# Partisanship as a Self-Fulfilling Prophecy*

Anke Kessler† Benno Bühler‡

January 2025

**Abstract**

We provide a new perspective on why politicians adopt partisan positions. In a simple model where office-holders have better information on which policy is best, we show that partisanship emerges even though everyone has identical non-partisan preferences. If voters associate politicians' identities (party affiliations) with their future policy choices, both sides can get caught in a 'partisan trap': voters' expectations of how a candidate's identity shapes their political actions generates incentives for politicians maintain a partisan stance, concealing that circumstances have changed. Expectations are thus self-fulfilling. The resulting equilibrium is one of political failure, policy persistence, and incumbency advantage.

*Keywords:* Partisanship, Self-fulfilling Expectations, Elite Polarization, Policy Inertia

*JEB classification:* D78, P16, H11

*We wish to thank Dan Bernhardt, 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 the Canadian Institute for Advanced Research and the Social Science and Humanities Research Council of Canada (Anke Kessler) and the German Science Foundation and German Academic Exchange Service (Benno Bühler) 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, voters do not like partisan politics.[1] Yet, partisanship has been on the rise: as Poole and Rosenthal (2007), McCarty, Poole and Rosenthal (2006) and others have extensively documented, voting in the U.S. Congress can largely be reduced to a single dimension which mirrors party affiliationand polarization has been steadily increasing since the 1970s. The phenomenon of elite polarization is neither exclusive to the U.S. (Goldberg and de Vreese, 2020) 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 observed even when there should be a general agreement among the electorate on issues such as crime, corruption, or foreign policy.In the U.S. Congress, for example, support for the president on matters of foreign policy and defence has largely been along party lines ever since the Vietnam War (Meernik, 1993).

Why, then, do officeholders consistently choose the same partisan and 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 provide a novel perspective on this question. 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, simply because voters' *expect* them to. Referring back to the above quote by James Harvey Robinson, it is the assumption of partisanship that generates it in the first place.

We consider a simple dynamic model where office holders have information on an underlying state that determines the efficacy of policy alternatives, while voters do not. Voters form expectations about a) which policy candidates are likely to implement once in office and b) which policy is most likely to succeed. To develop our argument in the strongest manner possible, we assume that everyone, candidates and voters alike, derive the same utility from the policy measure.[2] Hence, candidate identity (partisan affiliation, historic voting record, gender, geographic location) is payoff-irrelevant. Nevertheless, we show that candidate identity matters: different office holders adopt a different partisan policy and maintain it over time. The argument is one of self-fulfilling expectations and runs as follows. Suppose voters associate different policy alternatives with different parties or ideologies (there is a 'liberal' and a 'conservative' policy) and connect those with the identity of the candidates in their expectations on they will act once in office: they assume that candidates remain 'true to their colors' in implementing the respective policies affiliated with their party or ideology (historical voting record). Given these expectations, voters have a straightforward incentive to elect the representative whose perceived partisan policy (ideology) corresponds to what they think is

[1]A recent (2024) Gallup poll finds U.S. Congress approval ratings to be at a historic 13 percent low, with "party gridlock/bickering/not compromising" cited as the number one reason for disapproval.

[2]It should be emphasized that our insight also applies to partisan issues and preferences. There already is extensive literature on these types of settings, 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 their best interest based on their current information. We show that this effect induces candidates to 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 favor her opponents' partisan position, making her re-election more likely if voters expect partisan behavior in the future.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 cement the electorate's belief and improve electoral prospects in the next election. The result is political failure in the sense that the equilibrium partisan policy outcomes are payoff dominated for the electorate. 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 a '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 our stylized framework highlights, 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 need not be '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 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 partisan policies independent of the prevailing state in equilibrium, our model also implies policy inertia: despite new circumstances, inefficient policies are likely to persist.

*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. Indeed, there is considerable literature on why elite polarization emerges in the first place and why it may increase over time, which we survey below. We seek to complement this literature by illustrating that voters' expectations of partisan behavior can be self-fulfilling.[3]

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 and Snyder, 2000), politically motivated candidates Besley and Coate (1997); Osborne and Slivinski (1996), valence advantages (Aragones and Palfrey, 2002; Buisseret and Van Weelden, 2022), rent-seeking officeholders (Van Weelden, 2013), party discipline (Canen, Kendall and Trebbi, 2020), and incomplete information among candidates or voters[4]Those explanations generally build on preexisting differences in policy preferences among candidates or voters.

There is a small literature on how polarization emerges in non-partisan settings. Ashworth and Bueno de Mesquita (2009) show that candidates take partisan stances to soften competition in other dimensions such as campaign spending. In Carrillo and Castanheira (2008) candidates diverge to increase future (self-)commitment

[3]For a discussion of the large empirical literature on elite polarization and further references, see e.g. McCarty, Poole and Rosenthal (2006).

