WORKING PAPER · NO. 2023-145

Partisan Traps

Ethan Bueno de Mesquita and Wioletta Dziuda

NOVEMBER 2023

Becker Friedman Institute
FOR ECONOMICS AT UCHICAGO

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

PARTISAN TRAPS*

Ethan Bueno de Mesquita[†] Wioletta Dziuda[‡]

November 1, 2023

Abstract

The desire to stifle political competition may lead elected officials to eschew common-interest reforms and focus instead on zero-sum partisan conflict. By forgoing opportunities for common-interest reforms, incumbents may convince their constituents that such reforms are rarely feasible, so that policymaking is primarily about choosing partisan sides. Voters with such beliefs vote based on ideological alignment, rather than factors such as competence or honesty. This is electorally beneficial for incumbents, who are typically ideologically aligned with their constituents. We capture this logic in an infinite horizon model and characterize the resulting dynamics of politics and policymaking. Equilibrium exhibits partisan traps–voters are pessimistic about common-interest opportunities, and hence elect ideologically aligned incumbents, and incumbents respond by behaving in a purely partisan manner that shuts down voter learning. Partisan traps often occur in equilibrium even when common-interest reforms are in fact frequently feasible. The model shows how elite and mass polarization are intertwined, with politicians engaging in strategically polarized and polarizing behavior which leads to pessimistic beliefs among voters, who then vote in partisan fashion. *JEL codes: D00, D7, P*

*We've received helpful feedback from the seminar audiences at Caltech, the University of Chicago, Harvard University, Princeton University, NYU, UCSD, the University of Michigan and the conference audience of Benelux PECO. We are grateful to Nageeb Ali, Renee Bowen, Matt Gentzkow, Pietro Ortoleva, Galina Zudenkova, and Jesse Shapiro for helpful suggestions.

[†]Harris School of Public Policy, University of Chicago, 1307 E 60th St, Chicago, 60637, bdm@uchicago.edu

[‡]Harris School of Public Policy, University of Chicago, 1307 E 60th St, Chicago, 60637, and NBER, wdziuda@uchicago.edu (corresponding author)

Elected officials often face a choice between pursuing partisan policies that benefit one group at the expense of another or pursuing broadly beneficial, common-interest policies. Frequently, it seems, politicians choose partisanship. As Fiorina, Abrams and Pope (2011) write:

> Most citizens want a secure country, a healthy economy, safe neighborhoods, good schools, affordable health care, and good roads, parks, and other infrastructure. These issues do get discussed, of course, but a disproportionate amount of attention goes to issues like abortion, gun control, the Pledge of Allegiance, medical marijuana, and other narrow issues that simply do not motivate the great majority of Americans.

What are the incentives that drive this behavior, and what do they imply for the dynamics of politics and policymaking over the long run?

We argue that politicians pursue partisanship because doing so helps them convince their constituents that opportunities for implementing common-interest policies are rare, and instead that most policymaking is about choosing partisan sides. Most incumbent politicians are ideologically aligned with their constituents—in a choice among zero-sum, partisan policies they share the same preferences as their voters (Fowler, 2016). It follows that if voters believe that most policymaking is about choosing between partisan alternatives, they will be reluctant to replace their ideologically-aligned incumbents with ideologically-misaligned challengers even if the latter dominate on some other dimension (e.g., honesty, quality, or valence). By contrast, if voters believe that policymaking presents many opportunities to implement common-interest reforms, they will be less concerned about whether they are ideologically aligned with their representatives. Thus, if voters believe that policymaking is primarily about zero-sum, partisan choices, incumbents that are ideologically aligned with their constitutents are electorally insulated.

We capture this idea in an infinite-horizon model of elections and policymaking. In each period, voters choose between candidates from two parties: one that is ideologically aligned with the voters and one that is ideologically misaligned with them.[1] In each period, either candidate may have a valence (e.g., quality, honesty) advantage that is valued by the voters independent of policy choice. The elected candidate becomes the incumbent and implements a policy, which can be either partisan or common interest. In terms of pure policy preferences (ignoring electoral consequences), partisan policy benefits the party that implements it (i.e., the party in power) and harms the other party. Partisan policy implemented by the aligned party benefits the voters while partisan policy implemented by the misaligned party harms them. Common-interest policy benefits everyone and, indeed, is preferred (again, putting aside strategic considerations) by voters and politicians over any other policy.

The crucial assumption in our model involves asymmetric information between politicians and voters. For political or technological reasons, common-interest policies are not available every period. Voters are uncertain about whether common-interest policy opportunities arise rarely or frequently—that is, whether the *policymaking environment* is *unfavorable* or *favorable* for common-interest reforms. Voters also do not observe whether a common-interest policy is available in any given period unless a common-interest policy is in fact implemented in that period. By contrast, politicians know the policymaking environment, and in each period they observe whether a common-interest policy is available.[2]

After the incumbent implements a policy, voters observe the policy type, update their beliefs about the policymaking environment, observe the valence of the candidates in the next election, and choose whom to elect for the next period.

The model yields several key insights.

First, equilibrium exhibits *partisan traps.* When voters are moderately pessimistic about

[1] Voters in our model stand for a series of representative voters.

[2] The assumption that politicians know the policymaking environment is made mainly for simplicity, but since policymakers observe more than the voters, they learn about the policymaking environment faster, so at some point in the game such a strategic environment is well approximated by our model.

the policymaking environment (i.e., their beliefs lean somewhat toward the policymaking environment being unfavorable), they elect ideologically-aligned incumbents. But since they are only moderately pessimistic, seeing common-interest policy implemented makes them sufficiently optimistic about the policymaking environment that they become willing to consider replacing the incumbent with a high-valence misaligned opposition candidate. In such circumstances, we show, aligned incumbents forgo common-interest opportunities, implementing only partisan policies. When this happens, voters can no longer learn about the policymaking environment. As a result they are trapped—permanently moderately pessimistic about the policymaking environment, never experiencing common-interest policies, and never replacing the ideologically-aligned incumbent.

Second, the political system always operates under the risk of a partisan trap. No matter how optimistic the voters are about the policymaking environment, as long as their beliefs are not degenerate, there exists a finite sequence of unlucky periods in which common-interest opportunities are not available after which the voters become pessimistic enough to fall into a partisan trap. Hence, partisan traps occur with positive probability even if the voters start with optimistic beliefs about the policymaking environment.

Third, politicians behave differently in office depending on whether or not they are ideologically aligned with their voters. As we've discussed, politicians who are ideologically aligned with their voters face a trade-off—for them, pursuing common-interest ideas is good policy but bad politics. Politicians who are ideologically misaligned with the voters do not face this same trade-off—for them, common-interest policies are both good policy and good politics. Therefore, we expect more common-interest policies when voters elect misaligned politicians to represent them. Of course, for the voters, this comes at the cost of having to accept partisan policies they disagree with much of the time.

Finally, the model shows one way in which elite and mass polarization are intertwined. In the model, elites (i.e., politicians) frequently engage in strategically polarized behavior, pursuing partisan policies that divide citizens and politicians, rather than common-interest reforms on which everyone agrees. They do so to insulate themselves electorally by causing voters to have pessimistic beliefs about the policymaking environment—voters conclude that policymaking is primarily partisan. In this sense, elite behavior is not just polarized, it is polarizing. By making voters pessimistic, elite behavior drives voters to become loyal supporters of the party that they are ideologically aligned with, even if that party is inferior on other dimensions (e.g., it is corrupt or has low valence). This finding resonates with empirical studies of polarization, which conclude that polarization appears to be an elite-driven phenomenon. Elites engage in partisan, polarized policymaking, but voters hold moderate views (Fiorina, Abrams and Pope, 2011). Nevertheless, voters show an increasing allegiance to their parties and are unwilling to vote across party lines (Iyengar et al., 2019; Abramowitz and Webster, 2016).

In addition to these specific implications, our model offers a novel framework for thinking about some salient empirical patterns of political behavior and policymaking.

For instance, the existence of partisan traps offers a new perspective on the relationship between accountability and the politics of polarization. When caught in a partisan trap, voters behave as if they were rewarding the party that they are ideologically aligned with for behaving in a purely partisan manner—always reelecting an incumbent who always pursues partisan policy. This might lead observers to conclude that the voters are themselves hyperpartisans. But the voters in the model are not in fact intransigent partisans: they prefer moderate, common-interest policies to partisan policies. Voters fail to punish elite partisanship electorally not because they are themselves extremists, but out of fear that the policymaking environment is unfavorable, so that policy under the misaligned party would be even worse. Aligned incumbents engage in partisan behavior precisely to stoke this rational fear on the part of voters because such fear is electorally insulating.

Our model thus suggests that districts where one party has a strong fundamental advantage should experience more common-interest policymaking. And, indeed, this is consistent with some empirical evidence showing that parties in the United States vote in a less polarized way in less electorally competitive states (Hinchliffe and Lee, 2016). Our model, however, reveals that the theoretical relationship between electoral competitiveness and policy is more nuanced, which may explain why there seems to be substantial heterogeneity in parties' behavior. In particular, we point to another reason (besides a large fundamental advantage for one party) that an electoral district might have uncompetitive elections, because it is caught in a partisan trap. When an electoral district is caught in a partisan trap, the dominant party is dominant precisely because the voters are pessimistic about the policymaking environment and the dominant party keeps the voters that way by engaging in partisan policymaking. On this account, if the dominant party were to pursue common-interest policy, it would cost them electorally. Thus, our model does not predict that we should always expect a high level of common interest policymaking in electorally uncompetitive polities. Some uncompetitive polities are uncompetitive for equilibrium reasons that forestall common-interest policy. This might help explain the observation that in the contemporary United States, state legislators are increasingly elected in uncompetitive elections, legislatures appear to be engaged in increasingly partisan policymaking, and voters increasingly support party-aligned candidates, even when those candidates' policy preferences are extreme relative to the voters (Handan-Nader, Myers and Hall, 2023).

Polarization is, of course, a complex political phenomenon with many causes. Nonetheless, we believe that our model sheds light on a number of real-world phenomena, at both the micro- and macro-scales.

At the micro-scale, recent debates over firearms safety following mass shootings illustrate the tension politicians face in our model. There may indeed be bipartisan compromise measures that the majority of voters of both parties would prefer to the status quo.[3] However,

[3]Most American voters of both parties favor limiting gun access for those with diagnosed mental illness and support requiring background checks for guns purchased in private sales or at gun shows (Pew Research

voters may also face genuine uncertainty about whether a common-interest compromise exists—for instance, conservative voters may not be able to confidently assess whether limited background checks are in their interest or whether allowing background checks would create a regulatory slippery slope leading to restrictions they oppose. Consistent with our model, politicians have typically not pursued bipartisan compromise, instead retreating their partisan corners, with Democrats calling for expansive gun regulations and Republicans resisting any federal regulation—signaling to voters that there is no politically feasible, common-interest policy available. This suits incumbents from both parties, most of whom represent districts with whose voters they are ideologically aligned. In particular, it leaves voters pessimistic about the possibility of common-interest reform and thus inclined to elect a politician who is on their partisan side. Notably though, if voters become sufficiently pessimistic (as, arguably, they became after years of inaction in the United States), politicians may become willing to pursue the occasional common-interest reform precisely because they know that voters' pessimism about the overall policymaking environment implies that there is little electoral risk associated with the occasional bipartisan vote.

At a larger historical scale, our dynamic account may help make sense of some broad historical patterns of political polarization. Consider two important periods from twentieth-century American political history: the New Deal and the Contract with America.

The period from the end of Reconstruction through the Great Depression saw significant polarization and conflict over racial and economic policy in the United States. Seen through the lens of our model, this was a time of broad pessimism about the possibility of implementing national common-interest policies—national politics appeared to be a zero-sum contest among voters' ideological interests. It is precisely in such circumstances, according to our model, that politicians should be willing to embrace common-interest reforms. And one can understand the wave of policies that came to be known as the first New Deal as politi-

Center, 2021).

cians seizing just such an opportunity. The major legislation of the early New Deal passed Congress with significant bipartisan support.[4] These early New Deal policies were broadly popular and led to increasing public optimism that beneficial reforms were feasible (Schickler and Caughey, 2011). However, after the first wave of reforms, politicians returned to partisan politics (Katznelson, Geiger and Kryder, 1993). The period running from the late 1930s through World War II saw "the frustration of liberal hopes for an expansive 'cradle-to-grave' welfare state" and increasing public skepticism that more beneficial reforms were coming (Schickler and Caughey, 2011, p. 163). This pattern of back-and-forth, whereby common-interest policymaking coexisted with increasing mass optimism but suddenly both come to a halt, is consistent with our model's predictions.

Our model also provides an account of increasing polarization in U.S. politics over the last 30 years. The 1994 elections swept Republicans into control of the House of Representatives for the first time in a generation. As Jacobson (1996) notes, over the course of the preceding thirty years, the Democrats who ultimately lost in 1994 had become increasingly ideologically misaligned with their voters (most represented districts that leaned Republican in presidential elections) but were perceived to be more qualified than their Republican opponents. Consistent with our model, faced with ideologically misaligned voters, Democratic congressional representatives focused on policies that, within a congressional district, were viewed as common interest—namely, pork barrel projects. They also engaged in legislative compromises, such as the 1990 deficit reduction effort. As Gentzkow, Shapiro and Taddy (2019) show, during this period of relatively common-interest policymaking, congressional representatives also employed unusually non-partisan language. However, by the

[4]For instance, the Agricultural Adjustment Act passed the House of Representatives with the support of 272/296 Democratic votes and 38/111 Republican votes (https://www.govtrack.us/congress/votes/73-1/h7); the National Industry Recovery Act passed with the support of 267/292 Democratic votes and 53/103 Republican votes (https://www.govtrack.us/congress/votes/73-1/h44); the Federal Emergency Relief Act of 1933 passed with the support of 252/325 Democratic votes and 73/103 Republican votes (https://www.govtrack.us/congress/votes/73-1/h18); and the Social Security Act of 1935 with 287/300 Democratic votes and 77/95 Republican votes (https://www.govtrack.us/congress/votes/74-1/h39).

early 1990s, as a result of the Clinton Administration's struggles advancing major policy reform even under unified government, voters began questioning the Democrats' competence. This, combined with Republicans fielding high-quality candidates, resulted in Republicans winning the House in 1994, which increased the overall ideological alignment between the districts and their representatives (Jacobson, 1996). And, consistent with our model, parties reacted to this realignment by retreating to their partisan corners (Gentzkow, Shapiro and Taddy, 2019).

The paper is organized as follows. Section 1 discusses related literature. Section 2 introduces the formal model. Section 3 describes a particular equilibrium in which voters use cutoff strategies. Section 4 studies the dynamics of politics, policymaking, and polarization within this equilibrium—introducing the idea of a partisan trap and explicating several substantive implications. Section 5 provides an existence result for the equilibrium described in Section 3 and derives comparative statics. Section 6 characterizes all cutoff equilibria and provides conditions under which a cutoff equilibrium exists and under which all equilibria are cutoff equilibria. Section 7 offers concluding remarks on future research.

1 Related Literature

There is an extensive empirical literature on political polarization and its relationship to governance (for reviews see Lee, 2015; McCarty, 2019). The empirical analysis most closely related to our theoretical model is Matsusaka and Kendall (2022). They structurally estimate a model in which policy has both a common-interest component and a zero-sum component. They assume that voters are imperfectly informed about the common-interest component and conclude that there would be less polarization if voters were perfectly informed. This is consistent with our model's predictions: in our model, the incentives for elite partisanship come precisely from politicians' desire to manipulate voters' beliefs about how often common-interest policy is available. More broadly, we share a conceptual perspective with Matsusaka and Kendall (2022). Some of the literature views mass polarization in terms of ideological polarization (Abramowitz, 2010; Abramowitz and Saunders, 2008), or as a social phenomenon in which voters increasingly refuse to vote across party lines in order to express their social identity (Iyengar et al., 2019; Abramowitz and Webster, 2016) or even bias their perceptions of reality in the direction of their ideology (Alesina, Miano and Stantcheva, 2020). However, we think of polarization differently. In our conceptualization, there are some issues on which voters broadly agree (common-interest) and others on which they broadly disagree (partisan). Mass polarization increases, in this conceptualization, as voters come to believe that a greater share of issues are partisan rather than common-interest and hence are unwilling to vote across ideological lines.

Our focus on voters' beliefs that policymaking is a zero-sum game is similar to Chinoy et al. (2023) (see also citations within). Chinoy et al. (2023) demonstrate that a large fraction of the U.S. population hold zero-sum beliefs, whereby they perceive that gains to some imply losses to others, and show how those beliefs are shaped by personal and ancestral history. Our model shows that such zero-sum beliefs may benefit incumbents, and hence that incumbents may attempt to shape such beliefs by foregoing common-interest policies. Andrews Fearon (2023) document that zero-sum thinking correlates with higher hostility to political outgroup members. Our model abstracts from issues such as hostility, but predicts a correlation between zero-sum beliefs and voters' willingness to elect members of a political outgroup to office.

