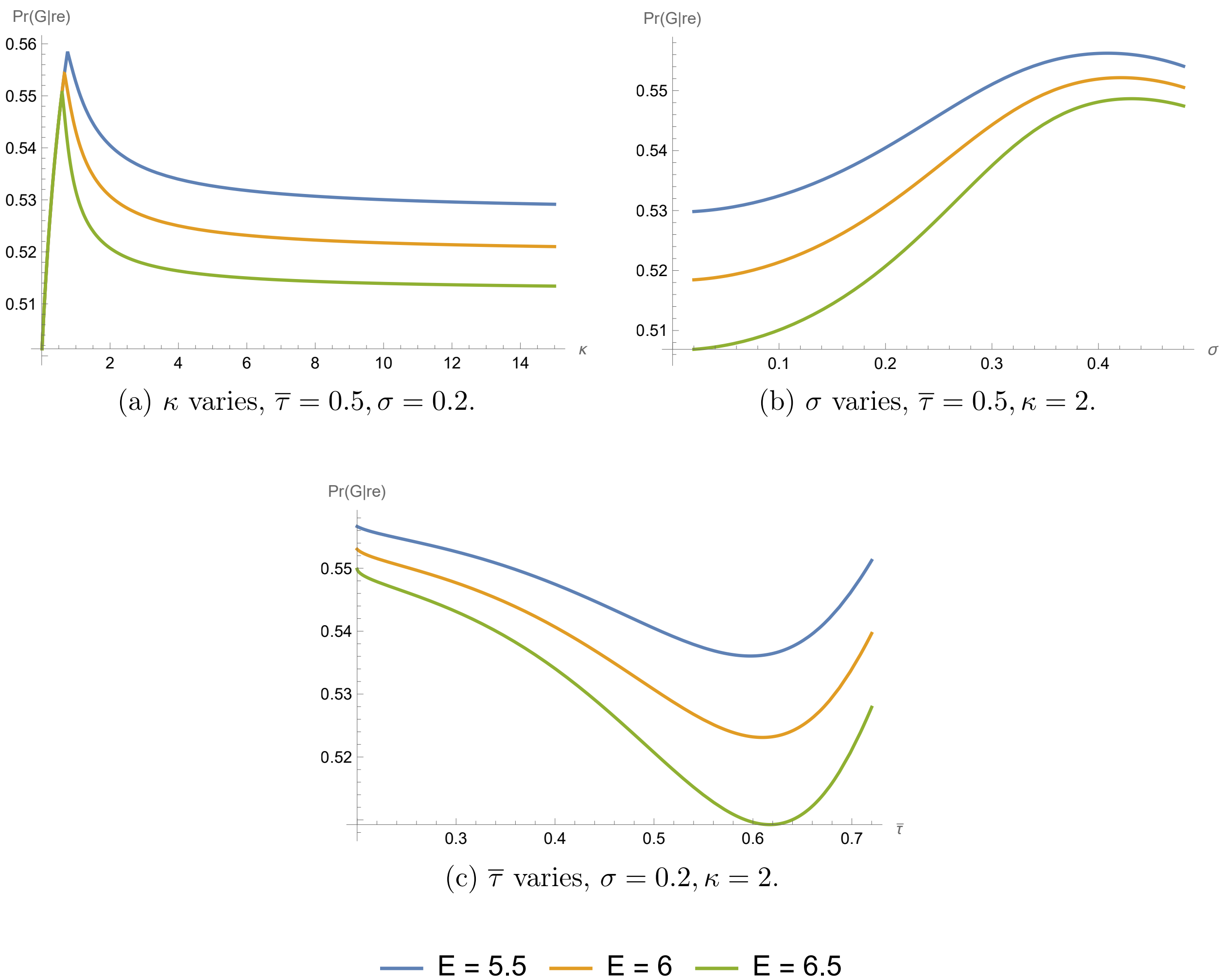
Figure 7 presents the results. The effects of κ, σ, and τ_i on selection can each be either positive or negative, reflecting a trade-off between a direct effect and a feedback effect. To illustrate, consider an increase in κ while holding discipline fixed at α^*. A higher κ reduces the magnitude of learning through $r(0)$. When $\alpha^* = 0$, this improves the effectiveness of selection, which explains the initial increase in selection observed in Figure 7a. This region also corresponds to the parameter range in which $\alpha^* = 0$. However, as κ increases further, the equilibrium level of discipline becomes strictly positive. The decline in $r(0)$ perturbs equation (9) away from equilibrium and, by Proposition 3, equilibrium can only be restored through an increase in α^*. We refer to this as the feedback effect. The increase in α^* weakens the signaling value of choosing $a = 1$ and reduces the re-election reward from remaining disciplined, as captured by $r(1)$.

The peak in Figure 7a arises at the point where the feedback effect begins to emerge, generating a sudden decline in the effectiveness of selection. Consequently, the overall effect of κ on selection depends on the balance between the improvement in learning induced by higher κ and the increase in pandering induced through higher equilibrium discipline. Figure 7a therefore illustrates a case in

[22]The results do not depend qualitatively on the choice of parameters, unlike the welfare analysis: see below

which the feedback effect through higher equilibrium discipline becomes sufficiently strong relative to the direct learning effect, leading to a gradual decrease in selection in the region where $\alpha^* \in (0, 1)$.

Figure 7: Effects of Parameter Changes on Selection



(a) κ varies, $\overline{\tau} = 0.5, \sigma = 0.2$.

(b) σ varies, $\overline{\tau} = 0.5, \kappa = 2$.

(c) $\overline{\tau}$ varies, $\sigma = 0.2, \kappa = 2$.

7 Candidate Entry

So far, we have simply followed the political agency literature by assuming that the probability that a politician drawn from the pool is good is fixed at π. To endogenize candidate entry, we introduce some additional structure to the model. We now assume two parties, A and B. Voter j's non-policy valence $\varepsilon_{jt} + z_t$ is now interpreted as preference for party A, not the incumbent. Both parties contest every election. In elections where the incumbent is not eligible (i.e. term-limited), then both parties choose a candidate via a selection procedure, described below. In elections with an eligible incumbent from one party, the other party chooses a candidate via the same selection procedure. To keep the analysis tractable, we assume $p = 1$ for simplicity, so that the good type does not pander when in office.

We model the selection procedure by supposing that each party has two potential candidates $k, l = 1, 2$, who are of good type with probability 0.5 (w.l.o.g), and who are independent draws from the same distribution. The outside options (i.e. payoffs if they do not enter the procedure) of the candidates are normalized to zero. We model the selection procedure as a *Tullock contest* (Klumpp and Polborn, 2006; Barbieri and Serena, 2025). Each candidate $k = 1, 2$ can exert effort e_k in the selection procedure, which is interpreted as time and resources devoted to campaigning. Candidate k wins with probability

$$w(e_k, e_l) = \frac{e_k^\beta}{e_k^\beta + e_l^\beta},\ \beta > 0 \tag{14}$$

So, a higher β amplifies the probability of winning of the higher-effort candidate. We assume for tractability that the voters do not observe the efforts made in the selection procedure. This assumption is also realistic; in almost all countries, candidates are selected via internal party procedures. The obvious exception is the US, where primaries are used, so arguably, in that case, e_k could be observed by voters.[23]

In the case of an election when the incumbent is term-limited, the order of events within an electoral cycle is as follows.

1. For each party, the types of the two candidates are determined and then observed by both candidates.

2. Within each party, the two candidates choose e_1, e_2 and candidates win with probabilities as in (14).[24]

3. The winner of the contest from party A then fights the election with the winner of the contest from party B.

In the case of an election when the incumbent is not term-limited, the order of events is similar, except that the party of the incumbent does not run a selection procedure, but just fields the incumbent for re-election.

To analyze this model, let V_G, V_B be the ex ante equilibrium payoffs from office for good and bad incumbents for a fixed $(\kappa, \boldsymbol{\lambda}, \boldsymbol{\tau})$. For simplicity, as well as assuming $p = 1$, we also assume that $\delta \to 1$, in which case the limiting payoffs are:[25]

$$\begin{aligned} V_G &= 1 + E + \Pr(\text{re}\,|G)(1+E) = 1 + E + r(1)(1+E) \\ V_B &= 1 - \alpha^* + E + \Pr(\text{re}\,|B)(1+E) = 1 - \alpha^* + E + \left[\alpha^* r(1) + (1-\alpha^*)r(0)\right](1+E) \end{aligned} \tag{15}$$

[23]In that case, e_k would also play a signalling role, which introduces a significant further complication.