[4]See e.g. Callander (2008); Castanheira (2003); Kartik and McAfee (2007); McMurray (2022).

to invest in the quality of their policies. Howell, Krasa and Polborn (2020) show how preference differences over the value of a long-term public project can generate political conflict even if everyone agrees on project implementation. In Izzo, Martin and Callander (2023), policy-motivated parties develop ideologies that affect what voters infer from observed outcomes. Winning the ideological contest thus confers a long-term advantage and polarization always occurs. Dziuda and Loeper (2016) consider (legislative) bargaining when the previous agreement becomes the next status quo, and show that this exacerbates the players' conflict of interest (polarization), resulting in gridlock. Finally, Callander and Carbajal (2022) note a strategic incentive for parties to diverge when voter preferences change with voting behavior; a feedback loop emerges with (progressive) polarization among both elites and the electorate that matches many aspects of observed polarization trends.

The idea that elected officials may take distortionary actions to prevent voters from learning has been explored in previous work. Aghion and 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, Bueno de Mesquita and Friedenberg (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 effort (higher effective electoral accountability) can lead to less informed re-election choices and decreased welfare when politicians' actions and characteristics are local substitutes. Our argument is also related to, but distinct from, existing work on political signalling and policy failures in scenarios where politicians are better informed than voters. Heidhues and Lagerlöf (2003), Maskin and Tirole (2004) Stasavage (2007), Callander (2008) and more recently Morelli and 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.

Finally, the self-fulfilling nature of the partisanship equilibrium is reminiscent of theories that were developed to explain bubbles in financial markets (Tirole, 1985) or persistent booms or recessions in the macroeconomy (e.g., Azariadis, 1981). The latter literature, in particular, studies rational expectations equilibria in which outcomes are perfectly correlated with non-fundamentals, i.e. 'sunspots' (see Azariadis and Guesnerie, 1986; Cass and Shell, 1983). In our model, voters associate a candidate's identity – ethnicity, gender, geographic residence, prior voting record 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 sunspots that pin down expectations.

# 2 The Model

## 2.1 Basic Framework

We consider an infinite-horizon economy in discrete time, 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.Without loss of generality, we also assume that the underlying state is binary, $s_t \in \{b, r\}$, and that it is optimal to match the policy to the state. Specifically,

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

Thus, if the policy choice aligns with the state, the policy is successful with probability one.[5] The representative citizen-voter knows the set of feasible policies but is uncertain about the underlying $s_t$, which evolves according to a stationary, persistent Markov process with $Prob\{s_{t+1} = s_t\} = \gamma > \frac{1}{2}$. His expected utility in $t$ is $E \sum_{j=0}^{\infty} \beta^j v\left(a_{t+j}, s_{t+j}\right)$ where $\beta < 1$ is the discount factor.

The voter elects an officeholder in each period, who selects and implements the policy alternative $a_t$. Unlike voters, politicians observe the state $s$. An officeholder's informational advantage may reflect their greater expertise, better access to resources, or their greater incentive to acquire information[6] 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. Politicians derive the same utility from the policy $a$ as their constituents, regardless of type $i$. The issue is thus entirely *non-partisan* and, a priori, a politician's type is completely unconnected to 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, politicians derive utility from a rent $\phi > 0$ that they receive while in office. The utility of an $i$-type incumbent in period $t$ is

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

Once defeated, a candidate never returns to holding office; instead, the current officeholder faces a new challenger in the subsequent election. When not in office, politicians receive a continuation utility of zero.[7]

The timing of the period $t$ game is as follows. First, elections are held in which the representative voter decides whether to re-elect the incumbent or to replace him with a challenger with a different party affiliation. Then, nature draws $s_t$, which is immediately revealed to the officeholder but not to ordinary citizens. Next, the officeholder 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.* We restrict attention to pure strategy, stationary, and symmetric Markov perfect equilibria; in particular, players use the same strategy for any history that results in the same payoff relevant information. For the electorate, the latter can be summarized by its belief $\mu_t$ about the state at time $t$. A strategy for a representative citizen-voter specifies the probability $P(\mu_t) \in [0, 1]$ with which the incumbent in place at the end of $t-1$ is re-elected, based on beliefs $\mu_t$ at the beginning of the period (we will see below that we can

[5]The simplifying assumptions of our model are made for analytical convenience only. In particular, the results that follow do not hinge on the fact that the decision is binary or 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.

[6]See e.g. Kessler (2005).

[7]Thus a politician does not benefit from the legacy of policies that her successor enacts. 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.

define $\mu_t$ independent of $i$). When voters are indifferent between incumbent and challenger, we assume that either stands an equal chance of (re-)election. A strategy for a type-$i$ incumbent $a^i(\mu_t, s_t)$ maps the current state $s_t$ into a policy choice $a_t \in \{b, r\}$, given voter's beliefs $\mu_t$ which together with the success of failure of the policy determine $\mu_{t+1}$. In equilibrium, strategies must be mutual best responses and beliefs along the equilibrium path evolve according to Bayes rule. Off equilibrium, we will use well-known refinements to pin down beliefs.

