

Partisanship as a Self-Fulfilling Prophecy*

Anke Kessler[†]

Benno Bühler[‡]

January 2025

Abstract

This paper provides a new perspective on why political leaders may adopt partisan positions and maintain them over time, even in the face of conflicting evidence. We study a stylized dynamic framework in which politicians are better informed than the voting public about an underlying state of the economy that determines the desirability of a given policy measure. Importantly, the issue itself is entirely non-partisan, that is, everyone has the same policy preferences. Still, assuming that voters attach ideological labels to both candidates (e.g. through party affiliation) and available policy alternatives, we show that both sides may be caught in an *partisan trap* of self-fulfilling expectations: because voters expect the identity of office holders to shape their political actions, politicians are tempted to act according to their perceived partisan identities, i.e., the latter act like ‘sunspots’ that determine subsequent policy choices only because voters think they do. We show that the resulting partisan equilibrium is one of political failure and policy persistence. We also consider some extensions to show robustness of the findings.

Keywords: Partisanship, Elite Polarization, Policy Persistence, Dynamic Political Competition

JEB classification: D 78, P16, H11

*We wish to thank Tim Besley, Helios Herrera, Martin Peitz, Torsten Persson, François Salanie, Al Slivinski, Klaus Schmidt, Jean Tirole and various participants of the CIFAR meeting in Calgary and seminars at UQUAM and Kellogg for valuable comments. Financial support from German Science Foundation and German Academic Exchange Service (Benno Bühler) and the Canadian Institute for Advanced Research as well as SSHRC (Anke Kessler) is gratefully acknowledged. An earlier version of this paper was circulated under the name of “Ideologues: Explaining Partisanship and Persistence in Politics (and Elsewhere)”, CEPR Discussion Paper # 7724.

[†]Corresponding Author. Simon Fraser University. Address of Correspondence: Simon Fraser University, Department of Economics, 8888 University Drive, Burnaby, B.C. V5A 1S6, Canada. E-mail: akessler@sfu.ca

[‡]Charles River Associates, Avenue Louise 143, 1050 Brussels, Belgium. E-mail: bbuehler@crai.com. The views expressed in this article are personal, they do not necessarily represent the opinions of CRA, or any client of CRA.

Partisanship is our great curse. We too readily assume that everything has two sides and that it is our duty to be on one or the other.

James Harvey Robinson (American historian, 1863-1936)

1 Introduction

Consistent with the above quote, there is evidence that voters do not like partisan politics.¹ Nonetheless, partisan politics is generally what voters get. As Poole & Rosenthal (2007), McCarty et al. (2006) and others have extensively documented, voting in Congress can largely be reduced to a single dimension which mirrors party affiliation: the labels ‘conservative’ (Republican) and ‘liberal’ (Democrat) accurately predict a politician’s stance on most policy issues, and elite polarization has been increasing over time. The phenomenon of rising partisanship in politics is neither exclusive to the U.S. () nor is it confined to moral or cultural issues, or policies where voter preferences naturally vary with, e.g. socio-economic status, race, gender, or religion. Partisan politics are a frequent phenomenon even in situations where there should be a general agreement among the electorate such as crime, corruption, or foreign policy. In the U.S. Congress, for example, support for the president on matters of foreign policy and defense has largely been along party lines ever since the Vietnam War (Meernik, 1993).

Why then do officeholders consistently choose the same persistent policy approaches, even in the face of changing circumstances, especially when voters indicate such strong preferences for compromise and flexibility in their representatives? The present paper seeks to shed light on this question from a novel perspective. Using a simple dynamic model that ties observable characteristics of political representatives to voters’ expectations, we show that political elites adopt a partisan policy and stick to it even in the face of contradicting evidence, simply because voters *expect* them to. Referring back to the above quote by James Harvey Robinson, it is the assumption of partisanship that generate it in the first place. The theory implies, for instance, that a female Democrat from California is likely to take a liberal stance on a variety of issues from taxes, healthcare and education to reproductive rights, family values, and gender issues, not because her true preferences or her belief system necessarily reflects those positions, but because her constituents expect a female Democrat from California to be a ‘Liberal’ (and elected her for this very reason).

We start with the observation that voters don’t always know which policy is best. Officeholders, in contrast, have better ideas about what makes a good policy due to their greater expertise, access to resources and greater incentives to become informed.² Specifically, incumbents learn about an underlying state that determines the efficacy of policy alternatives while voters do not. Given beliefs about the prevailing state, voters form expectations about a) which policy candidates are likely to implement once in office, and b) which of those is most likely to succeed. To develop our argument in the strongest manner possible, we assume that candidates running for office have no partisan preferences: they each derive the exact same utility from the

¹A recent (2024) Gallup poll finds US Congress approval ratings to be at an historic 13 percent low, with “party gridlock/bickering/not compromising” cited as the number one reason for disapproval.

²See e.g. Cukierman & Tommasi (1998) or Maskin & Tirole (2004). For a model of endogenous policy expertise, see Kessler (2005) and Callander et al. (2008)

policy measure, which is also identical to utility of the electorate at large.³ Hence, information on candidate identity (partisan affiliation, gender, geographic location, etc.) is payoff-irrelevant.

Our main finding is that, nevertheless, candidate identity matters: different candidates have, once in office, an incentive to adopt a different partisan position in their policy choice and maintain it over time. The argument is one of self-fulfilling expectations and runs as follows. Suppose voters base their expectations how a candidate will act once in office on her identity, for example, party affiliation or historical ‘ideology’. Specifically, all voters assume that candidates will act partisan once in office, i.e., they associate different policy choices with different parties or ideologies (there is a ‘liberal’ and a ‘conservative’ choice) and expect candidates to remain ‘true to their colors’ in implementing the respective policies affiliated with their party or ideology as perceived by the voting public. Given these expectations, voters have a straightforward incentive to elect the representative whose perceived partisan policy (ideology) corresponds to what they think is in their best interest based on their current information. This, in turn, can suffice to induce candidates to actually act partisan, i.e., according to their observed identity, in the first place. The specific motivation is one of signal-jamming: an incumbent who sticks to her partisan policy avoids revealing that current circumstances would favour her opponents’ partisan position, making her re-election more likely if voters expect partisan behaviour in the future.⁴

By implementing her partisan policy even in states where it is suboptimal, a sufficiently office-motivated incumbent rationally demonstrates confidence that present circumstances warrant her policy choice. A lucky policy outcome will then cement the electorate’s belief and improves electoral prospects in the next election. The result is political failure in the sense that the equilibrium partisan policy outcomes are Pareto dominated. Thus, the model can explain policy bias and divergence based solely on the fact that voters *perceive* policies to be ideologically tinted and *expect* candidates to act partisan. Both sides are caught in an ‘partisan trap’ of self-fulfilling expectations: because voters expect the ideology of office holders to determine their political actions, an official’s (re-)election chances will vary with his or her perceived ideology. In their desire to influence the outcome of the election, these expectations induce the officials to act partisan. As we show in our stylized framework, the issue itself can be perfectly non-partisan, meaning that neither voters nor politicians have to display any intrinsic preferences for either policy: a leader does not have to be a “true believer” to be partisan. Instead, partisan identity is purely a social perception or label, based on observable attributes such as gender, electoral district, or their party affiliation. Since a candidates’ stance on other issues is one of those attributes, the model also generates issue bundling: a politician who is strongly pro-life may also have an incentive to campaign against government-provided health care for no other reason than voters’ expectations tying the two issues together. Finally, because incumbents will tend to enact the partisan policy independent of the prevailing state in equilibrium, our analysis has another interesting implication. It explains why incumbent politicians will maintain their ideology and deny conflicting evidence, resulting in policies that are likely to persist.

³It should be emphasized that our insight also applies to partisan (positional) issues. There already is an extensive literature on these type of policies, however, which provides a range of complementary explanations for why candidates diverge in platforms and voting records. We refer to this literature in more detail below.

⁴Alesina & Cukierman (1990) study an environment in which voters are unsure about the ideological position of candidates (as opposed to the state of the economy as in our paper). Akin to the signal-jamming effect we find, they show that politicians may want to deliberately choose ‘ambiguous’ policies in order to conceal their true preferences, thereby keeping their ideological advantage.

Related Literature.

Our model is highly stylized and as such, cannot possibly account for the observed richness of partisanship in the U.S. or elsewhere. A considerable amount of work has produced theoretical models that help explain various aspects of why elite polarization emerges in the first place and why it may increase over time. Rather, we seek to complement this literature by illustrating that voters' expectations on partisan behavior can be self-fulfilling.⁵

Specifically, our results add to the body of research studying political conflict among candidates and polarization. Factors that have been identified as contributing to policy divergence include the multi-dimensional issues (Ansolabehere & Snyder, 2000), the threat of third-party entry (Palfrey, 1984), politically motivated candidates (Besley & Coate (1997); Osborne & Slivinski (1996)), valence advantage of one candidate over another (Aragones & Palfrey, 2002; Buisseret & Van Weelden, 2022; Groseclose, 2001), rent-seeking office holders (Van Weelden, 2013), party discipline (Canen et al., 2020), and incomplete information among candidates or voters (Callander (2008); Castanheira (2003); Kartik & McAfee (2007); Ghosh & Tripathi (2012); Martinelli (2001); McMurray (2022)).⁶ Those explanations generally build on pre-existing differences in policy preferences among candidates and/or the electorate. Other contributions exploit the idea that (strategic) interaction between a non-partisan issue and other considerations can lead to platform diverges. Ashworth & Bueno de Mesquita (2009) show that candidates may want to take partisan stances to soften competition in other dimensions such as campaign spending. Howell et al. (2020) show how preference differences over the value of the public good and the induced disagreement over how much of the public good to provide can naturally generate a long-term electoral advantage of one candidate or party over another. As a result, political conflict is inevitable if the value of holding office is sufficiently large even if all voters and candidates *a priori* agree on project implementation. Callander & Carbajal (2022) consider a dynamic model where voters' preferences change in response to their voting behavior and show that this creates a strategic incentive for parties to diverge; a feed-back loop emerges with (progressive) polarization among both elites and the electorate that matches many aspects of observed polarization trends. In Izzo et al. (2023), policy-motivated parties develop ideologies that once adopted by voters influence what they infer from observed outcomes. Winning the ideological contest thus confers a long term electoral advantage and ideological polarization always emerges in equilibrium.