Our paper relates to a literature studying the politics of divisive issues. Buisseret and Van Weelden (2022) show how parties may exploit divisive issues using referenda, but their mechanism relies on those issues cutting across parties, while we focus on divisive issues between parties. Ash, Morelli and Van Weelden (2017) study how incumbent politicians allocate effort between a common-interest issue and a divisive issue. Voters are unsure about incumbents' preferences, which creates incentives for the incumbent to signal alignment with the median voter by focusing excessively on the divisive issue. In our model, the candidates' preferences are known; politicians focus on divisive issues because of a desire to signal that common-interest reforms are rarely feasible.

In terms of substantive focus, the most closely related theoretical model is Callander and Carbajal (2022). Like us, they study a dynamic model of the interplay of elite and mass polarization. However, the mechanisms at work in the two models are quite different. In Callander and Carbajal (2022) polarization occurs because preferences change dynamically—when a voter votes for a candidate, their ideal point moves slightly in the direction of that candidate's party platform. Over time, this results in a paucity of voters with preferences in the center of the policy space and, as a consequence, leads strategic parties to polarize their policy positions. By contrast, in our model, preferences are fixed and strategic parties polarize their behavior in order to manipulate voters' beliefs by signaling that common-interest policy opportunities are rare.

In this way, our model shares some features of pandering models of electoral accountability (Canes-Wrone, Herron and Shotts, 2001; Maskin and Tirole, 2004; Acemoglu, Egorov and Sonin, 2013). In those models, like in ours, politicians implement policies they disagree with in order to manipulate voters into reelecting them. However, the nature of the manipulation is quite different in our model than in pandering models. In pandering models, incumbents seek to convince voters that they are high quality by choosing the policy that voters expected to be right ex ante. In our model, all of the politicians' characteristics are common knowledge—they are seeking to manipulate voters' beliefs about the underlying policymaking environment.

Like in pandering models, Levy and Razin (2022) study a dynamic model of polarization where voters seek to learn about politicians' competence. In that model, voters can only do so when politicians disagree with one another. Moreover, voters only have short-term memory. As a result, if there has been political consensus in recent periods, voters are unable to assess which party is more competent. Consequently, voters vote according to their pure partisan preferences. Hence, voters polarize but only insofar they were moderating their behavior relative to their preferences in earlier periods. By contrast, in our model, polarization is the result of strategic political manipulation of voters who are learning about the nature of the policymaking environment, not about the characteristics of politicians.

Our model also relates to Ali, Mihm and Siga's (2018) analysis of redistributive voting. In that model, uncertainty about the possibility of a negative correlation in voters' policy payoffs leads to adverse selection. Because voters worry that a policy that benefits others likely hurts them, they reject many policies that would in fact be to their benefit. Our model shows that politicians have endogenous incentives to keep voters uninformed in a way that gives rise to this adverse selection. In particular, in our model, politicians' desire to keep voters believing that there are relatively few common-interest opportunities is precisely a desire to make voters believe that there is a negative correlation in policy payoffs.

Finally, our paper relates to the strand of literature that studies the conditions under which widely beneficial reforms fail to be implemented (Fernandez and Rodrik, 1991; Strulovici, 2010; Dziuda and Loeper, 2016, 2018; Austen-Smith et al., 2019; Dziuda and Loeper, 2023; Bils and Izzo, 2023). These papers, however, do not rely on informational asymmetry between the voters and the parties, and have little to say about partisan policymaking.

2 The Model

The game is played over an infinite number of periods, $t = 0, 1, 2, \ldots$. There are two parties—the *aligned* party (a) and the *misaligned* party (m)—each with an infinite number of politicians, $\{a_k, m_k\}_{k=1}^{\infty}$. There are an infinite number of voters. Both the voters and each party's politicians are indexed by $k = 1, 2, \ldots$. There are three types of policies: two party-specific *partisan* policies, P_a and P_m, and a *common-interest* policy C. The meaning of the names of the parties and policies will become clear once we define the payoffs.

In $t = 0$, nature selects $q \in \{q_l, q_h\} \subset \mathbb{R}$ with $\Pr(q = q_h) = \theta_0$. This choice is fixed throughout and we refer to it as the *policymaking environment*, which can be *favorable* ($q = q_h$) or *unfavorable* ($q = q_l$). The parties observe q, but the voters do not.

Let $j_t \in \{a_t, m_t\}$ denote the incumbent politician in period t and $x_t \in \{C, P_a, P_m\}$ the policy implemented in t. Every $t > 0$ has the following stages:

1. Incumbent $j_t \in \{a_t, m_t\}$ is in power;
   - We set $j_1 = a_1$, the aligned party starts in power;
2. Nature makes C available with probability q. Let $c_t \in \{y, n\}$ denote whether C is available or not;
   - Nature's choice is observed by the politicians but not the voters;
3. If C is not available, the incumbent j_t from party j implements $x_t = P_j$. If C is available, the incumbent chooses between $x_t = P_j$ and $x_t = C$;
   - Voters observe x_t;
4. Nature determines m_{t+1}'s valence for the next period: $\nu_{t+1} \in \{-v, v\} \subset \mathbb{R}$ with $v > 0$ and $\nu_{t+1} = v$ with probability $p \in (0, 1]$;
   - All players observe ν_{t+1};
5. Voter t chooses between politicians a_{t+1} and m_{t+1}, determining who will be in power in $t + 1$: $j_{t+1} \in \{a_{t+1}, m_{t+1}\}$.

If in a given period the misaligned candidate's valence is v, we say that the *misaligned candidate has a valence advantage.* If instead the misaligned candidate's valence is $-v$, we say that the *aligned candidate has a valence advantage.*

Payoffs are as follows. All players care about policy, but the voters care also about valence. The voters agree with the aligned party on partisan policy. A player's payoff from a partisan policy they agree with is 1. A player's payoff from a partisan policy they disagree with is -1. A player's payoff from a common-interest policy is $b > 1$. So the common-interest policy delivers a higher payoff to a player than even the player's preferred partisan policy.

Player k gets payoffs only in periods $t = k, k+1$ and there is no discounting between periods. In all other periods, player k's payoff is 0. So voter k's payoffs in periods $t = k, k+1$ are

$$u_t^k = \begin{cases} 1 & \text{if } j_t = a_t \text{ and } x_t = P_a \\ b & \text{if } j_t = a_t \text{ and } x_t = C \\ -1 + \nu_t & \text{if } j_t = m_t \text{ and } x_t = P_m \\ b + \nu_t & \text{if } j_t = m_t \text{ and } x_t = C, \end{cases}$$

and politician $j_k \in \{a_k, m_k\}$'s payoffs in period $t = k, k+1$ are:

$$w_t^{j_k} = \begin{cases} 1 & \text{if } x_t = P_j \\ -1 & \text{if } x_t = P_{-j} \\ b & \text{if } x_t = C. \end{cases}$$

We make a technical assumption to focus on the interesting part of the parameter space:

Assumption 1

1. $v \in (2(1-q_h), 2 - q_l(b+1))$;

2. $b < 1 + \frac{2p(1-q_h)}{1+q_h}$.

Assumption 1.1 ensures that the voters' decisions are not trivial. First, the valence shock is sufficiently large so that if the voters knew that the policymaking environment was favorable ($q = q_h$) and expected the misaligned party to implement common-interest policy whenever available, they would elect a valence-advantaged misaligned party independent of the expected behavior of the aligned party. Second, the valence advantage is sufficiently small so that the if the voters knew that the policymaking environment was unfavorable ($q = q_l$) and expected the aligned party to never implement the common-interest policy, they would still elect the aligned party regardless of the valence and the expected behavior of the misaligned party. Assumption 1.2 says that electoral incentives matter to politicians—in particular, a politician prefers to choose partisan over common-interest policy if the former leads to her party retaining power and the latter leads to her party losing power to a valence-advantaged competitor, even if in the future, the aligned party is expected to choose only partisan policies and the competitor is expected to implement common-interest policies when available.

Our solution concept is an extension of the standard stationary Markov Perfect Equilibrium of Maskin and Tirole (2001) adapted to games with incomplete information: incumbents condition their policymaking only on the policymaking environment and the current voters' beliefs, and the voters condition their voting choices only on their current beliefs and valence of the misaligned candidate.[5] Let h_t denote the history of the game observed by voters up to the period t election: $h_t = (\{j_n\}_{n=1}^{t}, \{x_n\}_{n=1}^{t}, \{v_n\}_{n=1}^{t+1})$, and let (s^a, s^m) be any (not necessarily Markovian) strategy profile in the game. Let θ_t be the voters' belief that the policymaking environment is favorable ($q = q_h$) formulated using Bayes' rule whenever possible just before the period t election under the conjecture that parties use the strategy profile (s^a, s^m)—that is,

$$\theta_t = \Pr(q = q_h \mid h_t, (s^a, s^m)).$$

Given our equilibrium concept, we write $r(\theta, \nu)$ for the probability that any voter who

[5] Our approach follows Kreps and Wilson (1982) and Fudenberg, Levine and Tirole (1985). See Appendix A for a precise definition of our equilibrium concept.

holds belief θ and observes valence ν votes for the misaligned party. We write $s_i^j(\theta)$ for the probability that an incumbent from party j in policymaking environment $q_i \in \{q_l, q_h\}$ chooses C conditional on policy C being available and voters' holding the belief θ. Abusing terminology slightly, we refer to r and s^j as strategies, and to $(r, s^a, s^m) = (r, s_l^a, s_h^a, s_l^m, s_h^m)$ as a strategy profile.

We additionally restrict attention to a subset of the stationary Markov Perfect Equilibria in which voters use cutoff strategies (we discuss the extent to which this restriction is binding in Section 6): the voter in period t elects a valence-advantaged misaligned candidate if and only if θ_t is above some threshold (for a formal statement see Definition 1).[6]

In what follows, we write $\Pr(q = q_h \mid \theta_{t-1}, x_t, s^j, j_t)$ for the voters' posterior belief at the period t election if at the beginning of t they observe a history resulting in belief θ_{t-1}, see incumbent j_t implement x_t, and conjecture that the incumbent uses strategy s^j. So $\theta_t = \Pr(q = q_h \mid \theta_{t-1}, x_t, s^j, j_t)$. We say that party j plays *sincerely* at θ if $s_l^j(\theta) = s_h^j(\theta) = 1$. We say that a voter in t is *more optimistic* (resp. *pessimistic*) the higher (resp. lower) θ_t is.

2.1 Comments on the Model

Before turning to the analysis, several assumptions of the model merit further explanation.

First, we assume that voters are active and collect payoffs for only two periods, yet they have beliefs that reflect the full history of play and are sufficient statistics for behavior. This assumption is intended to represent a situation in which there are overlapping generations of short-lived voters. Old voters receive payoffs, but do not care about the future and do not vote. Young voters do not actually observe the entire history of play, but instead inherit beliefs from their parents (which justifies our restriction to Markov Perfect equilibria). The voters anticipate receiving payoffs in the future (when they are old) and, thus, vote with an eye toward the future (as do the voters in our model). Alternatively, the setup is also

[6] We show in Appendix B that in a two-period version of the model voters use cutoff strategies in all equilibria.

equivalent to a model with one long-lived but myopic voter. We make this assumption for simplicity, but it is not innocuous. If voters were long-lived and patient, they could experiment with different parties to learn about the policymaking environment. Our results should survive qualitatively, however, as long as voters are not too patient. The assumption that parties are represented by a string of short-lived politicians is made to simplify the model: an incumbent needs only to worry about the consequences of her policy choice on tomorrow's policymaking, instead of taking into account the entire future trajectory of beliefs and the resulting policies. Most of our findings, however, survive when we assume long-lived parties, each represented by one politician.

Second, it is important to ask what the real-world analogues to the common-interest policies in our model are. We offer three possibilities. Formally, in our model, common-interest policies are short-run Pareto improvements—putting aside strategic considerations, they deliver a higher instantaneous payoff to all actors than even their most-preferred ideological policy. Given this, one might think of a common-interest policy as a technocratic solution to a problem that allows for a Pareto improvement. This might look like a Pigouvian tax with compensatory transfers, for example. Second, one can think of the common-interest policy as representing a public good (say, investment in public health or infrastructure). To see how this relates to our setup, imagine that in each period the incumbent has a budget of 1 that they can spend on distributive politics (i.e., give the budget to their allies) or that they can invest in a public good of value b. The value of the public good varies from period to period. In some periods, voters value the public good more than receiving distributive benefits ($b > 1$) so that common-interest policy is available, and in some periods voters value receiving distributive benefits more than the public good ($b < 1$). No matter what, voters dislike other voters receiving distributive benefits. Finally, one might think of a common-interest policy as a stable bipartisan compromise to a contentious issue that all sides prefer to the instability of fluctuations between two partisan extremes.

Under all of these interpretations, it seems natural to assume that politicians have more information than voters regarding the policy-making environment. For the first two interpretations, politicians surely have access to more expertise than voters do about the feasibility of technocratic Pareto improvements or the value of public goods in many policy domains. For the final interpretation, politicians presumably know more than voters about when and how often the moment is politically ripe for bipartisan compromise and when such compromise is instead politically infeasible.

Third, we assume that politicians can only implement one policy per period. This is meant to represent, in stylized form, the politicians' time constraint when governing (Curry and Lee, 2020). There is only so much that can be achieved in a given legislative session or term in office. Thus, politicians face decisions, on the margin, about what kind of policy to emphasize—partisan or common interest. We capture this in stark form by allowing politicians to do only one thing per period.[7]

Finally, one or the other politician in an election always has a valence advantage—that is, a characteristic that is valued by voters and is orthogonal to policy alignment. Such valence advantages are a standard assumption in models of elections and are typically thought to reflect electoral considerations like charisma, propensity for corruption, name recognition, partisan tides, or access to campaign resources (Stokes, 1963; Ansolabehere and Snyder Jr, 2000; Groseclose, 2001; Aragones and Palfrey, 2002; Schofield, 2007).

[7]Regarding time constraints, Curry and Lee (2020) quote James C. Wright, the 48th Speaker of the House, as follows:

> My two biggest competitors are the clock and the calendar. There are so many things I would like to do... The trouble is you have only so many weeks in the legislative year, and so many days in the legislative week, so many hours in the legislative day.

.

3 Preliminaries

3.1 Equilibrium

As mentioned in the introduction, we focus on equilibria in which the voters use cutoff strategies: they elect misaligned candidates if and only if the misaligned candidate is valence advantaged and the voters are sufficiently optimistic that the policymaking environment is favorable. We start by defining a cutoff equilibrium:

Definition 1 *Let $\theta_c \in [0,1]$. An equilibrium is a θ_c-cutoff equilibrium with strategies (s^m, s^a, r) if for some $r_{\theta_c} \in [0,1]$, r satisfies*

$$r(\theta,\nu) = \begin{cases} 0 \text{ if } \nu = -v \text{ or } \theta < \theta_c \\ r_{\theta_c} \text{ if } \nu = v \text{ and } \theta = \theta_c \\ 1 \text{ if } \nu = v \text{ and } \theta > \theta_c. \end{cases} \tag{1}$$

Our focus on cutoff equilibria is driven by a variety of considerations. In the model, the voters suffer from having a misaligned politician in power only when a common-interest policy is not implemented. So if the politicians played sincerely, the expected loss from having a misaligned politician in power would be decreasing in voter optimism, and the voters would necessarily play a cutoff strategy. We show that formally in Appendix B where we solve a two-period version of our model.

In what follows, we graphically present one cutoff equilibrium, describe its properties (Section 3.2), and discuss the dynamics of politics and policymaking in this equilibrium (Section 4). In Section 5, we present a formal result on this equilibrium's existence. In Section 6, we establish the properties of all cutoff equilibria for a subset of parameters. All cutoff equilibria are qualitatively similar to the one considered in this section, and hence focusing on it is without loss of generality. In Section 6, we also discuss what we know about the existence of equilibria that are not in cutoff strategies.

3.2 Equilibrium

Figure 1 illustrates a θ_c-cutoff equilibrium. In this equilibrium, voters' beliefs can be divided into four categories delineated by three thresholds (defined later), $\theta_*, \hat{\theta}$ and $\tilde{\theta}$. The four categories are *very pessimistic* ($\theta < \theta_*$), *moderately pessimistic* ($\theta \in (\theta_*, \hat{\theta})$), *moderately optimistic* ($\theta \in (\hat{\theta}, \tilde{\theta})$, and *very optimistic* ($\theta > \tilde{\theta}$).

In this equilibrium, voters never elect a misaligned politician who is not valence advantaged. Voters elect valence-advantaged misaligned politicians if and only if they are optimistic. When elected (on the equilibrium path), misaligned politicians play sincerely: they take all common-interest opportunities that arise ($s^m = (1, 1)$). The behavior of the aligned politicians, however, is more complex. The aligned politicians play sincerely ($s^a = (1, 1)$) when voters are very pessimistic or very optimistic. When voters are moderately optimistic, aligned politicians forgo some common-interest opportunities. In particular, they forgo all such opportunities in the unfavorable policymaking environment ($s^a_l = 0$), and forgo some of them in the favorable policymaking environment ($s^a_h \in (0, 1)$). Specifically, they play a strategy such that following P_a, voters hold belief $\hat{\theta}$. Finally, when voters are moderately pessimistic (i.e., $\theta_t \in (\theta_*, \hat{\theta})$), the aligned politicians implement only partisan policies, even when common-interest policies are available ($s^a = (0, 0)$).

