Ideological Competition

Federica Izzo*   Gregory J. Martin†   Steven Callander‡

August 17, 2022

Abstract

We propose a model of political competition not over policy programs, but over ideologies: models of the world that organize voters' experiences and guide the inferences they draw from observed outcomes. Policy-motivated political parties develop ideologies, and voters choose the ideology that best explains their observations. Preferences over policies are then induced by the adopted ideology. Parties thus care about winning the ideological battle as it confers an advantage in the electoral arena. We show that in equilibrium political parties always propose different models of the world. This divergence extends to all features of the environment, not just policy dimensions. A lower degree of policy extremism in the past increases the divergence on the policy dimension, thus leading to higher ideological polarization.

---

*University of California, San Diego. Email: `fizzo@ucsd.edu`.
†Stanford Graduate School of Business. Email: `gjmartin@stanford.edu`.
‡Stanford Graduate School of Business. Email: `sjc@stanford.edu`.

Introduction

How do citizens form policy preferences? The question has been a central one in political science since the discipline's behavioral turn in the mid-20th century. Indeed, democratic citizenship confronts the mass public with a difficult problem: how to evaluate policy alternatives with which they may lack familiarity, experience, or basic understanding.

A half-century of research effort has produced a rich and varied set of answers. Scholars have variously argued that citizens have policy preferences that are temporally unstable, if they can be said to exist at all (Converse, 1964), formed on the spot when prompted (Zaller, 1992), formed by following elite cues (Brady and Sniderman, 1985; Lenz, 2013), or considered in reaction when politics intrudes into life unbidden (Klar and Krupnikov, 2016).

The classical formal treatments of political competition under complete information (Downs, 1957) on the other hand, sidestep the question of the origin of policy preferences, instead taking them as exogenous and fixed model primitives. This approach has proved fruitful for deriving predictions about the relationship of outcomes to the distribution of preferences in the population. But it considerably narrows the scope of political inquiry, and renders opaque much of the day-to-day work that political practitioners invest in campaigning, organizing, and honing rhetorical arguments.

In this paper we aim at bridging this gap by offering a formal model of policy preference formation embedded in a political competition framework. Our model is a hierarchical one, in which citizens first choose ideologies — systems of belief about the underlying process that generates social outcomes — and then apply those ideologies to decide which policy inputs they prefer. Ideology thus provides a narrative that allows citizens to "navigate and orient themselves in the sea of politics" (Sartori, 1969). By providing a theory of how the world works, ideologies enable citizens to form expectations about the consequences of the various policy choices, which in turn inform their policy preferences. By their nature, such ideologically *induced* preferences are contingent and flexible, and subject to the kinds of temporal shifts and responses to elite signals that the behaviorists have documented.

This conception of ideology as distinct from and prior to policy preferences delivers a very different view of the nature of political competition among party leaders, which contrasts sharply

with canonical formal treatments. Where the traditional approach views parties as strategically choosing platforms to assemble a winning electoral coalition, our model highlights the role of parties in developing political narratives, with the aim of manipulating and directing the masses.[1] Importantly, the parties' ability to persuade voters does not come from an information asymmetry, but rather from their role as ideological entrepreneurs: parties have monopoly power over the development of ideologies.

Our model thus reverses the logic of standard formal models of elections. In our world, the parties' strategic problem is not solely to choose a policy platform that appeals to exogenous citizen preferences. Rather, political parties try to generate a favorable electoral environment by inducing citizens to adopt an ideology that translates into preferences aligned with the party's own policy program. The parties compete over ideology, not because they care about ideology per se but because ideology shapes the subsequent policy competition. A party that can win the ideological battle, so to speak, attains a favorable position in the policy competition. Ideology will thus be an equilibrium outcome of our model, rather than one of its primitives.

OPERATIONALIZING POLITICAL IDEOLOGY

Political ideology is a concept that is at once familiar yet difficult to pin down. As Sartori (1969) drily notes, "the growing popularity of the term has been matched, if anything, by its growing obscurity." Gerring (1997, p. 980) contends that the element common to all definitions of ideology is coherence, that a citizen's values, her understanding of the world, and the policy preferences she holds are "bound together". Narrower definitions of ideology take values as fixed and focus on how ideology translates those values to policy preferences (Freeden, 2001), whereas broader definitions encompass the construction of values as well (Knight, 2006).

In this paper we operationalize ideology in the narrower sense. We take a citizen's objective—and, thus, her values—as given and think about ideology as translating values into policy prefer-

---

[1]This is akin to what Gramsci (1929) refers to as cultural hegemony: the elite construct ideological world-views to impose their belief systems to the people, and thus achieve and maintain control even without the use of coercive means.

ences.[2] At this level, an ideology is a system of beliefs about the underlying process that generates social outcomes and that citizens use to map their values into concrete policy preferences.

