The Ideology Trap[*]

Benno Buehler
European Commission

Anke Kessler[†]
Simon Fraser University

January 2016

Abstract

This paper sheds light on why political leaders may want to adopt ideological positions and maintain them over time even in the face of conflicting evidence. We study a dynamic framework in which politicians are better informed than the voting public about an underlying state of nature that determines the desirability of a given policy measure. The issue itself is non-partisan (everybody has the same policy preferences) but voters attach ideological labels to both candidates and available policy alternatives. We show that both sides may be caught in an *ideological trap* of self-fulfilling expectations: because voters expect the perceived ideology of office holders to determine their political actions, politicians are tempted to act according to their perceived ideology, resulting in partisanship, policy persistence, and political failure.

Keywords: Partisanship, Ideology, Policy Persistence, Political Competition, Political Economy
JEL classification: D 78, P16, H11

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

1 Introduction

Political leaders often define themselves in terms of a set of beliefs and values that they adhere to, and consistently base their political action on that set. Such leaders, who place greater weight on ideology as a collection of ideas about how society should work and the best way to achieve this goal, can be referred to as ideological leaders or ideologues. Some well-known public figures can be placed into this category: from Charles de Gaulle to Margaret Thatcher, from Vladimir Lenin to Mohandas Gandhi, many historical leaders derived their power from ideological principles and their ability to convince others that one can accomplish a lot by adhering to that particular ideology.

Politics today is no exception. On a smaller scale, for instance, one has to look no further than to contemporary American politics to find plenty of ideologues: politicians routinely use labels such as "liberal", "conservative", "moderate", "leftist" to describe themselves and their opponents. Of course, one may wonder what's in a name. Surprisingly much as it turns out. As documented in the empirical work on Congressional voting behavior of Poole and Rosenthal (2007), McCarty et al. (2006) and others, the belief systems of political elites can often largely be captured with a single dimension, their *ideology*, which almost always mirrors party affiliation: with just the label "conservative" (Republican), for example, one can fairly accurately predict a politician's stance on policy issues as disparate as taxes, gun control, affirmative action, health care, and abortion. Moreover, ideological positions of individual members are remarkably stable. That is, based upon the roll call voting record, once elected to Congress, members adopt an ideological position and maintain that position throughout their careers.[1] As Poole (2007, p. 435) puts it, "members of Congress die in their ideological boots." Clearly, this phenomenon is neither exclusive to the U.S., nor is it confined to positional (divisive) issues that voters have different preference over, depending on their socio-economic status, race, gender, or religion. Partisan politics are a frequent phenomenon even regarding so-called valence issues for which there should be a common agreement among the electorate (such as crime, foreign policy, corruption and economic growth).[2]

The observation that ideological labels seem meaningful after all does not answer the question of why they are adopted and why they are played out in partisan politics, especially on policies where voters would prefer their representative to seek common ground.It has long been argued that, at least in the United States, political parties are more polarized than the voters they seek to govern. A different, but

---

[1]What is more, members of Congress seem to remain ideologically consistent even in the face of changing personal or electoral conditions: members' voting records remain essentially the same, regardless of whether they plan to retire, plan to run for a higher office, serve in a higher office, or have their districts redrawn. [see Poole and Rosenthal (1997) and Poole (2007) and the references therein].

[2]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)]. Indeed, evidence supports the view that partisanship of political representatives often does not simply mirror equally divided constituents. Rather than representing the district voters, a representative's own ideology is the primary determinant of roll-call voting patterns [see Lee et al. (2004) and Levitt (1996)]. Besides, voter polarization is presumably a lesser danger for valence issues. Polling data on foreign policy confirm this presumption. Two recent pools conducted by the Program on International Policy Attitudes (PIPA) and the Chicago Council on Foreign Relations (CCFR) found that Americans share common views on a wide array of foreign policy issues, and would prefer that Democrats and Republicans seek common ground [for details, see the website of Partnership for a Secure America (http://www.psaonline.org/), an organization dedicated to recreating the bipartisan center in American national security and foreign policy.]

related, question is why political elites tend to maintain their positions over time, i.e. why ideological views are so persistent, even in the face of changing circumstances to the point where they are at odds with the facts.

The present paper seeks to shed light on these issues from a novel perspective. Using a dynamic model that ties observable characteristics of political representatives to voters' expectations, we find that show that political elites have to adopt an ideological label and stick to it even in the face of contradicting evidence, simply because simply because voters' *expect* them to. The theory implies, for instance, that a female Democrat from California is likely to take a liberal stance on most issues, not because her true preferences or her belief system necessarily reflects this stance, but because her constituents expect a female Democrat from California to be a liberal (and elected her for this very reason).

We consider a world where voters are imperfectly informed about how policy instruments map into policy outcomes, i.e., are uncertain whether a specific policy is more desirable than another policy. Given their beliefs about the prevailing state, voters form expectations about which policy candidates are likely to implement once in office, and which of those is most likely to succeed. Importantly, voters attach ideological labels *both* to the various policy alternatives that are available *and* to the political candidates running for office. To develop our argument in the strongest manner possible, we assume that this association, i.e., the perceived positioning of policies and office holders in the political spectrum, is completely arbitrary; in other words, candidates derive the exact same utility from the policy measure as the electorate at large, [3] and their ideological characterization is truly nothing more than a label.

Our main finding is that, nevertheless, policy holders have an incentive to adopt a particular ideological position in their policy choice and maintain it over time. The argument is one of self-fulfilling expectations and runs as follows. Suppose voters expect political candidates to act partisan once in office, i.e., to remain 'true to their colors' in implementing policies that are 'close' to their own 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 perceived ideology, in the first place. The specific motivation is one of signal-jamming: an incumbent who sticks to his partisan policy avoids revealing that current circumstances would favour his opponents' partisan position, making his re-election more likely if voters expect partisan behaviour in the future.[4] By implementing his partisan policy, a sufficiently office motivated incumbent will demonstrate confidence in his own ideology. As even inefficient policies may turn out to be successful,

---

[3] It should be emphasized that our insight also applies for non-valence (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. In a companion paper with a different focus [Bernhardt et al. (2016)], we soften this stark assumption and instead assume a candidate's ideology to have a real affect which policy outcomes he or she is likely to implement.