[24]Both candidates are always willing to enter the contest as their outside options are zero, as in equilibrium, the payoffs from the contest net of effort are strictly positive for both types. The second fact follows from the fact that both V_G, V_B defined in (15) are both strictly positive.

[25]Taking this limit simplifies the first-period payoff to the bad incumbent in the event of not pandering.

where α^* is defined in equation (A.12).

We can now solve backwards in the case of an election when the incumbent is term-limited. As voters do not observe internal party selection, they believe that both party candidates are equally likely to be good, so each party wins office with equal probability. So, the expected payoff in the Tullock contest for a candidate of type $\nu = G, B$ is $\frac{V_\nu}{2}w - e$. Then, by standard results, the equilibrium winning probabilities of the good and bad candidates are as follows. In contests where both candidates are of the same type, each wins with probability 0.5. In contests where one is good and one is bad, the probability that the good candidate wins is

$$w_G = \frac{(\frac{V_G}{2})^\beta}{(\frac{V_G}{2})^\beta + (\frac{V_B}{2})^\beta} = \frac{(V_G)^\beta}{(V_G)^\beta + (V_B)^\beta} \tag{16}$$

So, for either party, the probability that the selected candidate is good is

$$\pi^* = \frac{1}{4} + \frac{1}{2}\frac{(V_G)^\beta}{(V_G)^\beta + (V_B)^\beta} = \frac{1}{4} + \frac{1}{2}\left[1 + \left(\frac{V_B}{V_G}\right)^\beta\right]^{-1} \tag{17}$$

This is because with probability $\frac{1}{4}$ both candidates are good, with probability $\frac{1}{4}$, both candidates are bad, and with probability $\frac{1}{2}$, one is good and one bad, and the good candidate wins with probability w_G. Note that because a good candidate is more likely to succeed in the selection procedure, $\pi^* > 0.5$, and is higher, the higher is the relative reward for winning the election for the good type, $\frac{V_G}{V_B}$.

When the incumbent is not term-limited, assume without loss of generality that he is from party A. Assume that candidates from party B know the type ν of the incumbent.[26] Then both party B candidates expect to win with probability $1 - \Pr(re\,|\nu)$ where $\Pr(\mathrm{re}\,|\nu)$ is defined in (13) above. Then, the win probability for the good candidate facing a bad candidate is as in (16), except that $1 - \Pr(\mathrm{re}\,|\nu)$ replaces $\frac{1}{2}$. So, it is clear from (16) that the term $1 - \Pr(\mathrm{re}\,|\nu)$ cancels.

Crucially, this means that w_G is *the same as in the term-limited election.* So, in both types of election, the equilibrium π^* is given by (17). Note that $r(0)$, $r(1)$, α^* depend on $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ and κ so there is *feedback from trust to quality of the pool* π^*. On the other hand, from (8), (9) note that in equilibrium, α^* does not depend on π^*. Overall, the structure of the equilibrium is recursive; both V_G, V_B are determined by $\alpha^*, r(0)$, $r(1)$ which are determined by $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ and κ and then π^* is also determined by $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ and κ via the same route. We can now establish some results about the effects of trust parameters on the quality of the pool.

Proposition 8 *For $\alpha^* < 1$, (a) π^* is strictly increasing in precision κ, (b) π^* is weakly decreasing in trust heterogeneity, and strictly so if $\alpha^* > 0$, (c) π^* is strictly increasing in trust τ_i.*

To understand this result, first note that when $p = 1$, a good politician always chooses $a = 1$ once elected. Consequently, any parameter change affects equilibrium policy choices only through

[26] This is consistent with the earlier assumption that candidates observe each other's types.

its impact on α^*, the probability that the bad politician is disciplined. From Proposition 3 and Proposition 6, we know that α^* is strictly increasing in both τ_i and κ. Hence, as τ_i (or κ) increases, α^* rises, reducing the re-election probability for all incumbents and further lowering the policy payoff of bad politicians once in office. This decreases the payoff ratio V_B/V_G and thereby improves the overall quality of the political pool. By contrast, trust heterogeneity affects π^* only indirectly through α^*. Once we fix $\alpha^* = 0$, variations in trust heterogeneity no longer affect the equilibrium quality of a randomly selected candidate.

However, at $\alpha^* = 1$, the relationship between π^* and trust τ_i reverses and we have the following result:

Proposition 9 *For $\alpha^* = 1$, (a) π^* does not change in precision κ and trust heterogeneity, (b) π^* is strictly decreasing in trust τ_i.*

The intuition is as follows. At $\alpha^* = 1$, observing the policy choice conveys no information about the quality of the incumbent, and voters in group i therefore base their voting decisions solely on their prior belief about incumbent quality, τ_i. As τ_i increases, the re-election probabilities of both good and bad politicians rise, increasing both V_B and V_G. However, the increase in payoff is relatively larger for bad politicians, implying that the payoff ratio V_B/V_G is strictly increasing in τ_i. Consequently, in a pooling equilibrium, the equilibrium quality of candidates is strictly decreasing in τ_i and independent of both κ and trust heterogeneity.

8 Welfare Implications

8.1 Voter Welfare

Here, we evaluate the effects of heterogeneity and precision of beliefs on voter welfare in our setting. We then discuss the implications of possible government interventions, such as civics education. Following an established approach in behavioral welfare economics, we calculate lifetime voter welfare given the objective quality of the candidate pool, not the subjective beliefs about quality (Chetty, 2015; Mullainathan and Schwartzstein, 2015). So, we define welfare as experienced utility, rather than decision utility.

We calculate a welfare formula as follows. We have already calculated the policy payoff to a term-limited incumbent as W_T, given by equation (12). Also, the expectation of the non-policy payoff $\varepsilon_{jt} + z_t$ across all citizens is zero, so averaging across all citizens, the average overall payoff from a term limited incumbent is W_T. Next, consider the payoff to citizens from a non-term-limited incumbent. Objectively, a first-term incumbent is always a random draw from the candidate pool and is thus good with probability π. Thus, the policy-related payoff to any citizen from a first-term incumbent is given by

$$\begin{aligned} W_N &= \pi\left[p+(1-p)(1-\alpha^*)\right]+(1-\pi)\left[p\alpha^*+(1-p)(1-\alpha^*)\right] \\ &= 1-p(1-\pi)+\left[(2-\pi)p-1\right]\alpha^*. \end{aligned}$$

Averaging across all citizens, we can ignore the non-policy related payoff, so the average payoff from a non-term limited incumbent is W_N. Finally, note that W_N is increasing in α^* i.e. the discipline effect dominates the pandering effect if

$$p > \frac{1}{2-\pi} \tag{18}$$

It remains to calculate the overall average lifetime welfare. To do this, note that the state variable of this model is the term limit status of the current incumbent i.e. either term-limited or not. The transition between the two states is as follows. If the incumbent is term-limited, with probability 1, in the next period, the incumbent will be non-term limited. If however, the incumbent is non-term limited, he contests the election, and wins with probability r, in which case, next period, he is term-limited. It is therefore straightforward to calculate that in the steady state, the fraction of periods with a term-limited incumbent is $\frac{r}{1+r}$. So, aggregate welfare is

$$\begin{aligned} W &= \frac{r}{1+r}W_T+\frac{1}{1+r}W_N \\ &= \frac{1}{1+r}\Big[r(1-p)+pr\Pr(G\,|\mathrm{re})+1-p(1-\pi)+[(2-\pi)p-1]\alpha^*\Big]. \end{aligned}$$