## 2.2 Equilibrium Analysis

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

### 2.2.1 The Non-Partisan Equilibrium

As a benchmark, we first construct an equilibrium in which officeholders choose efficient policies in every period $t$. Suppose an incumbent's strategy is $a_t^i = s_t$ in every period $t$, , irrespective of $i$ or electoral beliefs $\mu_t$. Since both types of politicians always 'do the right thing' and implement identical policies, voters hold no preference for the incumbent or the challenger. Hence, they (re-)elect either with probability 1/2. The officeholder now strictly prefers the non-partisan (efficient) alternative because

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

for any $\pi < 1$, irrespective of type $i$.

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 a Markov perfect equilibrium. This equilibrium has the highest payoff to the electorate (though not necessarily to the incumbent, see below). 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 $s_t$.

**Proposition 1.** *[Non-Partisan Equilibrium] There always exists an equilibrium in which elected officeholders 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, it is not the only possible outcome. As the following section demonstrate, partisan politics can be supported in equilibrium and may yield a higher payoff for incumbents.

### 2.2.2 The Partisan Equilibrium

To simplify notation, let $M_t$ denote a situation where the period $t$ state 'matches' the observed identity of the incumbent standing for re-election in $t$.Similarly, let $m_t$ denote a 'mismatch' between state and incumbent identity in $t$.Now define the electorates' beliefs $\mu_t$ as the probability it attaches to the $M_t$ in $t$. Under partisan politics, voters will observe sure success in state $M_t$ and success with probability $\pi$ in state $m_t$. Along the equilibrium path, beliefs $\mu_t$ are therefore

$$\begin{aligned}\mu_{t+1} = & \quad \gamma Pr(M_t|v_t) + (1-\gamma)Pr(m_t|v_t) \\ = & \begin{cases} \gamma\frac{\mu_t}{\mu_t+(1-\mu_t)\pi} + (1-\gamma)\frac{(1-\mu_t)\pi}{\mu_t+(1-\mu_t)\pi} & \text{if } v_t = v, \\ 1-\gamma & \text{if } v_t = 0, \end{cases}\end{aligned} \tag{2}$$

noting that $Pr(m_t|v_t = 0) = 1$ and that the state switches (persists) with probability $1-\gamma$ $(\gamma)$. Since partisan policies reveal no new information about the current state, the electorate only learns by observing whether the policy has been successful or not. A success reinforces beliefs in $M$, $\mu_{t+1} > \mu_t$ due to $\gamma > 1/2$ and $\pi < 1$. Failure, in contrast, indicates that $m$ has occurred with certainty and beliefs start at their lowest possible value $1-\gamma < 1/2$. Off the equilibrium path, i.e., when the electorate observes a non-partisan policy $a_t^B = r$ or $a_t^R = b$, we impose the natural assumption that a mismatch occurred so $\mu_{t+1} = 1-\gamma$ (unless $v = 0$ in which case the state must be $M$). 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 to ensure that office holders implement both alternatives with positive probability in equilibrium and show that the conclusions are unaltered.

The optimal citizen-voter strategy is now straightforward. Since candidate's identities match their policy choices given the incumbent's assumed equilibrium strategy of partisan politics, it is strictly optimal to re-elect the incumbent if the odds of a matching state are greater than even ($\mu_t > 1/2$), and elect the challenger otherwise ($\mu_t < 1/2$).[8] The only other possibility would be that the voter deliberately disposes of an incumbent for beliefs $\mu_t > 1/2$ in order to 'experiment', i.e., get more precise information about the state through a likely mismatch. This can never be optimal, however, since the most precise information voters can hope for happens if the policy fails, in which case they would even be more convinced that the current incumbent was the best choice (and not re-electing her was a mistake).[9]

Turning to officeholders, note first that it is strictly optimal for an incumbent to implement the efficient policy $a_t = s_t$ in state $M_t$, i.e. if $s_t = b$ and $i = B$ or $s_t = r$ and $i = R$: the period-$t$ payoff is higher and re-election is guaranteed. If identity does not match the state ($m_t$) there is a trade-off: implementing the inefficient partisan policy results in certain re-election if successful (due to $\mu_{t+1} > \mu_t > 1/2$) while deviating to the efficient non-partisan policy reveals the non-match and forgoes electoral success as voters have a strict preference for the challenger in this case. 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) update their beliefs and anticipate subsequent partisan play. Further, since voters are perfectly informed about politicians' types (in particular, there are no 'partisan' or 'non-partisan' types), incumbents cannot 'signal' an increased likelihood that they would play non-partisan in the future by deviating.

Off-equilibrium, it is possible that an incumbent finds herself in office even though $\mu_t < 1/2$. For $\mu_t$ is low enough, even a lucky draw with the partisan policy would then not convince the electorate that next periods

[8]We can ignore $\mu_t = 1/2$ as it is only possible in $t = 0$ from (2).