[4] Alesina and 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.

this behavior potentially allows to hold up the electorates' belief in the incumbent's ideology. 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 from the fact that voters *perceive* policies to be ideologically tinted and *expect* candidates to act partisan. Both sides are caught in an *ideology trap* of self-fulfilling expecations: 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. Importantly, the issue itself can be 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 an ideologue. Instead, ideology is purely a social perception based on observable attributes of candidates such as their gender, their 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 strongly disapproves of gay marriage may also have an incentive to campaign against government-provided health care for no other reason than because voters' expectations tie 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.[5]

Our argument is related to three different strands of the literature. First, we contribute to the literature on political failure. In a model similar to ours, Cukierman and Tommasi (1998) show that if voters are also imperfectly informed about an incumbent's preferences, his 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 and Tirole (2004)m Stasavage (2007), and Callander (2008) emphasize policy failure when candidates pander to public opinion. In these models, candidates are primarily interested in manipulating the electorates' beliefs about their preferences or ability, not about which policy is more appropriate at any given point in time. We also emphasize how the inefficiency can stem solelyfrom voters' *expectations* about the effects of a candidate's future policy intentions, rather than on a true discrepancy between the ideal policy of a candidate and that of the electorate at large. An entirely different explanation of ideological politics is proposed by Ghosh and Tripathi (2012). In a dynamic model of where information on policy desirability may arrive after the election, they show that since the pivotal voter may find herself in a fragile majority that is easily overturned, electing the ideologue may be a way to commit to the (ex ante preferred) future course of action. In our framework, in contrast, voters would prefer to commit *not* to elect candidates who pursue ideologically fixed policy since this would eliminate the latter's incentives to act partisan in the first place.

Second, our result contributes to the question of why political parties and politicians seeking office diverge

---

[5]The resilience of economic policies that benefit (target) a specific groups of voters has been studied by Coate and Morris (1999) who use a dynamic model to formalize the intuition that implementation of a policy increases the political effectiveness of its beneficiaries in lobbying. As in our model, this persistence gives rise to political failure in the sense that equilibrium policy sequences can be Pareto dominated. The main difference between Coate and Morris (1999) and our approach is that we focus on *non-partisan* (valence) issues, which do not target specific groups.

in their positions on critical issues, contrary to what the Downsian model would predict. In the past two decades, scholars in economics and political science have identified a number of factors that facilitate policy divergence, including the multi-dimensional issues [Ansolabehere and Snyder (2000)], the threat of third-party entry [Palfrey (1984)], citizen candidates [Osborne and Slivinski (1996), Besley and Coate (1997)], improved electoral control [Van Weelden (2009)], and an electorate that is imperfectly informed about candidates' types [Kartik and McAfee (2007)]. Those explanations, however, require *partisan preferences* of candidates. Indeed, we are not aware of a single contribution that is able to explain polarized and partisan politics on matters where voters and candidates commonly agree.[6] Moreover, since enacted policies in these models directly reflect the preferences of the electorate, they are silent on why policies can persist over time even in the face of new (and conflicting) evidence.

Finally, there is a growing economic literature on the origins of ideologies as a collection of ideas and firmly held beliefs. Bénabou and Tirole (2006) and Bénabou (2008) study voters' perceptions about a fundamental property of the underling economy, and show that maintaining beliefs that contradict reality can be an equilibrium phenomenon. In forming their beliefs, individuals optimally trade off the benefit of being able to motivate themselves (or their children) toward effort and the costs of misinformed decisions. While these papers can explain ideology as a collectively held belief system, our contribution focuses on leaders and political elites who *publicly act* upon – rather than genuinely entertain – certain beliefs in order to maintain their power and leadership role.[7]

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 A dynamic Model of Partisanship

2.1 Preferences and Economic Environment

Consider an infinite-horizon economy in discrete time. The economy is populated by an infinite number of risk-neutral consumer-voters who derive the same per-period benefit $b_t = b(a_t, s_t) \in \{0, b\}$ from a policy decision a_t. For simplicity, we take a_t to be binary; in particular, there is a 'left-wing' alternative $a_t = l$

---

[6]Another line of research has focused on explaining the prevailing polarization on '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).

[7]At the same time, our setting is not ideology free, since we require the electorate to attach ideological labels to policies and politicians alike, e.g, the Military Commissions Act (which effectively excluded U.S. prisoners of war from protection of the Geneva Conventions) is universally perceived to be "conservative", as is a male Republican candidate from Texas.

and a 'right-wing' alternative $a_t = r$.[8] Consumers know the set of feasible policies (and have common views on which they perceive as being left-wing and right-wing, respectively) but are uncertain about the underlying state of the economy $s_t \in \{l, r\}$. As an example, take the issue of state versus market provision of public services (such as health care and education): here, the underlying state s_t captures the relative efficacy of government provision and the policy decision is whether or not the service is publicly provided, where public provision is commonly viewed as the "left-wing" alternative and private provision is universally perceived as a "right-wing" policy.

Voters' per period payoff stochastically depends on the unobserved state s_t as follows:

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

In other words, if the policy choice matches the state, the policy is 'successful' with probability one and voters receive a certain payoff of b. Otherwise, the policy 'fails' with probability $1 - \pi > 0$ in which case we normalize payoffs to zero. [9]

The state of the economy evolves over time according to a symmetric transition function

$$Prob\{s_{t+1} = s_t\} = \gamma = 1 - Prob\{s_{t+1} \neq s_t\}, \tag{1}$$

independent of the policy chosen. We assume that the state is persistent, in the sense that $\gamma \in (0.5, 1)$. Letting μ_t denote the likelihood voters attach to the left-state $s_t = l$, we can write individual preferences as in period t

$$E \sum_{j=0}^{\infty} \beta^j b_{t+j} = E \sum_{j=0}^{\infty} \beta^j b\left(a_{t+j} | s_{t+j}\right). \tag{2}$$

where $\beta < 1$ is the discount factor. Note that, by construction, the issue is *non-partisan* (ideologically neutral) in the sense that all voters unanimously agree on the best alternative: if they knew the state to be s, they unanimously preferred the policy that is appropriate for the state, i.e., $a = s$. Since they do not know s but share a common belief μ, voters prefer policy l over policy r in any given period t if and only if $\mu_t \geq \frac{1}{2}$.