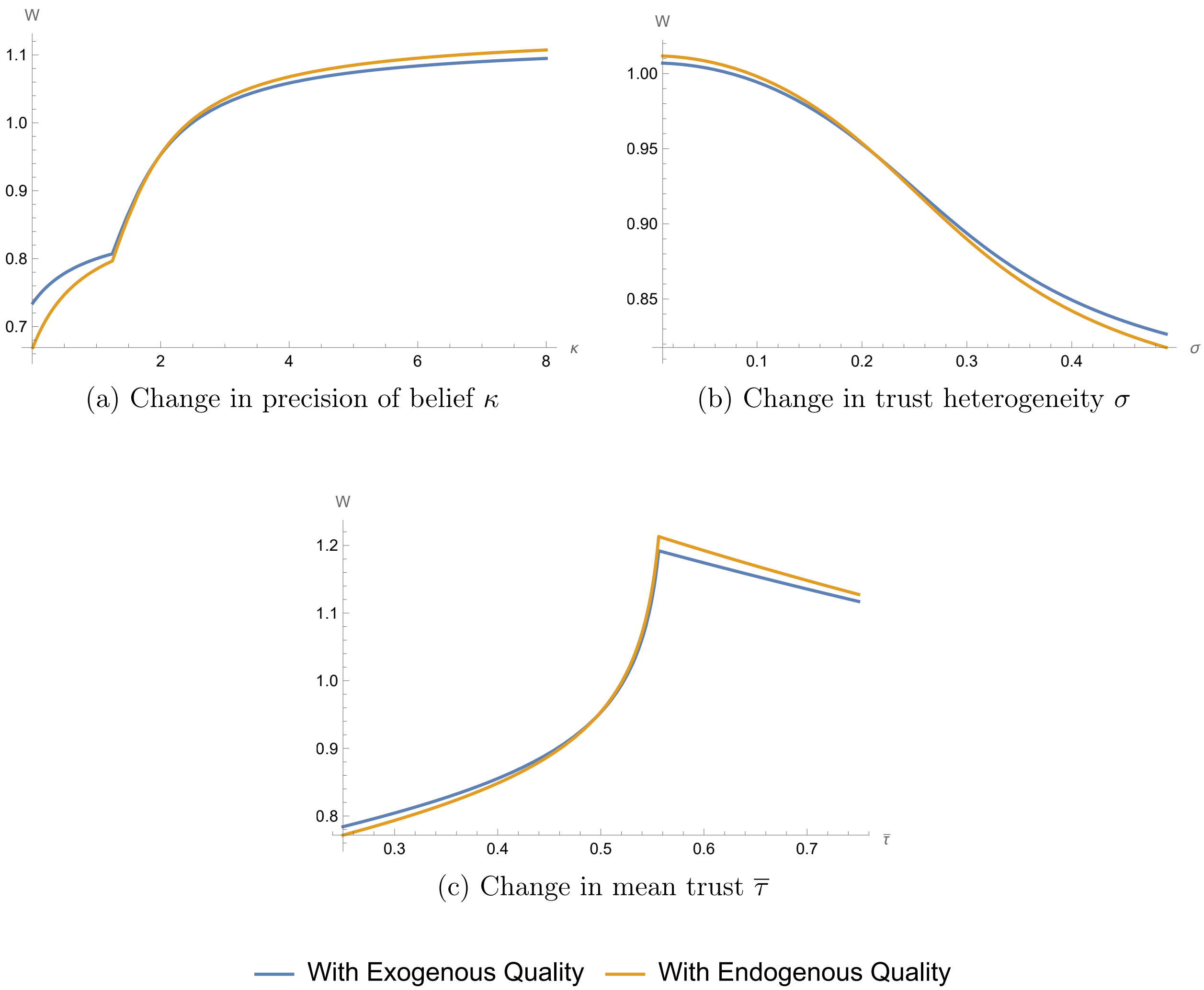
So, we see that welfare depends on trust parameters through three channels: equilibrium discipline α^*, re-election r and the quality of selection given by $\Pr(G\,|\mathrm{re})$. In particular, for a fixed r, welfare is increasing in the accuracy of selection, and is also increasing in α^* as long as the welfare effect of discipline dominates the welfare loss due to pandering, that is, equation (18) holds. In addition, if π is treated as endogenous as in Section 7, trust parameters have an additional effect through π^*.

To investigate further, we turn to numerical analysis, using the same set-up as in the case of selection. We want to compare results with and without an endogenous π^*, so we set $p=1$. Therefore, from (18), discipline dominates pandering. Figure 8 below shows the effects of parameter changes on welfare; welfare is increasing in precision of beliefs and decreasing in heterogeneity as shown in Figure 8a and 8b.

To interpret this, note that from Propositions 3 and 4 above, discipline is increasing in precision and decreasing in heterogeneity. Moreover, when $p=1$, as we have assumed, from inspection of (18), an increase in α^* leads to an increase in welfare, holding re-election r and selection $\Pr(G|\mathrm{re})$ constant. So, our results can be interpreted as showing that whatever the effects of these parameter changes on r and the selection effect, the discipline effect is dominant.

Figure 8c also illustrates the non-monotonic relationship between welfare and the average trust level of the electorate. The intuition is as follows. In a partially pooling equilibrium where $\alpha^* \in (0,1)$ and $p=1$, an increase in discipline has a positive effect on welfare, which explains the initial rise

Figure 8: Effect of Parameter Changes on Welfare



(a) Change in precision of belief κ

(b) Change in trust heterogeneity σ

(c) Change in mean trust $\overline{\tau}$

in welfare as $\overline{\tau}$ increases. However, from Proposition 6, we know that as mean trust continues to rise, the re-election incentive becomes stronger. This incentivizes bad politicians to remain disciplined and choose $a = 1$ with higher probability. Once mean trust becomes sufficiently large, the equilibrium converges to $\alpha^* = 1$, at which point the direct effect of α^* on welfare disappears. This explains the kink in Figure 8c. Beyond this point, further increases in mean trust affect welfare only through their impact on learning, re-election probabilities r, and selection $\Pr(G \mid \text{re})$. In a fully pooling equilibrium, higher trust increases the probability of re-election, and since welfare under a non-term-limited politician is strictly higher than under a term-limited politician, this leads to a decline in voters' average welfare.

The welfare results with endogenous candidate quality are not qualitatively different from those obtained under exogenous quality. However, the figures clearly illustrate the feedback effect of trust on politician quality. It is straightforward that welfare is strictly increasing in the quality of the candidate pool. From Propositions 8 and 9, the endogenous quality of the candidate pool, π^*, is increasing in κ and decreasing in trust heterogeneity, with the relationship holding strictly whenever $\alpha^* < 1$. Hence, an increase in κ (respectively, trust heterogeneity) improves (respectively, worsens) the quality of the candidate pool. As a result, for sufficiently high κ and low trust heterogeneity, the feedback effect operating through candidate quality has a positive effect on voter welfare. We observe a similar pattern in the feedback effect of trust on candidate quality. However, notice that the gap between welfare under exogenous and endogenous quality begins to shrink in the region where $\alpha^* = 1$. This is because, in a pooling equilibrium, the relationship between trust and the quality of the candidate pool becomes strictly decreasing.

8.2 Civic Education

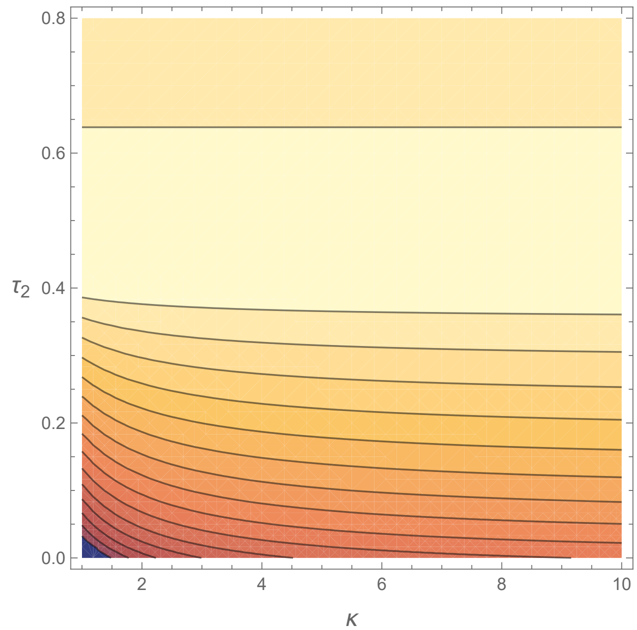
Over recent years, many education systems across the OECD have increasingly prioritised civic education at secondary school level to promote democratic citizenship in response to falling political participation rates (Borhan, 2025). A small literature shows that civic education initiatives have positive effects. For example, Claes and Hooghe (2017), using the Belgian Political Panel Survey of late adolescents and young adults, find that citizenship education (i.e., classroom instruction, being a member of a school council, and an open-classroom climate) are positively related to political trust. For the US, Hart et al. (2007) find that compulsory community service in high school is a strong predictor of adult voting and volunteering.