[9]Put differently, the electorate's beliefs $\mu_{>}1/2$ would simply be reinforced if the partisan policy of the challenger 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. A formal proof of this argument is available from the authors upon request.

state is a likely match, i.e., $v_t = v$ does not imply $\mu_{t+1} > 1/2$. In such a situation, a mismatched incumbent would strictly prefer the non-partisan policy, since defeat at the polls is always certain, and the voter is indifferent between candidates. To keep our analysis focused on partisan behavior, we assume

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

to ensure that by acting partisan, incumbents are able to improve their re-election chances by a sufficient margin even in the worst case where they were erroneously elected for $\mu_t = 1-\gamma$. Under (A1), a success guarantees re-election irrespective of beliefs $\mu_t$. Incumbents' value functions then only depend on $m_t$ or $M_t$. Given the optimal electorate's strategy, and exploiting stationarity using (1) gives

$$\begin{aligned} V(M) =& v + \phi + \beta\left[\gamma V(M) + (1-\gamma)V(m)\right] \\ \text{and} \qquad V(m) =& \pi v + \phi + \pi\beta\left[\gamma V(m) + (1-\gamma)V(M)\right], \end{aligned} \tag{3}$$

The payoff from the efficient non-partisan choice in the unmatched state is $v+\phi$. Comparing this payoff to $V(m)$, the partisan choice dominates if and only if the expected future value from remaining in office exceeds the short-term gains from the efficient policy,

$$\beta\pi\left[\gamma V(m) + (1-\gamma)V(M)\right] \geq (1-\pi)v. \tag{4}$$

Using (3) we can explicitly compute $V(M)$ and $V(m)$ to read

$$\begin{aligned} V(M) &= \left\{\left[1-\beta\pi(2\gamma-1)\right]v + \phi\left[1-\beta\gamma+\beta(1-\gamma\pi)\right]\right\}/D \\ V(m) &= \left\{\left[1-\beta(2\gamma-1)\right]\pi v + \phi\left[1-\beta\gamma+\beta\pi(1-\gamma)\right]\right\}/D \end{aligned}$$

where $D = 1-\beta\gamma(1+\pi)+\beta^2\pi(2\gamma-1) > 0\ \forall\gamma > 1/2$Note that $V(M)-V(m) = (1-\pi)(v+\beta\phi)/D > 0$ with $V(m) \to V(M)$ as $\pi \to 1$ as expected. Substituting for $V(M)$ and $V(m)$ in (4) yields

$$\frac{v+\beta\phi}{v} \geq \frac{1-\beta\gamma(1+\pi)+\beta^2\pi(2\gamma-1)}{\pi\left(1-\beta(2\gamma-1)\right)}. \tag{A2}$$

Thus, and not surprisingly, partisan behavior is more likely to arise whenever the future value of holding office ($\phi$) is high or the opportunity cost of partisanship ($v$ or $1-\pi$) is low. Assuming candidates are very patient ($\beta \to 1$) for example, (A2) holds for every value of $\phi, v$, and $\gamma$ if $\pi \geq \frac{1}{3}$. Figure 1 in the appendix illustrates the $(\pi,\gamma)$ parameter space for which conditions (A1) and (A2) hold.Is shows in particular that that a more persistent state is not conducive to partisan behavior. 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.

We summarize the discussion in

**Proposition 2.** *[Partisan Equilibrium] For parameter values that satisfy conditions* (A1) *and* (A2)*, a Markov Perfect Equilibrium exists in which elected officeholders act partisan. 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 ($\phi$ high), the environment is volatile ($\gamma$ low) and whenever inappropriate policies are unlikely to fail ($\pi$ 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, as in e.g. Maskin and Tirole (2004), Callander (2008), and Morelli and Weelden (2013). In our framework, incumbents have no incentive to follow the most popular course of action. Instead, they insist on maintaining a partisan policy so as not to reveal that 'times have changed'.

Specifically, the blame for the policy bias can be squarely laid on partisanship as a (social) perception of likely behavior. Voters *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 they affiliate with the policy that is most likely to succeed, induce incumbents to act partisan in the first place with the objective to jam the electorate's inference problem. In equilibrium, the electorate's expectations are *self-fulfilling*. Voters and representatives are caught in a partisan trap: because voters expect the ideology of officeholders 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, officials act partisan to demonstrate confidence in circumstances that warrant a policy that - in the eyes of voters - is associated with their identity.

*Properties of the Partisan Equilibrium*

Since a partisan policy choice does not vary with the underlying state, the model implies inefficient policy persistence: along the equilibrium path, a given policy is maintained to conceal any change in circumstance, until failure prompts voters to oust a politician from office.[10] Moreover, incumbents who 'stick to their political colors' and enact inefficient policies do not experience lower election chances. To the contrary, an officeholders chances of winning another term in office are strictly higher than in the non-partisan equilibrium (see appendix for a formal proof).