Political decisions are not taken in direct democratic vote. Instead, voters elect an office holder as their representative in each period, who selects and implements the policy alternative a_t. Unlike voters,

---

[8] Provided the policy issue is one-dimensional, the binary assumption could easily be relaxed. Assuming a binary political decision also has some appeal in that voters may find it difficult to make subtle distinctions between policies, e.g., they may only take note of whether government spending goes up or down. In this sense, policies may be quite broadly defined and fit well into the ideological spectrum of 'left' and 'right'. The presumption of one-dimensionality is supported by empirical evidence from the US Congress: in well-known study using data on roll-call votes from the House and the Senate, (Poole and Rosenthal, 1997, 2007) show that more than 80 percent of representatives' voting records over the past 40 years can be explained solely on the basis of the one-dimensional variable (i.e., their 'ideology').

[9] These 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.

politicians observe the state s, which may simply reflect their greater expertise, better access to resources, or their greater incentive to become informed.[10]

There are two observable types of politicians, left-wing L and right-wing R. We interpret the type $i \in \{L, R\}$ as politicians' 'ideology' or 'party affiliation', but any other observable characteristic such as the candidates' gender, their home district, or their previous position on a different (unrelated) policy issue would work equally well. Consistent with our notion that the issue is non-partisan, politicians derive the *same* utility from the policy a given state s as the voters, *independent* of their type i. However, they also care about holding office. We formalize this second motive in the usual fashion by a rent ϕ that politicians receive from being elected to office in period t. In summary, the per-period utility of an incumbent of type i in period t when the state is s_t is

$$u_t^i = b(a_t, s_t) + \phi. \tag{3}$$

When not in office, politicians receive a continuation utility of zero.[11] We thus assume that not being re-elected is an absorbing state, i.e., a once defeated incumbent never returns to holding office.

The timing of the stage game is as follows. First, nature draws the state s_t, which is immediately revealed to politicians but not to ordinary citizens. Next, elections are held in which voters decide whether to re-elect the incumbent or whether to newly elect the challenger for office (a period defines a term of office). Throughout, we restrict attention to the case where the challenger has a different ideology or party-affiliation than the incumbent. Once elected, the office holder chooses a policy alternative a_t. Finally, voters and politicians observe whether the policy was a success ($b_t = b$) or a failure ($b_t = 0$).

2.2 Equilibrium Definition

As is common in these types of models, we will restrict attention to pure strategy, stationary and symmetric Markov perfect equilibria of this game. In those equilibria, players ignore all details of the history (including its length) and condition their strategies only on the pay-off relevant information. Note that because there is no link between periods other than the information revealed by politicians about the underlying state and the evolution of that state, the latter can be summarized for the electorate by its belief μ_t at time t. A strategy for a representative voter specifies the probability $P^i(\mu_t) \in [0, 1]$ with which candidate i is elected, based on μ_t, with $P^L(\mu_t) + P^R(\mu_t) = 1$.[12] When voters are indifferent

---

[10]The natural assumption that politicians are generally better informed than the electorate at large is often evoked in the literature. See, e.g., Cukierman and Tommasi (1998) or Maskin and Tirole (2004). Kessler (2005) provides an analysis where officials to endogenously acquire competence on the issues they oversee and specialize in policy formation.

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

[12]There will be unanimity among electorate, of course, but since no single (infinitesimally small) voter can influence the outcome of an election, every voting strategy is consistent with equilibrium. To eliminate this artificial multiplicity, we will throughout consider a representative voter whose optimal strategy maximizes (4) below, i.e, a strategy that would be optimal in case the voter was decisive (the unique weakly undominated strategy if there is a finite number of citizens).

between two candidates, we assume either stands equal chances of winning the election. Similarly, a strategy for a type-i candidate $a^i(\mu_t, s_t)$ maps voters' beliefs μ_t (and hence, election outcomes) as well as the current state s_t into a policy choice $a \in \{l, r\}$. In equilibrium, strategies must be mutual best responses and beliefs evolve in a way consistent with Bayes rule whenever possible. Strategies are optimal if they maximize the value functions of candidates and voters. The value function for a representative voter can be written as

$$U(\mu_t) = \max_{P^i(\mu_t)} \ E\left[\sum_i P^i(\mu_t) b(a^i(\mu_t, s_t), s_t) + \beta U(\mu_{t+1})\right] \tag{4}$$

where the expectation is taken over b_t and s_t given current beliefs μ_t. Note that in general, beliefs μ_{t+1} at time $t+1$ will depend on the elected candidate, the equilibrium strategy, the implemented policy and the success or failure of the policy in t. The value function of a type i candidate is

$$V^i(\mu_t, s_t) = \max_{a^i(\mu_t, s_t)} \ P^i(\mu_t) E\left[b(a^i(\mu_t, s_t), s_t) + \phi + \beta V^i(\mu_{t+1}, s_{t+1})\right], \tag{5}$$

where the expectation is over b_t and s_{t+1}, given s_t.

3 Equilibrium Analysis

In the following we will 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, involves politicians selecting the alternative that corresponds to their ideology, i.e., $a_t = l$ if $i = L$ and $a_t = r$ if $i = R$, irrespective of the state s_t. Recall from (3) that an office holder's per-period utility is independent of her ideology or party affiliation. Consequently, the sole channel through which ideology can possibly influence the choice of policy is through voters' *expectations*, which for the politicians will translate into the likelihood they are (re-)elected to office. It is this link between actual policy choices and voters' expectations about candidates' post-election behavior – partisan or non-partisan – we are most interested in. What matters, as we will see below, are solely voters' perceptions as to a) what constitutes a left-wing and a right-wing policy alternative, and b) who is a left-wing and a right-wing politician. To highlight the interdependencies, we have eliminated all other well-studied determinants of partisan politics (partisan voters, partisan politicians etc.), not because we consider them implausible but simply because they would only serve to disguise the true effects at work here.