Our argument is also related to, but nevertheless distinct from, existing work on political signalling and policy failures in scenarios where candidates or parties are better informed than voters. In such a situation, Cukierman & Tommasi (1998) show that candidates electoral prospects may increase the more atypical is the policy he proposes to implement. Harrington (1993), Canes-Wrone et al. (2001) and more recently, Maskin & Tirole (2004) Stasavage (2007), and Callander (2008), Morelli & Weelden (2013) consider models where candidates pander to public opinion, ignoring socially valuable information that goes against the ex ante preferred action of the majority. These models focus on distortions which are rooted in the politician's incentive to signal 'congruence' with voters' preferences or their own competence, while we emphasize

⁵For a discussion of the large empirical literature on elite polarization and further references, see e.g. McCarty et al. (2006).

⁶Another line of research has focused on explaining the prevailing polarization on cultural or moral issues, such as abortion or gay marriage. Glaeser et al. (2005) identify a form of strategic extremism, which helps politicians to induce their core constituents to vote (or make donations). Enke et al. (2022) show that if voters' values are luxury goods, rising incomes will lead to more party polarization on social issues.

policy failure stemming from the incentives to jam voters’ inference problem – instead of pandering to public opinion, they manipulate public opinion through a persistent, partisan policy choice. The idea that elected officials may take distortionary actions to prevent voters learning has been explored in previous work. Aghion & Jackson (2016) investigate optimal tenure and term limit mechanisms in a model where incompetent leaders seek to hide their type by taking uninformative (conservative) actions. In Ashworth et al. (2017), the success of a policy serves as a signal of the incumbent’s quality, and the signal depends on the effort she put into implementing the policy. They show that higher level of effort (higher effective electoral accountability) can lead to less informed re-election choices and decreased welfare when politicians’ actions and characteristics are local substitutes.

Finally, the self-fulfilling nature of the partisanship equilibrium is reminiscent of arguments that have been put forward to explain bubbles in financial markets (Tirole, 1985) or persistent booms or recessions in the macro economy (e.g., Azariadis, 1981). The latter literature in particular studies so-called sunspot equilibria which are defined as rational expectations equilibria in which outcomes are perfectly correlated with non-fundamentals, i.e. ‘sunspots’ (see Azariadis & Guesnerie, 1986; Cass & Shell, 1983). In our model, (all) voters make an association between a candidates identity – based on observables such as ethnicity, gender, geographic residence, or party affiliation – and the policy that they expect the candidate to pursue once in office. This association is arbitrary because it is not based on payoffs, and is the driving force behind the self-fulfilling partisanship that emerges. In a sense, therefore, candidates’ identities are similar to (non-random) sunspots that pin down expectations.

The remainder of the paper is organized as follows. The basic framework is developed in Section 2. Section 3 provides an in-depth analysis of the model, and shows that both partisanship and non-partisanship can arise in equilibrium. Section 4 considers two extensions. We first demonstrate that our model uniquely predicts which of these equilibria occurs if candidates have arbitrary small biases towards their partisan policy. Second, we show that partisan behavior becomes even more plausible if the prospects of inefficient policies are themselves uncertain. Section 5 concludes.

2 The Model

2.1 Basic Framework

We consider an infinite-horizon economy in discrete time, which is populated by a representative risk-neutral consumer-voter whose period- t payoff v_t stochastically depends on the state of the economy s_t and the policy choice a_t of an elected official. For simplicity, we assume that the policy decision is binary, $a_t \in \{b, r\}$, with policy alternatives ‘red’ (r) and ‘blue’ (b) respectively.⁷ Without loss of generality, we also assume that the

⁷Provided the policy issue is one-dimensional, the binary assumption could easily be relaxed. One-dimensionality assumes that any given policy alternative can be placed in the blue-red (left-right) spectrum. (Poole & Rosenthal, 1997, 2007) show, over 80 percent of representatives’ voting records over the past 40 years can be explained solely on the basis of a one-dimensional variable (i.e., their position in the spectrum).

underlying state is binary, $s_t \in \{b, r\}$, and that it is optimal to match the policy to the state. Specifically,

$$v(a_t, s_t) = \begin{cases} v & \text{with probability 1 if } a_t = s_t \\ v & \text{with probability } \pi \\ 0 & \text{with probability } 1 - \pi \end{cases} \quad \text{if } a_t \neq s_t$$

Thus, if the policy choice aligns with the state, the policy is successful with probability one.⁸ Voters know the set of feasible policies but are uncertain about the underlying s_t , which evolves according to a stationary, persistent Markov process with $Prob\{s_{t+1} = s_t\} = \gamma > \frac{1}{2}$. The representative voter's expected utility in t is $E \sum_{j=0}^{\infty} \beta^j v(a_{t+j}, s_{t+j})$ where $\beta < 1$ is the discount factor. Note that, by construction, the issue is *non-partisan* (ideologically neutral) in the sense that all voters agree on the best alternative: if voters knew the state to be s_t , they unanimously preferred the policy that is appropriate for the state, i.e., $a_t = s_t$. Since they do not know s , they form beliefs $\mu_t = Prob\{s_t = b\}$ based on the available information at time t and prefer policy b over policy r if and only if $\mu_t \geq \frac{1}{2}$.

The representative citizen-voter elects an office holder in each period, who selects and implements the policy alternative a_t . Unlike voters, politicians observe the state s . An office holder's informational advantage may reflect their greater expertise, better access to resources, or their greater incentive to acquire information⁹ There are two observable types of politicians, a blue politician B and Red politician R . We interpret the type $i \in \{B, R\}$ as politicians' party affiliation, but any other observable characteristic such as their gender, their home district, or their observed position on a different (unrelated) policy issue would work equally well. Consistent with our notion that the issue is non-partisan, politicians' preferences over policies are fully aligned with voters: they derive the same utility from the policy a given state s as their constituents, regardless of type i . A priori, therefore, a politician's type is completely unconnected to any given policy choice. As we will see, though, a connection emerges in equilibrium, justifying our related notation for policy labels and candidates' types. In addition to policies, candidates care about holding office, which we formalize by a rent $\phi > 0$ that politicians receive from being elected to office in period t . The per-period utility of an incumbent of type i in period t when the state is s_t is

$$u_t^i = v(a_t, s_t) + \phi. \quad (1)$$

Once defeated, a candidate never returns to holding office; instead, the current office holder faces a new challenger in the subsequent election. When not in office, politicians receive a continuation utility of zero.¹⁰

The timing of period t game is as follows. First, an elections are held in which the representative voter decides whether to re-elect the incumbent or to replace her with a challenger with a different party affiliation. Then, nature draws the s_t which is immediately revealed to the office holder but not to ordinary citizens. Next, the

⁸These simplifying assumptions of our model are made for analytical convenience only. In particular, the results that follow do not hinge on the fact that a failure perfectly reveals an inefficient policy choice. Similarly, our conclusions would be qualitatively unaffected if we introduced a small probability that voters observe the state of the world at the end of each period. Details are available from the authors upon request.

⁹See Kessler (2005) for an analysis where officials endogenously acquire competence on the issues they oversee and specialize in policy formation.

¹⁰Thus a politician does not benefit from the legacy of policies that are enacted by a successor. Under the alternative assumption that politicians remain policy-motivated after having been ousted from office (that is $u_t^i = v(a_t, s_t)$), our results remain qualitatively unchanged. In particular, the partisan equilibrium described below exists if the office rent ϕ is large enough.

office holder enacts a policy alternative $a(t)$ and payoffs are realized. Having observed whether the policy was a success ($v_t = v$) or a failure ($v_t = 0$), voters update their beliefs and a new period begins.

Equilibrium. Since there is generally an infinite number of equilibria in this kind of dynamic games, we restrict attention to pure strategy, stationary, and symmetric Markov perfect equilibria; in particular, players ignore all details of the history and condition their only on the pay-off relevant information. For the electorate, the latter can be summarized by its belief μ_t about the state at time t . A strategy for a representative citizen-voter specifies the probability $P^i(\mu_t) \in [0, 1]$ with which candidate i is elected, based on μ_t , with $P^B(\mu_t) + P^R(\mu_t) = 1$. When voters are indifferent between two candidates, we assume either stands equal chances of winning the election.

A strategy for a type- i candidate $a^i(\mu_t, s_t)$ maps the voter's current beliefs μ_t as the current state s_t into a policy choice $a_t \in \{l, r\}$. In equilibrium, strategies must be mutual best responses and beliefs evolve in a way consistent with Bayes rule whenever possible. Off equilibrium path, beliefs may not be pinned down by Bayes rule in which case we will use well-known refinements. Strategies are optimal if they maximize the value functions of candidates and voters, taking as given the strategies of others and beliefs. The value function for a representative voter can be written as

$$U(\mu_t) = \max_{P^i(\mu_t)} \sum_{i=B,R} P^i(\mu_t) E [v(a^i(\mu_t, s_t), s_t) + \beta U(\mu_{t+1})]. \quad (2)$$

The value function of a type i incumbent is

$$V^i(\mu_t, s_t) = \max_{a^i(\mu_t, s_t)} E [v(a^i(\mu_t, s_t), s_t) + \phi + \beta P^i(\mu_{t+1}) V^i(\mu_{t+1}, s_{t+1})]. \quad (3)$$

Note that beliefs μ_{t+1} at time $t + 1$ will depend on beliefs in period t , the elected representative and her strategy, the implemented policy, and the success or failure of that policy in t .

2.2 Equilibrium Analysis

The problem for the representative citizen-voter is that he does not observe the current state, only the policy chosen a_t^i and his resulting payoff v_t . The voter then uses that information to forecast which state of the economy is most likely next period. Given that forecast, voters may prefer to replace a type i incumbent if she is expected to implement the wrong policy the next period. The problem for the current office-holder is thus that the choice of a_t influences voters' beliefs about s_{t+1} and, possibly, her chances for re-election. Since we have deliberately ruled out partisan preferences for office holders, however, the sole channel through which party affiliation i can affect electoral success through voters' *expectations* about how a candidate's type i relates to her subsequent choice of a_t when in office. It is this link between today's policy choices, voters' expectations about candidates' future post-election behavior, and subsequent electoral success we are interested in. To isolate this effect and bring it to the forefront, we have eliminated all other well-studied determinants of partisan politics (partisan voters, partisan politicians etc.).

In what follows, we use the term *non-partisan* politics to characterize the Pareto-optimal policy choice, i.e., the office holder implements $a_t = s_t$, regardless of her type i . *Partisan politics*, in contrast, involve politicians selecting the alternative that corresponds to their party affiliation instead, i.e., $a_t = b$ in state $s_t = r$ if $i = B$ and $a_t = r$ in state $s_t = b$ if $i = R$.