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 Inertia], and*

*c) the long run probability that an incumbent wins another term in office is strictly greater than one half [Incumbency Advantage].*

The last part of Proposition 3 trivially implies that because of the continued accumulation of office rents, an incumbent's expected payoff in the partisan will exceed her payoff in the non-partisan equilibrium for sufficiently high value of $\pi$, $\phi$, or $\beta$. While voters thus always strictly prefer non-partisan play, *officeholders often prefer partisan politics.* The findings of policy persistence and incumbency advantage distinguish our framework from many other models of polarization and are consistent with what we observe in democratic

[10] The policy inertia in the model is thus different from Coate and Morris (1999) where it stems from a struggle between interest groups and the public at large. Instead, inefficiency through inertia is similar to the inefficiency through inaction that has been identified in a signal-jamming framework by Aghion and Jackson (2016). In general, of course, policies can also be inefficient if they are too volatile (see e.g. Battaglini, 2014).

two-party systems. A plethora of studies of evidence on the US Congress in particular show positioning and polarization along party lines (McCarty, Poole and Rosenthal, 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.

A notable feature of the partisan equilibrium is that each candidate (despite identical preferences) pursues a distinct policy, which is the socially optimal policy only for a specific distinct state. The one-to-one correspondence that emerges between the identity of the politician, her policy choice, and the state in which the policy is optimal justifies our notation choices: without loss of generality, an $J$-type candidate sticks to policy $j$ in equilibrium. At the same time, how voters tie the identity of politicians to expectations about their future policies is based on observable characteristics that are payoff irrelevant, similar to 'sunspots' in macroeconomic models of business cycles. Looking at our economy from the outside, one would conclude that candidates either have different ideologies (their policy choices are fixed and inefficiently unresponsive to changes in the economy) or separate world views (e.g. they believe that the state is always blue and choose their policies accordingly). Neither conclusion is correct because incumbents are *not* true believers. Instead, their partisan stance is purely rooted in expectations based on identity: if voters expect a female 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' expectations tie candidates' policy intentions to their observed characteristics (their position on other salient issues).

# 3 Robustness

One possible misgiving against the result in Proposition 2 is that nonpartisan politics is never observed on the equilibrium path.[11] If voters unexpectedly see an officeholder act non-partisan, we have assumed that they infer that the state must be unfavorable to her ideological position. As a result, it is strictly optimal to elect the opponent in the next election given they expect continued partisan behavior from the officeholder. These expectations are pinned down by the definition of a Markov Perfect Equilibrium, but because they are critical to sustaining partisan play and occur off the equilibrium path, one may wonder whether the above reasoning continues to hold if non-partisan policies are sometimes observed in equilibrium.

While purposefully kept our model very simple to isolate the role of self-fulfilling expectations, it is straightforward to extend the model to allow for non-partisan policy choices on the equilibrium path. Suppose in particular that the likely consequences of the wrong policy choice $\pi$ are stochastic, i.e., sometimes imple-

[11] Another objection would be that voters benefit from non-partisan policies. This argument ignores that politicians generally do not. Moreover, non-partisan politics are fragile in the sense that they cannot survive if citizens' expectations about officeholders' behavior are subject to (small) uncertainty: it is easy to see that if the electorate puts some positive probability $\epsilon$ on partisan choices, it always has a strict incentive to match the identity of the officeholder to the likely state, breaking the indifference and generating the same incentive to play partisan, even if $\epsilon$ is very small.

menting the wrong policy is more damaging than other times. Clearly, if the prospects of the partisan policy are sufficiently dim ($\pi_t$ sufficiently low), officeholders will implement the efficient (non-partisan) policy in $t$, regardless of its effect on re-election. Voters observing non-partisan play on the equilibrium path eliminates out-of equilibrium beliefs and the need to assume a specific refinement to pin down those beliefs.

For simplicity, assume $\pi_t$ is drawn i.i.d as follows: 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 because inefficient policy choices never succeed. To fix ideas, we will refer to such a period as a 'crisis'. Let $\epsilon > 0$ be the probability of a crisis in $t$, independent of $s_t$. A normal period with $\pi$ thus occurs with probability $1 - \epsilon$. Assume officeholders learn $s_t$ and $\pi_t$ once elected, before choosing $a_t$ and that voters do not observe $\pi_t$.[12] Since a crisis does not persist, voters' beliefs over $\pi_t$ are the same each period, and we can w.l.o.g. condition the election probabilities exclusively on the belief over the state $\mu_t$, 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 non-partisan in a crisis is also supported. To see this, suppose consider a blue officeholder in a crisis period. A partisan policy $a_t = b$ will surely fail while the non-partisan choice $a_t = r$ is successful for sure. For any$\mu_{t+1}$ therefore, non-partisan politics are strictly optimal in a crisis. Observing $a_t = r$, the electorate rationally infers that $s_t = r$ and that a crisis occurred. Given incumbents play partisan in normal times, the optimal response to either policy failure or non-partisan policy choice is still to *not* re-elect the incumbent, however. In normal times, on the other hand, office holders whose policy was successful are optimally be re-elected for the same reason as above: voters correctly infer that either the officeholder's identity matched the state or that a mismatch happened but there was no crisis, but beliefs remain favorable.