3.1 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 their representatives – have no preferences for either type of politician. Thus, suppose incumbents always choose $a_t^i = s_t$, irrespectively of their ideology or party affiliation i. Since both types of politicians

implement the same Pareto efficient alternative in every period, voters hold no preference for the incumbent or the challenger and elect either with probability $1/2$. Let $U(i, \mu_t)$ be voter's utility from electing an i-type candidate in period t along the equilibrium path. We have

$$U(L, \mu_t) = U(R, \mu_t) \quad \text{and} \quad P^i(\mu_t) = \frac{1}{2} \qquad \forall \mu_t, t, i.$$

The implementation of an efficient policy alternative – precisely because it is necessarily conditional on the current state – provides voters with additional information about s_t. Indeed, since the choice of $a_t = s_t$ perfectly reveals s_t, the only uncertainty about the underlying economy stems from the fact that the conditions may change from one period to the next according to (1). For any initial belief μ_0, beliefs in this equilibrium therefore evolve according to

$$\mu_{t+1}(a_t, \mu_t) = \begin{cases} \gamma & \text{if } a_t = l \\ 1 - \gamma & \text{if } a_t = r \end{cases} \qquad \forall \mu_t, t.$$

In what follows, we will for notational simplicity focus on left-wing politicians $i = L$, dropping the index i whenever possible. The argument for right-wing politicians $i = R$ is analogous. Recalling that $b_t \equiv b$ if $a_t = s_t$ the value function of an incumbent politician if he or she implements the efficient alternative is

$$V(s_t) = \frac{1}{2} \left\{ b + \phi + \beta \ E\left[V(s_{t+1})\right] \right\}.$$

Note that $V(s_t)$ is independent of μ_t, because given the electorate's voting rule any incumbent faces equal chances of being re-elected and defeated, respectively, regardless of beliefs. If the incumbent deviates by choosing $a_t \neq s_t$ in some t, the value function becomes

$$\hat{V}(s_t) = \frac{1}{2} \left\{ \pi b + \phi + \beta E\left[V(s_{t+1})\right] \right\},$$

which by inspection is strictly less than $V(s_t)$ for any $\pi < 1$. Hence, $a_t = s_t$ is indeed the utility-maximizing choice for incumbents in each period. 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,

$$U^{\max} = \sum_{t=0}^{\infty} \beta^t b = \frac{1}{1 - \beta} b.$$

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 Pareto-dominates all other equilibria for the voters, it is not the only possible outcome. In the following sections, we will not only demonstrate that partisan politics can be supported in equilibrium as well, but also that non-partisan politics are fragile in the sense that they cannot survive if citizens' expectations about office holders' behavior are subject to (small) uncertainty.

9

3.2 The Partisan Equilibrium

We next study the possibility of a partisan equilibrium. Intuitively, suppose voters' expect office holders to play partisan and choose $a_t = i$ in every period, independent of the current state s_t. The key to observe is that voters are no longer indifferent across politicians with distinct ideologies. In particular, if a voter knew the state to be $s_t = l$, he or she would *strictly prefer* a type-L candidate to a type-R candidate, because only the former's partisan behavior coincides with the efficient policy choice in period t. A direct consequence of this strict preference ordering is that period-t incumbents now face a dilemma whenever their ideology does not match the state. A type-L 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. Similarly, 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, will *conceal* 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.

To this end, consider a type-i candidate whose strategy is to choose the partisan policy whenever in office in period t. Given $\mu_0 \in [1 - \gamma, \gamma]$, the voters' belief along the equilibrium path then evolves as follows

$$
\mu_{t+1}^L(a_t = l, \mu_t) = \begin{cases} 1 - \gamma + (2\gamma - 1)\frac{\mu_t}{\mu_t + (1 - \mu_t)\pi} & \text{if policy } a_t = l \text{ was a success} \\ 1 - \gamma & \text{if policy } a_t = l \text{ was a failure} \end{cases} \tag{6}
$$

$$
\mu_{t+1}^R(a_t = r, \mu_t) = \begin{cases} \gamma - (2\gamma - 1)\frac{1 - \mu_t}{1 - \mu_t + \mu_t \pi} & \text{if policy } a_t = r \text{ was a success} \\ \gamma & \text{if policy } a_t = r \text{ was a failure.} \end{cases}
$$

Note that the office holders' policy choice reveals no new information about the current state on the equilibrium path since the implemented policy always corresponds to the politcians' affiliation. Formally, the beliefs satisfy the property $E[\mu_{t+1}^L | a_t = l, \mu_t] = E[\mu_{t+1}^R | a_t = r, \mu_t] = \gamma \mu_t + (1 - \gamma)(1 - \mu_t)$. Thus, the electorate only learns by observing whether the policy has been successful or not.

As usual, beliefs are not defined off the equilibrium path, i.e., when the electorate observes the non-partisan policy being implemented. Off equilibrium, we make the natural assumption that non-partisan politics are perfectly revealing

$$
\mu_{t+1}^L(a_t = r) = 1 - \gamma \quad \text{and} \quad \mu_{t+1}^R(a_t = l) = \gamma, \tag{7}
$$

i.e., if the electorate unexpectedly observes a left-wing office holder to select $a_t = r$, it assumes that the non-partisan state $s_t = r$ must have occurred, and vice versa.[13]

Now suppose voters elect the left-wing (right-wing) candidate for beliefs $\mu_t > 1/2$ ($\mu_t < 1/2$) and give

---

[13]After adapting the Cho&Kreps intuitive criterion to our dynamic framework, it is easy to verify that this out of equilibrium belief is the unique belief satisfying the corresponding concept of equilibrium dominance, on which the intuitive criterion is based.

both candidates equal chances of winning for $\mu_t = 1/2$. The value function of the electorate is then

$$U(\mu_t) = \begin{cases} (\mu_t + (1 - \mu_t)\pi)\left(b + \beta U(\mu_{t+1}^L)\right) + (1 - \mu_t)(1 - \pi)\beta U(1 - \gamma) & \mu_t \geq 0.5 \\ (1 - \mu_t + \mu_t\pi)\left(b + \beta U(\mu_{t+1}^R)\right) + \mu_t(1 - \pi)\beta U(\gamma) & \mu_t < 0.5. \end{cases} \quad (8)$$

Closer inspection of (8) reveals that $U(\mu_t)$ is increasing in μ_t for values $\mu_t > 1/2$ and decreasing in μ_t otherwise (at $\mu_t = 1/2$, the function assumes a minimum). Intuitively, 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 receive more precise information about the state.[14] 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. Put differently, the electorate would dispose of a more accurate belief only if the implemented policy goes awry. In the 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.