The equilibrium depicted in Figure 1 captures the intuition that electorally-minded aligned politicians may refrain from common-interest policymaking. Moreover, it shows that such partisan bias can take an extreme form. When voters are moderately pessimistic, the system is caught in what we call a *partisan trap*: aligned politicians implement partisan policy, voters learn nothing, and because voters are pessimistic, they continue to reelect the aligned party.

We discuss how beliefs and policymaking evolves in this equilibrium as well as the steady states of this equilibrium in Section 4. Before doing so, we explain why the strategies depicted in Figure 1 are part of an equilibrium. To this end, we first define the belief

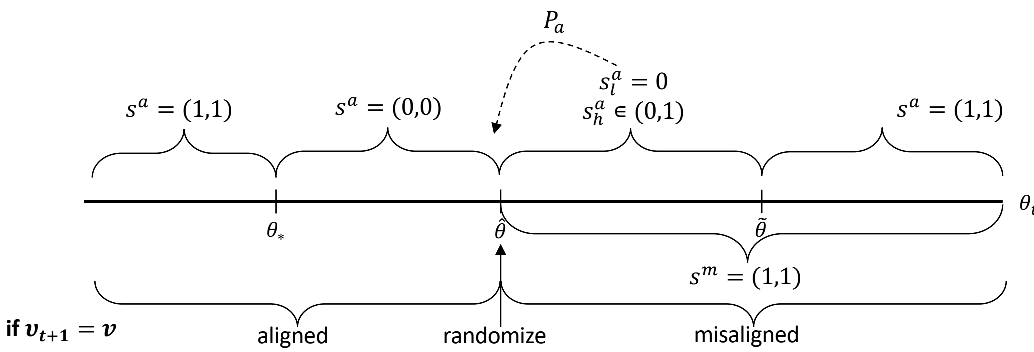

Figure 1: The structure of an equilibrium where voters use a cutoff strategy.

thresholds used in Figure 1.

The equilibrium depicted in Figure 1 is a θ_c-cutoff equilibrium with $\theta_c = \hat{\theta}$, where $\hat{\theta}$ is the belief at which voters are indifferent between the aligned and valence-advantaged misaligned candidates when they expect the aligned candidate to only implement partisan policies (i.e., $s^a(\hat{\theta}) = (0,0)$) and the misaligned candidate to play sincerely (i.e., $s^m(\hat{\theta}) = (1,1)$). In this case, the aligned party is expected to provide a payoff of 1 in the next period, and the misaligned party is expected to provide a payoff of $b + v$ when common-interest policy is available and a payoff of $-1 + v$ when common-interest policy is not available. Thus $\hat{\theta}$ satisfies:

$$\left(\hat{\theta}q_h + (1-\hat{\theta})q_l\right) b - \left(1 - \left(\hat{\theta}q_h + (1-\hat{\theta})q_l\right)\right) + v = 1,$$

which implies

$$\hat{\theta} = \frac{2 - v - q_l\,(b+1)}{(q_h - q_l)\,(b+1)}. \tag{2}$$

Two additional critical thresholds, θ_* and θ^*, are defined as follows:

$$\Pr(q = q_h \mid \theta_*, C, s^j = (1,1), j) = \hat{\theta}, \tag{3}$$

$$\Pr(q = q_h \mid \theta^*, P_j, s^j = (1,1), j) = \hat{\theta}. \tag{4}$$

Hence, θ_* is the prior belief such that if the voters expect an incumbent from party j to play sincerely and observe her implement C, their posterior is $\hat{\theta}$. Analogously, θ^* (which will be used in our definition of $\tilde{\theta}$) is the prior belief such that if voters expect an incumbent from party j to play sincerely and observe her implement P_j, their posterior is $\hat{\theta}$.

Let $\bar{\theta}$ be the belief at which the voters are indifferent between the aligned and valence-advantaged misaligned candidates when they expect both candidates to play sincerely (i.e., $s^a(\bar{\theta}) = s^m(\bar{\theta}) = (1,1)$). In this case, the aligned candidate is expected to provide a payoff of b when common-interest policy is available and a payoff of 1 when common-interest policy is not available, and the misaligned candidate is expected to provide a payoff of $b + v$ when common-interest policy is available and a payoff of $-1 + v$ when common-interest policy is not available. Thus, the expected payoff from electing the aligned candidate is

$$(\bar{\theta}q_h + (1-\bar{\theta})q_l)b + (1 - (\bar{\theta}q_h + (1-\bar{\theta})q_l)),$$

and the expected payoff from electing the misaligned candidate is

$$(\bar{\theta}q_h + (1-\bar{\theta})q_l)b - (1 - (\bar{\theta}q_h + (1-\bar{\theta})q_l)) + v.$$

So

$$\bar{\theta} = \frac{1 - q_l - \frac{v}{2}}{q_h - q_l}. \tag{5}$$

Note that $\hat{\theta} < \bar{\theta}$ and that under Assumption 1.1, both $\hat{\theta}, \bar{\theta} \in (0,1)$. Finally, we define the last threshold as $\tilde{\theta} = \max\{\theta^*, \bar{\theta}\}$.

To understand why the strategies depicted in Figure 1 are part of an equilibrium, consider first the incentives faced by an aligned incumbent in a period in which a common-interest policy is available. On the one hand, the incumbent's immediate policy payoff is higher if she implements C. One the other hand, C and P_a may lead to different voter beliefs, and hence different electoral outcomes. An aligned incumbent at time t cares about the electoral outcome because the party in power in the subsequent period chooses policy in that period. In expectation, an aligned incumbent prefers the policies that will be implemented if the aligned party remains in power.

If both policies lead to posterior voter beliefs resulting in the same electoral outcome (either both posteriors are below $\hat{\theta}$ or both are above $\hat{\theta}$), the aligned incumbent's incentives are driven entirely by immediate policy payoffs and so she strictly prefers to play sincerely.[8] This is the case if voters are either very pessimistic ($\theta_t < \theta_*$) or very optimistic ($\theta_t > \tilde{\theta}$).

By contrast, the aligned incumbent faces a tradeoff if P_a leads to a belief below $\hat{\theta}$ and C leads to a belief above $\hat{\theta}$, which is the case for moderate beliefs if the voter expects sincere play (technically, this is true for all $\theta_t \in (\theta_*, \theta^*)$; we will discuss later what happens for $\theta_t \in (\theta^*, \tilde{\theta})$). On the one hand, C provides the aligned incumbent a better immediate policy payoff than P_a. On the other hand, C creates the risk of an electoral loss and worse future policy for the aligned incumbent, while P_a does not. Assumption 1.2 ensures that in theses circumstances, the aligned incumbent is willing to forgo the immediate policy benefit in order to win the election. Hence, in equilibrium, when faced with moderate beliefs, aligned politicians must either choose only partisan policies, or P_a and C must have electoral consequences that are less stark than they are under sincere play.

Consider what happens with moderately optimistic beliefs, i.e., $\theta_t \in (\hat{\theta}, \theta^*)$. It cannot be part of an equilibrium for aligned incumbents to always choose P_a in these circumstances. If the aligned incumbent always plays P_a, then the choice of policy provides no information. As such, voters do not update upon seeing P_a. This means that the voters' posterior belief after P_a would equal their prior: $\theta_{t+1} = \theta_t > \hat{\theta}$. At these beliefs, the voter replaces the aligned incumbent when the misaligned challenger is valence advantaged, which is the worst possible electoral consequence. This implies that aligned incumbents are better off choosing

[8] The same electoral consequences imply the same future payoff only if the subsequent behavior of the replacement politician doesn't depend on the policy chosen. This is the case in the equilibrium described in Figure 1, but that is not given, which complicates the proof of uniqueness. We discuss this in more detail in Section 6.

C whenever possible, as they obtain higher immediate payoffs and face at worst the same electoral consequences, contradicting the conjectured equilibrium.

Hence, for moderately optimistic beliefs, an equilibrium will require that P_a have less stark electoral consequences. The only way to achieve this is for voters to randomize upon observing P_a, which in a cutoff equilibrium can only happen at the cutoff belief where the voter is indifferent. So in equilibrium the voters' belief after observing P_a must be $\hat{\theta}$, which requires that P_a be more informative than it is under the strategy $s^a = (0, 0)$ but less informative than it is under the strategy $s^a = (1, 1)$. To achieve this, the aligned party must randomize.

What do these mixed strategies look like? In the proposed equilibrium, the misaligned party plays sincerely on the equilibrium path, hence the downside of losing the election for the aligned party is smaller in the favorable policymaking environment than in the unfavorable environment. As a result, an aligned incumbent is more willing to make her party electorally vulnerable in the favorable policymaking environment than in the unfavorable one. Thus, if the incumbent is indifferent between C and P_a at θ_t in the favorable policymaking environment, then she strictly prefers playing P_a in the unfavorable policymaking environment. This means that, in a mixed strategy equilibrium, the aligned party always plays partisan policy in the unfavorable environment and mixes in the favorable environment—i.e., $s^a_l(\theta_t) = 0$ and $s^a_h(\theta_t) \in (0, 1)$ for all $\theta_t \in (\hat{\theta}, \theta^*)$. The exact mixing probability that the aligned party uses at each $\theta_t \in (\hat{\theta}, \theta^*)$ is pinned down by Bayes' rule and is derived formally in Appendix A. To induce this mixing by the incumbent, the voter must choose a reelection probability $r_{\hat{\theta}}$ at belief $\hat{\theta}$ that makes an aligned incumbent in the favorable policymaking environment indifferent between choosing C and then losing the election if the misaligned candidate is valance advantaged or choosing P_a and defeating a valence advantaged opponent with probability $r_{\hat{\theta}}$.

The discussion above suggests that for moderately optimistic beliefs $\theta_t \in (\theta^*, \tilde{\theta})$, if that interval is nonempty, both sincere play and randomization are sequentially rational, which will be one of several sources of multiplicity we discuss in Section 6. However, for the voter to prefer a valence-advantaged misaligned party, the aligned party cannot play sincerely, hence aligned incumbents must randomize.

Finally, consider moderately pessimistic beliefs, $\theta_t \in (\theta_*, \hat{\theta})$. As discussed above, aligned politicians must either choose only partisan policies, or P_a and C must lead to less stark electoral consequences than under sincere play. Unlike with moderately optimistic beliefs, with moderately pessimistic beliefs choosing only partisan policies is consistent with equilibrium. If the aligned incumbent always plays P_a, the voters' posterior beliefs equal their priors: $\theta_{t+1} = \theta_t < \hat{\theta}$. Since the voters are moderately *pessimistic*, this leads to the aligned party being reelected for certain. Playing C is off the equilibrium path, but under the natural assumption that voters update to $\theta_{t+1} > \hat{\theta}$ after observing C, playing C leads the voters to elect the misaligned candidate if she is valence advantaged. Thus, under this conjecture, the aligned party playing partisan policy is a best response in both policymaking environments.[9]

Given the voters' and the aligned politicians' behavior, it is optimal for the misaligned incumbents to play sincerely on the equilibrium path. Playing sincerely not only delivers a higher policy payoff, but it also increases voters' beliefs, which can only benefit the misaligned party electorally.[10]

And finally, it is easy to see that given the behavior or the politicians, it is indeed optimal for the voters to play a cutoff strategy with $\theta_c = \hat{\theta}$.

[9]There also cannot be a mixed strategy equilibrium in which C results in the belief $\hat{\theta}$. In order to sustain randomization by the aligned party, the voter has to randomize at $\hat{\theta}$ to make the aligned incumbent indifferent starting at a prior below $\hat{\theta}$. But the voter is already randomizing at $\hat{\theta}$ to make the aligned party indifferent starting at a prior above $\hat{\theta}$.

[10]Off the equilibrium path, there are beliefs for which the misaligned party does not play sincerely. We describe the behavior of the misaligned party for all beliefs in Section 5.

4 Dynamics of Politics, Policy, and Polarization

We now turn to what the equilibrium illustrated in Figure 1 predicts in terms of the dynamics of politics, policy, and polarization.

4.1 Partisan Traps

A key finding of our model is the existence of equilibrium partisan traps. If in some period t voters are moderately pessimistic, $\theta_t \in (\theta_*, \hat{\theta})$, voters are trapped in a permanent vicious cycle in which they always elect the aligned party, the aligned party always implements partisan policies, and voters learn nothing about the policymaking environment. Thus, even if the policymaking environment is favorable to common-interest policies ($q = q_h$), the voter remains permanently pessimistic and the aligned party's behavior remains permanently partisan.

Importantly, when the system is caught in a partisan trap, voters behave as if they were rewarding the party they are ideologically aligned with for behaving in a purely partisan manner. This might lead an observer to conclude that the voters are themselves hyper-partisans. But, in the model, voters are not intransigent partisans. Indeed, they would like the aligned party to engage in common-interest policymaking. Voters decline to electorally punish aligned incumbents for not doing so out of fear that the policymaking environment is unfavorable, so that policy will be even worse under the misaligned party. The aligned party behaves in a partisan manner to reinforce this rational fear on the part of voters because such fear is electorally insulating.

4.2 Policymaking with Pessimistic Priors

Consider some period, t, in which the voters start with very pessimistic prior beliefs about the policymaking environment ($\theta_t < \theta_*$). In this case, the aligned party engages in sincere policymaking, and the evolution of the political system depends on both the true fundamentals of the policymaking environment (i.e., whether $q = q_h$ or $q = q_l$), and on the idiosyncrasies of how frequently common-interest opportunities happen to arise.

First, suppose that the policymaking environment is favorable. Given that politicians are playing sincerely at this belief, voters' beliefs will eventually drift upwards towards the truth. That is, because politicians are taking common-interest opportunities when available, and those opportunities are frequent, over time the voters will start to become more optimistic. But as soon as the voters' beliefs become a little too optimistic, $\theta_t > \theta_*$, the system falls into a partisan trap. Voters elect the aligned party, the aligned party behaves in a partisan manner, and voters stop learning new information about the policymaking environment. Thus, if priors start pessimistic but the truth is that the policymaking environment is favorable, the system is guaranteed to end up in a partisan trap in the long run.

If, instead, the policymaking environment is unfavorable, then one of two things can happen. Given that the politicians are playing sincerely, it is possible for the voters' beliefs to converge toward the truth, so that voters are permanently pessimistic and the aligned party permanently plays sincerely. This is the best case scenario for the voters in an unfavorable policymaking environment. The alternative is a partisan trap. Even though the policymaking environment is unfavorable, it is possible that there will be enough "lucky" draws in which common-interest policies are available that the voter ends up with a belief $\theta_t > \theta_*$. At that point, the system has entered a partisan trap so all common-interest policymaking and voter learning ceases.

Two somewhat surprising take aways follow from this analysis. First, voters are better off if they start with very pessimistic than with moderately pessimistic beliefs. Extreme voter pessimism at least temporarily frees the aligned party from electoral consequences and allows it to engage in sincere policymaking, whereas moderate pessimism leads immediately to a partisan trap. Second, if voters have pessimistic beliefs, they are at least weakly better of in the long run in an unfavorable rather than a favorable policymaking environment. In an unfavorable policymaking environment, it is possible that the aligned politicians play sincerely in the long run and there is common-interest policy in q_l share of periods. In a favorable policymaking environment, a partisan trap is guaranteed in the long run and there is never common-interest policy.

4.3 Policymaking with Optimistic Priors

A related dynamic occurs when the voter starts with optimistic prior beliefs about the policymaking environment ($\theta_t \geq \hat{\theta}$). The system can fluctuate for a substantial period of time between partisan and common-interest policy and between aligned and misaligned parties. During that period, voters' beliefs may become more or less optimistic as information is revealed in each period. However, in the long run, if the policymaking environment is unfavorable, then eventually the system will end up in a partisan trap, as voters' beliefs are inevitably drawn toward the truth with revelation of more and more information. If, instead, the policymaking environment is favorable, two things can happen. First, voters' beliefs can converge toward the truth, and the system can end up in a situation where both parties play sincerely and the voters vote entirely based on valence. Second, if there are enough unlucky draws in a row, voters' beliefs can end up below $\hat{\theta}$ and the system can be stuck in a partisan trap. Thus, there is always a risk of ending up in a partisan trap, even when beliefs are optimistic and the policymaking environment is favorable.