2.2.1 The Non-Partisan (Efficient) Equilibrium

As a benchmark, we first construct an equilibrium in which candidates choose policies in a Pareto efficient manner along the equilibrium path, and voters – because they correctly expect non-partisan behavior from every type of representative – have no preferences for either type of politician. Thus, suppose the strategy of candidates is to choose $a_t^i = s_t$ in every period t when in office, irrespective of their party affiliation i or electoral beliefs μ_t . Since both types of politicians implement identical policies given their superior knowledge of the state, voters hold no preference for the incumbent or the challenger and elect either with probability $1/2$.

Given voter strategy, the incumbent office holder has a strict incentive to implement the non-partisan (efficient) alternative because

$$V^i(a_t = s_t) = v + \phi + \beta E \left[\frac{1}{2} V(s_{t+1}) \right] > V^i(a_t \neq s_t) = \pi v + \phi + \beta E \left[\frac{1}{2} V(s_{t+1}) \right],$$

for any $\pi < 1$, irrespective of type i . Note that $P^i(\mu_{t+1})$ is independent of μ_t given the electorate's voting rule under which any incumbent faces equal chances of being re-elected and defeated, respectively, regardless of beliefs or identity.

We can thus conclude that non-partisan politics and an electoral rule that assigns equal election chances to incumbents and challengers in all periods form an equilibrium. In fact, it is the Markov perfect equilibrium with the highest payoff to the electorate, $U^{\max} = \sum_{t=0}^{\infty} \beta^t v = \frac{1}{1-\beta} b$. Moreover, since implementation of an efficient policy alternative perfectly reveals s_t ; the only uncertainty about the underlying economy stems from the Markov process governing s_t .¹¹

Proposition 1. *[Non-Partisan Equilibrium] There always exists an equilibrium in which elected office holders act non-partisan and are re-elected with probability $1/2$. In this equilibrium, voters have full information about the prevailing state following the policy choice in each period, and receive the highest possible utility.*

While the non-partisan equilibrium always exists and payoff-dominates all other equilibria for the voters (though not necessarily for office holders, see below), it is not the only possible outcome. As the following section demonstrate, partisan politics be supported in equilibrium. Moreover, non-partisan politics are fragile in the sense that they cannot survive if citizens' expectations about office holders' behavior are subject to (small) uncertainty.

2.2.2 The Partisan Equilibrium

To see how partisanship can arise here, suppose voters' expect office holders to choose the partisan policy, matching the policy to their party affiliation i in every period instead of the state s_t . Formally, suppose the blue incumbent's strategy is to set $a_t = b$, irrespective of the state, and similarly for the red incumbent who sets $a_t = r$ whenever elected. Given this strategy, voters are no longer indifferent across politicians with distinct types: clearly, if the representative voter knew the state to be $s_t = b$ he would now *strictly prefer*

¹¹For any initial belief μ_0 , beliefs in this equilibrium therefore evolve according to $\mu_{t+1}(\cdot) = \gamma$ if $a_t = b$ and $\mu_{t+1}(\cdot) = 1 - \gamma$ if $a_t = r$.

a type- B candidate to a type- R candidate, and vice versa. A direct consequence of this strict preference ordering is that period- t incumbents now face a dilemma whenever their type does not align with the state. A type- B office holder who selects the non-partisan choice of $a_t = r$ would reveal the state to be $s_t = r$, and would not be re-elected given voters' expectations about future partisan politics. Likewise, a type- R incumbent who implemented the efficient left-wing alternative $a_t = l$ because the state was $s_t = l$ would face certain defeat. A partisan choice of $a_t \equiv i \neq s_t$, on the other hand, conceals the true state and thus may ensure – conditional on the observed success of the policy – re-election. It is then intuitive that this effect can induce partisan behavior provided politicians care sufficiently strong about their (re-)election prospects. The remainder of this section establishes this result formally.

Consider a blue politician incumbent $i = B$ whose strategy is to choose the partisan policy whenever in office in period t (the argument for a red politician $i = R$ is analogous). Given $\mu_0 \in [1 - \gamma, \gamma]$, the voters' beliefs along the equilibrium path evolve according to

$$\begin{aligned}\mu_{t+1}^B(a_t = b, v_t; \mu_t) &= \begin{cases} 1 - \gamma + (2\gamma - 1)\frac{\mu_t}{\mu_t + (1 - \mu_t)\pi} & \text{if } v_t = v \text{ (policy success)} \\ 1 - \gamma & \text{if } v_t = 0 \text{ (policy failure)} \end{cases} \\ \mu_{t+1}^R(a_t = r, v_t; \mu_t) &= \begin{cases} \gamma - (2\gamma - 1)\frac{1 - \mu_t}{1 - \mu_t + \mu_t\pi} & \text{if } v_t = v \text{ (policy success)} \\ \gamma & \text{if } v_t = 0 \text{ (policy failure)} \end{cases}\end{aligned}\quad (4)$$

In equilibrium, the office holders' policy choice reveals no new information about the current state and the electorate only learns by observing whether the policy has been successful or not.

Off the equilibrium path, i.e., when the electorate observes $a_t = r$ from a $i = B$ incumbent, we impose the natural assumption that non-partisan politics are perfectly revealing: $\mu_{t+1}^B(a_t = r, v_t = v; \mu_t) = 1 - \gamma$. That is, if the electorate unexpectedly observes a blue office holder to select a red policy, it beliefs that $s_t = r$ has occurred (unless $v_t = 0$, in which case the state must be $s_t = b$ with certainty). This belief is also the unique out-of-equilibrium belief that satisfies the concept of equilibrium dominance on which the Cho & Kreps intuitive criterion is based. In an extension of the model discussed below, we introduce a small modification that ensures every policy choice to be observed in equilibrium (there are no off-equilibrium path policy choices) and show that the conclusions are unaltered.

To analyze voter behavior, assume (as will be true in equilibrium) that the blue (red) candidate is elected for beliefs $\mu_t > 1/2$ ($\mu_t < 1/2$), with equal chances of winning for $\mu_t = 1/2$. The period- t value function of the electorate can be written as

$$U(\mu_t) = \begin{cases} \mu_t v + (1 - \mu_t)\pi v + \beta U(\mu_{t+1}^B) & \mu_t \geq 0.5 \\ (1 - \mu_t)v + \mu_t\pi v + \beta U(\mu_{t+1}^R) & \mu_t < 0.5. \end{cases}\quad (5)$$

where μ_{t+1}^i is given by (4). By inspection, voter payoff is increasing in μ_t for values $\mu_t > 1/2$, decreasing in μ_t otherwise, and assumes a minimum at $\mu_t = 1/2$: more extreme beliefs increase the benefit of electing the appropriate politician. Related to this property is that voters would never want to “experiment”, i.e., elect a candidate who subsequently is less likely to implement the efficient policy in order to gather more precise information about the state. Intuitively, doing so would only increase the chances of a policy failure, in which case voters would be even more convinced that the elected candidate was not appropriate.¹² Hence,

¹²Put differently, the electorate would dispose of a more accurate belief only if the implemented policy goes awry. In the

the citizen-voters' best response to blue candidate's strategy of partisan policy in every period is

$$P(\mu_t) = \begin{cases} 1 & \text{if } \mu_t > 0.5 \\ 0.5 & \text{if } \mu_t = 0.5 \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

It remains to see under which circumstances partisan politics are an optimal strategy for office holders. Assume without loss of generality that a *B*-type politician is in office in period t . The payoff from a partisan choice $a_t = b$ in every t irrespective of $s_t \in \{b, r\}$

$$V(\mu_t, s_t) = \begin{cases} v + \phi + \beta E P(\mu_{t+1}) [\gamma V(\mu_{t+1}, b) + (1 - \gamma)V(\mu_{t+1}, r)] & \text{if } s_t = b \\ \pi v + \phi + \beta E P(\mu_{t+1}) [(1 - \gamma)V(\mu_{t+1}, b) + \gamma V(\mu_{t+1}, r)] & \text{if } s_t = r \end{cases} \quad (7)$$

where μ_{t+1} evolves according to (4) given μ_t and the realization of v_t , and $P(\mu_{t+1})$ is given by (6)

Now consider a blue incumbent in state $s_t = r$ who deviates by setting the non-partisan (efficient) policy $a_t = r$. This will reveal to the electorate that the true state is r with corresponding beliefs $\mu_{t+1} = 1 - \gamma < \frac{1}{2}$, resulting in certain defeat and a utility normalized to zero. The payoff following a deviation from partisan play is therefore

$$\hat{V}(\mu_t, s_t = r) = b + \phi. \quad (8)$$

That rational voters should not reward efficient, non-partisan policy choices in equilibrium may seem counter-intuitive at first but it is strictly optimal, given they (correctly) anticipate subsequent partisan play. Since voters are perfectly informed about politicians' types (in particular, there are no 'partisan' or 'non-partisan' types), incumbents can also not signal an increased likelihood that they would play non-partisan in the future by deviating. Finally, note that no incumbent would ever want to select an opponent's partisan policy in a state where in fact her own partisan policy is myopically optimal: the strategy $a_t = r$ is dominated in state $s_t = b$ for a blue incumbent.

Comparing (7) and (8), we see that the partisan choice in every period is optimal if and only if

$$\beta E [\gamma V(\mu_{t+1}, b) + (1 - \gamma)V(\mu_{t+1}, r)] \geq (1 - \pi)v. \quad (9)$$

On the right-hand side of (9) are the short-term gains from deviating, as reflected in the additional expected benefit from the optimal non-partisan choice over the sub-optimal partisan choice. The left-hand side captures the utility lost by the corresponding electoral defeat: it is the expected future value from remaining in office. Noting that along the equilibrium path, a blue candidate only finds herself in office if $\mu_t \geq \frac{1}{2}$ and will always (respectively, never) be re-elected following a policy success (respectively, failure), it is easy to see that $V(\mu_{t+1}, s_t)$ is constant across beliefs and depends on s_t only.¹³ Formally, we have $V(\mu_t, s_t) \equiv \bar{V}(s)$

unlikely case of success on the other hand, the resulting belief is less precise than the one that would have resulted from having the appropriate candidate successfully implement his partisan policy. In the appendix, we formally establish how $U(\mu_t)$ depends on μ_t and show that experimentation does not improve voters' payoffs.