At this point, we can offer a tentative assessment of the welfare effects of civic education using our model. The empirical literature on the relationship between civic education and political trust remains relatively limited. Nevertheless, within our framework, it is plausible that civic education increases both the average level of trust, $\overline{\tau}$, and the precision of voters' beliefs about the quality of the candidate pool, captured by κ. Civic education may also reduce trust heterogeneity by exerting a larger effect on individuals with initially low levels of trust.

We capture these effects by considering an increase in the trust level of the lowest-trust group together with an increase in belief precision. Importantly, raising the trust of the lowest-trust group simultaneously increases average trust and reduces heterogeneity within the electorate. As before, we consider an electorate composed of two equally sized voter groups with $\tau_1 > \tau_2$. The contour plot in Figure 9 illustrates how welfare changes as both the trust level of the lower-trust group, τ_2, and the precision parameter, κ, increase jointly.

The figure shows that welfare is strictly increasing in the precision of beliefs, κ, but remains non-monotonic in trust, even when $p = 1$. The intuition follows directly from the earlier discussion: although higher trust improves discipline of bad politicians, excessively high trust also raises the incumbent's probability of re-election regardless of his type, which can reduce welfare on average. The implication is that civic education may not always raise voter welfare, especially if it works to raise trust amongst the lowest-trust groups, rather than reduce uncertainty in beliefs.

Figure 9: Effects of increases in both lowest trust and precision on welfare



$\theta = 0.45$, $\pi = 0.5$, $\delta = 0.5$. Lighter color denotes higher welfare.

9 Conclusions

This paper has developed a model of electoral accountability that incorporates three empirically well-documented features of political trust: incorrect and heterogeneous citizen beliefs about the quality of the politician pool, and the malleability of those beliefs, particularly during early adulthood. By embedding these features into an otherwise standard political agency framework, we have shown that the structure of equilibria and the welfare implications of elections depend critically on the precision and distribution of voter beliefs.

Several extensions remain for future work, for example, endogenizing the formation of trust through repeated electoral experience, allowing for strategic communication by politicians seeking to shape voter beliefs, and incorporating heterogeneity in voters' responsiveness to civic education interventions. We leave these for future research.

References

Ashworth, Scott, and Ethan Bueno de Mesquita. 2014. "Is voter competence good for voters? Information, rationality, and democratic performance." *American Political Science Review*, 108(3): 565–587.

Barbieri, Stefano, and Marco Serena. 2025. "Fair Representation in Primaries: Heterogeneity and the New Hampshire Effect." *Journal of Theoretical Politics*, 37(4): 290–318.

Besley, Timothy. 2004. "Paying politicians: theory and evidence." *Journal of the European Economic Association*, 2(2-3): 193–215.

Besley, Timothy. 2006. *Principled agents?: The political economy of good government.* Oxford University Press on Demand.

Besley, Timothy, Christopher Dann, and Sacha Dray. 2026. "Growth experiences and trust in Government." *The Quarterly Journal of Economics*, qjaf056.

Borhan, Hannah. 2025. "Civic education as a pathway to inclusive societies: Exploring the role of education in fostering civic and social engagement." *OECD Education Working Papers*.

Bracco, Emanuele, Federica Liberini, Ben Lockwood, Francesco Porcelli, Michela Redoano, and Daniel Sgroi. 2021. "The Effects of Social Capital on Government Performance and Turnover: Theory and Evidence from Italian Municipalities." Quantitative and Analytical Political Economy Research Centre.

Bracco, Emanuele, Federica Liberini, Ben Lockwood, Francesco Porcelli, Michela Redoano, and Daniel Sgroi. 2026. "The Effects of Social Capital on Government Performance and Turnover: Evidence from Italian Municipalities." Quantitative and Analytical Political Economy Research Centre.

Canes-Wrone, Brandice, Michael C. Herron, and Kenneth W. Shotts. 2001. "Leadership and pandering: A theory of executive policymaking." *American Journal of Political Science*, 45(3): 532–550.

Caselli, Francesco, and Massimo Morelli. 2004. "Bad politicians." *Journal of public economics*, 88(3-4): 759–782.

Chetty, Raj. 2015. "Behavioral economics and public policy: A pragmatic perspective." *American Economic Review*, 105(5): 1–33.

Claes, Ellen, and Marc Hooghe. 2017. "The effect of political science education on political trust and interest: Results from a 5-year panel study." *Journal of Political Science Education*, 13(1): 33–45.

Dal Bó, Ernesto, Frederico Finan, Olle Folke, Torsten Persson, and Johanna Rickne. 2017. "Who becomes a politician?" *The Quarterly Journal of Economics*, 132(4): 1877–1914.

Devine, Daniel, Will Jennings, George Gaskell, Gerry Stoker, and Viktor Valgarðsson. 2024. "Stability and Change in Political Trust: Evidence and Implications from Six Panel Studies." *European Journal of Political Research*, 63(2): 450–472.

Eichengreen, Barry, Orkun Saka, and Cevat Giray Aksoy. 2024. "The Political Scar of Epidemics." *The Economic Journal*, 134(660): 1683–1700.

Fudenberg, Drew, and Jean Tirole. 1991. *Game theory.* MIT press.

Hart, Daniel, Thomas M Donnelly, James Youniss, and Robert Atkins. 2007. "High school community service as a predictor of adult voting and volunteering." *American educational research journal*, 44(1): 197–219.

Haugsgjerd, Atle, and Staffan Kumlin. 2020. "Downbound spiral? Economic grievances, perceived social protection and political distrust." *West European Politics*, 43(4): 969–990.

Herrera, Helios, and Federico Trombetta. 2024. "Alternative Worldviews, Distrust, and Populism." *CEPR Discussion Paper No. 19674. CEPR Press, Paris & London.*

Klumpp, Tilman, and Mattias K. Polborn. 2006. "Primaries and the New Hampshire Effect." *Journal of Public Economics*, 90(6–7): 1073–1114.

Li, Anqi, Davin Raiha, and Kenneth W Shotts. 2022. "Propaganda, alternative media, and accountability in fragile democracies." *The Journal of Politics*, 84(2): 1214–1219.

Maskin, Eric, and Jean Tirole. 2004. "The politician and the judge: Accountability in government." *American Economic Review*, 94(4): 1034–1054.

Messner, Matthias, and Mattias K Polborn. 2004. "Paying politicians." *Journal of Public Economics*, 88(12): 2423–2445.

Mullainathan, Sendhil, and Joshua Schwartzstein. 2015. "Learning and attention in behavioral economics." *Journal of Economic Perspectives*, 29(4): 3–28.

Persson, Torsten, and Guido Enrico Tabellini. 2002. *Political economics: explaining economic policy.* MIT press.

Schultz, Christian. 1996. "Polarization and inefficient policies." *The Review of Economic Studies*, 63(2): 331–344.

Testa, Cecilia. 2012. "Is polarization bad?" *European Economic Review*, 56(6): 1104–1118.

Trombetta, Federico. 2020. "When the light shines too much: Rational inattention and pandering." *Journal of Public Economic Theory*, 22(1): 98–145.

Valgarðsson, Viktor, Will Jennings, Gerry Stoker, Hannah Bunting, Daniel Devine, Lawrence McKay, and Andrew Klassen. 2025. "A crisis of political trust? Global trends in institutional trust from 1958 to 2019." *British Journal of Political Science*, 55: e15.

Wolton, Stephane. 2019. "Are biased media bad for democracy?" *American Journal of Political Science*, 63(3): 548–562.

Appendix

Derivation of equation (4). A voter of group i's expected policy payoff from voting for the incumbent conditional on observing a is

$$V_i^I(a) = P_i(a) + (1-p)[1 - P_i(a)]. \tag{A.1}$$

Similarly, a voter of group i's expected policy payoff from voting for the challenger conditional on observing a

$$V_i^C(a) = E[\pi_i^e \,|a]\, v_G + (1 - E[\pi_i^e \,|a])\, v_B \tag{A.2}$$

where v_G, v_B are defined in equation (5).