Turning now to candidates, we will without loss of generality again consider the behavior of left-wing candidates, omitting the index L whenever possible. Anticipating the voting behavior of the electorate, the equilibrium value of acting partisan for a left-wing candidate is

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

where the expectation is taken over b_t (and, consequently, μ_{t+1}) given s_t, and

$$P(\mu_t) = \begin{cases} 1 & \text{if } \mu_t > 0.5 \\ 0.5 & \text{if } \mu_t = 0.5 \ . \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

A candidate who deviates by setting $a_t = r$ in period t, in contrast, would reveal the true state to be $s_t = r$. Voters' beliefs at the beginning of the next period are therefore $\mu_{t+1} < 1/2$, resulting in certain defeat and a utility normalized to zero. Hence, we can write the office holder's utility $\hat{V}(\mu_t, s_t)$ from such a deviation as

$$\hat{V}(\mu_t, s_t) = \begin{cases} P(\mu_t)\left\{\pi b + \phi\right\} & \text{if } s_t = l \\ P(\mu_t)\left\{b + \phi\right\} & \text{if } s_t = r. \end{cases}$$

Obviously, no rational incumbent would ever want to select an opponent's partisan policy in a state where in fact her own partisan policy is myopically optimal. Thus, the strategy $a_t = l$ is trivially utility maximizing in the 'partisan' state $s_t = l$. It remains to study when politicians are willing to sacrifice the utility from the Pareto-optimal choice of $a_t = r$ by choosing $a_t = l$ in state $s_t = r$. Comparing $V(\mu_t, r)$ with $\hat{V}(\mu_t, r)$, we see that the answer is yes if $V(\mu_t, r) \geq \hat{V}(\mu_t, r)$ or

$$\beta E\left[\gamma V(\mu_{t+1}, l) + (1 - \gamma)V(\mu_{t+1}, r)\right] \geq (1 - \pi)b. \quad (10)$$

---

[14]See Lemma A1 in the Appendix, which formally establishes how $U(\mu_t)$ depends on μ_t and shows that experimentation does not improve voters' payoffs.

On the right-hand side of (10) are the short-term gains from deviating, as reflected in the additional expected benefit from the optimal non-partisan choice over the suboptimal partisan choice. The left-hand side captures the utility lost by facing certain defeat in this case; it is the future value from remaining in office, which naturally increases in the discount factor β and office rents ϕ (see below). But another, and perhaps less apparent, factor also plays a crucial role: by acting partisan, the candidates must also be able to *improve* their (re-)election chances by a sufficient margin. For the remainder of this section, we will therefore assume that the success probability π of a sub-optimally chosen partisan policy is small enough, such that an office holder who chooses the partisan policy has a chance of being re-elected for any belief $\mu \in [1 - \gamma, \gamma]$. In other words, even for $\mu_t = 1 - \gamma$, the electorate's updated belief satisfies $\mu_{t+1} > 0.5$, which is equivalent to

Assumption 1.

$$\pi < \frac{1 - \gamma}{\gamma}. \tag{A1}$$

Under Assumption 1, a success guarantees re-election (and failure results in sure defeat) irrespective of the state s_t or of the belief μ_t. In this case, $V(\mu_t, s_t)$ assumes a particularly simple form. It is constant (and equal to zero) for beliefs $\mu_t \in [1 - \gamma, \frac{1}{2})$ where the candidate is not elected in equilibrium, takes on a single intermediate value for $\mu_t = \frac{1}{2}$, and is constant again for all higher beliefs $\mu_t \in (\frac{1}{2}, \gamma]$, where the candidate is elected with probability one. Formally, $\forall \mu_t \in (0.5, \gamma]$ we have $P(\mu_t) = 1$ and $\mu_{t+1} > \frac{1}{2}$ if the policy was successful and $\mu_{t+1} = 1 - \gamma < \frac{1}{2}$ otherwise. $V(\mu_t, s_t) \equiv \bar{V}(s_t)$ for all values in this interval. Similarly, $\forall \mu_t, \in [1 - \gamma, \frac{1}{2})$, $P(\mu_t) = 0$, implying $V(\mu_t, s_t) \equiv 0$. Selecting the non-partisan policy in state r then will not be optimal if

$$b + \phi \le \pi b + \phi + \pi \beta [(1 - \gamma) \bar{V}(l) + \gamma \bar{V}(r)]$$

or

Assumption 2.

$$(1 - \pi) b \le \pi \beta [(1 - \gamma) \bar{V}(l) + \gamma \bar{V}(r)] \tag{A2}$$

where $\bar{V}(r)$ and $\bar{V}(l)$ can explicitly be computed to read

$$\bar{V}(r) = \frac{b\pi(1 + \beta(1 - 2\gamma\beta)) + (\pi\beta(1 - \gamma) + 1 - \beta\gamma)\phi}{\pi\beta(\beta(2\gamma - 1) - \gamma) + 1 - \beta\gamma}$$
$$\bar{V}(l) = \frac{b(\pi\beta(1 - 2\gamma) + 1) + (1 - \beta(\pi\gamma + \gamma - 1))\phi}{\pi\beta(\beta(2\gamma - 1) - \gamma) + 1 - \beta\gamma}. \tag{11}$$

We can conclude:

Proposition 2. *[Partisan Equilibrium] Under (A1) and (A2), there exists an equilibrium in which elected office holders act partisan regardless of the state. In this equilibrium, politicians are re-elected with*

12

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).

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) and Maskin and Tirole (2004) investigate this phenomenon, which turns the accountability role of elections on its head. The authors show that, 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. While similar in its consequences, the policy choice in a partisan equilibrium does not follow the most popular course of action. Instead, incumbents in our model stick to their once enacted policies 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-congruency: ideology is a social perception not an innate characteristic of the candidates.

In particular, comparing Proposition 1 and 2, the blame for the policy bias can be squarely laid on the fact that voters *perceive* policies to be ideologically tinted and *expect* candidates to act partisan. If any one of these conditions is missing, i.e., policies are perceived to be ideologically neutral or candidates are expected to act non-partisan, even the most office-minded politician has no incentive to deviate from what is optimal for the electorate [Proposition 1]. Only if voters expect partisan politics in the future will they have an incentive to elect the candidates whose perceived position corresponds to what they think is in their best interest given their current information. And it is the voters' expectations, in turn, which induce candidates to actually act partisan, i.e., according to their ideology, in the first place. Put differently, voters and representatives are caught in an *ideology trap*: 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. Shifts from non-partisan politics to partisan politics confirm the electorate's assessed likelihood of the latter, cementing the polarization even further. Ideologues emerge who are *not* true believers. Instead, ideology is purely a social perception based on observable characteristics of candidates: if voters expect a female representative from California who supports gun control to also favor big government, then this is what she will do in equilibrium. Thus, issue bundling occurs not because preferences are bundled, but because voters' expectations tie candidates' policy intentions to their observed characteristics (such as their party affiliation or their position on other issues).

Similarly, the model also should be contrasted with the widely-used adverse selection approach of reputation in repeated games, initially formalized by Kreps et al. (1982) and Kreps and 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.[15]

---

[15] In 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).