¹³Off equilibrium, it is possible that voters did not elect the politician whose type correspond to the state they consider most likely, e.g., if voters elected a blue incumbent for $\mu_t < \frac{1}{2}$. See the appendix for a full characterization of out-of-equilibrium behavior in this case.

$\forall \mu_t \in (0.5, 1]$ where s is the current (period- t) state. We can then explicitly compute $\bar{V}(b)$ and $\bar{V}(r)$ to read

$$\begin{aligned}\bar{V}(r) &= \frac{v\pi(1 + \beta(1 - 2\gamma\beta)) + (\pi\beta(1 - \gamma) + 1 - \beta\gamma)\phi}{\pi\beta(\beta(2\gamma - 1) - \gamma) + 1 - \beta\gamma} \\ \bar{V}(b) &= \frac{v(\pi\beta(1 - 2\gamma) + 1) + (1 - \beta(\pi\gamma + \gamma - 1))\phi}{\pi\beta(\beta(2\gamma - 1) - \gamma) + 1 - \beta\gamma}.\end{aligned}$$

Substituting for $\bar{V}(r)$ and $\bar{V}(b)$ in (9) yields

$$\frac{v}{\phi + v} \leq \frac{\pi\beta(1 - \beta(2\gamma - 1))}{(1 - \pi)(1 - \beta\gamma)}. \quad (10)$$

Thus, and not surprisingly, partisan behavior is more likely to arise whenever the future value of holding office (ϕ) is high or the opportunity cost of doing so ($v, 1 - \pi$) is low. For example, assuming candidates are very patient, $\beta \rightarrow 1$, it is easy to verify that (10) holds for every value of ϕ, v , and γ if $\pi \geq \frac{1}{3}$. Figure 1 in the appendix illustrates the (π, γ) parameter space for which condition (10) holds, for given values of v, β , and ϕ . It shows in particular that a more persistent state is not conducive to partisan behavior, all else equal. The intuition here is that since the incumbent faces the trade-off between re-election and efficiency only if the state is unfavorable, a more persistent state lowers the chances that the partisan policy will become efficient in the near future. The prospect of repeatedly having to implement inefficient policies lowers the expected value from staying in the office when the state is more persistent.

We summarize the discussion in

Proposition 2. *[Partisan Equilibrium] For parameter values that satisfy condition (10), there exists an equilibrium in which elected office holders act partisan regardless of the state. In this equilibrium, politicians are re-elected with probability one if their implemented policy was a success and face certain defeat if it was a failure, and voters receive no information about the prevailing state from the choice of policy (other than ex post from its success or failure).*

The partisan equilibrium is more likely to exist whenever the motive of office holding is strong (ϕ high), the environment is volatile (γ low) and whenever inappropriate policies are unlikely to fail (π high).

It is important to contrast the equilibrium behavior in Proposition 2 to the well-known danger of office-motivated representatives “pandering to public opinion”. Harrington (1993), Maskin & Tirole (2004), Callander (2008), Morelli & Weelden (2013) and others mentioned in the introduction consider situations where – because the electorate is unable to evaluate the official’s actions directly – the desire to be (re-)elected may lead representatives to pursue the most popular rather than the welfare maximizing, course of action. In our framework, incumbents in a partisan equilibrium have no incentive to follow the most popular course of action. Instead, they insist on maintaining a policy so as not to reveal that ‘times have changed’. Moreover, what is at the heart of the resulting policy bias is a perceived – as opposed to a real – non-congruence: partisanship is a (social) perception of likely behavior, not an innate characteristic of the candidates.

Specifically, comparing Proposition 1 and 2, the blame for the policy bias can be squarely laid on the fact that voters *expect* candidates to act partisan, that is, the expect politicians with different party affiliations to select different policies when in office and, if re-elected, to continue to pursue those policies in the future. The anticipated partisanship causes voters to strictly prefer the candidate who is expected to choose the

policy that is most likely to succeed given their information about the current situation. And it is the voters' expectations, which induce candidates to actually act partisan in the first place, to manipulate the electorate's inference problem. In equilibrium, therefore, these expectations are *self-fulfilling*. Voters and representatives are caught in a partisan trap: because voters expect the ideology of office holders to determine their political actions, an official's (re-)election chances will vary with her partisan affiliation. In their desire to influence the outcome of the election, these expectations induce the officials to act partisan.

The idea formalized by the model should also be contrasted with the widely-used adverse selection approach of reputation in repeated games, initially formalized by Kreps et al. (1982) and Kreps & Wilson (1982). In these models, small amounts of imperfect information regarding their payoff can induce players to attempt to build a reputation for being of a certain type, as to trigger more favorable responses from others.¹⁴ Translated into our framework, this approach would assume that politicians can be of two unobservable (payoff) types, a "partisan" type and a "non-partisan" type, where the latter is strictly preferable to the electorate. In such a world, candidates with partisan preferences would be tempted to implement an efficient policy so as to appear non-partisan. Clearly, this line of argument cannot explain sub-optimal partisanship. Moreover, candidate types' in our framework are common knowledge; they therefore cannot build a reputation for acting non-partisan. Implementing a deviation to an efficient policy choice in the partisan equilibrium cannot serve as a signal for being a non-partisan type. Rather, the electorate is unsure about the current state of the world, and an incumbent who implements a non-partisan policy will signal that the current state is unfavorable to his intended future policy, which endangers her re-election prospects.

Properties of the Partisan Equilibrium

As explained above, the specific motivation for acting partisan given voters' expectations is one of 'signal-jamming' (rather than signaling itself). An efficient policy choice conveys information about the state of the world, making it less likely that the incumbent office holder is re-elected if he is expected to act partisan in the future. To improve his chances of re-election, the incumbent thus 'jams' the voters' inference problem by instead using the partisan policy, which is both inefficient and less responsive to current circumstances.

The latter fact is noteworthy, not only because it can explain the emergence of partisanship as a self-fulfilling prophecy but also because, by definition, an a partisan incumbent's preferred policy choice does not vary with the underlying state. Thus the model generates inefficient policy persistence: along the equilibrium path, there will not be a deviation from a given policy unless voters oust a politician from office. Moreover, the probability that the policy (ideology of the office holder) varies with the state and changes from one period to the next is smaller than in the non-partisan equilibrium.

Finally, despite the fact that incumbents who 'stick to their political colors' enact inefficient policies, the political failure does not result in lower election chances. In fact, it is easy to show that – relative to the efficient equilibrium – incumbents enjoy an advantage in the partisan equilibrium: their chances of winning another term in office are strictly higher than even.¹⁵

¹⁴In a recent application of this approach to a related question, Morris (2001) for example assumes that political advisers can be either good or bad. A priori, both types of adviser would like being perceived as good, which may prompt them to keep their advice "politically correct" (against better knowledge).

¹⁵One may object to this assertion that since voters are indifferent between candidates in the non-partisan equilibrium, any probability of re-election is consistent with equilibrium behavior (including perfect incumbency advantages with re-election probabilities equal to one). Note, however, that such outcomes would require voters to co-ordinate their voting strategies, an

These observations are summarized in

Proposition 3. *In a partisan equilibrium:*

- a) voters receive strictly less utility than in the non-partisan equilibrium [Policy Failure]*
- b) incumbents' policies do not vary with the current state and in the long run, policies are less likely to be changed than would be efficient [Policy Persistence], and*
- c) the long run probability that an incumbent wins another term in office is strictly greater than one half [Incumbency Advantage].*

The implication of policy inertia is notable not only because it gives rise to political failure. Rather than resulting from a struggle between interest groups and the public at large as in Coate & Morris (1999), persistence here is driven by the fact that, in a world on partisanship, office holders are reluctant to admit that new circumstances warrant a new policy and, therefore, new leaders in the eyes of the electorate.

The equilibrium characteristics of policy persistence and incumbency advantage distinguish our model from many existing explanations of policy divergence (such as the citizen-candidate model) and are consistent with what we observe in democratic two-party systems. A plethora of studies of evidence on the US Congress in particular show positioning and polarization along party lines (McCarty et al., 2006) and that, once elected, members adopt a consistent ideological position and maintain it over time (Poole, 2007). In spite of (or perhaps even because of) their stubborn behavior, re-election rates for senators and House members are regularly above 80 percent. In the 2022 House election, for example, 435 districts were up for election and only 25 incumbents running for re-election were defeated.

Finally, note that in equilibrium, each candidate pursues a distinct policy which is the socially optimal policy only for a specific distinct state. A one-to-one correspondence between the identity of the politician, her policy choice, and the state in which the policy is optimal justifies our notations choices: without loss of generality, an I -type candidate sticks to policy i in equilibrium. The ‘starting point’ of this connection is how voter’s – arbitrarily (i.e. payoff irrelevant) – tie the identity of candidates to expectations about the policies they would implement once elected. It is this feature of the model that is most closely connected to the sunspot literature in macroeconomics. The ‘sunspot’,

More importantly, an outside observer would conclude that candidates either have separate ideologies (their policy choices are fixed and inefficiently unresponsive to changes in the economy) or have separate world views (e.g. they believe that the state is always blue and never red, and choose their policies accordingly). This conclusion would be incorrect because incumbents are *not* true believers. Instead, their partisan stance is purely based on expectations based on observable characteristics of candidates: if voters expect a Democrat from California to always support “big government”, then this is what she will do when in office. Interestingly, since the observable candidate characteristic can also be a declared preference over unrelated policies, the model can also provide a new perspective on issue bundling: if the electorate knows a candidate to be pro-life and against gun-control, and for that reason expects her to favor lower taxes, then these expectations can be self-fulfilling. Issue bundling occurs not because candidate’s preferences are bundled, but because voters’

implausible scenario when the electorate is large.

expectations tie candidates' policy intentions to their observed characteristics (their position on other salient issues).