So, a voter j in group $i \in N$ will vote for the incumbent only if

$$\varepsilon_j + z + V_i^I(a) - V_i^C(a) \geq 0 \iff \varepsilon_j \geq -z - \left[V_i^I(a) - V_i^C(a)\right]$$

So, the probability that a voter j in group $i \in N$ votes for the incumbent, conditional on z and a is

$$s_i(a) = \Pr\left[\varepsilon_j \geq -z - \left[V_i^I(a) - V_i^C(a)\right]\right] = \frac{1}{2} + \rho\left[z + V_i^I(a) - V_i^C(a)\right]. \tag{A.3}$$

Note that equation (A.3) holds as we have group i voter payoffs $V_i^I(a), V_i^C(a)$ are between zero and one and

$$\theta - \rho > 2\theta\rho \iff \frac{1}{2\rho} - \frac{1}{2\theta} > 1 \implies -\frac{1}{2\rho} < z + V_i^I(a) - V_i^C(a) < \frac{1}{2\rho}.$$

So, given that there are a continuum of voters in each group, this is also the incumbent's vote share among group i voters, conditional on z and a. So, the total vote share for the incumbent conditional on z and a is

$$s(a) = \sum_{i \in N} \lambda_i s_i(a) = \frac{1}{2} + \rho\left[z + \sum_{i \in N} \lambda_i \left[V_i^I(a) - V_i^C(a)\right]\right].$$

So, if $\theta < 0.5$, the probability that the incumbent is re-elected conditional on a is[27]

$$r(a) = \Pr\left(s(a) \geq \frac{1}{2}\right) = \Pr\left(z \geq -\sum_{i \in N} \lambda_i \left[V_i^I(a) - V_i^C(a)\right]\right) = \frac{1}{2} + \theta \sum_{i \in N} \lambda_i \left[V_i^I(a) - V_i^C(a)\right]. \tag{A.4}$$

[27] The assumption $\theta < 0.5$ is required as voter payoffs $V_i^I(a), V_i^C(a)$ are between zero and one, implying that $\sum_{i \in N} \lambda_i \left[V_i^I(a) - V_i^C(a)\right]$ is between -1 and 1. These assumptions are standard in the probabilistic voting literature, and rule out uninteresting corner cases where $r(a) \in \{0, 1\}$.

So, from (A.1), (A.2) and (A.4) this is:

$$\begin{aligned} r(1)-r(0) &= \theta\sum_{i\in N}\lambda_i\left[\left(V_i^I(1)-V_i^I(0)\right)-\left(V_i^C(1)-V_i^C(0)\right)\right] \\ &= \theta\sum_{i\in N}\lambda_i p\left[P_i(1)-P_i(0)\right]-\theta(v_G-v_B)\sum_{i\in N}\lambda_i\left(E[\pi_i^e|1]-E[\pi_i^e|0]\right) \\ &= \theta\sum_{i\in N}\lambda_i p\left[P_i(1)-P_i(0)\right]-\theta\frac{v_G-v_B}{\kappa+1}\sum_{i\in N}\lambda_i\left[P_i(1)-P_i(0)\right] \end{aligned} \tag{A.5}$$

where we have used (A.4) and the definitions $V_i^I(a), V_i^C(a)$ in the first line, (A.1), (A.2) in the second line, and finally (2), (3) in the last line.

Proof that there is no electoral equilibrium where $r(1)\leq r(0)$. Suppose, by contradiction, that $r(1)\leq r(0)$. Then we have $\alpha_B=0$ and $\alpha_G^0=0$. Substituting into equation (5) gives us $v_G=p\alpha_G^1$ and $v_B=1-p$. From equation (A.5), we have

$$r(1)-r(0)=\theta p\underbrace{\left(1-\frac{\alpha_G^1}{\kappa+1}+\frac{1-p}{p(\kappa+1)}\right)}_{>0}\sum_{i\in N}\lambda_i\left[P_i(1)-P_i(0)\right]\leq 0,$$

which implies that there exists $i'\in N$ such that $P_{i'}(1)\leq P_{i'}(0)$.

First, we consider the case $r(1)=r(0)$. Then, from equation (7), we have $\alpha_G^0=0, \alpha_G^1=1$ and $\alpha_B=0$. Then using Bayes' rule, we have $P_i(1)=1$ and $P_i(0)<1$ for all $i\in N$, contradiction.

Next, consider the case $r(1)<r(0)$, equation (7) implies $\alpha_G^0=0$ and $\alpha_B=0$. If $\alpha_G^1>0$, then again, using Bayes' rule, we have $P_i(1)=1$ and $P_i(0)<1$ for all $i\in N$, contradiction. Suppose that $\alpha_G^1=0$. Then by Bayes' rule, $P_i(0)=\tau_i$ and as $r(1)<r(0)$, it must be that $P_i(1)<P_i(0)=\tau_i$ for all $i\in N$. But under the restriction on off-equilibrium belief, it must be that $P_i(1)\geq E[\pi_i^e]=\tau_i$ for all $i\in N$, contradiction. Hence, there is no electoral equilibrium where $r(1)<r(0)$. $\square$

Derivation of equation (8). Given $\alpha_G^0=\alpha_B\equiv\alpha, \alpha_G^1=1$, we can use Bayes' rule to derive

$$\begin{aligned} P_i(1) &= \frac{\mathbb{P}[\nu=G,a=1]}{\mathbb{P}[a=1]}=\frac{p\tau_i+(1-p)\alpha\tau_i}{p\tau_i+(1-p)\alpha\tau_i+\alpha(1-\tau_i)}, \\ P_i(0) &= \frac{\mathbb{P}[\nu=G,a=0]}{\mathbb{P}[a=0]}=\frac{(1-p)(1-\alpha)\tau_i}{(1-p)(1-\alpha)\tau_i+(1-\alpha)(1-\tau_i)}=\frac{(1-p)\tau_i}{1-p\tau_i}. \end{aligned} \tag{A.6}$$

Then we have

$$P_i(1)-P_i(0)=\frac{p\tau_i(1-\tau_i)}{[p\tau_i+(1-p\tau_i)\alpha](1-p\tau_i)}. \tag{A.7}$$

Also, substituting $\alpha_B^1=\alpha_B^0=\alpha_B=\alpha$ in (5), and simplifying, we get

$$v_G-v_B=p+(1-p)(1-\alpha)-(p\,\alpha+(1-p)(1-\alpha))=p(1-\alpha) \tag{A.8}$$

Then substituting (A.7), (A.8) into (A.5), we have the result. $\square$

Proof of Proposition 1. (i) First we show that $\frac{1-\delta}{\Delta(0)} < 1+E < \frac{1+\delta}{\Delta(1)}$ can hold. Note that

$$\Delta(0) = \theta p \left(\frac{\kappa}{\kappa+1}\right) \sum_{i \in N} \lambda_i \frac{(1-\tau_i)}{(1-\tau_i p)}, \quad \Delta(1) = \theta p \sum_{i \in N} \lambda_i \frac{\tau_i (1-\tau_i) p}{(1-\tau_i p)} \tag{A.9}$$

So,

$$\Delta(0) - \Delta(1) = \theta p \sum_{i \in N} \lambda_i \frac{(1-\tau_i)}{\tau_i p (1-\tau_i p)} \left(\frac{\kappa}{\kappa+1} - p\tau_i\right) \tag{A.10}$$

So, as $\frac{\kappa}{\kappa+1} > p \max_{i \in N} \tau_i$, from (A.10), $\Delta(0) > \Delta(1)$, which in turn implies that $\frac{1-\delta}{\Delta(0)} < \frac{1+\delta}{\Delta(1)}$ so there are values of E and other parameters for which $\frac{1-\delta}{\Delta(0)} < 1+E < \frac{1+\delta}{\Delta(1)}$ holds.