Politicians in practice devote considerable effort to shaping ideological beliefs. These efforts manifest as arguments for how the world works and why it looks the way it does, and are often devoid of specific policy content. Illustrative examples come from the political rhetoric of Margaret Thatcher on one side of the political spectrum, and Bernie Sanders on the other. In 1987, then-Prime Minister and leader of the UK Conservative party Margaret Thatcher famously stated her disbelief in the existence of society: "[Too many people] are casting their problems on society and who is society? There is no such thing! There are individual men and women and there are families." The statement contains no reference to any particular policy or element of the Tory platform. It instead invites the audience to adopt a particular model of the causal forces that underlie the variation in outcomes they observe in the world: if some are rich, and others poor, it must be because of immutable differences in individual abilities or talents and not societal factors amenable to democratic control.

Almost thirty years later, Sanders' rhetoric offered a moral vision almost a mirror image of the one offered by Thatcher: "I believe... we are in this together. These are not just words. The truth is on some level when you hurt, when your children hurt, I hurt. And when my kids hurt, you hurt." The statement makes what amounts to an empirical claim: that our well-being (or the absence of "hurt," in Sanders' formulation) is not driven solely by our own income or consumption but by the well-being of everyone we interact with. The Vermont Senator thus offers a theory of the world but, again, no specific policy prescription.

Sophisticated observers often lament that policy-free statements like these reflect a superficial or dumbed-down approach to politics. By taking seriously the role of ideology, we argue that much of this rhetoric is, in fact, highly policy relevant. The policy stakes of such rhetoric can be extremely stark: consider Bonilla-Silva's (1997) account of racial ideology, which argues that it is only after society accepts the placement of people into racial categories that racist policy

---

[2]We therefore set aside, at least for the moment, questions of where those values originate from and how elites can shape values. We take up these ideas briefly in the concluding discussion.

preferences can emerge.[3] By shaping how citizens view the world, politicians can indirectly shape their policy preferences, and do so more persuasively than by appealing to policy directly.

MODELING IDEOLOGICAL COMPETITION

Building on this operationalization of ideology, our contribution is to understand what ideologies and policies emerge from political competition. Our model has three essential elements. First, we conceive of political parties as ideological entrepreneurs, creating and competing on ideology in addition to the traditional competition over policy. Specifically, each party develops and offers to the public a theory of the world. The goal is to have the public accept a party's ideology, to adopt its system of beliefs and view of the world and, thus, to shape the policy preferences that the public forms. Winning the ideological competition puts a party in a favorable position for the electoral competition stage that follows.

The second essential element of our model is that citizens are initially "ideologically innocent" (Kinder and Kalmoe, 2017), in the sense that they have no ideology of their own, and cannot or will not do the work of developing a theory of the world. In other words, parties have monopoly power over the construction of new ideologies. Citizens can evaluate an ideology if one is offered to them, but they cannot construct their own.

While this technological advantage gives parties the chance to influence citizens' (induced) preferences, this power is not unconstrained. The third key element of our model is a common-knowledge history of outcomes observed in the past, against which citizens will test the plausibility of parties' models. Our citizens adopt the model of the world that best resonates with their own lived experience and the past experience they absorb from the media and their schooling. This shared historical memory constrains and disciplines parties' ability to persuade.

OVERVIEW OF RESULTS

In this setting, we show that ideological competition matters for political outcomes. In equilibrium, the two competing parties offer distinct ideologies, with each offering a view of the world

---

[3]Bonilla-Silva (1997) argues that "The placement of groups of people in racial categories stemmed initially from the interests of powerful actors in the social system" (p. 473), and that "Categories such as 'Indians' and 'Negroes' were invented in the sixteenth and seventeenth centuries to justify the conquest and exploitation of various peoples." (p. 471)

in which its preferred policies are more attractive to voters. In particular, we show that the true model of the world never emerges in equilibrium. Although the parties are constrained by history, and voters evaluate ideologies with open eyes, the finiteness of history provides freedom to the parties to offer an explanation of the world that is more favorable to them. The limitations of history provide the space for the emergence of radically different ideologies, such as Margaret Thatcher's individualism or Bernie Sanders' collectivistic view.

Thus, political ideologies are polarized in equilibrium just as are policy preferences. We show that the difference between the ideologies offered in equilibrium is a function of the history of policy choices: ideological polarization emerges from the narrowness of policy experience. The intuition is that historical data generated under moderate policies are less informative, and this allows parties to persuade voters with more extreme ideologies farther from the best fitting model in the data. Our model shows, therefore, that it is exactly following periods of moderation in policy that we should expect political parties to espouse radically different ideologies.