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. Obviously, one could not possibly explain ideologically tinted behavior with this line of argument. In contrast, there is no uncertainty about the candidates' type in our model. Thus, implementing efficient policies in the partisan equilibrium cannot serve as a signal for being an efficient type. Rather, the electorate is unsure about the current state of the world, and an incumbent who implements a non-partisan policy will at most signal that a certain state prevails, which in turn makes it desirable to out him from power.

Note that the the existence of the partisan equilbrium 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.[16]

There are two possible misgivings one could have against this line of reasoning. First, voters are strictly better off in the non-partisan equilibrium than in the 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. As we will see, both concerns are rooted in the simple nature of the model and can easily be addressed. We do so in Section 4 below, where 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]

We close this section by studying the set of parameters that supports partisan behavior as an equilibrium phenomenon. First, note that Assumption 1 is satisfied for small values of either π or γ, or both. Ceteris paribus, partisan behavior is thus more likely to arise if either i) the electorate is sufficiently uncertain about the underlying state or ii) the success and failure of policies is a sufficiently accurate signal of the state. Intuitively, these conditions ensure that challengers do not credibly deviate to non-partisan behavior (which in turn would make their election optimal for voters). If the state persists over long time horizons ($\gamma \to 1$) or if the signal of a policy's success or failure is very inaccurate ($\pi \to 1$), a challenger who unexpectedly (i.e., off the equilibrium path) won an election would have no incentive to act partisan because even if her partisan choice was successful, the electorate would not be sufficiently convinced of an underlying state change to re-elect her.

---

[16] A 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 (12) still holds for sufficiently high ϕ/b.

[17] In 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 $\beta = 1$.

Second, to better understand the restrictions embodied in Assumption 2, we can substitute for $\bar{V}(r)$ and $\bar{V}(l)$ in condition (A2) using (11), which yields

$$\frac{b}{\phi + b} \leq \frac{\pi\beta\left(1 - \beta(2\gamma - 1)\right)}{(1 - \pi)\left(1 - \beta\gamma\right)} \tag{12}$$

Thus, and not surprisingly, partisan behavior is more likely to arise whenever politicians have a strong office holding motive: their rent from holding onto power ϕ, relative to the the payoff b they forgo by not choosing the correct policy must be sufficiently high. Moreover, the incumbent will be more inclined to play partisan for high values of π, i.e., whenever the efficiency cost of inappropriate policies is low because they are still likely to succeed (note the tension to (A2) though, which requires π to be low enough for a successful partisan policy to be convincing). Less obviously, the left hand side of (12) decreases in γ. Intuitively, since the incumbent faces the trade-off between reelection and efficiency only if the state is unfavorable ($s_t = r$), 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 can thus conclude:

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

3.3   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 "ideologues" but also because, by definition, an ideologue'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' and do not change policies 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.[18]

---

[18]One may object to this assertion that since voters are indifferent between candidates in the non-partisan equilibrium,

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 persistence is particularly interesting for two reasons. First, it shows that policies may be resilient not just because they are targeted and thus allow for the formation of powerful interest groups who subsequently lobby for their continued enactment as in Coate and Morris (1999). Persistence may also be a problem for non-targeted (valence) issues, simply because incumbent politicians may be reluctant to abandon their previously enacted policies so as to not openly admit that "times have changed". Second, this persistence gives rise to political failure. Rather than the result of a struggle between powerful interest groups and the public at large, the inefficient inertia in the political process 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.

Both policy persistence and incumbency advantage distinguish our model from other models of policy divergence (such as the citizen-candidate model) and can potentially be tested for empirically. While a full-fledged empirical analysis of these phenomena is beyond the scope of the present paper, we confine ourselves to point out that these implications are consistent with empirical observations regarding democratic two-party systems. As stated in the Introduction, studies of voting behavior in the U.S. Congress in particular confirm our theoretical predictions of ideological positioning and polarization along party lines [McCarty et al. (2006)]. Using data from roll call voting records, Poole (2007) presents a variety of evidence showing that, once elected, members adopt a consistent ideological position and maintain it over time. Moreover, 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 2002, for instance, 398 House members ran for reelection, of which only 16 were defeated. In the Senate, a mere three out of 26 senators running for reelection lost.

---

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 implausible scenario when the electorate is large.

4 Extensions: Voter Uncertainty

4.1 Candidate Behavior

As mentioned above, one possible objection to the partisan equilibrium is that it is Pareto dominated by the non-partisan equilibrium for the voters (though not for the politicians). Arguably, this could make sub-optimal partisan behavior less likely to be observed: if the electorate collectively benefits from expecting representatives to act in its best interest, then why should it expect otherwise? We will show in this section that there are compelling arguments in favor of the partisan equilibrium. Specifically, the non-partisan equilibrium is fragile (unstable) 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 with some small probability $\epsilon > 0$.[19]

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 follow their ideology in each period. Then, the partisan equilibrium continues to exist and generically there is no equilibrium in which each candidate plays 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] Conversely, a seemingly extraneous act such as a public appeal

---

[19]One explanation for why voters could expect partisan behavior to arise with positive probability is party pressure [see Cukierman and Tommasi (1998)]. The possibility of a "partisan shock" could then 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., $B > (1 - \pi)b$.

[20]For 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.

[21]Using 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, is has become much more infrequent afterwards.

for non-partisanship could revert voters' expectations, thus helping political actors to coordinate on the efficient equilibrium. For this reason, the result is also consistent with – and can possibly account for – occurrences of within-party polarization and convergence, such as the split between Southern and Northern Democrats during the Civil War area and its diminishing importance in the past decades.

4.2   Policy Prospects

In this section we allow voters to be uncertain as to the prospect of an inefficiently chosen policy. Apart from capturing reality, the extension serves two purposes. First, since candidates will prefer to implement efficient (non-partisan) policies whenever their partisan policy is unlikely to succeed, voters will observe non-partisan behavior on the equilibrium path, eliminating out-of equilibrium beliefs. Second, the partisan equilibrium will exist for a wider range of parameters.

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 $s_t \in \{r, l\}$. 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 again elect the left-wing (right-wing) 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 begin with, suppose the left-wing candidate has been elected in a crisis period and $s_t = r$. A partisan policy $a_t = l$ 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.

As before, a candidate is only willing to implement the partisan policy if this assures reelection in case of success; in particular this must be true if the electorate holds the worst possible beliefs, $\mu_t = 1 - \gamma$. However, since the partisan policy is less often implemented than in the base model, observing a successful partisan policy now contains more information and therefore has a larger effect on the posterior belief.