4.4 Steady State of Policymaking

The discussion above reveals that the game can converge to three different steady states. First, if the policymaking environment is unfavorable and voters start with very pessimistic beliefs, voters' beliefs may converge to the truth, in which case they always elect the aligned party and the aligned party engages in common-interest policymaking whenever possible. Second, if the policymaking environment is favorable and voters start with optimistic be-

	pessimistic prior $\theta_0 < \hat{\theta}$	**optimistic prior** $\theta_0 > \hat{\theta}$
unfavorable policymaking environment q_l	partisan trap or permanent pessimism with sincere policy	partisan trap
favorable policymaking environment q_h	partisan trap	partisan trap or permanent optimism with sincere policy

Table 1: Possible equilibrium steady states as a function the true policymaking environment and the voters' priors.

liefs, voters' beliefs may converge to the truth, in which case they vote entirely based on valence and politicians engage in common-interest policymaking whenever possible. Finally, the system ends up in a partisan trap if voters start (i) with moderately pessimistic beliefs, (ii) with beliefs mismatched to the policymaking environment (i.e., pessimistic beliefs in a favorable policymaking environment or optimistic beliefs in an unfavorable policymaking environment), or (iii) with optimistic beliefs and a series of unlucky draws. Table 1 summarizes these possibilities.

4.5 Relationship to Affective Polarization

A lot has been written about increasing political polarization in the United States in the recent years. On the empirical side, evidence suggests that polarization is an elite-driven phenomenon. The literature has documented that voting in the U.S. Congress is increasingly divided along party lines (McCarty, 2019), and political elites engage in polarizing discourse that paints policymaking as a zero-sum fight and the opposition as out-of-touch (Gentzkow, Shapiro and Taddy, 2019). At the same time, voters appear to have moderate preferences (Fowler et al., 2023) and to care about policy (Orr, Fowler and Huber, Forthcoming). Moreover, those preferences do not seem to have polarized dramatically in recent decades (Hill and Tausanovitch, 2015). Multiple studies have documented, however, strong affective polarization and party loyalty: voters exhibit animosity toward the opposing party and both expressed and actual reluctance to vote across ideological lines (Iyengar et al., 2019;

Abramowitz and Saunders, 2008).

Our analysis helps make sense of such dynamics. In our model, the voters' policy preferences are moderate, but elite policymakers engage in strategically partisan behavior. As a consequence, voters become more pessimistic about the policymaking environment and hence become more electorally tied to the party they are ideologically aligned with. In the model, there is only one voter active in each period, but one can envision a population of voters ordered on the ideological spectrum. Some of them would be aligned with one party and some of them with the other. As long as the ideological alignment of the median voter is known and stable, the equilibria of our model would persist, with the aligned party understood to be the party ideologically aligned with the median voter. In response to partisan policymaking, each voter would become more pessimistic, but their pessimism would translate into unwavering support for different parties, depending on their individual ideology. Thus, increasing partisanship in our model is naturally interpreted as increasing polarization.

4.6 Relationship to Introductory Examples

Partisan policymaking is a complex phenomenon and no single model can provide a fully satisfactory account of it. Nevertheless, it is useful to see to what extent the dynamics of our model help make sense of some historical patterns of policymaking and political polarization. In this spirit, we now revisit the two examples we discussed in the introduction to compare them to the equilibrium dynamics of our model.[11]

The period from the end of Reconstruction through the Great Deal can be thought of as a period of extremely pessimistic beliefs ($\theta_t < \theta_*$). In this circumstance, our model predicts that the aligned party will engage in common-interest policymaking when oppor-

[11] These examples involve policymaking by multiple politicians, each of which faces different voters. Nevertheless, as we argue in the Conclusions, the lessons from our model would extend to a model with multiple policymakers.

tunities arise, consistent with the bipartisan reforms enacted early in the New Deal. The model, however, also predicts that such common-interest reforms necessarily lead to voters becoming more optimistic about the policymaking environment, consistent with findings from public opinion research about the New Deal period (Schickler and Caughey, 2011). As voters become more optimistic, electoral consequences start to loom large in the minds of incumbent politicians, leading to a partisan trap. That is, after a period of common-interest reforms, representatives that are aligned with their constituencies should refuse to engage in any further common-interest policymaking. Instead, they should pursue policies that divide politics along partisan lines, as indeed the latter New Deal reforms did (Katznelson, Geiger and Kryder, 1993).

The period of Democratic control of the House before 1994 can be understood as a period of ideological misalignment between voters and their representatives. Democrats tended to field higher quality candidates, but were misaligned with the voters in increasingly conservative southern Democratic districts (Jacobson, 1996). Consistent with our model, these misaligned incumbents engaged in common-interest policymaking, focusing on pork barrel projects that benefited all of their constituents. In the elections of 1994, Republicans recruited a new crop of high-quality challengers. This swept away the Democrats' valence advantage and led to the replacement of misaligned southern Democrats by aligned Republicans. According to our model, aligned incumbents find common-interest policy less attractive than misaligned incumbents. Thus, our model predicts that Republicans had an incentive to shore up their electoral standing by making sure that voters perceived policymaking and politics as primarily a zero-sum, partisan competition. And, consistent with this prediction, elites do appear to have retreated to their partisan corners, pursuing partisan policies and using increasingly polarizing language to describe political disagreements (Gentzkow, Shapiro and Taddy, 2019).

4.7 Red versus Blue States

Partisan policymaking arises in our model because convincing the voters that common-interest policies are rare benefits aligned incumbents electorally. Hence, in uncompetitive environments, aligned incumbents are less incentivized to resort to purely partisan policymaking. And indeed, in our model the equilibrium described above ceases to exist when the probability that a misaligned candidate is valence advantaged (p) is sufficiently low or when the magnitude of the valence advantage is sufficiently small (v less than the lower bound of Assumption 1.1), both reasonable representations of an uncompetitive electoral district.

One story thus suggests that districts where one party has a strong fundamental advantage should experience more common-interest policymaking. And, indeed, this is consistent with some empirical evidence showing that parties in the United States vote in a less polarized way in less electorally competitive states (Hinchliffe and Lee, 2016). Our model, however, reveals that the theoretical relationship between electoral competitiveness and policy is more nuanced, which may explain why there seems to be substantial heterogeneity in parties' empirical behavior. In particular, we point to another reason (besides a large valence advantage for one party) that a district might have uncompetitive elections—it may be caught in a partisan trap. When an electoral district is caught in a partisan trap, the dominant party is dominant precisely because the voters are pessimistic about the policymaking environment, and the dominant party keeps the voters that way by engaging in partisan policymaking. On this account, were the dominant party to pursue common-interest policy, it would cost them electorally. Thus, our model does not predict that we should always expect a high level of common-interest policymaking in apparantely electorally uncompetitive polities. Some uncompetitive polities are uncompetitive for equilibrium reasons that forestall common-interest policy. This might help explain the observation that in the contemporary United States, state legislators are elected increasingly in uncompetitive elections, legislatures appear to be engaged in increasingly partisan policymaking, and voters increasingly support party-aligned candidates, even when those candidates's policy preferences are extreme relative to the voters (Handan-Nader, Myers and Hall, 2023).[12]

5 Existence and Comparative Statics

In this section, we formally state the result on the existence of the equilibrium characterized in Section 3 and analayze comparative statics with respect to the parameters b, v and p.

5.1 Existence

To state the existence result, the following thresholds will be useful. Let θ_{**} and θ^{**} satisfy:

$$\Pr(q = q_h \mid \theta_{**}, C, s^j = (1,1), j) = \theta_*, \tag{6}$$

$$\Pr(q = q_h \mid \theta^{**}, P_j, s^j = (0,1), j) = \hat{\theta}. \tag{7}$$

Definition 2 formally defines the strategy that leads voters to hold belief $\hat{\theta}$ after observing a partisan policy.

Definition 2 *Let $\hat{s} : [\hat{\theta}, \theta^{**}] \to [0,1]^2$ satisfy $\hat{s}_l(\theta) = 0$ and $\Pr(q = q_h \mid \theta, P_j, \hat{s}(\theta), j) = \hat{\theta}$ for all $\theta \in [\hat{\theta}, \theta^{**}]$.*

Proposition 1 *Fix p and v satisfying Assumption 1.1. There exist $\overline{b}(v,p) > 1$ such that for all $b \in (1, \overline{b}(v,p))$ there exists a $\hat{\theta}-$equilibrium with*

$$r_{\hat{\theta}} = 1 - (b-1)\frac{1 + q_h(1-p)}{p(2-(1+b)q_h)}, \tag{8}$$

and the following properties.

[12] On the rise of partisan policymaking focused on matters like culture war issues, see, for example, the unprecedented increase in bills related to transgender care and rights ("Historic surge in bills targeting transgender rights pass at record speed," Washington Post, April 17, 2023).

1. *In each t, an aligned party incumbent:*

   (a) *for any $\theta_t \in (0, \theta_*)$, plays sincerely;*

   (b) *for any $\theta_t \in [\theta_*, \hat{\theta}]$, plays $s_l^a(\theta_t) = s_h^a(\theta_t) = 0$;*

   (c) *for any $\theta_t \in (\hat{\theta}, \max\{\bar{\theta}, \theta^*\}]$, plays $s^a(\theta_t) = \hat{s}(\theta_t)$;*

   (d) *for any $\theta_t \in (\max\{\bar{\theta}, \theta^*\}), 1]$, plays sincerely.*

2. *In each t, a misaligned party incumbent:*

   (a) *for any $\theta_t \in (0, \theta_{**})$, plays sincerely;*

   (b) *for any $\theta_t \in [\theta_{**}, \theta_*]$, plays $s_l^m(\theta_t) = s_h^m(\theta_t) = 0$;*

   (c) *for any $\theta_t \in (\theta_*, 1]$, plays sincerely.*

Proposition 1 formalizes the description of the equilibrium depicted in Figure 1 and completes the description of the behavior of the misaligned candidates. As discussed before, the incentives for the misaligned party tend to go together and favor playing C. For some pessimistic beliefs, however, they do not. Consider $\theta_t \in [\theta_{**}, \theta_*]$. For such beliefs, the misaligned party loses power in the next period independent of the policy implemented. Implementing C, however, leads to a belief at which the aligned replacement behaves in a purely partisan way, while implementing P_m leads to a belief at which the aligned replacement behaves sincerely. Since the latter delivers a higher payoff, the misaligned politician is willing to forgo the policy benefit from C in order to obtain better policymaking in the next period.

Proposition 1 establishes that the equilibrium exists if b is not too large, i.e., if $b < \bar{b}(v, p)$. The restriction comes from the following consideration. The larger is b, then greater is the distance between $\hat{\theta}$ and $\bar{\theta}$: the threshold at which the voter stops preferring a partisan-behaving aligned party goes down. That means that there exists a set of beliefs close to $\bar{\theta}$ for which any strategy of the aligned party must result in both policies leading to the same electoral consequences. That implies that the aligned party plays sincerely. But if the aligned party plays sincerely for beliefs to the left of $\bar{\theta}$, then by the definition of $\bar{\theta}$, the voter prefers the aligned party, which contradicts the hypothesis that we are in a cutoff equilibrium.

5.2 Comparative statics

Here we consider how the equilibrium described in Proposition 1 changes with the parameters of the model. In particular, we are interested in how behavior and outcomes are affected by the value of common-interest policies (b), the size of the valence shock (v), and the probability that the misaligned candidate is valence advantaged (p).

Proposition 2 *Consider the equilibrium strategy profile (s^m, s^a, r) characterized in Proposition 1. Let θ_*, $\hat{\theta}$, θ^*, $\bar{\theta}$, be the corresponding belief thresholds. Then*

1. $\frac{d\theta_*}{db} < 0$, $\frac{d\hat{\theta}}{db} < 0$, $\frac{d\theta^*}{db} < 0$, $\frac{d\bar{\theta}}{db} = 0$,
2. $\frac{d\theta_*}{dv} < 0$, $\frac{d\hat{\theta}}{dv} < 0$, $\frac{d\theta^*}{dv} < 0$, $\frac{d\bar{\theta}}{dv} < 0$,
3. $\frac{d\theta_*}{dp} = \frac{d\hat{\theta}}{dp} = \frac{d\theta^*}{dp} = \frac{d\bar{\theta}}{dp} = 0$.

The first part of Proposition 2 states that if the payoff from common-interest policies increases, all equilibrium belief thresholds (weakly) decrease. This is because when common-interest policies are more valuable to the voters, a misaligned incumbent who plays sincerely becomes more valuable relative to an aligned incumbent who plays insincerely. Hence, the belief at which the voters are indifferent between those two, $\hat{\theta}$, decreases. Since θ_* and θ^* are defined as priors that (for a fixed strategy) lead to $\hat{\theta}$ after C and P_a respectively, those thresholds decrease as well. This implies that, when common-interest policy is more valuable, electoral incentives kick in at lower beliefs and therefore partisan traps require greater pessimism.

This result suggests that the political and policy consequences of common-interest reforms becoming more valuable are contingent on the voters' beliefs. When voters are not too pessimistic, that is, when $\theta > \theta_*$, an increase in b has either no or positive consequences for the voters: they may remain stuck in the partisan trap, but they may also be pushed out of it. The latter possibility occurs because voters are more inclined to elect a misaligned politician in order to enjoy the benefits of common-interest policies when b is higher. However, when voters are very pessimistic, that is, when $\theta < \theta_*$, an increase in b either moves the system into a partisan trap or increases the probability of ending up in a partisan trap. Thus, for very pessimistic voters, increasing the value of common-interest policies worsens the political environment.

The second part of Proposition 2 states that increasing the size of valence shocks (v) has the same qualitative effect as increasing the value of common-interest policy (b). Increasing v makes the valence-advantaged misaligned candidate more attractive, which means that policymaking has electoral consequences at more pessimistic beliefs. Therefore partisan behavior arises for more pessimistic beliefs and all of the same implications follow.

Finally, increasing the frequency with which misaligned candidates are valence advantaged (p) does not affect the relevant equilibrium thresholds. It does, however, affect the evolution of beliefs. Perhaps surprisingly, increasing p has a complicated effect on the probability that the voters end up in a partisan trap. To see why, note that when starting with moderately optimistic beliefs, voters may end up in a partisan trap in that period only if the misaligned politician is in office. Since misaligned politicians are elected for such beliefs only if they are valence-advantaged, they are more likely to be elected as p increases. So when misaligned politicians are likely to be advantaged, voters receive more common-interest policies in the short run, but are more likely to end up in the partisan trap in the long run.

6 Uniqueness

The discussion in Section 3 hinted at the possibility of equilibrium multiplicity even within the class of cutoff equilibria. For example, it is straightforward to see that one can enlarge the set of beliefs sustaining a partisan trap by replacing θ_* with a lower belief, all the way down to 0. The aligned party playing only partisan policies for any pessimistic belief can be sustained by a natural off-equilibrium belief that the policymaking environment is favorable following C. In fact, for any $\theta \in (0, \theta_*)$, playing sincerely or playing $s^a(\theta) = (0, 0)$ can be part of an equilibrium.[13] Similarly, for beliefs $\theta_t \in (\theta^*, \min\{\theta^{**}, \bar{\theta}\})$ anticipation of sincere play makes the voters form relatively optimistic belief after seeing P_a, but anticipation of $s^a(\theta) = (0, 1)$ makes the voters form relatively pessimistic belief after seeing P_a. So playing sincerely could be an equilibrium, but playing $\hat{s}$ could be as well. All these variations on the equilibrium from Proposition 1 have less sincere policymaking on the equilibrium path than the equilibrium from Proposition 1 but share the same qualitative features. Hence, the discussion from Section 4 applies to all cutoff equilibria.

The proposition below describes the properties of all cutoff equilibria for a restricted set of parameters and reveals that multiplicity is restricted to the two cases just described.

Proposition 3 *Restrict attention to θ_c-cutoff equilibria in which the voters update weakly more positively after C than after P_j for every $j \in \{a, m\}$.*

For any v satisfying Assumption 1.1, there exists $\underline{p}(v) < 1$ and $\underline{b}(p)$ such that for any $p > \underline{p}(v)$, $\underline{b}(p) < \bar{b}(v, p)$ and for all $b \in (\underline{b}(p), \bar{b}(v, p))$, every equilibrium in this class has the following properties:

1. *$\theta_c = \hat{\theta}$ and $r_{\hat{\theta}}$ satisfies (8);*

2. *The aligned party incumbent plays:*

    (a) *for any $\theta_t \in (0, \theta_*)$, either $s^a_l(\theta_t) = s^a_h(\theta_t) = 0$ or sincerely;*

[13] The existence condition on b can be weakened if one replaces θ_* with 0 in Proposition 1.