2.3 Robustness and Extensions

There are two major misgivings one could have against the basic insight of self-fulfilling expectations of partisanship that is formalized in Proposition 2. First, voters are strictly better off in the non-partisan equilibrium and thus there may *a priori* be little reason to expect partisan behavior to prevail. Second, non-partisan behavior is not observed on the equilibrium path in the partisan equilibrium: by assumption, if voters unexpectedly see candidates acting non-partisan, they infer that the state must be unfavorable to their ideological position. One can argue that any alternative assumption on voter beliefs following an unexpected non-partisan policy choice makes the partisan equilibrium very fragile: if a candidate deviates from partisanship, shouldn't voters believe she would continue to act non-partisan in the future and therefore, re-elect her?¹⁶ Both concerns are rooted in the very simple nature of our model and can easily be addressed. Specifically, in this section we develop a) a straightforward refinement that selects the partisan equilibrium whenever it exists, and b) a natural extension of the model in which incumbents act non-partisan on the equilibrium path.¹⁷

Note first that the the existence of the partisan equilibrium does not depend on our assumption that there is no uncertainty in the voting behavior of the electorate, which makes competition between candidates especially fierce. In particular, a standard probabilistic voting model where candidates face uncertain electoral prospects and cater to the swing voter would yield identical conclusions.¹⁸

Candidate Behavior. One possible objection to the partisan equilibrium is that it is Pareto dominated by the non-partisan equilibrium for the voters (though not for incumbent politicians). We should therefore expect the electorate to favor the non-partisan equilibrium. Here, we argue that the non-partisan equilibrium is fragile in the sense that it does not survive small perturbations in voters' expectations. Formally, suppose that the electorate expects the office holder to choose the partisan policy *only* with some small probability $\epsilon > 0$. Put differently, voters do not expect incumbents to act partisan most of the time, but are not 100% sure that will do so all of the time.¹⁹ We have

Proposition 4. *Consider a set of parameters for which the partisan equilibrium exists according to Proposition 2, and suppose there is an arbitrarily small and i.i.d. probability $\epsilon > 0$ that office holders act partisan*

¹⁶Note, however, that there are no partisan or non-partisan types; thus candidates here cannot develop a reputation for acting non-partisan.

¹⁷In general, the model may have further equilibria. Assuming myopic voters, however, it is possible to show that generically in any symmetric pure strategy Markov equilibrium both parties either always act partisan or their actions converge to efficient play for $\beta = 1$.

¹⁸A proof for Proposition 2 where $P(\mu_t)$ is an arbitrary increasing function of μ_t is available from the authors upon request. Our results are equally robust to the possibility that voters occasionally observe the state of the world: while introducing a small probability that s_t is commonly observed will make partisan-behavior less attractive, *ceteris paribus*, condition (10) still holds for sufficiently high ϕ/b .

¹⁹For example, one explanation for why voters could expect partisan behavior to arise with a small but positive probability is party pressure Cukierman & Tommasi (1998); Canen et al. (2020). The possibility of a "partisan shock" could then be formalized by a probability ϵ with which the office holder realizes an additional benefit $B^i(a) \equiv B$ whenever he chooses the policy a corresponding to her ideology or party affiliation i , and assuming that the per-period payoff from a partisan choice is sufficient to compensate for the expected loss from not choosing the efficient alternative, i.e., $B > (1 - \pi)b$. Another possibility is that candidates sometimes make a mistake in their assessment of the state, but there are more likely to make a partisan mistake because they will rely on information from their party rather than neutral sources.

in each period. Then, the partisan equilibrium continues to exist and generically there is no equilibrium in which candidates play non-partisan (with probability $1 - \epsilon$) along the equilibrium path.

Proposition 4 shows that a small amount of voter uncertainty regarding candidate behavior suffices to select the inefficient, partisan equilibrium. Intuitively, non-partisan behavior is unstable because everyone is equally good as long as he or she is expected to act non-partisan. In such a situation even small amounts of uncertainty regarding candidates' subsequent behavior will make voters strictly prefer the candidate whose ideological position is more likely to succeed given their beliefs about the current state.²⁰

While we use the result in Proposition 4 primarily to select among equilibria, the fragility of non-partisan equilibria has obvious implications concerning how shifts in voters' expectations translate into policy changes. Consider a situation where non-partisanship has historically prevailed along the equilibrium, so voters have no reason to suspect politicians to enact (inefficient) ideological policies. Yet, a relatively small change in the perception of voters concerning an increased likelihood of partisan behavior would be sufficient to trigger a major trend towards partisanship and polarization. On matters of foreign policy, for example, partisanship as measured by the lack of support for the President by members of the U.S. congress increased dramatically following the Vietnam war (an event that may well have changed peoples' expectations about partisan behavior).²¹

Policy Prospects. Consider an extension of our model where voters are uncertain with respect to the likelihood of success of an inefficiently chosen policy. If the prospects of the partisan policy are sufficiently dim, candidates will prefer to implement efficient (non-partisan) policies. Thus, voters will observe non-partisan behavior on the equilibrium path, eliminating out-of equilibrium beliefs and the need to assume a specific refinement to pin down those beliefs. Below, we show that the Partisan equilibrium is robust to this natural modification.

Specifically, assume that the probability of success of an inefficient policy choice π_t evolves stochastically over time in the following way: in each period t , it is either $\pi > 0$, as before, or zero. The latter case captures a situation where it is very important to pick the right policy: inefficient policy choices never succeed and, consequently, the electorate always learns when the wrong policy was implemented. To fix ideas, we will refer to such a period as a *crisis*. Let q be the probability of a normal period (with success probability π), so a crisis occurs with probability $1 - q$, independent of the state $s_t \in \{b, r\}$. Candidates learn π_t at the beginning of each period, together with the state of the world. Voters do not observe π_t .²² Since a crisis doesn't persist by assumption, voters' beliefs over π_t are the same each period, and we can w.l.o.g. condition the election probabilities exclusively on the belief over the state, as before.

Turning to equilibria, observe first that the non-partisan equilibrium still exists since deviating to a partisan policy is even less attractive in a crisis. As in the baseline model, though, a partisan equilibrium where politicians act partisan in normal times and efficient in a crisis is also supported. In this equilibrium, voters

²⁰For the same reason, even in an equilibrium where the incumbents implement non-partisan policies with probability $1 - \epsilon$, the voting behavior now depends on the voters' beliefs, in contrast to the equilibrium discussed in Section 3.1.

²¹Using data on foreign policy and defense roll-call votes in the U.S. House and Senate, Meernik (1993) documents that the Vietnam War had a significant impact on bipartisan presidential support: whereas substantial consensus existed prior to the War, it has become much more infrequent afterwards.

²²The assumption that voters do not observe the success probability at all is made to simplify matters. Our qualitative argument remains valid as long as there is some residual uncertainty with regard to π_t .

again elect the blue (red) candidate for beliefs $\mu_t > 1/2$ ($\mu_t < 1/2$) and give both candidates equal chances of winning for $\mu_t = 1/2$. To see this, suppose the blue candidate has been elected in a crisis period and $s_t = r$. A partisan policy $a_t = b$ will surely fail, leading to a current payoff of ϕ and next period's belief of $\mu_{t+1} = 1 - \gamma$. A non-partisan choice $a_t = r$ on the other hand will be successful, yielding a higher current payoff of $b + \phi$ with the same next period's belief $\mu_{t+1} = 1 - \gamma$. Therefore, non-partisan politics are optimal in a crisis.

Next, let $V^c(\mu_t, s_t)$ denote the blue candidate's expected discounted value if state s_t occurs, the electorate has belief μ_t and he follows the equilibrium strategy for the rest of the game. We can adapt the condition (9) of the base model that supports partisan behavior in any *normal* period,

$$\beta E[\gamma V^c(\mu_{t+1}, b) + (1 - \gamma)V^c(\mu_{t+1}, r)] \geq (1 - \pi)v \quad (11)$$

where the value functions are slightly modified to account for the additional uncertainty induced by π_t :

$$V^c(\mu_t, s_t) = \begin{cases} v + \phi + \beta [\gamma V^c(\mu_{t+1}, l) + (1 - \gamma)V^c(\mu_{t+1}, r)] & \text{if } s_t = b \\ q\pi v + (1 - q)v + \phi + q\beta\pi [(1 - \gamma)V^c(\mu_{t+1}, l) + \gamma V^c(\mu_{t+1}, r)] & \text{if } s_t = r \end{cases}$$

It is easy to show that $V^c(\mu_t, s_t) = 0$ for $\mu_t < 0.5$ and $V^c(\mu_t, s_t) \equiv \bar{V}^c(s_t)$ for $\mu_t > 0.5$, as in section 2.2.2.

Simple algebra shows that condition (10) is unchanged. To see why, assume for the moment that (10) is satisfied with equality. Then, the office holder is indifferent between implementing his inefficient partisan policy and the efficient one whenever $\pi_t = \pi$. In this case, the value of being in office in the non-partisan state and following the equilibrium strategy equals that of implementing the efficient policy after observing $\pi_t = 0$ (and not getting reelected afterwards): $\bar{V}^c(r) = \phi + b$. Since the value in state r equals that of the base model, and both the strategy and the payoff in state l remain as in section 2.2.2, we must also have $\bar{V}^c(l) = \bar{V}^c(l)$. Now suppose that (A2) holds with strict inequality, which renders holding office more attractive. By the preceding paragraph, both in the base model and in this section, an incumbent would prefer to implement the partisan policy whenever $\pi_t = \pi$. Since incumbents implement the efficient policy if $\pi_t = 0$, the possibility of a crisis *ceteris paribus* decreases the value of office holders in the partisan equilibrium whenever it exists, i.e. $\bar{V}^c(l) \leq \bar{V}(l)$ and $\bar{V}^c(r) \leq \bar{V}(r)$.

Proposition 5. *Under (A1') and (A2), there exists an equilibrium in which elected office holders act partisan in normal times and efficient in times of crisis. In this equilibrium, politicians are re-elected with probability one if their implemented policy was a success and face certain defeat if it was a failure or they implemented the non-partisan policy.*

In summary, we find that the possibility of a crisis renders the partisan equilibrium more plausible. Intuitively, if the electorate is uncertain about the prospects of inefficient policies, it expects the candidates sometimes to implement the non-partisan policy. If voters observe that a politician has abandoned his ideology, they know that he did so to avoid a certain political failure - as a result, they (correctly) do not interpret this behavior as a sign of honesty and therefore do not draw inferences regarding the politician's future strategy. Finally observe that the partisan equilibrium continuously converges to the equilibrium in the basic framework as $q \rightarrow 0$, thereby justifying the off-equilibrium beliefs of section 2.2.2: upon observing the non-partisan policy being implemented, the electorate assumes that the incumbent has been forced to abandon his ideology, simply because the conflicting evidence was too strong.

3 Concluding Remarks

This paper proposes a new insight on partisanship. We have shown that there are circumstances under which elected officials may adopt ideologically opposed positions for no reason other than the electorate expects them to. In equilibrium, voters' expectations are therefore self-fulfilling and result in inefficient policies even in areas where everyone should agree on optimal courses of action. A "partisan trap" emerges because voters associate alternative policy measures with partisan types, and expect incumbents to remain true to their type which itself is a social perception grounded in observable characteristics (such as party affiliation, gender, or their position on a different policy issue).