---

[22]The 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.

Specifically, (A1) becomes

$$\pi < \frac{(1-\gamma)}{\gamma q}. \tag{A1'}$$

Next, let $V^c(\mu_t, s_t)$ denote the left-wing 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 (10) of the base model that supports partisan behavior in any *normal* period,

$$\beta E \left[\gamma V^c(\mu_{t+1}, l) + (1-\gamma)V^c(\mu_{t+1}, r) \right] \geq (1-\pi)b \tag{A2'}$$

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

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

It is easy to show that $V^c(\mu_t, s_t) = 0$ for $\mu_t < 0.5$ and $V^c(\mu_t, s_t) \equiv \overline{V}^c(s_t)$ for $\mu_t > 0.5$, as in section 3.2.

Simple algebra shows that (A2') is equivalent to (A2) from section 3.2. To intuitively understand why (A2) remains unchanged, assume for the moment that (A2) 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): $\overline{V}^c(r) = \phi + b$. 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 3.2, we must also have $\overline{V}(l) = \overline{V}^c(l)$. 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. $\overline{V}^c(l) \leq \overline{V}(l)$ and $\overline{V}^c(r) \leq \overline{V}(r)$.

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 $q \to 0$, thereby justifying the off-equilibrium beliefs of section 3.2: 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.

5 Discussion and Concluding Remarks

This paper proposes a theory of ideology for public leaders. We have shown that there are circumstances under which elected officials may adopt ideologically opposed positions, resulting in inefficient partisan policies even in areas that are generally perceived to be non-partisan. In contrast to existing explanations of partisanship, equilibrium polarization can emerge in our model despite the fact that voters and their representatives are in complete agreement as to which is the optimal course of action. The problem the parties face can be viewed as an 'ideology trap' of self-fulfilling expectations, which emerges because voters perceive alternative policy measures to be ideologically tinted, and expect candidates to remain 'true to their ideology' which itself is a social perception grounded in observable characteristics (such as their gender, their party affiliation, or their position on a different policy issue).

The basic argument is simple: if voters expect political candidates to act partisan once in office, they have an incentive to elect the a representative whose perceived partisan policy (ideology) corresponds to what they think is in their best interest based on their current information. As we show, this may suffice to induce candidates to actually act partisan in the first place, thereby confirming the expectations of the electorate. This is because choosing the efficient (non-partisan) 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, a sufficiently office-motivated incumbent thus 'jams' the voters' inference problem by instead using the partisan policy, which is less responsive to current circumstances.

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 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 polices 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 leaders in the eyes of the electorate. [23]

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 ideology trap. This is true, of course. Suppose for example that there is a (possibly small) probability of a non-partisan 'type' of politician to occur, one that will always implement the efficient, non-partisan policy. Politicians then may have an incentive to deviate from their ideological stance to signal that they are this type of candidate. As the result in Proposition 3 shows, however, if a deviation

---

[23]Using the US relations to Iraq as an example, take George W. Bush's reluctance to admit that his strategy in Iraq failed. According to our model, it is not the gain from appearing competent (or the loss from appearing incompetent) that causes the political failure. Instead, admitting mistakes would imply that the Democrats' strategy to deal with the situation in Iraq was preferable, which in turn implies that a Democrat could do better when in office.

from partisanship may also be the result of an error (rather than as a deliberate equilibrium choice), voters may still prefer a (known) partisan candidate. At least in the short run, the state of the world is revealed in a deviation, which makes voters prefer the ideologue devil they know over a potentially non-partisan candidate, as long as they cannot be one hundred percent sure about the latter's idealism. The nature of our equilibrium is thus somewhat 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.

Specifically, the key insight from our analysis provides a plausible explanation for a range of empirical regularities that, collectively, the previous literature on polarization cannot account for. In particular, the theory shows why ideology plays a role on matters that should be non-partisan (e.g., national security), why differences in observed characteristics such as party affiliation, gender, or electoral district can lead to differences in the political platforms of candidates that otherwise share similar policy preferences, why these differences can lead to issue bundling, and why bad, ideology driven policies can persist. The model is also consistent with the observation that the ongoing polarization in the U.S. (which is largely driven by increased extremism within the Republican Party) is not mirrored by an equal shift in the American public. Our view that this development need not reflect a trend in the true preferences of the political elites seems to be shared by some prominent observers of American politics. Commenting on the Republican stance on ObamaCare, for instance, David Frum, who is a former speech writer to President George W. Bush and fellow of the American Enterprise Institute, said in a resent CBC radio interview that Republican candidates " know that unless they strike certain very strong notes they will not be acceptable in a world of talk and Fox News. That will cut them off from potential support. People who could be real world leaders, actual representatives, say things they don't think and take positions they don't believe in."[24]

Finally, we believe that our model could fruitfully be applied to other settings where leadership and ideology 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 live up to his supporters expectations. If those expectations are ideologically biased, then ideologues will emerge irrespective of whether the context is one of political, religion, or ethnic affiliation.

---

[24]Comments made in the show "The current" on CBC radio, April 1st 2010. Recording available from the CBC website (see http://www.cbc.ca/thecurrent/podcast.html) or from the authors upon request.

References

Alesina, A. and A. Cukierman (1990). The Politics of Ambiguity. *The Quarterly Journal of Economics*, 105(4), pp. 829–850.

Ansolabehere, S. and J. M. J. Snyder (2000). Valence Politics and Equilibrium in Spatial Election Models. *Public Choice*, 103(3-4), pp. 327–36.

Bénabou, R. (2008). Joseph Schumpeter Lecture: Ideology. *Journal of the European Economic Association*, 6(2-3), pp. 321–352.

Bénabou, R. and J. Tirole (2006). Belief in a Just World and Redistributive Politics. *The Quarterly Journal of Economics*, 121(2), pp. 699–746.

Bernhardt, D., B. Buehler, and A. Kessler (2016). The Partisan Paradox. Mimeo, Simon Fraser University.

Besley, T. J. and S. Coate (1997). An Economic Model of Representative Democracy. *The Quarterly Journal of Economics*, 112(1), pp. 85–114.

Callander, S. (2008). Political Motivations. *Review of Economic Studies*, 75(3), pp. 671–697.

Canes-Wrone, B., M. Herron, and K. Shotts (2001). Leadership and pandering: A theory of executive policymaking. *American Journal of Political Science*, 45(3), pp. 532–550.