As before, we avoid possible multiplicity of equilibria by imposing parameter restrictions ensuring that elected candidates implement the partisan policy, even off equilibrium path when they erroneously find themselves in office and success would guarantee re-election, similar to (A1). Since the partisan policy is less often implemented than in the base model, however, observing a successful partisan policy now contains more information and therefore has a larger effect on the posterior belief, increasing the respective parameter space. Specifically, (A1) becomes

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

Next, let $\hat{V}(M)$ be the value function of an officeholder whose identity matches the state and $\hat{V}(m)$ be the value function of an officeholder whose identity does not match the state in a normal period ($\pi_t = \pi$), assuming players follow their equilibrium strategy. Because the non-partisan policy is chosen in a crisis, the payoff of the incumbent who finds herself in a crisis period is simply $v + \phi$. We have

$$\begin{aligned} \hat{V}(M) =& v + \phi + \beta \left[ \gamma \hat{V}(M) + (1-\gamma) \left( (1-\epsilon)\hat{V}(m) + \epsilon(v+\phi) \right) \right] \qquad (5) \\ \text{and} \qquad \hat{V}(m) =& \pi v + \phi + \pi\beta \left[ \gamma \left( (1-\epsilon)\hat{V}(m) + \epsilon(v+\phi) \right) + (1-\gamma)\hat{V}(M) \right], \end{aligned}$$

where the value functions are now slightly modified to account for the additional uncertainty induced by $\pi_t$.

[12]Our qualitative argument remains valid as long as there is some residual uncertainty with regard to $\pi_t$.

The partisan choice in a normal period is thus optimal if and only if $\hat{V}(m) \geq v + \phi$ or

$$\beta \left[ \gamma \left( (1-\epsilon)\hat{V}(m) + \epsilon(v+\phi) \right) + (1-\gamma)\hat{V}(M) \right] \geq (1-\pi)v. \tag{6}$$

In the appendix, we show that (6) is satisfied under the same condition as before, i.e., (A2) is unchanged. This result is very intuitive: assume for the moment that (A2) is satisfied with equality. Then, the incumbent is indifferent between implementing his inefficient partisan policy and the efficient one in a normal period. Her payoff in this case is also identical to the payoff she received in a crisis period, where the efficient policy is implemented and electoral defeat is certain, namely, $v + \phi$. Since the value in a state $m$ equals that of the base model, and the strategy in a state $M$ is unchanged, $V(M)$ remains as in section 2.2.2 as well. Hence, the same parameter values satisfying (A2) and hence (4) with equality will satisfy (6) with equality. Now suppose (A2) holds with strict inequality, which means playing partisan becomes more attractive in the base model. By the same argument, this implies that partisan behavior in a normal period becomes more attractive, $\hat{V}(m) > v + \phi$, implying an incumbent would prefer to implement the partisan policy whenever $\pi_t = \pi$.

**Proposition 4.** *Under* (A1') *and* (A2)*, there exists a Markov Perfect Equilibrium in which elected office-holders act partisan in normal times and non-partisan 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 if they implemented the non-partisan policy.*

In summary, we find that allowing for uncertainty over $\pi_t$ generates non-partisan politics along the equilibrium path but does not destroy the self-fulfilling nature of partisan politics. Intuitively, if the electorate is uncertain about the prospects of inefficient policies, it expects elected officials to sometimes to be non-partisan. Upon observing a politician who abandoned his partisan identity, the electorate assumes that the incumbent has been forced to abandon his partisan stance because the conflicting evidence was too strong; 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 $\epsilon \to 0$.

# 4 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 self-fulfilling and result in inefficient policies even in areas where everyone should agree on optimal courses of action. A 'partisanship trap' emerges because voters associate alternative policy measures with candidates' characteristics (party affiliation, gender, voting record), and expect incumbents to remain true to their partisan identity.