The basic argument is simple. Suppose voters are not fully informed about the external circumstances that make one policy preferable to another. Instead, they make inferences from payoffs, which noisily convey information about whether the policy was the right one. Office holders, in contrast, have more information about what makes a good policy than voters. Suppose further that voters expect politicians to be partial to a particular policy once elected. In such a situation, voters elect the representative whose perceived partisan policy corresponds to what they think is in their best interest based on their current information. Incumbents thus have incentives to conceal which policy should be pursued whenever their superior information indicates that their opponents' partisan policy would be preferable. This effect prompts re-election motivated incumbents to stick to their partisan positions regardless of how likely they view it to be optimal. To improve her chances of re-election, a sufficiently office-motivated incumbent thus 'jams' the voters' inference problem. Voters' expectations about partisan behavior are fulfilled in equilibrium.

The result is political polarization and policy failure. Thus, policy bias and divergence stem from the fact that voters *perceive* policies to be ideologically tinted and *expect* candidates to act partisan. Moreover, such partisan politics are persistent in the sense that equilibrium policies are less volatile and less responsive to changes in the underlying state than efficient policies. Importantly, the inertia is not driven by a fear of appearing incompetent. Rather, in a partisan world, leaders are reluctant to abandon previously enacted policies and admit that "times have changed" because new circumstances will warrant a new policy and, therefore, new (partisan) leaders in the eyes of the electorate.

Since the equilibrium we describe is one of self-fulfilling expectations, one may object to our argument is that it only takes a small shift in voter's expectations to break the vicious circle of expectations and escape the partisan trap. This is true, of course. However, we have also shown that small extensions of our very basic model make the partisan equilibrium more robust than one would think at first glance. Besides, even though equilibria that are self-fulfilling and expectation-driven may be temporary in nature, they can still be highly empirically relevant, as the literature explaining stock market bubbles and excess output volatility through sunspot equilibria has shown. While how much the forces similar to what we identify here matter in practice is an empirical question, we believe that our model provides insights on why party affiliation or other observable characteristics play a role even if candidates or voters should *prima facie* share similar policy preferences, and why candidates promote particular policies even though they are not 'true believers'. Instead, in our stylized world partisan identity is purely a social perception or label, based on observable attributes such as gender, electoral district, or their party affiliation. Notably, a candidate's position on a policy question other than what is currently the salient issue is one of those attributes. As a result, the

model can also shed light on the empirically relevant phenomenon of issue bundling: a politician who is strongly pro-life, for example, may have an incentive to campaign against government-provided health care for no other reason than these two issues being tied together in voters' perceptions.

Finally, we believe that our insight could fruitfully be applied to other settings where leadership and identity are tied together. While we have cast the discussion within the framework of policy formation in a representative democracy for obvious reasons, it is important to note that our basic line of argumentation is valid in a broader context: as long as a leader needs supporters to stay in power and is challenged in his leadership (implicitly and explicitly) on occasion, he will have an incentive to base actions on his supporters' expectations. If those expectations are systematically associated with observable characteristics, living up to expectations can be equilibrium behavior of a current incumbent in a leadership position, which in turn can result in sub-optimal decision making.

Appendix

A. Figures

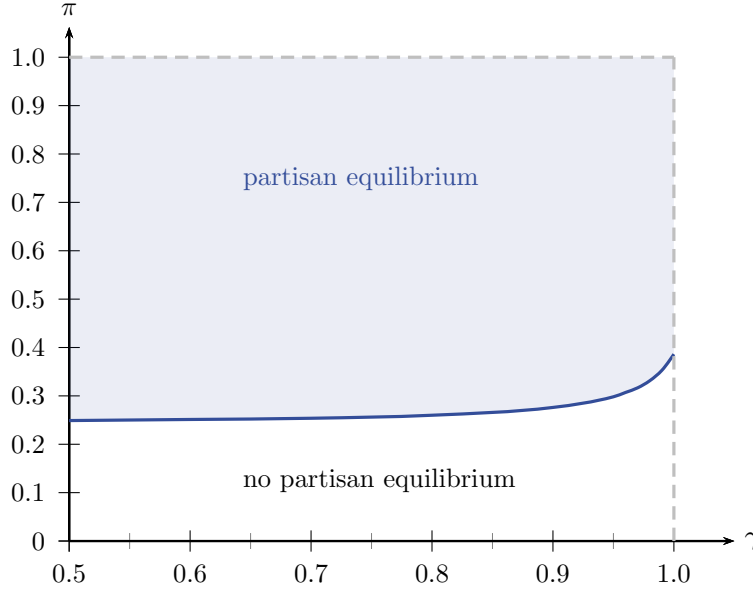


Figure 1: Existence of the Partisan Equilibrium in the (γ, π) parameter space.²³

B. Proofs

The following lemma establishes that there is no ‘experimenting’ in equilibrium.

Lemma. *Suppose that both candidates implement their partisan policy in each period and that $\beta < 1$. Then the electorate’s value function $U(\cdot)$ is unique and*

- i) is axially symmetric around 0.5, i.e $U(\mu_t) = U(1 - \mu_t)$ for $\mu_t \in [1 - \gamma, \gamma]$,*
- ii) satisfies $U'(\mu_t) \leq -(1 - \gamma)(1 - \pi)^2 v$ for $\mu_t < 0.5$ and $U'(\mu_t) \geq (1 - \gamma)(1 - \pi)^2 v$ for $\mu_t > 0.5$,*
- iii) satisfies $v + \beta U(\mu_t) - \beta U(1 - \gamma) \geq (1 - \gamma)(1 - \pi)v \quad \forall \mu_t \in [1 - \gamma, \gamma]$.*

The electorate’s optimal voting strategy is identical to that of a myopic electorate.

²³The curve corresponds to condition (9) assuming $b = 3$, $\phi = 2$, and $\beta = 0.95$. Increasing β and ϕ enlarges the region for which the partisan equilibrium exists, while increasing v does the opposite.

Proof. Let the function $\varphi^B(\mu_t) \equiv 1 - \gamma + (2\gamma - 1) \frac{\mu_t}{\mu_t + (1 - \mu_t)\pi}$ map the belief μ_t in period t into the belief that results in $t + 1$ when incumbent B successfully implements the policy l . Define $\varphi^R(\mu_t) \equiv \gamma - (2\gamma - 1) \frac{1 - \mu_t}{1 - \mu_t + \mu_t \pi}$ similarly for incumbent R .

Step 1: We prove uniqueness and properties i)-iii) by use of the Contraction Mapping Theorem: Define the functional operator $T : \mathcal{U} \mapsto \mathcal{U}$ that maps the space of bounded continuous functions $\mathcal{U}$ defined on $[1 - \gamma, \gamma]$ with range $\mathbb{R}^+$ into itself as follows:

$$TU(\mu) = \begin{cases} (\mu + (1 - \mu)\pi) (v + \beta U(\varphi^B(\mu))) + (1 - \mu)(1 - \pi)\beta U(1 - \gamma) & \mu \geq 0.5 \\ (1 - \mu + \mu\pi) (v + \beta U(\varphi^R(\mu))) + \mu(1 - \pi)\beta U(\gamma) & \mu < 0.5 \end{cases}$$

Since T is a contraction, there exists a unique electorate's value function $U(\cdot)$.²⁴ We will prove properties i), ii) and iii) of U by use of Corollary 1 of (Stokey & Lucas, 1989, Theorem 3.2). We have to show that if U satisfies these properties, then TU also satisfies them. Suppose that U satisfies properties i), ii) and iii).

i) Since $\varphi^B(0.5 + x) = 1 - \varphi^R(0.5 - x)$, TU also satisfies $TU(0.5 - x) = TU(0.5 + x)$ for $x \in [0, \gamma - 0.5]$.

ii) For $\mu_t > 0.5$,

$$\begin{aligned} TU'(\mu_t) &= (1 - \pi) (v + \beta U(\varphi^B(\mu_t)) - \beta U(1 - \gamma)) + (\mu_t + (1 - \mu_t)\pi) \beta U'(\varphi^B(\mu_t)) \varphi^{B'}(\mu_t) \\ &\geq (1 - \gamma) (1 - \pi)^2 v \end{aligned}$$

where the inequality is because of $b + \beta U(\mu_t) - \beta U(1 - \gamma) \geq (1 - \gamma) (1 - \pi) v$ by property iii) and because the second term is non negative by property ii). For $\mu_t < 0.5$, an analogous argument applies.

iii) For $\mu_t > 0.5$ we have $TU(\mu_t) = (\mu_t + (1 - \mu_t)\pi) (v + \beta U(\varphi^B(\mu_t)) - \beta U(1 - \gamma)) + \beta U(1 - \gamma)$ which implies

$$\begin{aligned} v + \beta TU(\mu_t) - \beta TU(1 - \gamma) &= v - (\gamma + (1 - \gamma)\pi) (v + \beta U(\varphi^B(\gamma)) - \beta U(1 - \gamma)) \\ &\quad + (\mu_t + (1 - \mu_t)\pi) (v + \beta U(\varphi^B(\mu_t)) - \beta U(1 - \gamma)) \\ &\geq v - (\gamma + (1 - \gamma)\pi) (v + \beta U(\varphi^B(\gamma)) - \beta U(1 - \gamma)) \\ &\geq (1 - \gamma) (1 - \pi) v \end{aligned}$$

where we used property i) repeatedly. The first inequality is due to property iii) and the last one due to property ii).