(b) for any $\theta_t \in (\theta_*, \hat{\theta}]$*,* $s_l^a(\theta_t) = s_h^a(\theta_t) = 0$*;*

(c) for any $\theta_t \in (\hat{\theta}, \max\{\bar{\theta}, \theta^*\})$*,* $s^a(\theta_t) = \hat{s}(\theta_t)$*;*

(d) for any $\theta_t \in (\max\{\bar{\theta}, \theta^*\}), \theta^{**})$*, either* $s^a(\theta_t) = \hat{s}(\theta_t)$ *or sincerely;*

(e) for any $\theta_t \in (\theta^{**}, 1)$*, sincerely;*

3. The misaligned party plays sincerely on the equilibrium path.

When $p = 1$*, then* $\underline{b}(p) = 1$*.*

The conditions for uniqueness stated in Proposition 3 are more stringent than those needed for existence in Proposition 1, in particular, b cannot be too small. These conditions for uniqueness in Proposition 3 are sufficient, but do not seem to be necessary; that said, we have not been able to prove a tighter result. The main complication in proving uniqueness is as follows. In Section 3, much of our argument relies on the observation that if for some belief, both policies lead to the same electoral consequences, then the aligned incumbent plays sincerely at that belief. This argument is correct when we consider the equilibrium described in Proposition 1, because in this equilibrium, the misaligned party plays sincerely on the equilibrium path, so not only the electoral consequences are the same, but the policymaking that follows in case of replacement is the same. Suppose, however, for the sake of argument, that the misaligned party does not play sincerely, and in fact plays C with a probability that is decreasing in the voters' beliefs. In that case, even if for some belief P_a and C make the aligned party lose power with the same probability, P_a and C still lead to different policy consequences because the misaligned replacement's behavior is different at the posterior beliefs formed after observing P_a and C. So we cannot claim that for the belief in question that the incumbent necessarily plays sincerely. In the equilibrium described in Proposition 1, we posited that the misaligned party plays sincerely on the equilibrium path, but we cannot assume that this holds in every equilibrium. The lower bound on b in Proposition 3 guarantees that the misaligned party plays sincerely on the equilibrium path independent of how the aligned party is expected to play. And for the misaligned party to play sincerely, it must be that it prefers playing C followed by partisan policymaking to playing P_m followed by sincere policymaking. That requires that the current benefit of b is sufficiently large.

The argument above reveals that if we could prove uniqueness for a larger set of parameters by restricting attention to equilibria in which the aligned party plays sincerely on the equilibrium path. In fact, it turns out that under that restriction, non-cutoff equilibria do not exist.[14]

7 Conclusions

Motivated by the politics of elite polarization, we developed a model in which politicians are driven by electoral considerations to engage in partisan policymaking even though voters are moderates. There are many possible extensions and elaborations. For example, our model features a unitary policy maker, which is perhaps most reasonable for applications to elected executives such as presidents, governors, or mayors. The main mechanism, however, is likely to extend to a model with checks and balances or legislative policymaking.

Consider, for example, a legislature composed of two districts (A and B), each of which has a median voter aligned with a different party. Suppose further that an agreement of the representatives of both districts is needed to implement any policy. As long as the voters in (say) district A expect that district B is represented by a politician aligned with district B's voters (and therefore misaligned with district A's voters), their problem looks similar to the problem of the voters in our model. Electing an aligned politician creates legislative gridlock on partisan policies, while electing a misaligned politician means that

[14] We can see that one can dispense with the lower bound of b from the proof of Proposition 3, where the lower bound on b is needed only when we establish the sincere behavior of the misaligned politicians for sufficiently optimistic beliefs. The proof that we can rule out other non-cutoff equilibria is available upon request.

some partisan policies that district A's median voter dislikes will be implemented. Hence, as long as the voter is sufficiently pessimistic about the policymaking environment, she prefers the aligned representative to a valence-advantaged misaligned one. This, in turn, creates the same electoral considerations for legislators as those faced by the politicians in our model. This logic suggests the main incentives for the players would remain unchanged in the extended model. Additional considerations, however, may enter into play. Whether electing a misaligned politician in one district leads to partisan policymaking or gridlock depends on the ideology of the representative of the other district, which is also determined in equilibrium. Hence, the voters' calculus needs to incorporate her expectations about the behavior of other voters. We leave a formal investigation of this and related issues for future work.

A Appendix: Preliminaries, Notation, and Proofs

Let h_t denote the history of the game observed by voters up to period t election: $h_t = (\{j_n\}_{n=1}^t, \{x_n\}_{n=1}^t, \{v_n\}_{n=1}^{t+1})$. Let $\hat{h}_t$ denote the history observed by politicians up to the period t policy choice: $\hat{h}_t = (q, \{j_n\}_{n=1}^t, \{x_n\}_{n=1}^{t-1}, \{v_n\}_{n=1}^t, \{c_n\}_{n=1}^t)$. Since the incumbent in t faces a nontrivial policy choice only if C is available, in what follows we use $\hat{h}_t$ to denote only those histories with $c_t = y$. Let H_t be the set of all feasible h_t and let $\hat{H}_t$ be the set of all feasible $\hat{h}_t$.[15] Let $\hat{H}_t^j$ be the subset of $\hat{H}_t$ at which party j is the incumbent. A strategy of voter k is a mapping from H_k to a probability distribution $r \in [0,1]$, where r is interpreted as the probability of voting for the misaligned party. A strategy of incumbent j_t is a mapping from $\hat{H}_t^j$ to a probability distribution $s^j \in [0,1]$, where s^j is interpreted as the probability the j_t implements C when available.

Let θ_t be the voters' belief that $q = q_h$ formulated using Bayes' rule (whenever possible) just before the period t election under the conjecture that parties use the strategy (s^a, s^m)—i.e., $\theta_t = \Pr(q = q_h \mid h_t, (s^a, s^m))$. We restrict attention to equilibria in which $r(h_k) = r(h_{k'})$ for all h_k and $h_{k'}$ for which $\theta_k = \theta_{k'}$ and $v_{k+1} = v_{k'+1}$, and $s^j(\hat{h}_t) = s^j(\hat{h}_{t'})$ for any $\hat{h}_t$ and $\hat{h}_{t'}$ with the same q and for which $\theta_{t-1} = \theta_{t'-1}$.

Given this equilibrium restriction, and abusing slightly notation, we write $r(\theta, \nu)$ for the probability that any voter who holds belief θ and observes valence ν votes for the misaligned party, and $s_i^j(\theta)$ for the probability that an incumbent from party j in policymaking environment $q_i \in \{q_l, q_h\}$ chooses C conditional on policy C being available and voters' belief at that time being θ. We write s^j to denote $\left(s_l^j, s_h^j\right)$. When we omit the superscript, we are referring to a generic strategy. Obviously, we allow the voters and the parties to use strategies that condition on the entire observable history (and so cannot be summarized using the above notation), but we are searching for equilibria in which the equilibrium strategies can be summarized using this notation.

[15] Feasibility requires that there are no histories in which for some t, $c_t = n$ but $x_t = C$.

Notation 1 *Let $\sigma = (s^a, s^m, r, \mu)$ be an assessment in which the players use strategies that condition only on beliefs and μ is voters' belief at any information set that is consistent with (s^a, s^m, r); that is, μ is derived using Bayes' rule whenever possible. Let $U_i^j(P_j, \theta, \sigma)$ be the expected utility of a politician from party j in policymaking environment q_i when this politician faces voters' belief θ at the start of her term, implements P_j and expects that σ is played thereafter. Let $U_i^j(C, \theta, \sigma)$ be defined analogously. When confusion is not an issue, we omit σ from those functions.*

Notation 2 *Let $\theta^-(\theta, s) \equiv \Pr(q = q_h | \theta, P_j, s, j)$, that is, it is the posterior belief of the voter who holds prior θ, observes P_j, and believes that the current incumbent j plays a strategy s. Let $\theta^+(\theta, s) \equiv \Pr(q = q_h | \theta, C, s, j)$ be the corresponding posterior when she observes C.*

Lemma 1 *Consider an equilibrium with a strategy profile (s^a, s^m, r). For any θ,*

$$\theta^-(\theta, s) = \frac{\theta}{\theta + (1-\theta)\frac{1 - q_l s_l(\theta)}{1 - q_h s_h(\theta)}}, \tag{9}$$

and for all $\theta \in [0,1]$ such that $s_h(\theta) \neq 0$,

$$\theta^+(\theta, s) = \frac{\theta}{\theta + (1-\theta)\frac{q_l s_l(\theta)}{q_h s_h(\theta)}}. \tag{10}$$

Proof. This follows from Bayes' rule. ■

We characterize now the belief thresholds used in Propositions 1 and 3. From (3), (4) and (7), we obtain

$$\theta_* = q_l \frac{2 - (1+b) q_l - v}{(q_h - q_l)((b+1)(q_h + q_l) - 2 + v)}, \tag{11}$$

$$\theta^* = (1 - q_l) \frac{2 - (1+b) q_l - v}{(q_h - q_l)(2 - (q_h + q_l - 1)(b+1) - v)}, \tag{12}$$

$$\theta^{**} = \frac{2-(1+b)\,q_l - v}{q_h\,(2-(1+b)\,q_h - v) + (b+1)\,(q_h - q_l)}. \tag{13}$$

Proof of Proposition 1

Lemma 2 *Define*

$$b^{**}(v) \equiv \frac{(vq_h - 2q_l\,(1-q_h))\,(2-v)}{(q_l\,(vq_h - 2q_l\,(1-q_h)) + (q_h - q_l)\,(1-q_h)\,(2\,(1-q_l) - v))} - 1. \tag{14}$$

For all v satisfying Assumption 1.1, $b^{**}(v) > 1$. Moreover, for any fixed v satisfying Assumption 1.1, if $b < b^{**}(v)$, then $\bar{\theta} < \theta^{**}$.

Proof. Using (5) and (13), we obtain that $\bar{\theta} < \theta^{**}$ requires $b < b^{**}(v)$. Simple algebra shows that $b^{**}(v) > 1$ for all v satisfying Assumption 1.1. ■

Lemma 3 *Fix v, and assume that $b < b^{**}(v)$. Suppose that for some $\theta \in (\hat{\theta}, \theta^{**}]$, $s^m(\theta) = (1,1)$ and $s^a(\theta) = \hat{s}(\theta)$, where $\hat{s}$ is as defined in Definition 2. Then $r(\theta, v) = 1$ is sequentially rational for the voter.*

Proof. Consider a $t-$voter holding belief $\theta_t = \theta$ at the time of voting and observing $\nu_{t+1} = v$. Her expected payoff from electing party a is

$$\theta\,(bq_h s_h^a(\theta) + (1 - q_h s_h^a(\theta))) + (1-\theta)\,(bq_l s_l^a(\theta) + (1 - q_l s_l^a(\theta))), \tag{15}$$

and her expected payoff from electing party m is

$$\theta\,(bq_h s_h^m(\theta) - (1 - q_h s_h^m(\theta))) + (1-\theta)\,(bq_l s_l^m(\theta) - (1 - q_l s_l^m(\theta))) + v. \tag{16}$$

Subtracting (16) from (15) and using $s^m(\theta) = (1,1)$, we obtain

$$\Delta(\theta, s^a, s^m) \equiv (\theta q_h s_h^a(\theta) + (1-\theta)\,q_l s_l^a(\theta))\,(b-1) - (\theta q_h + (1-\theta)\,q_l)\,(b+1) + 2 - v. \tag{17}$$

Definition 2 requires that for all $\theta \in (\hat{\theta}, \theta^{**}]$, $\hat{s}$ satisfies

$$\theta^{-}(\theta, \hat{s}) = \frac{\theta}{\theta + (1-\theta)\frac{1-q_l \hat{s}_l(\theta)}{1-q_h \hat{s}_h(\theta)}} = \hat{\theta}, \tag{18}$$

and $\hat{s}_l(\theta) = 0$. This delivers

$$\hat{s}_h(\theta) \equiv \frac{v - 2 + (b+1)((1-\theta) q_l + \theta q_h)}{q_h \theta ((1+b) q_h + v - 2)}. \tag{19}$$

Note that $\hat{s}_h(\theta) \in [0,1]$ for all $\theta \in \left[\hat{\theta}, \theta^{**}\right]$. Plugging $\hat{s}(\theta)$ from (19) into (17), we obtain

$$\Delta(\theta, \hat{s}, s^m) = -(v - (1-q_h)(b+1)) \frac{v + q_l + bq_l - 2 + \theta(q_h - q_l)(b+1)}{v - 2 + (1+b) q_h}. \tag{20}$$

Using the formula for $\hat{\theta}$ from (2), we obtain that $\Delta\left(\hat{\theta}, \hat{s}, s^m\right) = 0$, and differentiating (20) with respect to θ, we obtain

$$\frac{d\Delta(\theta, \hat{s}, s^m)}{d\theta} = \frac{(1-q_h)(b+1) - v}{v - 2 + (1+b) q_h}(b+1)(q_h - q_l).$$

Note that $v - 2 + (1+b) q_h > v - 2(1-q_h) > 0$, where the last inequality comes from Assumption 1.1. Using $b < b^{**}(v)$ we obtain $(1-q_h)(b+1) - v < 0$. So if $b < b^{**}(v)$ then $\frac{d\Delta(\theta, \hat{s}, s^m)}{d\theta} < 0$, which implies $\Delta(\theta, \hat{s}, s^m) < 0$. Hence, in equilibrium, $r(\theta, v) = 1$ for all θ for which $s^a(\theta) = \hat{s}(\theta)$. ■

Proof of Proposition 1. Let (s^a, s^m, r) be the strategies described in Proposition 1. We will show that there exists a belief system μ that is derived using Bayes' rule whenever possible such that for each player k, the behavior prescribed by (s^a, s^m, r) is sequentially rational given the conjecture that the remaining players will follow (s^a, s^m, r) and the voters will update their beliefs according to μ.

For any p and any v satisfying Assumption 1.1 define

$$\bar{b}(v,p) \equiv \min\left\{ b^{**}(v), 1 + \frac{2p(1-q_h)}{1+q_h}, \frac{1+q_l}{1-q_l} \right\}. \tag{21}$$

Using Lemma 3, it follows that $\bar{b}(v,p) > 1$.

Proof of Part 1 and equation (8).

Consider an aligned incumbent at time t and let $\theta_t = \theta$. We show below that playing according to s^a is sequentially rational for this incumbent for any value of θ.

Proof of Part 1a: If $\theta = 0$, then playing sincerely is sequentially rational. For any $\theta \in (0, \theta_*)$, we have $\theta^+(\theta, s^a = (1,1)) < \hat{\theta}$ and hence the aligned party is elected with certainty regardless of which policy is implemented. Depending on θ, the posterior belief may fall into $(0, \theta_*)$—in which case the next period aligned politician plays sincerely—or into $[\theta_*, \hat{\theta})$—in which case the next period aligned politician plays $s^a = (0,0)$. So

$$U_i^a(C, \theta) \geq b + 1.$$

For any $\theta \in (0, \theta_*)$, $\theta^-(\theta, s^a = (1,1)) \in (0, \theta_*)$, so

$$U_i^a(P_a, \theta) = 1 + q_i b + (1 - q_i) < b + 1.$$

The last inequality implies that for all $\theta \in (0, \theta_*)$, $s^a(\theta) = (1,1)$ is sequentially rational.

Proof of Part 1b and equation (8): For any $\theta \in [\theta_*, \hat{\theta})$, we have $\theta^-(\theta, s^a = (0,0)) = \theta$. Since C is off the equilibrium path, we can set $\theta^+(\theta, s^a) > \hat{\theta}$. For any such belief,

$$U_i^a(C, \theta) \leq b + p(bq_i - (1 - q_i)) + (1 - p)(bq_i + (1 - q_i)),$$

$$U_i^a(P_a, \theta) = 1 + 1.$$

So

$$U_i^a(C,\theta) - U_i^a(P_a,\theta) = (1+q_i)b - (1+q_i) - 2p(1-q_i) < 0,$$

where the last inequality comes from Assumption 1.2.

Consider now $\theta = \hat{\theta}$. We have

$$U_i^a(P_a,\theta) = 1 + \left(1 - pr_{\hat{\theta}}\right) + pr_{\hat{\theta}}\left(q_i b - (1-q_i)\right). \tag{22}$$

Since C is off the equilibrium path, we can set $\theta^+(\theta, s^a) > \max\left\{\bar{\theta}, \theta^*\right\}$, so that

$$U_i^a(C,\theta) = b + p\left(bq_i - (1-q_i)\right) + (1-p)\left(bq_i + (1-q_i)\right). \tag{23}$$

So

$$U_i^a(C,\theta) - U_i^a(P_a,\theta) \leq 0$$

if and only if $r_{\hat{\theta}} \leq 1 - (b-1)\frac{1+q_i(1-p)}{p(2-(1+b)q_i)}$. Since the last expression decreases in q_i, we have that when $r_{\hat{\theta}}$ satisfies (8), $U_h^a(C,\theta) = U_h^a(P_a,\theta)$ and $U_l^a(C,\theta) < U_l^a(P_a,\theta)$, so $s^a(\hat{\theta}) = (0,0)$ is sequentially rational. Moreover, by Assumption 1.2, $r_{\hat{\theta}} \in [0,1]$.