(ii) It remains to show that

$$\alpha^* = \frac{1}{2} + \frac{\Delta(\alpha^*)(1+E) - 1}{2\delta} \equiv h(\alpha^*)$$

has a unique solution $\alpha^* \in (0,1)$ if $\frac{1-\delta}{\Delta(0)} < 1+E < \frac{1+\delta}{\Delta(1)}$. First, it must admit at least one solution $\alpha^* \in (0,1)$, because $h(.)$ is continuous and

$$\frac{1-\delta}{\Delta(0)} < 1+E \Rightarrow h(0) > 0, \quad \frac{1+\delta}{\Delta(1)} > 1+E \Rightarrow h(1) < 1$$

so the intermediate value theorem applies. Next, taking first derivative of $h(\cdot)$ with respect to α gives

$$\frac{\partial h(\alpha)}{\partial \alpha} = \frac{(1+E)}{2\delta} \frac{\theta p}{\kappa+1} \sum_{i \in N} \lambda_i \tau_i (1-\tau_i) p \, \frac{(\kappa+1)\tau_i p - \kappa}{(1-\tau_i p)\left[\tau_i p + \alpha(1-\tau_i p)\right]^2} < 0 \tag{A.11}$$

where the last inequality applies as $p \max_{i \in N} \tau_i < \frac{\kappa}{\kappa+1}$ by assumption. □

Proof of Proposition 2. From (A.10), it is clear that if $p \min_{i \in N} \tau_i > \frac{\kappa}{\kappa+1}$, $\Delta(0) < \Delta(1)$. Then, there is clearly a $\hat{\delta}$ such that for $\delta < \hat{\delta}$, $\frac{1-\delta}{\Delta(0)} > \frac{1+\delta}{\Delta(1)}$ and so there can be no interior equilibrium. Then, the remainder of the Proposition follows by inspection of (9). □

Equilibrium In the Limit Case as $\delta \to 0$. Applying Propositions 1 and 2, we see that when (10) holds, then:

$$\alpha^* = \begin{cases} 1 & \text{if} \quad 1+E \geq \frac{1}{\Delta(1)} \\ \Delta^{-1}\left(\frac{1}{1+E}\right) & \text{if} \quad \frac{1}{\Delta(0)} < 1+E < \frac{1}{\Delta(1)} \\ 0 & \text{if} \quad 1+E \leq \frac{1}{\Delta(0)} \end{cases} \tag{A.12}$$

If on the other hand, beliefs are not precise, then

$$\alpha^* = \begin{cases} 1 & \Leftrightarrow \quad 1+E \geq \frac{1}{\Delta(1)} \\ 0 & \Leftrightarrow \quad 1+E \leq \frac{1}{\Delta(0)} \end{cases}$$

so there are multiple corner equilibria over the range $\frac{1}{\Delta(1)} < 1+E < \frac{1}{\Delta(0)}$. The interpretation of α^* in the limit is that it is a *mixed strategy* in the *limit game* where $u_{G,t} = u_{B,t} = 1$ i.e. are not random. Specifically, the bad politician randomizes between actions 1 and 0 with probabilities $\alpha^*, 1-\alpha^*$ respectively and the good politician does the same when $s=0$. These are not the only equilibria in the limit game, where there are a continuum of mixed-strategy equilibria, but this mixed-strategy equilibrium is of particular interest as it is consistent (via purification) with a model where politicians have identical perturbed payoffs from their short-run optimal actions. All comparative statics results in Section 6 above apply, appropriately restated, to this mixed-strategy equilibrium. $\square$

Proof of Proposition 3. (i) Note from (8) that $\Delta(0)$ is increasing in κ. Also, recall that the parameter restriction for the corner equilibrium at $\alpha^*=0$ is $1+E \leq \frac{1-\delta}{\Delta(0)}$. Combining these two facts gives the result.

(ii) An interior equilibrium, $\alpha^* \in (0,1)$ is characterized by the condition that

$$\alpha^* = \frac{\Delta(\alpha^*,\kappa)(1+E)-(1-\delta)}{2\delta} \tag{A.13}$$

So, totally differentiating (A.13) :

$$\frac{\partial \alpha^*}{\partial \kappa} = \frac{\partial \Delta/\partial \kappa}{2\delta/(1+E)-\partial\Delta/\partial\alpha} \tag{A.14}$$

We already know that $\partial\Delta/\partial\alpha < 0$ from (A.11). Moreover, from (8), the first derivative of $\Delta(\alpha^*)$ with respect to κ is :

$$\frac{\partial\Delta(\alpha^*,\kappa)}{\partial\kappa} = \theta p \left(\frac{1-\alpha}{(1+\kappa)^2}\right)\sum_{i\in N}\lambda_i \frac{\tau_i(1-\tau_i)p}{[\tau_i p+\alpha(1-\tau_i p)](1-\tau_i p)} \geq 0. \tag{A.15}$$

So, combining (A.14) and (A.15) gives the result.

(iii) Note from (8) that $\Delta(1)$ is independent of κ. Thus, the conditions for existence of a corner equilibrium $\alpha^*=1$, i.e. $1+E \geq \frac{1+\delta}{\Delta(1)}$, are unchanged. $\square$

Proof of Proposition 4. (i) Note that we can write $\Delta(\alpha)$ as

$$\Delta(\alpha;\boldsymbol{\lambda},\boldsymbol{\tau}) = \theta p\left(\frac{\kappa+\alpha}{\kappa+1}\right)\sum_{i\in N}\lambda_i g(\alpha;\tau_i) \text{ where } g(a;\tau) = \frac{\tau(1-\tau)p}{[\tau p+\alpha(1-\tau p)](1-\tau p)} \tag{A.16}$$

Also, throughout, let $(\boldsymbol{\lambda}',\boldsymbol{\tau}')$ be a mean-preserving spread of $(\boldsymbol{\lambda},\boldsymbol{\tau})$. By standard results, if $g(a;\tau)$ is strictly concave in τ for fixed a, then $\Delta(\alpha;\boldsymbol{\lambda}',\boldsymbol{\tau}') < \Delta(\alpha;\boldsymbol{\lambda},\boldsymbol{\tau})$. Taking second derivative of g with respect to τ gives

$$\frac{\partial^2 g(a;\tau)}{\partial\tau^2} = \frac{2p\Pi(\tau,\alpha)}{[\tau p+\alpha(1-\tau p)]^3(1-\tau p)^3}. \tag{A.17}$$

where

$$\Pi(\tau,\alpha) = (p-1)[\tau p+\alpha(1-\tau p)]^3 - \alpha(1-\alpha)(p+\alpha-p\alpha)(1-\tau p)^3.$$

As $p\in(0,1)$, we have $p-1<0$ and the first term on the RHS is strictly negative. Also, the second term is non-negative for all α. So, we conclude from (A.17) that $\frac{\partial^2 g(a;\tau)}{\partial \tau^2} < 0$ for all $\alpha \in [0,1]$ and thus $\Delta(\alpha; \boldsymbol{\lambda}', \boldsymbol{\tau}') < \Delta(\alpha; \boldsymbol{\lambda}, \boldsymbol{\tau})$.

(ii) From part (i), we have $\Delta(0; \boldsymbol{\lambda}', \boldsymbol{\tau}') < \Delta(0; \boldsymbol{\lambda}, \boldsymbol{\tau})$. Also, recall that the parameter restriction for the corner equilibrium at $\alpha^* = 0$ is $1 + E \leq \frac{1-\delta}{\Delta(0)}$. Combining these two facts gives the result in Proposition 4(a).

(iii) For a fixed $(\boldsymbol{\lambda}, \boldsymbol{\tau})$, an interior equilibrium, $\alpha^* \in (0,1)$ is characterized by the condition that

$$\alpha^* = \frac{\Delta(\alpha^*; \boldsymbol{\lambda}, \boldsymbol{\tau})(1+E) - (1-\delta)}{2\delta} \tag{A.18}$$

So, a mean-preserving spread of $(\boldsymbol{\lambda}, \boldsymbol{\tau})$ reduces $\Delta(\alpha^*; \boldsymbol{\lambda}, \boldsymbol{\tau})$ and thus the RHS of (A.18). Moreover, we already know that $\Delta_\alpha < 0$ from (A.11). So, α^* must fall to preserve the equality in (A.18), following the mean-preserving spread, proving Proposition 4(b).