Step 2: Now we show that it is indeed optimal to vote for the B party if $\mu_t > 0.5$ (an analogous argument holds for $\mu_t < 0.5$). Deviating once and electing the right party yields $\hat{U}(\mu_t) = (1 - \mu_t + \mu_t \pi) (v + \beta U(\varphi^R(\mu_t))) + \mu_t(1 - \pi)\beta U(\gamma)$. Hence

$$\begin{aligned} U(\mu_t) - \hat{U}(\mu_t) &= (2\mu - 1)(1 - \pi) (v + \beta U(\varphi^B(\mu_t)) - \beta U(\gamma)) \\ &\quad + (1 - \mu_t + \mu_t \pi) (\beta U(\varphi^B(\mu_t)) - \beta U(\varphi^R(\mu_t))) \geq 0 \end{aligned}$$

²⁴It can be easily verified that this operator is a *contraction* since it satisfies the Blackwell's sufficient conditions of *discounting* and *monotonicity* according to (Stokey & Lucas, 1989, Theorem 3.2). As $\mathcal{U}$ together with the *sup*-Norm is a complete metric space, the contraction mapping Theorem (Stokey & Lucas, 1989, Theorem 3.2) applies.

where the inequality follows because the first term is positive due to property iii) and the second is positive due to $\varphi^B(\mu_t) - 0.5 > |0.5 - \varphi^R(\mu_t)|$ and property ii). To see that $\varphi^B(\mu_t) - 0.5 > 0.5 - \varphi^R(\mu_t)$ when $\varphi^R(\mu_t) < 0.5$, inserting the formulas from above and rearranging yields

$$\frac{1}{1 + (1 - \mu_t)\pi\mu_t^{-1}} > \frac{1}{1 + (1 - \mu_t)^{-1}\pi\mu_t}$$

which is true for $\mu_t > 0.5$. □

Out-of-Equilibrium Play [Partisan Equilibrium].

We already established in the main text that if a blue incumbent deviates and implements a non-partisan policy choice, voters infer that the red state has occurred and beliefs for the following period are $\mu_{t+1} = 1 - \gamma$. Voters' strategy of voting for the red opponent is sequentially optimal, given the anticipated partisan choices of both types of candidates in the future (one-shot deviation). The only other remaining possibility of off equilibrium behaviour is if voters did not elect the politician whose type correspond to the state they consider most likely, e.g., if voters elected the blue type in t despite beliefs $\mu_t < \frac{1}{2}$. Note first that if μ_t is sufficiently close to $\frac{1}{2}$ (e.g, if either y is close enough to $1/2$ or if it is very unlikely that playing partisan results in success) then conditional on a successful partisan action, beliefs μ_{t+1} will exceed $1/2$, and analysis is unchanged (that is, voters re-elect the incumbent if and only her policies match her party affiliation and they are successful). Play with thus return to equilibrium after one period. If μ_t is small, however, then a successful partisan policy is not sufficient to raise beliefs to $\mu_{t+1} > \frac{1}{2}$. In this case, either a) the incumbent has a strict incentive to play non-partisan, resulting in sure defeat (independent of her party affiliation and policy success) and return to equilibrium play the following period, or plays partisan. In the latter case, succeeding at the partisan policy is a necessary, but not sufficient, condition for re-election, and voters will adopt a mixed re-election standard. To see this, observe that the re-election probability following a successful partisan policy must be strictly positive (otherwise it'll be optimal to implement the optimal non-partisan all the time). At the same time, if voters always re-elected politicians following a partisan policy success, then beliefs are $\mu_t > \frac{1}{2}$ contradicting optimal voter behaviour. Hence, for some values of π , the incumbent must do the wrong partisan action and beliefs must be $\mu_{t+1} = 1/2$. Together with Bayes Rule, the latter determines the cutoff probability $\hat{\pi}$ above which a successful partisan policy will result in beliefs $\mu_t = 1/2$. The re-election probability of the incumbent in this case is pinned down by the indifference of the officeholder at $\hat{\pi}$. Again, however, this off-equilibrium play only lasts one period after which we return to the equilibrium path as either $I_{t+1} = B$ and $\mu_{t+1} = 1/2$ or $I_{t+1} = R$ and $\mu_{t+1} \leq 1/2$ the following period.

Proof of Proposition 3

Part a) is trivial. To show part b), define the random variable $\tilde{s}_t \in \{m, n\}$ whose two realizations are 'match' $\tilde{s}_t = m$ when equilibrium play prescribes $a_t = s_t$ in a given period and 'non-match' n whenever $a_t \neq s_t$. In the partisan equilibrium, the transition probabilities between these 'states' are:

$$T = \begin{pmatrix} t_{mm} & t_{mn} \\ t_{nm} & t_{nn} \end{pmatrix} = \begin{pmatrix} \gamma & 1 - \gamma \\ (1 - \pi)\gamma + \pi(1 - \gamma) & (1 - \pi)(1 - \gamma) + \pi\gamma \end{pmatrix}$$

where the element t_{ij} of the transition matrix T denotes the transition probability from state i to state j . In the partisan equilibrium, a change in the implemented policy (i.e. $a_t \neq a_{t+1}$) only occurs if the implemented policy in period t was $a_t \neq s_t$ and failed. Hence the probability of a policy change between period t and $t + 1$ is $Pr(\tilde{s}_t = n)(1 - \pi)$. In the efficient equilibrium a policy change occurs whenever the true state changes, i.e. with probability $1 - \gamma$. By definition, the partisan equilibrium involves more persistence in a given period t whenever the probability of a change in policies between period t and $t + 1$ is lower than the probability of change in the efficient equilibrium which is $1 - \gamma$. This condition is satisfied whenever $Pr(\tilde{s}_t = n)(1 - \pi) \leq 1 - \gamma$.

We proceed to show that for any initial belief and state, the long run probability of having a non-match is small enough to satisfy this condition. The (generically unique) stationary distribution corresponds to the eigenvector which is associated to the unit eigenvalue of T' . It is $\bar{f}' = \left(\frac{-2\pi\gamma + \gamma + \pi}{1 - 2\gamma\pi + \pi}, \frac{1 - \gamma}{1 - 2\gamma\pi + \pi} \right)$, where the first (second) element denotes the stationary probability that a match (non-match) occurs. The long run probability that a non-match occurs is thus $\lim_{t \rightarrow \infty} Pr(\tilde{s}_t = n) = \frac{1 - \gamma}{1 - 2\gamma\pi + \pi}$. Due to $\gamma < 1$, we have

$$(1 - \pi) \lim_{t \rightarrow \infty} Pr(\tilde{s}_t = n) = \frac{(1 - \pi)(1 - \gamma)}{1 - (2\gamma - 1)\pi} < (1 - \gamma)$$

which completes the proof.

To show part c), recall that in the partisan equilibrium, an incumbent is not re-elected only in the event of a political failure. From the proof of part b), this occurs with probability $Pr(\tilde{s}_t = n)(1 - \pi)$, which is in the long run equal to

$$(1 - \pi) \lim_{t \rightarrow \infty} Pr(\tilde{s}_t = n) = \frac{(1 - \pi)(1 - \gamma)}{1 - 2\gamma\pi + \pi} < (1 - \gamma) < \frac{1}{2}$$

where the last inequality follows from $\gamma > \frac{1}{2}$.

Proof of Proposition 4

Since ϵ restricts the *minimum* probability for implementing the partisan policy, for all $\epsilon \in (0, 1)$, strategies and re-election probabilities *in the partisan equilibrium* are unchanged. Moreover, neither voters' nor office holders' payoffs are affected. Thus, partisan behavior continues to be an equilibrium under (A1) and (A2).

Turning to the *most efficient equilibrium* (or ϵ -efficient equilibrium, indicated by the superscript ϵE), recall that voters' optimally vote as if they were myopic by Bemma ???. Hence for any $\epsilon > 0$ the reelection probabilities are now

$$P^{\epsilon E}(\mu) = \begin{cases} 1 & \text{if } \mu > 0.5 \\ 0.5 & \text{if } \mu = 0.5 \\ 0 & \text{else} \end{cases}$$

and equal those of the partisan equilibrium. The evolution of beliefs in the non-partisan equilibrium is

$$\mu_{t+1}^B(a_t, \mu_t) = \begin{cases} 1 - \gamma + (2\gamma - 1) \frac{\mu_t}{\mu_t + (1 - \mu_t)\epsilon\pi} \equiv \varphi^{B, \epsilon E}(\mu_t) & \text{if } a_t = l \text{ was a success} \\ 1 - \gamma & \text{if } a_t = l \text{ failed or } a_t = r \end{cases}$$

The value for a left wing politician in the ϵ -efficient equilibrium is $V^{\epsilon E}(\mu_t) = 0$ for $\mu_t < 0.5$ and

$$\bar{V}^{\epsilon E}(s) \equiv \begin{cases} v + \phi + \beta [\gamma \bar{V}^{\epsilon E}(l) + (1 - \gamma) \bar{V}^{\epsilon E}(r)] & \text{if } s = l \\ (1 - \epsilon)v + \phi + \epsilon\pi [v + \beta (\gamma \bar{V}^{\epsilon E}(r) + (1 - \gamma) \bar{V}^{\epsilon E}(l))] & \text{if } s = r, \end{cases}$$

for $\mu_t > 0.5$ since B -type incumbents are not re-elected following the efficient choice of $a_t = r$ in state $s_t = r$.

Now suppose that the partisan equilibrium exists. Then, generically, (9) is satisfied with strict inequality,

$$(1 - \pi)v < \pi\beta[(1 - \gamma)\bar{V}^P(l) + \gamma\bar{V}^P(r)], \quad (11)$$

where $\bar{V}^P(l) > 0$ and $\bar{V}^P(r) > 0$ are candidate B 's values in the partisan equilibrium [see the proof of Proposition 2]. Because the reelection probabilities are the same in the ϵ -efficient equilibrium as in the partisan equilibrium, a repeated deviation by playing $a_t = l$ in states $s_t = r$ guarantees an expected payoff of $\bar{V}^P(s)$. We want to show that whenever (11) holds, then $\bar{V}^P(s) > \bar{V}^{\epsilon E}(s)$, i.e. a repeated deviation is profitable. We use the same contraction argument as in Lemma (??) of the appendix. According to this reasoning, it suffices to show that if $\bar{V}^P(s) > \bar{V}^{\epsilon E}(s)$, $s \in \{l, r\}$ then also

$$\bar{V}^P(r) > (1 - \epsilon)v + \phi + \epsilon\pi[v + \beta(\gamma\bar{V}^{\epsilon E}(r) + (1 - \gamma)\bar{V}^{\epsilon E}(l))].$$

To see that this inequality is indeed satisfied, note that

$$\begin{aligned} \phi + \pi[v + \beta(\gamma\bar{V}^P(r) + (1 - \gamma)\bar{V}^P(l))] \\ &> (1 - \epsilon)(v + \phi) + \epsilon[\phi + \pi[v + \beta(\gamma\bar{V}^P(r) + (1 - \gamma)\bar{V}^P(l))]] \\ &> (1 - \epsilon)(v + \phi) + \epsilon[\phi + \pi[v + \beta(\gamma\bar{V}^{\epsilon E}(r) + (1 - \gamma)\bar{V}^{\epsilon E}(l))]] \end{aligned}$$

where the first inequality comes from (11) and the second from the hypothesis $\bar{V}^P(s) > \bar{V}^{\epsilon E}(s)$.