Proof of Part 1c: By definition of $\hat{s}$, for any $\theta \in (\hat{\theta}, \max\left\{\bar{\theta}, \theta^*\right\}]$, we have $\theta^-(\theta,\hat{s}) = \hat{\theta}$ and $\theta^+(\theta,\hat{s}) = 1$. So $U_i^a(P_a,\theta)$ satisfies (22) and $U_i^a(C,\theta)$ satisfies (23). So by what precedes, $U_h^a(C,\theta) = U_h^a(P_a,\theta)$ and $U_l^a(C,\theta) < U_l^a(P_a,\theta)$ for all $\theta \in (\hat{\theta}, \max\left\{\bar{\theta}, \theta^*\right\}]$, and hence $\hat{s}$ is sequentially rational. By Lemma 2, when $b < b^{**}(v)$, then $\bar{\theta} < \theta^{**}$ and from (4) and (7), $\theta^* < \theta^{**}$, so $\max\left\{\bar{\theta}, \theta^*\right\} < \theta^{**}$, and hence $\hat{s}$ is well defined for any $\theta \in (\hat{\theta}, \max\left\{\bar{\theta}, \theta^*\right\}]$.

Proof of Part 1d: For all $\theta \in (\max\left\{\bar{\theta}, \theta^*\right\}, 1]$, both C and P_a lead to posterior belief greater than $\hat{\theta}$, so $U_i^a(C,\theta)$ satisfies (23) and $U_i^a(P_a,\theta) \leq 1 + p\left(bq_i - (1-q_i)\right) + (1-p)\left(bq_i + (1-q_i)\right)$, which implies that playing sincerely is sequentially rational.

Proof of Part 2.

Consider a misaligned incumbent at time t and let $\theta_t = \theta$. We show below that playing according to s^m is sequentially rational for this incumbent for any value of θ. For any $\theta > \theta_*$, playing C delivers a higher instantaneous payoff than P_m. Playing C also leads to a higher posterior than P_m, which means that party m is more likely to be elected after implementing C. Moreover, for any θ, party m delivers a higher payoff to party m politicians than party a. And finally, s^a is such that for any posterior greater than θ_*, a higher posterior implies higher probability that a politician of party a implements C. These four facts together imply that playing sincerely is sequentially rational for a misaligned incumbent for any $\theta > \theta_*$. By definition of θ_{**}, for any $\theta < \theta_{**}$, a sincere strategy leads to $\theta^+ (\theta, s^m) < \theta^*$, so regardless of which policy the misaligned incumbent implements when faced with $\theta < \theta_{**}$, she is replaced with an aligned replacement who then plays sincerely. Hence $s^m(\theta) = (1, 1)$ is sequentially rational. Consider now $\theta \in [\theta_{**}, \theta_*]$. Playing P_m results in a posterior for which the aligned replacement plays sincerely while playing C results in a posterior for which the aligned replacement plays $(0, 0)$. So

$$U_i^m (C, \theta) = b - 1,$$

$$U_i^m (P_m, \theta) = 1 + bq_i - (1 - q_i).$$

Hence

$$U_i^m (C, \theta) - U_i^m (P_m, \theta) = (1 - q_i) b - (1 + q_i) < 0,$$

where the last inequality comes from $b < \bar{b}(v, p)$ and (21). So indeed $s^m(\theta) = (0, 0)$ is sequentially rational for $\theta \in [\theta_{**}, \theta_*]$ if the off equilibrium belief is chosen to be $\theta^+ (\theta, s^m = (0, 0)) \in \left(\theta_*, \hat{\theta}\right)$.

Proof that the voters' cutoff strategy is sequentially rational.

By definition of $\hat{\theta}$, $r(\theta, v) = 0$ is sequentially rational for all $\theta < \hat{\theta}$. Since $s^a(\hat{\theta}) = (0, 0)$ and $s^a(\hat{\theta}) = (1, 1)$, by definition of $\hat{\theta}$ voters are indifferent between the parties at

$\hat{\theta}$, so $r(\hat{\theta}, v) = r_{\hat{\theta}}$ is sequentially rational for the voters. Sequential rationality of r for $\theta \in (\hat{\theta}, \max\{\bar{\theta}, \theta^*\}]$ comes from Lemma 3. For any $\theta > \max\{\bar{\theta}, \theta^*\}$, $s^m(\theta) = (1,1)$ and by definition of $\bar{\theta}$, it is sequentially rational for the voters to play $r(\theta, v) = 1$. ■

Proof of Proposition 2

Proof of Proposition 2. The comparative statics easily follows from differentiating (2), (5), (11) and (12) ■

A.1 Proof of Proposition 3

We prove the uniqueness of the equilibria described in Proposition 3 among the class of θ_c-cutoff equilibria in which additionally the voters update more positively after C than after P, that is, for any θ and any $j \in \{a, m\}$, $\theta^-(\theta, s^j) \leq \theta^+(\theta, s^j)$. In the proof below when we refer to *equilibrium* we mean an equilibrium with these restrictions.

Proof of Proposition 3. Define

$$\underline{b}(p) \equiv \frac{1 + q_h - 2pq_h}{1 - q_h}. \tag{24}$$

We first use steps 1–5 to prove that $s^m(\theta) = (1,1)$ for all $\theta \geq \theta_c$. This helps establish that $\theta_c = \hat{\theta}$. The rest of the proof then follows.

Step 1: In any equilibrium, for any θ and any $j \in \{a, m\}$, we have $\theta^-(\theta, s^j) < \theta^+(\theta, s^j)$.

Proof: Suppose there exists θ and j such that $\theta^-(\theta, s^j) = \theta^+(\theta, s^j)$. Then there are no electoral consequences of the policy choice, so obviously the best response of j is $s^j(\theta) = (1,1)$, which leads to $\theta^-(\theta, s^j) < \theta^+(\theta, s^j)$.

Step 2: In any equilibrium, for all $\theta \geq \theta_c$, $\theta_c \leq \theta^-(\theta, s^a) < \theta^+(\theta, s^a)$ and $s^a(\theta_c) = (0,0)$.

Proof: By contradiction, suppose that for some $\theta \geq \theta_c$, we have $\theta^- (\theta, s^a) < \theta_c < \theta^+ (\theta, s^a)$. Then using voters' election strategy (1), we obtain

$$\begin{aligned} U_i^a (C, \theta) &= b + p \left(bq_i s_i^m \left(\theta^+ (\theta, s^a) \right) - (1 - q_i) s_i^m \left(\theta^+ (\theta, s^a) \right) \right) \\ &+ (1 - p) \left(bq_i s_i^a \left(\theta^+ (\theta, s^a) \right) + \left(1 - q_i s_i^a \left(\theta^+ (\theta, s^a) \right) \right) \right) \\ &\leq b + p (bq_i - (1 - q_i)) + (1 - p) (bq_i + (1 - q_i)), \end{aligned}$$

$$U_i^a (P_a, \theta) \geq 1 + 1,$$

so

$$U_i^a (C, \theta) - U_i^a (P_a, \theta) \leq b (1 + q_i) - 2p - q_i + 2pq_i - 1 < 0,$$

where the last inequality comes from $b < \bar{b}(v, p)$ and (21) (actually, it follows from Assumption 1.2 directly). So $U_i^a (P_a, \theta) > U_i^a (C, \theta)$, which means that the aligned party should always play P_a. But that means that the voter does not update her belief after seeing P_a, hence $\theta^- (\theta, s^a) = \theta$, which contradicts the supposition that $\theta^- (\theta, s^a) < \theta_c$.

Consider now θ_c. By what preceds, it must be that $\theta^- (\theta_c, s^a) = \theta_c$, so s^a is such that voters do not update upon seeing P_a. This can arise only if either $q_l s_l^a (\theta_c) = q_h s_h^a (\theta_c) > 0$ or $s^a (\theta_c) = (0, 0)$. Under the former, either policy leads to the same belief, which implies that a plays sincerely at θ_c. But that contradicts $q_l s_l^a (\theta_c) = q_h s_h^a (\theta_c) > 0$, which establishes that $s^a (\theta_c) = (0, 0)$.

Step 3: $s^m (\theta_c) = (1, 1)$ or $s^m (\theta_c) = (0, 0)$.

Proof: In equilibrium, $\theta^- (\theta_c, s^m) \leq \theta_c < \theta^+ (\theta_c, s^m)$. Suppose first that $\theta^- (\theta_c, s^m) <$

$\theta_c < \theta^+(\theta_c, s^m)$. Then using voter's election strategy (1), we obtain

$$\begin{aligned}
U_i^m(C, \theta_c) &= b + p\left(bq_i s_i^m\left(\theta^+(\theta, s^m)\right) + 1 - q_i s_i^m\left(\theta^+(\theta_c, s^m)\right)\right) \\
&+ (1-p)\left(bq_i s_i^a\left(\theta^+(\theta_c, s^m)\right) - \left(1 - q_i s_i^a\left(\theta^+(\theta_c, s^m)\right)\right)\right) \geq \\
&\geq b + p - (1-p),
\end{aligned}$$

and

$$\begin{aligned}
U_i^m(P_m, \theta_c) &= 1 + \left(bq_i s_i^a\left(\theta^-(\theta_c, s^m)\right) - \left(1 - q_i s_i^a\left(\theta^-(\theta_c, s^m)\right)\right)\right) \\
&\leq 1 + bq_i - (1 - q_i).
\end{aligned}$$

So

$$U_i^m(C, \theta_c) - U_i^m(P_m, \theta_c) \geq (1 - q_i)\, b + (2p - q_i - 1),$$

where the last inequality follows from $b > \underline{b}(p)$ and equation (24).

Suppose then that $\theta^-(\theta_c, s^m) = \theta_c$. By the same argument as in Step 2 this implies $s^m(\theta_c) = (0, 0)$.

Step 4: $s^m(\theta_c) = (1, 1)$.

Proof: By Step 2, $s^a(\theta_c) = (0, 0)$, and by Step 3, $s^m(\theta_c) = (1, 1)$ or $s^m(\theta_c) = (0, 0)$. Suppose, by contradiction, that $s^m(\theta_c) = (0, 0)$. Then $\theta^-(\theta_c, s^m) = \theta_c$ and by Step 2, $\theta_c < \theta^+(\theta_c, s^m)$. Hence

$$\begin{aligned}
U_i^m(C, \theta_c) &= b + p\left(bq_i s_i^m\left(\theta^+(\theta_c, s^m)\right) + \left(1 - q_i s_i^m\left(\theta^+(\theta_c, s^m)\right)\right)\right) \\
&+ (1-p)\left(bq_i s_i^a\left(\theta^+(\theta_c, s^m)\right) - \left(1 - q_i s_i^a\left(\theta^+(\theta_c, s^m)\right)\right)\right) \\
&\geq b + p - (1-p),
\end{aligned}$$

and

$$\begin{aligned} U_i^m(P_m,\theta_c) &= 1-(1-pr(\theta_c,v))+pr(\theta_c,v) \\ &\leq 1-(1-p)+pr<U_i^m(C,\theta), \end{aligned}$$

which contradicts $s^m(\theta_c)=(0,0)$.

Step 5: $s^m(\theta)=(1,1)$ for all $\theta\geq\theta_c$.

Proof: Since in equilibrium $\theta_c\leq\theta<\theta^+(\theta,s^m)$ for all $\theta\geq\theta_c$, then for all $\theta\geq\theta_c$ we have

$$\begin{aligned} U_i^m(C,\theta) &= b+p\left(bq_is_i^m\left(\theta^+(\theta,s^m)\right)+1-q_is_i^m\left(\theta^+(\theta,s^m)\right)\right) \\ &+ (1-p)\left(bq_is_i^a\left(\theta^+(\theta,s^m)\right)-\left(1-q_is_i^a\left(\theta^+(\theta,s^m)\right)\right)\right) \\ &\geq b+p-(1-p). \end{aligned}$$

Suppose first $\theta^-(\theta,s^m)<\theta_c<\theta^+(\theta,s^m)$. Then following the same proof for the case of $\theta=\theta_c$ as in Step 3, we obtain $s^m(\theta)=(1,1)$.

Suppose then that $\theta_c<\theta^-(\theta,s^m)<\theta^+(\theta,s^m)$ for some $\theta\geq\theta_c$. Then

$$\begin{aligned} U_i^m(P_m,\theta) &= 1+p\left(bq_is_i^m\left(\theta^-(\theta,s^m)\right)+1-q_is_i^m\left(\theta^-(\theta,s^m)\right)\right) \\ &+ (1-p)\left(bq_is_i^a\left(\theta^-(\theta,s^m)\right)-\left(1-q_is_i^a\left(\theta^-(\theta,s^m)\right)\right)\right) \\ &\leq 1+p(bq_i+(1-q_i))+(1-p)(bq_i-(1-q_i)), \end{aligned}$$

and hence

$$U_i^m(C,\theta)-U_i^m(P_m,\theta)\geq(1-q_i)b+(-q_i+2pq_i-1)>0,$$

where the last inequality is satisfied if $b>\underline{b}(p)$.

Finally, suppose $\theta^{-}(\theta, s^{m}) = \theta_c$. This means

$$\begin{aligned} U_i^m(P_m, \theta) &= 1 + (1 - pr_c)\left(bq_i s_i^a(\theta_c) - (1 - q_i s_i^a(\theta_c))\right) \\ &\quad + pr_c\left(bq_i s_i^m(\theta_c) + (1 - q_i s_i^m(\theta_c))\right) \\ &= 1 - (1 - pr_c) + pr_c\left(bq_i + (1 - q_i)\right) \\ &\leq 1 - (1 - p) + p\left(bq_i + (1 - q_i)\right), \end{aligned}$$

so

$$U_i^m(C, \theta) - U_i^m(P_m, \theta) \geq (1 - pq_i)(b - 1) > 0,$$

and hence $s^m(\theta) = (1, 1)$.

Proof of part 1 that $\theta_c = \hat{\theta}$: Suppose not. Then by definition of $\hat{\theta}$, it must be that $\hat{\theta} < \theta_c$. Consider the behavior of a at θ_c. From Step 2, $\theta_c = \theta^{-}(\theta_c, s^a)$ and $s^a(\theta_c) = (0, 0)$, and from Step 4, $s^m(\theta_c) = (1, 1)$. From the definition of $\hat{\theta}$, we obtain that at $\theta_c > \hat{\theta}$, the voter strictly prefers the valence advantaged misaligned party expected to play $s^m(\theta_c) = (1, 1)$ to the misaligned party expected to play $s^a(\theta_c) = (0, 0)$. So playing P_a leads to:

$$U_i^a(P_a, \theta_c) = 1 + p\left(bq_i - (1 - q_i)\right) + (1 - p),$$

and playing C leads to

$$\begin{aligned} U_i^a(C, \theta_c) &= b + p\left(bq_i - (1 - q_i)\right) + (1 - p)\left(bq_i s_i^a\left(\theta^{+}(\theta, s^a)\right) + (1 - q_i) s_i^a\left(\theta^{+}(\theta, s^a)\right)\right) \geq \\ &\geq b + p\left(bq_i - (1 - q_i)\right) + (1 - p). \end{aligned}$$

So

$$U_i^a(C, \theta_c) - U_i^a(P_a, \theta_c) \geq b - 1 > 0,$$

which contradicts that $s^a(\theta_c) = (0, 0)$.

Proof of part 2 c-e and the formula for $r_{\hat{\theta}}$ from Part 1

Consider $\theta \in \left(\hat{\theta}, 1\right)$. Since $\theta_c = \hat{\theta}$, Step 2 implies that $\hat{\theta} \leq \theta^{-}(\theta, s^a) < \theta^{+}(\theta, s^a)$. Suppose first that for some θ, $\hat{\theta} < \theta^{-}(\theta, s^a) < \theta^{+}(\theta, s^a)$, so any policy leads to the election of the misaligned party when the later is valence advantaged. So using Step 5, we have

$$\begin{aligned}U_i^a(C, \theta) &= b + p(bq_i - (1-q_i)) + (1-p)\left(bq_i s_i^a\left(\theta^{+}(\theta, s^a)\right) + \left(1 - q_i s_i^a\left(\theta^{+}(\theta, s^a)\right)\right)\right) \\ &\geq b + p(bq_i - (1-q_i)) + (1-p),\end{aligned}$$

$$\begin{aligned}U_i^a(P_a, \theta) &= 1 + p(bq_i - (1-q_i)) + (1-p)\left(bq_i s_i^a\left(\theta^{-}(\theta, s^a)\right) + \left(1 - q_i s_i^a\left(\theta^{-}(\theta, s^a)\right)\right)\right) \\ &\leq 1 + p(bq_i - (1-q_i)) + (1-p)(bq_i + (1-q_i)).\end{aligned}$$

Hence

$$U_i^a(C, \theta) - U_i^a(P_a, \theta) \geq b - 1 + (1-p)(1 - (bq_i + (1-q_i))) > 0,$$

so it is sequentially rational to choose $s^a(\theta) = (1,1)$. By definition of θ^{**}, (13), for all $\theta > \theta^{**}$, any s^a leads to $\theta^{-}(\theta, s^a) > \hat{\theta}$, so the above implies that $s^a(\theta) = (1,1)$ is the only sequentially rational strategy for all $\theta > \theta^{**}$. Moreover, for any $\theta \in (\theta^*, \theta^{**}]$, $\theta^{-}(\theta, (1,1)) > \hat{\theta}$, so $s^a(\theta) = (1,1)$ is sequentially rational if voters conjecture $s^a(\theta) = (1,1)$. Since $s^m(\theta) = (1,1)$ (by Step 5), if $s^a(\theta) = (1,1)$ for some $\theta \in (\theta^*, \bar{\theta})$, then the voters prefer a at θ, which cannot be the case in a $\hat{\theta}-$cutoff equilibrium. So $s^a(\theta) = (1,1)$ for all $\theta > \max\left\{\theta^*, \bar{\theta}\right\}$ and is sequentially rational for $\theta \in \left(\theta^*, \min\left\{\theta^{**}, \bar{\theta}\right\}\right)$ and voters' conjecture $s^a(\theta) = (1,1)$, but is not sequentially rational for any $\theta < \left(\theta^*, \max\left\{\theta^*, \bar{\theta}\right\}\right)$.