(iv) From part (i), we have $\Delta(1; \boldsymbol{\lambda}', \boldsymbol{\tau}') < \Delta(1; \boldsymbol{\lambda}, \boldsymbol{\tau})$. Also, recall that the parameter restriction for the corner equilibrium at $\alpha^* = 1$ is $1 + E \geq \frac{1+\delta}{\Delta(1)}$. Combining these two facts gives the result in Proposition 4(c). $\square$

Proof of Proposition 5. (i) Recall that the condition for a corner with $\alpha^* = 0$ is $1 + E \leq \frac{1-\delta}{\Delta(0)}$. From (A.5),

$$\Delta(0) = \theta p \left(\frac{\kappa}{\kappa+1}\right) \sum_{i \in N} \lambda_i \frac{(1-\tau_i)}{(1-\tau_i p)}$$

Then

$$\frac{\partial}{\partial \tau_i} \Delta(0) = \theta p \left(\frac{\kappa}{\kappa+1}\right) \lambda_i \frac{p-1}{(1-\tau_i p)^2} < 0. \tag{A.19}$$

So, $\frac{1-\delta}{\Delta(0)}$ is increasing in τ_i and the result in part (a) follows.

(ii) Recall that the condition for a corner with $\alpha^* = 1$ is $1 + E \geq \frac{1+\delta}{\Delta(1)}$. From (A.5),

$$\Delta(1) = \theta p \left(\frac{\kappa}{\kappa+1}\right) \sum_{i \in N} \lambda_i \frac{\tau_i(1-\tau_i)p}{(1-\tau_i p)} \implies \frac{\partial}{\partial \tau_i} \Delta(1) = \theta p^2 \left(\lambda_i \frac{1 - 2\tau_i + p\tau_i^2}{(1-\tau_i p)^2}\right). \tag{A.20}$$

The quadratic $1 - 2\tau_i + p\tau_i^2 = 0$ numerator admits one solution in $(0,1)$ which is given by $\tau^* = \frac{1-\sqrt{1-p}}{p}$. As $p > 0$, we see from (A.20) that $\Delta(1)$ is strictly increasing in $\tau_i < \tau^*$ and decreasing in $\tau_i \geq \tau^*$, and the result in part (c) follows.

(iii) Finally, consider the case where $\alpha^* \in (0,1)$, that is

$$\alpha^* = \frac{\Delta(\alpha^*; \tau)(1+E) - (1-\delta)}{2\delta}.$$

So, totally differentiating (9) :

$$\frac{\partial \alpha^*}{\partial \tau_i} = \frac{\partial \Delta / \partial \tau_i}{2\delta/(1+E) - \partial \Delta / \partial \alpha} \tag{A.21}$$

Now we know $\frac{\partial \Delta}{\partial \alpha} < 0$, so from (A.21), $\frac{\partial \alpha^*}{\partial \tau_i}$ has the sign of

$$\frac{\partial \Delta(\alpha^*)}{\partial \tau_i} = \theta p^2 \left(\frac{\kappa + \alpha^*}{\kappa + 1} \right) \lambda_i \frac{A(\tau_i)}{(1 - p\tau_i)^2 (p\tau_i + (1 - p\tau_i)\alpha^*)^2}. \tag{A.22}$$

where

$$A(\tau_i) = (2\alpha^* p + p^2 - \alpha^* p^2 - p)\tau_i^2 - 2\alpha^* \tau_i + \alpha^*$$

Furthermore we have

$$A(0) = \alpha^* > 0, \; A(1) = -(1 - p)\left(\alpha^*(1 - p) + p\right) < 0.$$

as $p < 1$. As $A(\tau_i)$ is a quadratic function of τ_i there exists a unique $\tau^*(\alpha^*) \in (0, 1)$ such that $A(\tau^*) = 0$. So, from (A.22), $\frac{\partial \Delta}{\partial \tau_i}$ is strictly positive for $\tau_i < \tau^*(\alpha^*)$ and strictly negative for $\tau_i > \tau^*(\alpha^*)$, which implies from (A.21) that $\frac{\partial \alpha^*}{\partial \tau_i}$ has the same properties. $\square$

Proof of Proposition 6. From inspection of (A.19) we see that when $p = 1$, $\frac{\partial \Delta(0)}{\partial \tau_i} = 0$ which proves part (a). Similarly, setting $p = 1$ in part (c) of Proposition 5 proves part (c). Finally, if $p = 1$, $A(\tau_i) = \alpha^*(1 - \tau_i)^2 > 0$ in the proof of part (b) of Proposition 5, which implies $\frac{\partial \Delta}{\partial \tau_i} > 0$ and thus $\frac{\partial \alpha^*}{\partial \tau_i} > 0$. $\square$

Proof of Proposition 7. In equilibrium, the good type will choose $a = 1$ with probability $p + (1 - p)\alpha^*$, and the bad type will choose $a = 1$ with probability α^*. So, the overall probability that $a = 1$ is chosen is $\pi p + \alpha^* (1 - \pi p)$. Then, it follows that

$$\begin{aligned} \Pr(\text{re}) &= r(0) + [\pi p + \alpha^* (1 - \pi p)]\Delta(\alpha^*), && \text{(A.23)} \\ \Pr(\text{re}|G) &= r(0) + [p + (1 - p)\alpha^*]\Delta(\alpha^*), && \text{(A.24)} \end{aligned}$$

using $\Delta(\alpha^*) = r(1) - r(0)$. Then substituting (A.23),(A.24) into (13), into we can write selection effectiveness as

$$\Pr(G|\text{re}) = \pi \frac{r(0) + (p + (1 - p)\alpha^*)\Delta(\alpha^*)}{r(0) + (\pi p + (1 - \pi p)\,\alpha^*)\Delta(\alpha^*)} \tag{A.25}$$

In a pooling equilibrium $\alpha^* = 1$. So changes in trust parameter has no effect on discipline in a pure pandering equilibrium.

(i) First notice that

$$\frac{d}{d\kappa} \Pr(G|\text{re}) = \underbrace{\frac{\partial \Pr(G|\text{re})}{\partial \kappa}}_{\text{direct effect}} + \underbrace{\frac{\partial \Pr(G|\text{re})}{\partial \alpha^*} \frac{d\alpha^*}{d\kappa}}_{\text{feedback effect through } \alpha^*}.$$

At the corner equilibrium $\alpha^* = 0$, an increase in τ_i does not change α^* and as the feedback effect is equal to zero, we have

$$\frac{d}{d\kappa} \Pr(G|\text{re}) = \frac{\partial}{\partial \kappa} \Pr(G|\text{re}),$$

that is, the effect of trust on selection effect at $\alpha^* = 0$ depends only on the sign of the direct effect.

Next, we evaluate the direct effect of κ on selection effect:

$$\frac{\partial}{\partial \kappa} \Pr\left(G \,|\text{re}\right) = p\pi(1-\pi)\frac{r(0)\frac{\partial \Delta(\alpha^*)}{\partial \kappa} - \Delta(\alpha^*)\frac{\partial r(0)}{\partial \kappa}}{[r(0) + (\pi p)\Delta(\alpha^*)]^2}.$$

From (A.15), $\Delta(\alpha^* = 0)$ is strictly increasing in κ. So we need to verify that $r(0)$ is strictly decreasing in κ. Taking partial derivative of $r(0)$ with respect to κ gives

$$\frac{\partial r(0)}{\partial \kappa} = \theta \sum_{i \in N} \lambda_i \left(\frac{\partial V_i^I(0)}{\partial \kappa} - \frac{\partial V_i^C(0)}{\partial \kappa} \right) \text{ where } \frac{\partial V_i^I(0)}{\partial \kappa} = 0, \quad \frac{\partial V_i^C(0)}{\partial \kappa} = p\frac{\tau_i - P_i(0)}{(\kappa+1)^2} > 0,$$

which implies $r(0)$ is strictly decreasing in κ. Hence, selection is strictly increasing in κ when $\alpha^* = 0$.