Proof of Proposition 5

Note that condition (11) is equivalent to $\bar{V}^R(r) \geq v + \phi$ where we use the same notation as in the base model, i.e. $\bar{V}^R(s_t) \equiv V^R(\mu_t, s_t)$ for $\mu_t > 0.5$. We have to show that this condition is satisfied if assumption (A2) holds.

At the same time, we show that assumption (A2) also implies $\bar{V}^R(l) \geq \frac{(1 + \beta(1 - \gamma))(v + \phi)}{1 - \gamma\beta}$. Applying a contraction argument similar to Lemma ??, we have to show that whenever $\bar{V}^R(r) \geq v + \phi$, $\bar{V}^R(l) \geq \frac{(1 + \beta(1 - \gamma))(v + \phi)}{1 - \gamma\beta}$ and assumption (A2) holds, then the following two inequalities are satisfied:

$$\begin{aligned} v + \phi + \beta[\gamma\bar{V}^R(l) + (1 - \gamma)\bar{V}^R(r)] &\geq \frac{(1 + \beta(1 - \gamma))(v + \phi)}{1 - \gamma\beta} \\ \pi v + \phi + q\beta\pi[(1 - \gamma)\bar{V}^R(l) + \gamma\bar{V}^R(r)] + (1 - q)(1 - \pi)v &\geq v + \phi \end{aligned}$$

To see that the first inequality is true note that our hypothesis implies:

$$\begin{aligned} v + \phi + \beta[\gamma\bar{V}^R(l) + (1 - \gamma)\bar{V}^R(r)] &\geq v + \phi + \beta\left[\gamma\frac{(1 + \beta(1 - \gamma))(v + \phi)}{1 - \gamma\beta} + (1 - \gamma)(v + \phi)\right] \\ &= \frac{(1 + \beta(1 - \gamma))(v + \phi)}{1 - \gamma\beta} \end{aligned}$$

The second inequality is equivalent to $\pi v + \phi + \beta\pi[(1 - \gamma)\bar{V}^R(l) + \gamma\bar{V}^R(r)] \geq v + \phi$. By our hypothesis, $\pi v + \phi + \beta\pi[(1 - \gamma)\bar{V}^R(l) + \gamma\bar{V}^R(r)] \geq \pi v + \phi + \beta\pi\left[(1 - \gamma)\frac{(1 + \beta(1 - \gamma))(v + \phi)}{1 - \gamma\beta} + \gamma(v + \phi)\right]$. Simplifying (A2) yields $\pi\beta\left[(1 - \gamma)\frac{(1 + \beta(1 - \gamma))(v + \phi)}{1 - \gamma\beta} + \gamma(v + \phi)\right] \geq (1 - \pi)b$. Putting both observations together confirms the second inequality.

4 References

References

- Aghion, P., & Jackson, M. O. (2016). Inducing leaders to take risky decisions: Dismissal, tenure, and term limits. *American Economic Journal: Microeconomics*, 8(3), 1–38.
URL <https://www.aeaweb.org/articles?id=10.1257/mic.20150083>
- Alesina, A., & Cukierman, A. (1990). The Politics of Ambiguity. *The Quarterly Journal of Economics*, 105(4), 829–850.
- Ansolabehere, S., & Snyder, J. M. J. (2000). Valence Politics and Equilibrium in Spatial Election Models. *Public Choice*, 103(3-4), 327–36.
- Aragones, E., & Palfrey, T. R. (2002). Mixed equilibrium in a downsian model with a favored candidate. *Journal of Economic Theory*, 103(1), 131–161.
- Ashworth, S., & Bueno de Mesquita, E. (2009). Elections with platform and valence competition. *Games and Economic Behavior*, 67(1), 191–216. Special Section of Games and Economic Behavior Dedicated to the 8th ACM Conference on Electronic Commerce.
URL <https://www.sciencedirect.com/science/article/pii/S089982560800208X>
- Ashworth, S., Bueno de Mesquita, E., & Friedenberg, A. (2017). Accountability and information in elections. *American Economic Journal: Microeconomics*, 9(2), 95–138.
URL <https://www.aeaweb.org/articles?id=10.1257/mic.20150349>
- Azariadis, C. (1981). Self-fulfilling prophecies. *Journal of Economic Theory*, 25(3), 380–396.
URL <https://www.sciencedirect.com/science/article/pii/0022053181900387>
- Azariadis, C., & Guesnerie, R. (1986). Sunspots and cycles. *The Review of Economic Studies*, 53(5), 725–737.
URL <http://www.jstor.org/stable/2297716>
- Besley, T. J., & Coate, S. (1997). An Economic Model of Representative Democracy. *The Quarterly Journal of Economics*, 112(1), 85–114.
- Buisseret, P., & Van Weelden, R. (2022). Polarization, valence, and policy competition. *American Economic Review: Insights*, 4(3), 341–52.
URL <https://www.aeaweb.org/articles?id=10.1257/aeri.20210169>
- Callander, S. (2008). Political Motivations. *Review of Economic Studies*, 75(3), 671–697.
- Callander, S., & Carbajal, J. C. (2022). Cause and effect in political polarization: A dynamic analysis. *Journal of Political Economy*, 130(4), 825–880.
- Callander, S., et al. (2008). A Theory of Policy Expertise. *Quarterly Journal of Political Science*, 3(2), 123–140.

Canen, N., Kendall, C., & Trebbi, F. (2020). Unbundling polarization. *Econometrica*, 88(3), 1197–1233.

URL <https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA16756>

Canes-Wrone, B., Herron, M., & Shotts, K. (2001). Leadership and Pandering: A Theory of Executive Policymaking. *American Journal of Political Science*, 45(3), 532–550.

Cass, D., & Shell, K. (1983). Do sunspots matter? *Journal of political economy*, 91(2), 193–227.

Castanheira, M. (2003). Why Vote for Losers? *Journal of the European Economic Association*, 1(5), 1207–1238.

URL <https://doi.org/10.1162/154247603770383433>

Coate, S., & Morris, S. (1999). Policy Persistence. *American Economic Review*, 89(5), 1327–1336.

Cukierman, A., & Tommasi, M. (1998). When Does It Take a Nixon to Go to China? *American Economic Review*, 88(1), 180–197.

Enke, B., Polborn, M., & Wu, A. (2022). Values as luxury goods and political polarization. Working Paper 30001, National Bureau of Economic Research.

URL <http://www.nber.org/papers/w30001>

Ghosh, S., & Tripathi, V. (2012). Ideologues Beat Idealists. *American Economic Journal: Microeconomics*, 4(2), 27–49.

Glaeser, E. L., Ponzetto, G. A. M., & Shapiro, J. M. (2005). Strategic Extremism: Why Republicans and Democrats Divide on Religious Values. *The Quarterly Journal of Economics*, 120(4), 1283–1330.

Groseclose, T. (2001). A model of candidate location when one candidate has a valence advantage. *American Journal of Political Science*, 45(4), 862–886.

Harrington, J. E. (1993). Economic Policy, Economic Performance, and Elections. *American Economic Review*, 83(1), 27–42.

Howell, W., Krasa, S., & Polborn, M. (2020). Political conflict over time. *American Journal of Political Science*, 64(3), 554–568.

URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12467>

Izzo, F., Martin, G., & Callander, S. (2023). Ideological Competition. *American Journal of Political Science*, 67(3), 687–700.

Kartik, N., & McAfee, R. P. (2007). Signaling Character in Electoral Competition. *American Economic Review*, 97(3), 852–870.

Kessler, A. S. (2005). Representative versus Direct Democracy: The Role of Informational Asymmetries. *Public Choice*, 122(1), 9–38.

Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational Cooperation in the Finitely Repeated Prisoners' Dilemma. *Journal of Economic Theory*, 27(2), 245 – 252.

URL <http://www.sciencedirect.com/science/article/B6WJ3-4CYGCVT-10P/2/2bc66a0a3d2cf7c9a74abe5e5164ec94>

- Kreps, D. M., & Wilson, R. (1982). Reputation and Imperfect Information. *Journal of Economic Theory*, 27(2), 253 – 279.
- Martinelli, C. (2001). Elections with privately informed parties and voters. *Public choice*, 108(1/2), 147–167.
- Maskin, E., & Tirole, J. (2004). The Politician and the Judge: Accountability in Government. *American Economic Review*, 94(4), 1034–1054.
- McCarty, N., Poole, K. T., & Rosenthal, H. (2006). *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge, Mass.: The MIT Press.
- McMurray, J. (2022). Polarization and pandering in common-interest elections. *Games and Economic Behavior*, 133, 150–161.
URL <https://www.sciencedirect.com/science/article/pii/S0899825622000471>
- Meernik, J. (1993). Presidential Support in Congress: Conflict and Consensus on Foreign and Defense Policy. *The Journal of Politics*, 55(3), 569–587.
- Morelli, M., & Weelden, R. V. (2013). Ideology and information in policymaking. *Journal of Theoretical Politics*, 25(3), 412–439.
- Morris, S. (2001). Political Correctness. *Journal of Political Economy*, 109(2), 231–265.
- Osborne, M. J., & Slivinski, A. (1996). A Model of Political Competition with Citizen-Candidates. *The Quarterly Journal of Economics*, 111(1), 65–96.
- Palfrey, T. R. (1984). Spatial Equilibrium with Entry. *Review of Economic Studies*, 51(1), 139–56.
- Poole, K. (2007). Changing Minds? Not in Congress! *Public Choice*, 131(3), 435–451.
- Poole, K. T., & Rosenthal, H. (1997). *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.
- Poole, K. T., & Rosenthal, H. (2007). *Ideology & Congress*. Transaction Pub.
- Stasavage, D. (2007). Polarization and Publicity: Rethinking the Benefits of Deliberate Democracy. *Journal of Politics*, 69(1), 59–72.
- Stokey, N. L., & Lucas, R. E. (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.
URL <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0674750969>
- Tirole, J. (1985). Asset bubbles and overlapping generations. *Econometrica*, 53, 1499–1528.
URL <https://api.semanticscholar.org/CorpusID:1793910>
- Van Weelden, R. (2013). Candidates, Credibility, and Re-election Incentives. *The Review of Economic Studies*, 80(4), 1622–1651.
URL <https://doi.org/10.1093/restud/rdt014>