For any $\theta \in \left(\hat{\theta}, \theta^*\right)$, $\theta^{-}(\theta, s^a = (1,1)) < \hat{\theta}$, which contradicts Step 2. So $s^a(\theta)$ must be such that $\hat{\theta} \leq \theta^{-}(\theta, s^a)$. By what precedes, if $\hat{\theta} < \theta^{-}(\theta, s^a = (1,1))$, then only $s^a(\theta) = (1,1)$ is sequentially rational, which contradicts $\hat{\theta} < \theta^{-}(\theta, s^a = (1,1))$. So it must be $\theta^{-}(\theta, s^a) = \hat{\theta}$ for all $\theta \in \left(\hat{\theta}, \theta^*\right)$. This means that $s^a(\theta)$ must satisfy (18), which requires that a randomizes at least in the favorable policymaking environment. Suppose first that a randomizes also in the unfavorable policymaking environment. Pick $\sup\left\{\theta > \hat{\theta} : s_l^a(\theta) \in (0,1) \text{ and } s_h^a(\theta) \in (0,1)\right\}$. Then by what precedes, $s^a\left(\theta^+\left(\theta, s^a(\theta)\right)\right) = (1,1)$. So

$$U_i^a(C,\theta) = b + p\left(bq_i - (1-q_i)\right) + (1-p)\left(bq_i + (1-q_i)\right), \tag{25}$$

$$U_i^a(P_a,\theta) = 1 + pr_{\hat{\theta}}\left(bq_i - (1-q_i)\right) + \left(1 - pr_{\hat{\theta}}\right)1.$$

So we have

$$U_i^a(C,\theta) - U_i^a(P_a,\theta) = \left(b + 2p - pr_\theta(b+1) - 1\right)q_i + \left(b - 2p + 2pr_\theta - 1\right), \tag{26}$$

which is strictly increasing in q_i. So if we choose $r_{\hat{\theta}}$ to make a_l indifferent between C and P_a, a_h strictly prefers C. So we have either $s_l^a(\theta) = 0$ or $s_h^a(\theta) = 1$. Consider θ just above $\hat{\theta}$. Then $\theta^-\left(\theta, \left(s_l^a(\theta), 1\right)\right) < \hat{\theta}$, so for such small θ, we must have $s_l^a(\theta) = 0$. But that means that $r_{\hat{\theta}}$ must be such that a is indifferent when $q = q_h$ (equation (26) satisfied with equality). That delivers (8) and by Assumption 1.2, we have $r_{\hat{\theta}} \in (0,1)$. Since (26) does not depend on θ, for such $r_{\hat{\theta}}$, when $q = q_h$ a is indifferent between P_a and C for all θ. So it must be that $s_l^a(\theta) = 0$ and $s_h^a(\theta) < 1$ for all $\theta \in \left(\hat{\theta}, \theta^*\right)$. Using this in (18) we obtain (19).

Proof of part 2 b: That $s^a\left(\hat{\theta}\right) = (0,0)$ comes from Step 2. Consider $\theta \in \left(\theta_*, \hat{\theta}\right)$. If $\theta^+(\theta, s^a) < \hat{\theta}$, then both P_a and C lead to the aligned party being elected. In that case, the aligned incumbent prefers to obtain the higher payoff from C today independent of how it expects a to play in the next period, so it is sequentially rational to play sincerely. But then by the definition of θ_*, $\theta^+(\theta, s^a = (1,1)) > \hat{\theta}$ for all $\theta > \theta^*$, which is a contradiction. So it must be that $\theta^+(\theta, s^a) \geq \hat{\theta}$ for all $\theta \in \left(\theta_*, \hat{\theta}\right)$. To have $\theta^+(\theta, s^a) = \hat{\theta}$, a would have to randomize in at least one policy environment, but generically $r_{\hat{\theta}}$ as defined in (8) does not make a indifferent between c and P_a in any q_i. So it must be that $\theta^+(\theta, s^a) > \hat{\theta}$. In that case, $U_i^a(C,\theta)$ is given by (25) and $U_i^a(P_a,\theta) \geq 1+1$, which

$$U_i^a(C,\theta) - U_i^a(P_a,\theta) \leq (1+q_i)\,b - (1+q_i+2p(1-q_i)) < 0,$$

where the last inequality comes from $b < \bar{b}(v,p) \leq 1 + \frac{2p(1-q_h)}{1+q_h}$. So a strictly prefers to play P_a, and hence $s^a(\theta) = (0,0)$ is the only sequentially optimal choice. It must be sustained by an off equilibrium belief $\theta^+(\theta, s^a) > \hat{\theta}$.[16]

Proof of part 2 a: that for all $\theta < \theta_*$, either $s^a(\theta) = (0,0)$ or $s^a(\theta) = (1,1)$.

That for all $\theta \in (0,\theta_*]$, $s^a(\theta) = (0,0)$ can be sustained as a part of an equilibrium using an off equilibrium belief $\theta^+(\theta, s^a) > \hat{\theta}$ follows from the argument in Step 4. Since $\theta^+(\theta,(1,1)) < \hat{\theta}$ for any $\theta < \theta_*$, clearly $s^a(\theta) = (1,1)$ can be sustained as well for all θ such that $s^a(\theta^+(\theta,(1,1))) = (1,1)$. Consider now θ for which $s^a(\theta^+(\theta,(1,1))) = (0,0)$. Then at such θ, we have

$$U_i^a(C,\theta) = b+1,$$

$$U_i^a(P_a,\theta) = 1 + bq_i + (1-q_i) < b+1,$$

so even for such θ, $s^a(\theta) = (1,1)$ is an equilibrium.

Proof that the set of parameters for uniqueness is nonempty.

We need to show now that there exists $p(v) < 1$ such that for all $p > p(v)$, $\underline{b}(p) < \bar{b}(v,p)$. Using (24) and (21), we get that $\underline{b}(p) < \frac{1+q_l}{1-q_l}$, and $\underline{b}(p) < \frac{1+q_h-2pq_h}{1-q_h}$ if

$$p > q_h \frac{1+q_h}{1-q_h+2q_h^2}. \tag{27}$$

And finally, $\underline{b}(p) < b^{**}(v)$ requires

$$\underline{b}(p) < b^{**}(v),$$

[16] For some parameters, an off equilibrium belief $\theta^+(\theta, s^a) = \hat{\theta}$ could also be part of the equilibrium. But since it is off-equilibrium and only works for a subset of the parameters, we ignore this possibility.

which translates into

$$p > \frac{(1-q_h)\, v^2 + 2\,(2q_h + 2q_l - q_h q_l - 2)\, v + 4\,(2q_l - 1)\,(q_h - 1)}{2\,(q_l\,(v q_h - 2q_l\,(1-q_h)) + (q_h - q_l)\,(1-q_h)\,(2\,(1-q_l) - v))}. \tag{28}$$

Set $\underline{p}\,(v)$ to be the maximum of the right-hand sides of (27) and (28). It remains to show that for any v satisfying Assumption 1.1 we have $\underline{p}\,(v) < 1$. But the right-hand side of (27) is always smaller than 1 and the the right-hand sides of (28) is smaller than 1 if

$$(1-q_h)\,(v - 2\,(1-q_h))\,(v - 2\,(1-q_l)) < 0,$$

which by Assumption 1.1 is always satisfied.

The fact that $\underline{b}\,(p=1) = 1$ follows directly from (24).

Proof of Part 3: This follows directly from Steps 4 and 5 and the fact that m is elected with positive probability only for $\theta \geq \hat{\theta}$.

Proof that in any equilibrium $r\,(\theta, \nu)$ must satisfy (1):

By definition of $\hat{\theta}$, $r\,(\theta, v) = 0$ is the only sequentially rational choice of the voters for all $\theta < \hat{\theta}$. Since $s^a(\hat{\theta}) = (0,0)$ and $s^a(\hat{\theta}) = (1,1)$, by definition of $\hat{\theta}$ voters are indifferent between the parties at $\hat{\theta}$, so $r(\hat{\theta}, v) = r_{\hat{\theta}}$ is sequentially rational for the voters and by what precedes, $r(\hat{\theta}, v)$ must equal $r_{\hat{\theta}}$ in any cutoff equilibrium. Sequential rationality of r for $\theta \in (\hat{\theta}, \max\{\bar{\theta}, \theta^*\}]$ comes from Lemma 3. For any $\theta > \max\{\bar{\theta}, \theta^*\}$, $s^m\,(\theta) = (1,1)$ and by definition of $\bar{\theta}$, it is sequentially rational for the voters to play $r\,(\theta, v) = 1$.

■

B Appendix: Two-Period Model

Consider a two-period version of the model with just one voter facing incumbent $j_1 = a_1$ and selecting incumbent j_2. The game ends after payoffs from the second period policy are realized.

Since there are no electoral incentives in the second period, any j_2 will implement common-interest policy if it is available. Given this, consider the behavior of the voter in the election at the end of the first period. If the voter's posterior belief after observing the first period's policy choice is θ_1, then the expected payoff from electing the aligned party is

$$(\theta_1 q_h + (1-\theta_1)q_l)b + (1 - (\theta_1 q_h + (1-\theta_1)q_l)),$$

and the expected payoff from electing the misaligned party is

$$(\theta_1 q_h + (1-\theta_1)q_l)b - (1 - (\theta_1 q_h + (1-\theta_1)q_l)) + \nu_1.$$

The voter prefers to retain the aligned party if

$$\theta_1 \leq \frac{1 - q_l - \frac{\nu_1}{2}}{q_h - q_l} \equiv \overline{\theta}(v_1).$$

Notice two intuitions. First, $\overline{\theta}(-v) > 1$, which means that when the misaligned candidate has valence disadvantage, the voter elects the aligned candidate for all θ_1. This is because in addition to valence advantage, the aligned party is better for the voter on policy grounds: in the second period, both parties will implement a partisan policy only when a common-interest policy is unavailable, and the voter agrees with the aligned party's partisan policies.

Second, Assumption 1.1 implies that $\overline{\theta}(v) \in (0, 1)$, so when the misaligned party is valence advantaged, the voter is willing to elect the aligned party only when θ_1 is sufficiently low. The voter prefers the aligned party over the misaligned party on policy grounds because of what the parties do when common-interest policy is not available. If common-interest policy is likely to be available, the misaligned candidate's deficit on partisan policy is more than compensated by its valence advantage. But if common-interest policy is unlikely to be available, the misaligned candidate's valence advantage cannot make up for the fact that this candidate will likely implement partisan policies that the voter dislikes.

Hence, the voter plays a *cutoff* strategy: for all beliefs smaller than $\overline{\theta}(v)$ the voter elects the aligned candidate, and for all beliefs larger than $\overline{\theta}(v)$, the voter elects the misaligned candidate if and only if the latter is valence advantaged. For beliefs exactly equal to $\overline{\theta}(v)$, the voter is indifferent between an aligned candidate and a valence advantaged misaligned candidate.

Since the voter only considers electing the misaligned candidate when she is valance advantaged, we simplify the rest of the equilibrium derivation in this two-period example by focusing on the case where the misaligned candidate is valence advantaged with probability 1. We also drop the argument in $\overline{\theta}(v)$ and simply write $\overline{\theta}$.

Let $s^a = (s^a_l, s^a_h)$ denote the equilibrium strategy of the first-period incumbent. To characterize this strategy, it will be useful to define the following critical beliefs, $\theta_*(\cdot)$ and $\theta^*(\cdot)$:

$$\Pr(q = q_h \mid \theta_*, C, s^a, j_1 = a_1) = \overline{\theta},$$

and

$$\Pr(q = q_h \mid \theta^*, P_a, s^a, j_1 = a_1) = \overline{\theta}.$$

That is, if $\theta_0 = \theta_*(s^a)$, then a voter who observes C and conjectures that the incumbent's strategy is s^a ends up with the posterior belief $\theta_1 = \overline{\theta}$. Analogously, if $\theta_0 = \theta^*(s^a)$, then a voter who observes P_a and conjectures that the incumbent's strategy is s^a ends up with the posterior belief $\theta_1 = \overline{\theta}$. Note that the policy choice is most informative at the strategy $s^a = (0, 1)$; that is $\theta^*(0, 1) = \max_{s^a} \theta^*(s^a)$.

We now characterize the behavior of the aligned incumbent in the first period as a function of the voter's prior beliefs.

Suppose first that $\theta_0 > \theta^*(0, 1)$; that is, the voter is quite optimistic about the policymaking environment. Then, regardless of the voter's conjecture about the incumbent's strategy, the voter's posterior belief following any policy choice is greater than $\overline{\theta}$. As such, no matter what the aligned incumbent does, the voter will elect the misaligned party. This implies that for any $\theta_0 > \theta^*(0,1)$, any choice of the aligned incumbent in the first period has the same electoral consequences. Given this, the aligned incumbent will implement common-interest policy when available; i.e., will use the strategy $s^a(\theta_0) = (1,1)$ for all $\theta_0 > \theta^*(0,1)$.

Now consider $\theta_0 \in (\overline{\theta}, \theta^*(1,1))$, so the voter is only moderately optimistic about the policymaking environment. Here electoral consequences are possible—in particular, if the voter becomes more pessimistic, the aligned incumbent can hope to retain office. We show that as a result, no pure strategy equilibrium is possible. To see this, consider two possible scenarios, illustrated in the two top pictures of Figure 2.

1. **Voter remains optimistic after** P_a: Suppose $s^a(\theta_0)$ is such that upon seeing P_a, the voter's posterior belief stays above $\overline{\theta}$. In that case, choosing P_a ensures the election of the misaligned candidate, which is the worst electoral consequence that the aligned incumbent can face. This implies that, whatever its electoral consequences, playing C delivers the highest payoff in either policymaking environment; hence, it must be that $s^a(\theta_0) = (1,1)$ for all $\theta_0 \in (\overline{\theta}, \theta^*(1,1))$. But by definition of $\theta^*(1,1)$, for any $\theta_0 \in (\overline{\theta}, \theta^*(1,1))$ that strategy for the aligned incumbent makes the voter pessimistic upon seeing P_a; P_a leads to a posterior $\theta_1 < \overline{\theta}$. This contradicts the hypothesis of this case, so it is not possible for there to be an equilibrium in which the voter remains optimistic after P_a when the voter's prior is $\theta_0 \in (\overline{\theta}, \theta^*(1,1))$.

2. **Voter becomes pessimistic after** P_a: Suppose $s^a(\theta_0)$ is such that upon seeing P_a, the voter's posterior belief falls below $\overline{\theta}$. In this case, Bayes' rule implies that the voter must become more optimistic upon seeing C. Hence, C leads to a posterior belief above the prior and hence above $\overline{\theta}$. So implementing C leads to the misaligned

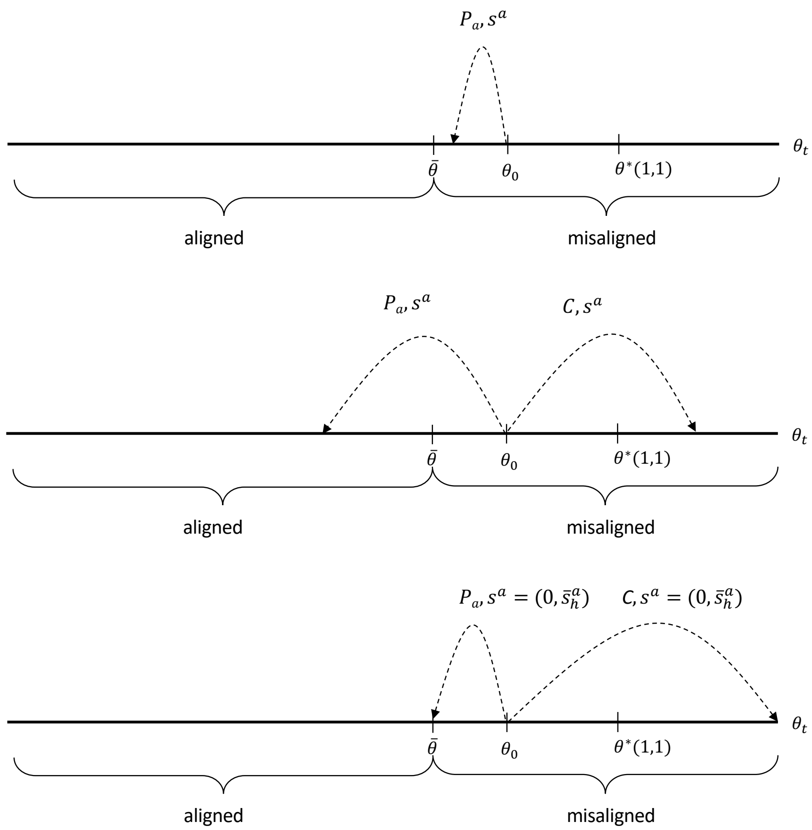

Figure 2: The dashed arrows represent what happens to posterior beliefs given the prior belief θ_0, the policy outcome and the conjectured aligned incumbent's strategy in the first period. The first two pictures illustrate two possibilities that cannot arise in equilibrium. The last picture shows what must happen in equilibrium.

candidate winning while implementing P_a lead to the aligned incumbent winning. By Assumption 1.2, the aligned incumbent strictly prefers the latter, which makes P_a the unique best response for the aligned incumbent, regardless of the policymaking environment. That is, the aligned incumbent must play the strategy $s^a(\theta_0) = (0,0)$ for any $\theta_0 \in (\overline{\theta}, \theta^*(1,1))$. But when $s^a(\theta_0) = (0,0)$, then P_a leaves the voter's prior unchanged, which contradicts the hypothesis of the case. So it is not possible for there to be an equilibrium in which the voter to remain optimistic after P_a when the voter's prior is $\theta_0 \in (\overline{\theta}, \theta^*(1,1))$.