(ii) Next we check how selection changes with trust heterogeneity when $\alpha^* = 0$. Consider a mean-spreading spread $(\boldsymbol{\lambda'}, \boldsymbol{\tau'})$ of the electorate trust distribution $(\boldsymbol{\lambda}, \boldsymbol{\tau})$. Holding $\alpha^* = 0$ and other parameters of the model constant, from (A.17), we have $\Delta(\boldsymbol{\lambda'}, \boldsymbol{\tau'}) < \Delta(\boldsymbol{\lambda}, \boldsymbol{\tau})$ at $\alpha^* = 0$. Next we will need to evaluate the effect of trust heterogeneity on $r(0)$. We can write

$$r(0) = \frac{1}{2} + \theta \sum_{i \in N} \lambda_i \left[V_i^I(0) - V_i^C(0) \right] = \frac{1}{2} + \theta \left(\frac{\kappa}{\kappa+1} \right) \sum_{i \in N} \lambda_i P_i(0) + \theta \left[1 - p - p\frac{\overline{\tau}\kappa}{\kappa+1} \right].$$

Finally for $\alpha^* = 0$ we have

$$P_i(0) = \frac{\tau_i(1-p)}{1-p\tau_i} \text{ which is a convex function of } \tau_i \text{ for } p \in (0,1).$$

Hence, we have $r(0)$ is higher for a mean spreading spread $(\boldsymbol{\lambda'}, \boldsymbol{\tau'})$ of $(\boldsymbol{\lambda}, \boldsymbol{\tau})$, which implies selection is strictly decreasing in trust heterogeneity.

(iii) At the corner equilibrium $\alpha^* = 0$, an increase in τ_i does not changes α^* and as the feedback effect is equal to zero, we have

$$\frac{d}{d\tau_i} \Pr\left(G \,|\text{re}\right) = \frac{\partial}{\partial \tau_i} \Pr\left(G \,|\text{re}\right),$$

that is, the effect of trust on selection effect at $\alpha^* = 0$ depends only on the sign of the direct effect. Next, we evaluate the direct effect of τ_i on selection effect:

$$\frac{\partial}{\partial \tau_i} \Pr\left(G \,|\text{re}\right) = p\pi(1-\pi)\frac{r(0)\frac{\partial \Delta(\alpha^*)}{\partial \tau_i} - \Delta(\alpha^*)\frac{\partial r(0)}{\partial \tau_i}}{[r(0) + (\pi p)\Delta(\alpha^*)]^2} \text{ where } \frac{\partial r(0)}{\partial \tau_i} = \theta \lambda_i \left(\frac{\partial V_i^I(0)}{\partial \tau_i} - \frac{\partial V_i^C(0)}{\partial \tau_i} \right).$$

From (A.1), (A.2), we have

$$\frac{\partial P_i(0)}{\partial \tau_i} = \frac{1-p}{(1-\tau_i p)^2} \implies \frac{\partial V_i^I(0)}{\partial \tau_i} = p \cdot \frac{\partial P_i(0)}{\partial \tau_i} > 0,$$
$$\frac{\partial V_i^C(0)}{\partial \tau_i} = \frac{\partial E[\pi_i^e \,|\, 0]}{\partial \tau_i}(v_G - v_B) = p(1-\alpha^*)(\frac{\kappa}{\kappa+1} + \frac{1}{\kappa+1} \cdot \frac{\partial P_i(0)}{\partial \tau_i}).$$

From Proposition 6, we also have that

$$\frac{\partial}{\partial \tau_i}\Delta(\alpha^* = 0) < 0 \;\text{ for } p \in (0,1) \;\text{ and }\; \frac{\partial}{\partial \tau_i}\Delta(\alpha^* = 0) = 0 \;\text{ for } p = 1.$$

Finally, notice that

$$\frac{\partial r(0)}{\partial \tau_i} > 0 \iff \frac{\partial}{\partial \tau_i}\left[V_i^I(0) - V_i^C(0)\right] > 0 \iff \tau_i > \frac{1-\sqrt{1-p}}{p}.$$

Thus, at $\alpha^* = 0$, for $p \in (0,1)$, the direct effect is positive and an increase in τ_i will increase the selection when trust is sufficiently high. Note that for $p = 1$, we have

$$\frac{\partial r(0)}{\partial \tau_i} < 0 \text{ as } \tau_i < 1, \quad \frac{\partial \Delta(\alpha^* = 0)}{\partial \tau_i} = 0,$$

which implies an increase in trust always decreases selection.

Proof of Proposition 8. First we simplify the expression V_B/V_G:

$$\frac{V_B}{V_G} = \frac{(1-\alpha^*+E) + [\alpha^* r(1) + (1-\alpha^*)r(0)](1+E)}{1+E+r(1)(1+E)} = 1 - \frac{\alpha^* + (1-\alpha^*)[r(1)-r(0)](1+E)}{[1+r(1)](1+E)}. \tag{A.26}$$

Case 1. Consider the case where $\alpha^* \in (0,1)$. As $\delta \to 0$, we have

$$\alpha^* = \Delta^{-1}\left(\frac{1}{1+E}\right) \implies \Delta(\alpha^*) \equiv r(1) - r(0) = \frac{1}{1+E}$$

and equation (A.26) can be simplified as

$$\frac{V_B}{V_G} = 1 - \frac{1}{[1+r(1)](1+E)}.$$

Under the assumption $p = 1$ we have

$$r(1) = r(1) - r(0) + r(0) = \frac{1}{1+E} + r(0) = \frac{1}{1+E} + \sum_{i \in N} \lambda_i \left[V_i^I(0) - V_i^C(0)\right].$$

From (A.1), (A.2), we have

$$V_i(0) = \underbrace{P_i(0)}_{=0} - (1-\alpha)\frac{\tau_i \kappa}{\kappa+1} - \alpha \implies r(1) = \frac{1}{1+E} - (1-\alpha^*)\frac{\tau_m \kappa}{\kappa+1} - \alpha^*.$$

Notice that $r(1)$ depends on trust heterogeneity only through α^* and

$$\frac{d\,r(1)}{d\tau_i} = \underbrace{\frac{\partial\, r(1)}{\partial \tau_i}}_{(-)} + \underbrace{\frac{\partial\, r(1)}{\partial \alpha}}_{(-)} \cdot \underbrace{\frac{d\,\alpha^*}{\partial \tau_i}}_{(+)} < 0, \quad \frac{d\,r(1)}{d\kappa} = \underbrace{\frac{\partial\, r(1)}{\partial \kappa}}_{(-)} + \underbrace{\frac{\partial\, r(1)}{\partial \alpha}}_{(-)} \cdot \underbrace{\frac{d\,\alpha^*}{\partial \kappa}}_{(+)} < 0.$$

For $\alpha^* \in (0,1)$, the payoff ratio V_B/V_G is strictly decreasing in τ_i, κ and strictly increasing in trust heterogeneity. Hence, π^* is strictly increasing in τ_i, κ and strictly decreasing in trust heterogeneity.

Case 2. Consider the case where $\alpha^* = 0$. Then we have

$$\frac{V_B}{V_G} = 1 - \frac{r(1) - r(0)}{1 + r(1)}.$$

Substituting $p = 1$ and $\alpha = 0$ into (8) gives

$$r(1) - r(0) = \frac{\theta\kappa}{\kappa + 1} \quad \text{and} \quad r(1) = \frac{1}{2} + \theta \sum_{i \in N} \lambda_i \left[V_i^I(1) - V_i^C(1)\right] = \frac{1}{2} + \theta(1 - \frac{\overline{\tau}\kappa}{\kappa + 1}).$$

It is straightforward to check that V_B/V_G is strictly decreasing in τ_i and independent of trust heterogeneity. Finally notice that $r(1)$ is strictly decreasing while $r(1) - r(0)$ is strictly increasing in κ, hence V_B/V_G is strictly decreasing in κ. □

Proof of Proposition 9. For $\alpha^* = 1$, we have

$$\frac{V_B}{V_G} = 1 - \frac{1}{[1 + r(1)](1 + E)}$$

Recall that for $p = 1$, we have

$$V_i^I(1) - V_i^C(1) = P_i(1) - 1 = \tau_i - 1 \implies r(1) = \frac{1}{2} + \theta \sum_{i \in N} \lambda_i \left[V_i^I(1) - V_i^C(1)\right] = \frac{1}{2} + \theta(\tau_m - 1),$$

which does not depend on κ and trust heterogeneity. As $r(1)$ is strictly increasing τ_i, the payoff ratio V_B/V_G is also strictly increasing in τ_i, which implies π^* is strictly decreasing in τ_i. □