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

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

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

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

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

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

Kessler, A. S. (2005). Representative versus direct democracy: The role of informational asymmetries. *Public Choice*, 122(1), pp. 9–38.

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, 27(2), pp. 245 – 252.

Kreps, D. M. and R. Wilson (1982). Reputation and imperfect information. *Journal of Economic Theory*, 27(2), pp. 253 – 279.

Lee, D. S., E. Moretti, and M. J. Butler (2004). Do Voters Affect or Elect Policies? Evidence from the U.S. House. *The Quarterly Journal of Economics*, 119(4), pp. 807–859.

Levitt, S. D. (1996). How Do Senators Vote? Disentangling the Role of Voter Preferences, Party Affiliation, and Senate Ideology. *American Economic Review*, 86(3), pp. 425–441.

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

McCarty, N., K. T. Poole, and H. Rosenthal (2006). *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge, Mass.: The MIT Press.

Meernik, J. (1993). Presidential Support in Congress: Conflict and Consensus on Foreign and Defense Policy. *The Journal of Politics*, 55(3), pp. 569–587.

Morris, S. (2001). Political Correctness. *Journal of Political Economy*, 109(2), pp. 231–265.

Osborne, M. J. and A. Slivinski (1996). A Model of Political Competition with Citizen-Candidates. *The Quarterly Journal of Economics*, 111(1), pp. 65–96.

Palfrey, T. R. (1984). Spatial Equilibrium with Entry. *Review of Economic Studies*, 51(1), pp. 139–56.

Poole, K. (2007). Changing minds? Not in Congress! *Public Choice*, 131(3), pp. 435–451.

Poole, K. T. and H. Rosenthal (1997). *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

Poole, K. T. and H. Rosenthal (2007). *Ideology & Congress*. Transaction Pub.

Stasavage, D. (2007). Polarization and Publicity: Rethinking the Benefits of Deliberate Democracy. *Journal of Politics*, 69(1), pp. 59–72.

Stokey, N. L. and R. E. Lucas (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.

Van Weelden, R. (2009). Candidates, Credibility, and Re-election Incentives. Mimeo, Yale University.

Appendix

The following lemma establishes that there is no 'experimenting' in equilibrium

Lemma A1. *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 $U'(\mu_t) \leq -(1 - \gamma)(1 - \pi)^2 b$ for $\mu_t < 0.5$ and $U'(\mu_t) \geq (1 - \gamma)(1 - \pi)^2 b$ for $\mu_t > 0.5$,

    iii) satisfies $b + \beta U(\mu_t) - \beta U(1 - \gamma) \geq (1 - \gamma)(1 - \pi) b \quad \forall \mu_t \in [1 - \gamma, \gamma]$.

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

Proof. Let the function $\varphi^L(\mu_t) \equiv 1 - \gamma + (2\gamma - 1)\frac{\mu_t}{\mu_t + (1 - \mu_t)\pi}$ map the the belief μ_t in period t into the belief that results in $t + 1$ when incumbent L 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:

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

Since T is a contraction, there exists a unique electorate's value function $U(\cdot)$.[25] We will prove properties i), ii) and iii) of U by use of Corollary 1 of Stokey and 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 $\varphi^L(0.5 + x) = 1 - \varphi^R(0.5 - x)$, TU also satisfies $TU(0.5 - x) = TU(0.5 + x)$ for $x \in [0, \gamma - 0.5]$.

ii) For $\mu_t > 0.5$,

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

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

---

[25] It 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 and Lucas (1989, Theorem 3.3). As $\mathcal{U}$ together with the *sup*-Norm is a complete metric space, the contraction mapping Theorem Stokey and Lucas (1989, Theorem 3.2) applies.

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

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

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 left party if $\mu_t > 0.5$ (an analogous argument holds for $\mu_t < 0.5$). Deviating once and electing the right party yields $\hat{U}(\mu_t) = (1 - \mu_t + \mu_t\pi)\left(b + \beta U(\varphi^R(\mu_t))\right) + \mu_t(1 - \pi)\beta U(\gamma)$. Hence

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

where the inequality follows because the first term is positive due to property iii) and the second is positive due to $\varphi^L(\mu_t) - 0.5 > |0.5 - \varphi^R(\mu_t)|$ and property ii). To see that $\varphi^L(\mu_t) - 0.5 > 0.5 - \varphi^R(\mu_t)$ 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 $\mu_t > 0.5$. $\qquad\square$

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 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 $\lim_{t\to\infty} Pr(\tilde{s}_t = n) = \frac{1-\gamma}{1-2\gamma\pi+\pi}$. Due to $\gamma < 1$, we have

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

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

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

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 Lemma A1. 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

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

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

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

for $\mu_t > 0.5$ since L-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, (10) is satisfied with strict inequality,

$$(1 - \pi)b < \pi\beta[(1 - \gamma)\bar{V}^P(l) + \gamma\bar{V}^P(r)], \tag{13}$$

where $\bar{V}^P(l) > 0$ and $\bar{V}^P(r) > 0$ are candidate L'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 (13) 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 (A1) 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

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

To see that this inequality is indeed satisfied, note that

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

where the first inequality comes from (13) and the second from the hypothesis $\bar{V}^P(s) > \bar{V}^{\epsilon E}(s)$.

Proof of Proposition 5

Note that condition (A2') is equivalent to $\overline{V}^R(r) \geq b + \phi$ where we use the same notation as in the base model, i.e. $\overline{V}^R(s_t) \equiv V^R(\mu_t, s_t)$ 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 $\overline{V}^R(l) \geq \frac{(1 + \beta(1-\gamma))(b+\phi)}{1-\gamma\beta}$. Applying a contraction argument similar to Lemma A1, we have to show that whenever $\overline{V}^R(r) \geq b + \phi$, $\overline{V}^R(l) \geq \frac{(1+\beta(1-\gamma))(b+\phi)}{1-\gamma\beta}$ and assumption (A2) holds, then the following two inequalities are satisfied:

$$b + \phi + \beta\left[\gamma\overline{V}^R(l) + (1 - \gamma)\overline{V}^R(r)\right] \geq \frac{(1 + \beta(1 - \gamma))(b + \phi)}{1 - \gamma\beta}$$

$$\pi b + \phi + q\beta\pi\left[(1 - \gamma)\overline{V}^R(l) + \gamma\overline{V}^R(r)\right] + (1 - q)(1 - \pi)b \geq b + \phi$$

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

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

27

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