The two cases above cannot occur in equilibrium because P_a leads to the incumbent's party losing power for sure in the first case and retaining power for sure in the second case. To sustain an equilibrium, P_a has to have less stark electoral consequences. But this requires the voter to randomize upon observing P_a which, in turn, requires voter indifference. The voter is only indifferent if her posterior belief, after observing P_a, is precisely $\overline{\theta}$. So in equilibrium, P_a must lead to the posterior belief $\overline{\theta}$ for every $\theta_0 \in (\overline{\theta}, \theta^*(1,1))$. For this to happen, P_a must be more informative than it is under the strategy $s^a = (0,0)$ but less informative than under the strategy $s^a = (0,1)$, which will require some randomization by the aligned incumbent.

What do these mixed strategies look like? Note that in both policymaking environments, the aligned incumbent expects the second-period incumbent to implement common-interest policies whenever they are available; hence, the downside of losing office is larger in the unfavorable policymaking environment. As a result, the incumbent is less willing to be replaced in an unfavorable policymaking environment. Therefore, if the incumbent is indifferent between C and P_a at θ_0 in the favorable policymaking environment, then she must strictly prefer playing P_a in the unfavorable policymaking environment. This means that, in a mixed strategy equilibrium, the aligned incumbent will always play partisan policy in the unfavorable environment and will mix in the favorable environment—i.e., $s^a_l(\theta_0) = 0$

and $s_h^a(\theta_0) \in (0,1)$ for all $\theta_0 \in (\overline{\theta}, \theta^*(1,1))$. The exact mixing probability that the aligned incumbent uses at each $\theta_0 \in (\overline{\theta}, \theta^*(1,1))$ is pinned down by Bayes' rule.

To derive the exact randomization, note that the continuation value of an aligned incumbent in policymaking environment q_i from having an aligned incumbent in office in period 2 is

$$V_i^a = q_i b + (1 - q_i),$$

and from having a misaligned incumbent in office in period 2 is

$$V_i^m = q_i b - (1 - q_i).$$

Let $\bar{r}$ be the probability with which the voter elects the misaligned party at posterior belief $\overline{\theta}$. Then, in policymaking environment q_i, the expected payoff to the aligned party from P_a is

$$1 + (1 - \bar{r})V_i^a + \bar{r}V_i^m,$$

and the expected payoff to the aligned party from C is:

$$b + V_i^m.$$

The incremental return from P_a is the difference between the two, that is,

$$1 - b + (1 - \bar{r})\left(V_i^a - V_i^m\right)) = 1 - b + 2(1 - \bar{r})(1 - q_i),$$

which is strictly decreasing in q_i.

For this randomization to be a best response, the aligned party must be indifferent between P_a and C in the favorable policymaking environment q_h. This requires that the incremental return be zero:

$$1 - b + 2(1-\bar{r})(1-q_h) = 0 \iff \bar{r} = 1 - \frac{b-1}{2(1-q_h)}.$$

And, in order for the randomization to leave the voter with posterior belief $\overline{\theta}$, we need $s^a(\theta_0) = \bar{s}^a(\theta_0)$ where $\bar{s}^a_l(\theta_0) = 0$ and $\bar{s}^a_h(\theta_0)$ satisfies:

$$\frac{(q_h(1-\bar{s}_h(\theta_0)) + 1 - q_h)\,\theta_0}{(q_h(1-\bar{s}_h(\theta_0)) + 1 - q_h)\,\theta_0 + 1 - \theta_0} = \overline{\theta} \iff \bar{s}_h(\theta_0) = \frac{\theta_0 - \overline{\theta}}{q_h\theta_0(1-\overline{\theta})}.$$

Consider now $\theta_0 \in (\theta^*(1,1), \theta^*(0,1))$. The discussion above implies that two strategies can be supported in equilibrium. First, if the incumbent plays sincerely, i.e., $s^a(\theta_0) = (1,1)$, then both policies lead to the posterior belief above $\overline{\theta}$, and the voter's plan to vote for the misaligned candidate for all beliefs above $\overline{\theta}$ makes this strategy sequentially rational. Second, if the incumbent plays $s^a(\theta_0) = (0, \bar{s}^a_h(\theta_0))$, then P_a leads to a posterior belief equal to $\overline{\theta}$, and again the voter's randomization at $\overline{\theta}$ derived above makes this strategy sequentially rational.

Now turn to prior beliefs below $\overline{\theta}$.

First, consider $\theta_0 \in (\theta_*(1,1), \overline{\theta}]$. For the same reason discussed above, the aligned party's strategy cannot, in equilibrium, be such that C leads to a belief above $\overline{\theta}$ and P_a leads to a belief below $\overline{\theta}$. There also cannot be a mixed strategy equilibrium in which C results in the belief $\overline{\theta}$. To sustain randomization by the aligned party, the voter must randomize at $\overline{\theta}$ to make the aligned incumbent indifferent starting at priors below $\overline{\theta}$. But the voter is already randomizing at $\overline{\theta}$ to make the aligned party indifferent starting at prior beliefs above $\overline{\theta}$. It is, however, an equilibrium for the aligned incumbent to always choose P_a— i.e., play $s^a(\theta_0) = (0,0)$. In this case, playing P_a does not change the voter's prior, and she votes for the aligned candidate. Playing C is off the equilibrium path, but under the natural assumption that the voter updates to $\theta_1 > \overline{\theta}$ after observing C, playing C leads

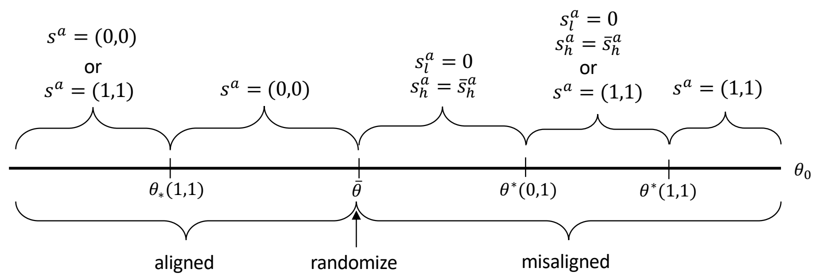

Figure 3: Equilibrium in the two-period model as a function of the voter's prior θ_0 assuming the misaligned candidate is valence advantaged.

the voter to elect the misaligned candidate. Thus, under this conjecture, playing partisan policy is a best response for the aligned party in both policymaking environments. Thus, in equilibrium the aligned party always plays partisan policy, the voter learns nothing, and because the voter is pessimistic ex ante, the voter retains the aligned party.

Finally, consider $\theta_0 < \theta_*(1,1)$. For the same reasons as above, $s^a(\theta_0) = (0,0)$ is an equilibrium here. However, so is the sincere strategy $s^a(\theta_0) = (1,1)$. If the aligned party plays this sincere strategy, then even when the outcome is C the voter's posterior is below $\overline{\theta}$ and the voter retains the aligned party. Thus, under this conjecture, there are no electoral consequences, and playing sincerely is sequentially rational. For the same reasons as above, no other strategy is consistent with equilibrium for $\theta_0 < \theta_*(1,1)$.

Figure 3 summarizes equilibrium in the two-period model as a function of the voter's prior, θ_0. The aligned party's first period strategy is described above the line and the voter's strategy is described below the line.

Consistent with the main model, we have assumed in this section that the aligned party is in power in the first period. Since in the main model the misaligned party may be in power in some periods, it is instructive to consider the case in which the misaligned party is in power in the first period. It is easy to see that playing sincerely is an equilibrium strategy. Under that strategy, C delivers a higher immediate payoff than P_m and it leads to more optimistic beliefs, which can only be beneficial for the misaligned party's electoral survival.

This two-period model delivers an interesting equilibrium structure: the voter plays a cutoff strategy, electing the valence-advantaged misaligned candidate only if sufficiently optimistic. Very optimistic beliefs lead to sincere behavior while the aligned party forgoes some common-interest policies when voter's beliefs are intermediate. In particular, for $\theta_0 \in (\theta_*, \bar{\theta})$, only partisan policies are implemented.

Even though the structure of the equilibria in the infinite-horizon model resembles the equilibria of the two-period model, the analysis above reveals that ex-ante, there were reasons to believe this may not be the case, and hence there is value in solving the infinite-horizon model. In the derivation of the first-period behavior in the two-period model above, we rely heavily on the fact that the last-period incumbent plays sincerely independent of her party affiliation and policymaking environment. It is the anticipation of such sincere play that delivers the cutoff structure of the first-period play. But since in the first period, the aligned party does not play sincerely, one cannot immediately conjecture that the particular cutoff structure would be preserved if there were a preceding period.

References

Abramowitz, Alan I. 2010. *The Disappearing Center: Engaged Citizens, Polarization, and American Democracy.* Yale University Press.

Abramowitz, Alan I and Kyle L Saunders. 2008. "Is polarization a myth?" *The Journal of Politics* 70(2):542–555.

Abramowitz, Alan I and Steven Webster. 2016. "The rise of negative partisanship and the nationalization of US elections in the 21st century." *Electoral Studies* 41:12–22.

Acemoglu, Daron, Georgy Egorov and Konstantin Sonin. 2013. "A Political Theory of Populism." *Quarterly Journal of Economics* 128(2):771–805.

Alesina, Alberto, Armando Miano and Stefanie Stantcheva. 2020. The polarization of reality. In *AEA Papers and Proceedings.* Vol. 110 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 pp. 324–328.

Ali, S Nageeb, Maximilian Mihm and Lucas Siga. 2018. "Adverse Selection in Distributive Politics." *Available at SSRN 3579095* .

Andrews Fearon, Patricia. 2023. Zero-Sum Mindset & Its Discontents PhD thesis University of Cambridge.

Ansolabehere, Stephen and James M Snyder Jr. 2000. "Valence Politics and Equilibrium in Spatial Election Models." *Public choice* 103(3-4):327–336.

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.

Ash, Elliott, Massimo Morelli and Richard Van Weelden. 2017. "Elections and divisiveness: Theory and evidence." *The Journal of Politics* 79(4):1268–1285.

Austen-Smith, David, Wioletta Dziuda, Bård Harstad and Antoine Loeper. 2019. "Gridlock and inefficient policy instruments." *Theoretical Economics* 14(4):1483–1534.

Bils, Peter and Federica Izzo. 2023. "Ideological Infection.".

Buisseret, Peter and Richard Van Weelden. 2022. "Pandora's Ballot Box: Electoral Politics of Direct Democracy." *arXiv preprint arXiv:2208.05535* .

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

Canes-Wrone, Brandice, Michael C. Herron and Kenneth W. Shotts. 2001. "Leadership and Pandering: A Theory of Executive Policymaking." *American Journal of Political Science* 45:532–550.

Chinoy, Sahil, Nathan Nunn, Sandra Sequeira and Stefanie Stantcheva. 2023. "Zero-sum thinking and the roots of US political divides." *Preprint at https://scholar. harvard. edu/files/stantcheva/files/zero_sum_us_political_divides. pdf* .

Curry, James M and Frances E Lee. 2020. *The limits of party: Congress and lawmaking in a polarized era.* University of Chicago Press.

Dziuda, Wioletta and Antoine Loeper. 2016. "Dynamic collective choice with endogenous status quo." *Journal of Political Economy* 124(4):1148–1186.

Dziuda, Wioletta and Antoine Loeper. 2018. "Dynamic pivotal politics." *American Political Science Review* 112(3):580–601.

Dziuda, Wioletta and Antoine Loeper. 2023. "Voters and the trade-off between policy stability and responsiveness." *Available at SSRN 3908921* .

Fernandez, Raquel and Dani Rodrik. 1991. "Resistance to reform: Status quo bias in the presence of individual-specific uncertainty." *The American Economic Review* pp. 1146–1155.

Fiorina, Morris, Samuel J Abrams and Jeremy C Pope. 2011. "Culture War? The Myth of a Polarized America.".

Fowler, Anthony. 2016. "What Explains Incumbent Success? Disentangling Selection on Party, Selection on Candidate Characteristics, and Office-Holding Benefits." *Quarterly Journal of Political Science* 11(3):313–338.

Fowler, Anthony, Seth J Hill, Jeffrey B Lewis, Chris Tausanovitch, Lynn Vavreck and Christopher Warshaw. 2023. "Moderates." *American Political Science Review* 117(2):643–660.

Fudenberg, Drew, David K. Levine and Jean Tirole. 1985. Infinite-horizon Models of Bargaining with One-Sided Incomplete Information. In *Game-Theoretic Models of Bargaining*, ed. Al Roth. West Nyack, NY: Cambridge University Press chapter 5.

Gentzkow, Matthew, Jesse M. Shapiro and Matt Taddy. 2019. "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech." *Econometrica* 87(4):1307–1340.

Groseclose, Tim. 2001. "A Model of Candidate Location when One Candidate has a Valence Advantage." *American Journal of Political Science* pp. 862–886.

Handan-Nader, Cassandra, Andrew C. W. Myers and Andrew B. Hall. 2023. "Polarization and State Legislative Elections." Stanford University Typescript.

Hill, Seth J. and Chris Tausanovitch. 2015. "A Disconnect in Representation? Comparison of Trends in Congressional and Public Polarization." *Journal of Politics* 77(4):1058–1075.

Hinchliffe, Kelsey L. and Frances E. Lee. 2016. "Party Competition and Conflict in State Legislatures." *State Politics & Policy Quarterly* 16(2):172–197.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra and Sean J. Westwood. 2019. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22(1):129–146.

Jacobson, Gary C. 1996. "The 1994 House Elections in Perspective." *Political Science Quarterly* 111(2):203–223.

Katznelson, Ira, Kim Geiger and Daniel Kryder. 1993. "Limiting Liberalism: The Southern Veto in Congress, 1933–1950." *Political Science Quarterly* 108(2):283–306.

Kreps, David M and Robert Wilson. 1982. "Reputation and imperfect information." *Journal of Economic Theory* 27(2):253–279.

Lee, Frances E. 2015. "How Party Polarization Affects Governance." *Annual Review of Political Science* 18:261–282.

Levy, Gilat and Ronny Razin. 2022. "Political Social-Learning: Short Term Memory and Cycles of Polarisation." CEPR Discussion Paper #16737.

Maskin, Eric and Jean Tirole. 2001. "Markov perfect equilibrium: I. Observable actions." *Journal of Economic Theory* 100(2):191–219.

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

Matsusaka, John G. and Chad Kendall. 2022. "The Common Good and Voter Polarization." USC typescript.

McCarty, Nolan. 2019. *Polarization: What Everyone Needs to Know.* Oxford University Press.

Orr, Lilla V., Anthony Fowler and Gregory A. Huber. Forthcoming. "Is Affective Polarization Driven by Identity, Loyalty, or Substance?" *American Journal of Political Science* .

Pew Research Center. 2021. Amid a Series of Mass Shootings in the U.S., Gun Policy Remains Deeply Divisive. Technical report.

Schickler, Eric and Devin Caughey. 2011. "Public Opinion, Organized Labor, and the Limits of New Deal Liberalism, 1936–1945." *Studies in American Political Development* 25(2):162–189.

Schofield, Norman. 2007. "The Mean Voter Theorem: Necessary and Sufficient Conditions for Convergent Equilibrium." *Review of Economic Studies* 74(3):965–980.

Stokes, Donald E. 1963. "Spatial Models of Party Competition." *American Political Science Review* 57:368–377.

Strulovici, Bruno. 2010. "Learning While Voting: Determinants of Collective Experimentation." *Econometrica* 78(3):933–971.