Further, we show that ideologies are not limited to policy variables. Acting as ideological entrepreneurs, parties will include non-policy variables, over which they have no control or preferences, into their ideologies. They do so in order to shift blame for policy failures onto unrelated factors. This force is strong enough that in equilibrium, ideologies are always total: *all* available dimensions, including those that are not inherently political, are drawn in to parties' models and become sites of ideological conflict. Hence, even if the policy choice can be described by a single dimension, the political domain is inherently multidimensional. This result resonates with work on affective polarization, such as Mason's (2018) study of the politicization of identity characteristics with no inherent connection to policy disagreements.

Our results illuminate the logic behind such social polarization, and allow us to make predictions about which identity attributes are likely to become woven into ideological disagreements, and when. We show that non-policy variables are incorporated in the parties' political narratives when their historical correlation with the policy dimension is sufficiently high. This allows parties to use the non-policy variables to improve the overall fit of their model, and thus offer even more favorable—and extreme—ideologies on the policy dimension.

In our model parties have no control over non-policy dimensions. Thus, if a strong enough historical correlation with the policy variable emerges, it is purely by chance. The results thus

6

highlight the importance of path dependence and random events in the genesis of political ideologies. Political parties, acting as ideological entrepreneurs, will exploit emergent correlations to construct complex models of the world that support their favored policy position. This process generates a relationship between complexity and ideological polarization: the higher the number of politically salient dimensions (i.e., dimensions included in the proposed model of the world by at least one of the parties), the higher is polarization on the policy dimension in equilibrium.

PERSPECTIVES ON OUR APPROACH

In the formal theory literature, ideology and policy preferences are typically collapsed into a single characteristic. This is not a requirement of formal theory per se, but of the often embedded assumption that voters are fully rational information processors and decision makers. Under this assumption, all information from history, from voters' experience, from the various political actors, and so on can be processed and collapsed into a final policy preference. We argue that ideology plays a distinct role in this process.

Our approach accords with the foundational work of Simon (1955) and the behavioralist tradition. That literature has emphasized the difficulty voters have in understanding politics and forming preferences over political objects. The mapping between policies and outcomes is fundamentally uncertain (Callander, 2011), and even voters who are well-informed about policies that have been tried in the past face a challenge of using this information to develop preferences about what policies to implement the future.

Concepualizing ideologies as worldviews, our work connects with a recent literature in behavioral economics that explores the role of models in decision making.[4] In this literature, economic agents, as voters do in our model, use data to select a model through which to view the world. The behavioral limitations of agents in these models differ from those that affect voters in our model and we also differ in our focus on the competition among elites in shaping how citizens view the world (see, for example, Mullainathan et al. (2008), Ortoleva (2012), and Schwartzstein and Sunderam (2021)).

---

[4]The notion of ideology as shaping voters' interpretation of the observed reality distinguishes our work from Hafer and Landa (2007). Hafer and Landa (2007) also see ideology and beliefs as closely connected, but conceptualize ideology as a person's innate propensity to be persuaded by a left-wing or a right-wing argument.

We are also distinct from the strand of that literature that connects to political economy. Eliaz and Spiegler (2020) focus on causal models that form narratives about politics. Benabou and Tirole (2006) explore citizens' view of how just the world is, developing a model of motivated reasoning in which voters distort or ignore information that doesn't fit with their preferred view of the world. Closest to our work is Levy et al. (2020), but there the distinction is between people with either complex or simple views of the world and not between elites and the masses.

Indeed, we follow the strand of the behavioral literature that emphasizes the role leaders and elites play in helping voters make sense of the political world, providing a specific channel—the formation of ideologies—through which this occurs. Our formulation reflects the power of elites while also acknowledging its limits. Ample evidence exists that citizens process information and events themselves and that while they are influenced by elites, perhaps heavily, they do not do so blindly (Bullock, 2011). Voters are led, to be sure, but only willingly and with open eyes.

The Baseline One-Dimensional Model

The model describes a one-period, two-stage game. The players are two policy-motivated parties (R and L), and a representative voter. At the beginning of the game, the players observe a public record of T policies (z) and associated outcomes (y), indexed by $t \in \{-T, ..., -1\}$. As the notation suggests, we think of these as a history of past policy-outcome pairs, but this interpretation is not essential; the record might instead represent cross-sectional information from policies and outcomes in different countries or states. In addition, the voter privately observes the outcome of the (exogenous) policy that is in place in the first stage (z_0).

Voters in the model have no intrinsic preferences over z. Instead, they care exclusively about outcomes. Preferences over policies will then be fully determined by voters' beliefs about the relationship of policy to outcome.