The result is political polarization, policy failure, and policy inertia. Importantly, the latter 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 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, it is easy to see how simple extensions of the model could add robustness to the partisan equilibrium by making it strictly optimal for voters to keep (oust) a 'matching' ('noon-matching) incumbent even in situations where they expect future non-partisan play, e.g., if candidates have some expertise in their partisan policy or there is some exogenous probability of partisan play due to lobbying. Besides, even though equilibria that are self-fulfilling and expectation-driven may be temporary in nature, they can be relevant in practice, 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, perceived partisan identity based on observable attributes can matter even if it is payoff irrelevant. Notably, a candidate's position on policy questions other than what is currently on the table is one of those attributes. The model can thus 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, our insight could be fruitfully 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, 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

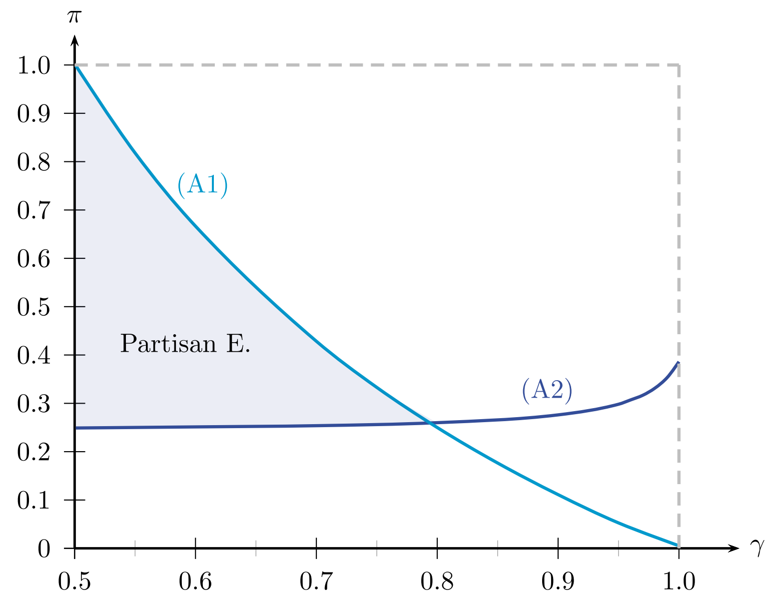


Figure 1: Existence of the Partisan Equilibrium in the $(\gamma, \pi)$ parameter space. The light and dark blue curves represent (A1) and (A2), respectively, where the latter assumes parameters $v = 3$, $\phi = 2$, and $\beta = 0.95$. Increasing $\beta$ and $\phi$ shifts (A2) down and enlarges the shaded region for which the partisan equilibrium exists, while increasing $v$ does the opposite.

**Proof of Proposition 3**

Part a) is trivial. To show part b), define $\tilde{s}_t \in \{M, m\}$ with $\tilde{s}_t = M$ when $a_t = s_t$ and $\tilde{s}_t = m$ when $a_t \neq s_t$.

$$T = \begin{pmatrix} t_{MM} & t_{Mm} \\ t_{mM} & t_{mm} \end{pmatrix} = \begin{pmatrix} \gamma & 1-\gamma \\ (1-\pi)\gamma + \pi(1-\gamma) & (1-\pi)(1-\gamma) + \pi\gamma \end{pmatrix}$$

where $t_{ij}$ denotes the transition probability from state $i$ to state $j$. Note that $a_{t+1} \neq a_t$ only if $\tilde{s}_t = m$ and $v_t = 0$. Hence $Pr(a_t \neq a_{t+1}) = Pr(\tilde{s}_t = m)(1-\pi)$. In the non-partisan equilibrium, a policy change occurs with probability $1-\gamma$. Thus, the condition for policy inertia is $Pr(\tilde{s}_t = m)\,(1-\pi) \leq 1-\gamma$.

We proceed to show that the long run probability of $\tilde{s}_t = m$ satisfies this condition. The (generically unique) stationary distribution corresponds to the eigenvector which is associated with 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 denote the probability of $\tilde{s}_t = M$ ($\tilde{s}_t = m$). The long run probability that $m$ occurs is thus $\lim_{t\to\infty} Pr(\tilde{s}_t = m) = \frac{1-\gamma}{1-2\gamma\pi+\pi}$. Due to $\gamma < 1$, we have

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

Part c) now follows because an incumbent is not reelected only in the event of a political failure, which from part b) we know is less that $1-\gamma$, which in turn is less than $1/2$ due to $\gamma > \frac{1}{2}$.

**Proof of Proposition 4**

Consider parameters for which $\hat{V}(m) \geq v + \phi$ (e.g., sufficiently high $\pi$). Using (5), this implies

$$\hat{V}(M) \geq v + \phi + \beta\left[\gamma \hat{V}(M) + (1-\gamma)(v+\phi)\right] \quad \Leftrightarrow \quad \hat{V}(M) \geq \frac{1+\beta(1-\gamma)}{1-\beta\gamma}(v+\phi)$$

Substituting the right-hand side into $\hat{V}(m)$ from(5) and again using $\hat{V}(m) \geq v + \phi$, we obtain

$$\hat{V}(m) \geq \pi v + \phi + \beta\left[(1-\gamma)\frac{1+\beta(1-\gamma)}{1-\beta\gamma}(v+\phi) + \gamma(v+\phi)\right]$$

Condition (6) is thus implied if the right-hand side exceeds $v + \phi$ or

$$\beta\pi\left[(1-\gamma)\frac{1+\beta(1-\gamma)}{1-\beta\gamma}(v+\phi) + \gamma(v+\phi)\right] \geq (1-\pi)v \tag{7}$$

Simple algebra shows that (7) is equivalent to (A2). $\square$

## References


**Aghion, Philippe, and Matthew O. Jackson.** 2016. "Inducing Leaders to Take Risky Decisions: Dismissal, Tenure, and Term Limits." *American Economic Journal: Microeconomics*, 8(3): 1–38.

**Ansolabehere, Stephen, and James M Jr Snyder.** 2000. "Valence Politics and Equilibrium in Spatial Election Models." *Public Choice*, 103(3-4): 327–36.

**Aragones, Enriqueta, and Thomas R Palfrey.** 2002. "Mixed Equilibrium in a Downsian Model with a Favored Candidate." *Journal of Economic Theory*, 103(1): 131–161.

**Ashworth, Scott, and Ethan Bueno de Mesquita.** 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.

**Ashworth, Scott, Ethan Bueno de Mesquita, and Amanda Friedenberg.** 2017. "Accountability and Information in Elections." *American Economic Journal: Microeconomics*, 9(2): 95–138.

**Azariadis, Costas.** 1981. "Self-fulfilling prophecies." *Journal of Economic Theory*, 25(3): 380–396.

**Azariadis, Costas, and Roger Guesnerie.** 1986. "Sunspots and Cycles." *The Review of Economic Studies*, 53(5): 725–737.

**Battaglini, Marco.** 2014. "A dynamic theory of electoral competition." *Theoretical Economics*, 9(2): 515–554.

**Besley, Timothy J., and Stephen Coate.** 1997. "An Economic Model of Representative Democracy." *The Quarterly Journal of Economics*, 112(1): 85–114.

**Buisseret, Peter, and Richard Van Weelden.** 2022. "Polarization, Valence, and Policy Competition." *American Economic Review: Insights*, 4(3): 341–52.

**Callander, Steven.** 2008. "Political Motivations." *Review of Economic Studies*, 75(3): 671–697.

**Callander, Steven, and Juan Carlos Carbajal.** 2022. "Cause and Effect in Political Polarization: A Dynamic Analysis." *Journal of Political Economy*, 130(4): 825–880.

**Canen, Nathan, Chad Kendall, and Francesco Trebbi.** 2020. "Unbundling Polarization." *Econometrica*, 88(3): 1197–1233.

**Carrillo, Juan D., and Micael Castanheira.** 2008. "Information and Strategic Political Polarisation." *The Economic Journal*, 118(530): 845–874.

**Cass, David, and Karl Shell.** 1983. "Do sunspots matter?" *Journal of political economy*, 91(2): 193–227.

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

**Coate, Stephen, and Stephen Morris.** 1999. "Policy Persistence." *American Economic Review*, 89(5): 1327–1336.

**Dziuda, Wioletta, and Antoine Loeper.** 2016. "Dynamic Collective Choice with Endogenous Status Quo." *Journal of Political Economy*, 124(4): 1148–1186.

**Goldberg, Andreas C., Erika J. van Elsas, and Claes H. de Vreese.** 2020. "Mismatch? Comparing elite and citizen polarisation on EU issues across four countries." *Journal of European Public Policy*, 27(2): 310–328.

**Heidhues, Paul, and Johan Lagerlöf.** 2003. "Hiding information in electoral competition." *Games and Economic Behavior*, 42(1): 48–74.

**Howell, William, Stefan Krasa, and Mattias Polborn.** 2020. "Political Conflict over Time." *American Journal of Political Science*, 64(3): 554–568.

**Izzo, Federica, Greg Martin, and Steven Callander.** 2023. "Ideological Competition." *American Journal of Political Science*, 67(3): 687–700.

**Kartik, Navin, and R. Preston McAfee.** 2007. "Signaling Character in Electoral Competition." *American Economic Review*, 97(3): 852–870.

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

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

**McCarty, Nolan, Keith T. Poole, and Howard Rosenthal.** 2006. *Polarized America: The Dance of Ideology and Unequal Riches.* Cambridge, Mass.:The MIT Press.

**McMurray, Joseph.** 2022. "Polarization and pandering in common-interest elections." *Games and Economic Behavior*, 133: 150–161.

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

**Morelli, Massimo, and Richard Van Weelden.** 2013. "Ideology and information in policymaking." *Journal of Theoretical Politics*, 25(3): 412–439.

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

**Poole, Keith.** 2007. "Changing Minds? Not in Congress!" *Public Choice*, 131(3): 435–451.

**Poole, Keith T., and Howard Rosenthal.** 2007. *Ideology & Congress.* Transaction Pub.

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

**Tirole, Jean.** 1985. "Asset Bubbles and Overlapping Generations." *Econometrica*, 53: 1499–1528.

**Van Weelden, Richard.** 2013. "Candidates, Credibility, and Re-election Incentives." *The Review of Economic Studies*, 80(4): 1622–1651.