We model the data generating process in a simple way: outcomes y_t are a linear function of the chosen policy z_t and a noise term ε_t.[5]

---

[5]The policy dimension here is abstract and need not map one-to-one onto a real-world policy instrument. The important feature for our purposes is that it captures a dimension on which the parties disagree and over which they will have control once in power.

$$y_t = \beta_0 z_t + \varepsilon_t \tag{1}$$

Where ε_t is i.i.d. from a normal distribution with mean 0 and variance σ^2. β_0 here is a scalar representing the (expected) change in outcomes resulting from a one-unit change in the policy variable z. While the baseline model is one-dimensional, i.e., assumes that outcomes are only a function of a single (policy) variable, an extension presented below generalizes equation (1) to include a (possibly multi-dimensional) set of non-policy covariates.

The ideological contest. The first stage of the model is the ideological contest. The two parties move simultaneously, each proposing an ideology, i.e., a model of the world. In this one-dimensional setup, a model of the world is simply a scalar indicating the slope of the policy-mapping function. We denote the ideology proposed by party i as β_i. The voter faces some uncertainty about the data generating process underlying outcomes. Formally, we assume that the voter does not know the slope of the mapping function, i.e., the coefficient β_0.

We make two restrictions on the way that voters can use information from the past history in evaluating the parties' proposals. First, we do not specify the voter's complete beliefs over the true model of the world (i.e., the true β_0). Rather than the usual assumption of Bayesian voters, ours are frequentists. They will evaluate point models in terms of their likelihood, but will not hold a complete probability distribution over the full space of possible models.

Second, we do not allow the voter to freely choose her own ideology from $\mathcal{R}$. Instead, she must adopt one of the two ideologies proposed by the parties: the choice is a discrete rather than continuous one. This limitation is motivated by the classic activist's observation that political consciousness rarely forms spontaneously, but requires leadership to be expressed,[6] and the more modern social-scientific observation that most citizens invest little effort in paying attention to politics or in developing their own political ideas. Ideologies offered by the parties are free, from the voter's perspective, while developing an alternative requires effort. We assume the cost of such effort is high enough that the voter will not opt to pay it.

---

[6]E.g., Marx in *The Eighteenth Brumaire of Louis Bonaparte* analogized the French agrarian peasantry to a "sack of potatoes," atomized and unaware of their shared class position and material interests.
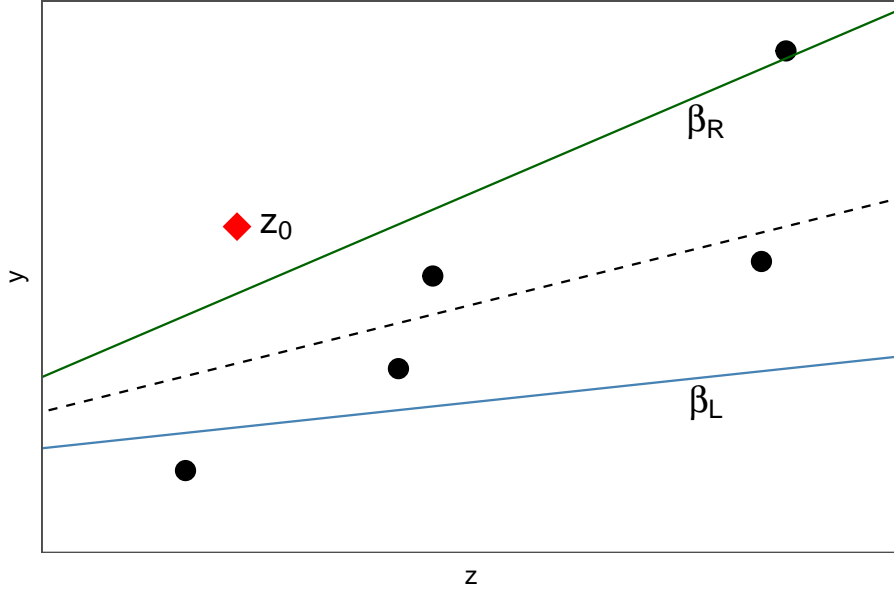
Figure 1: A visual depiction of the ideological contest in one dimension. Voter and both parties observe the policy-outcome pairs represented by the black dots; the voter privately observes an additional outcome realization at z_0, the policy in place at time 0 (the red diamond). Parties L and R offer ideologies, in one dimension representable by scalars β_R and β_L (the slope of the line connecting policy and outcome) that center around the expected OLS fit (the dashed line).

This combination implies that the voter will adopt the ideology, among those offered by the parties, which has the highest likelihood given the data she observes.[7] Given the assumption that the noise term is normally distributed, this implies that the voter adopts the model of the world proposed by R if and only if

$$-\sum_{t=-T}^{0}(z_t\beta_R - y_t)^2 > -\sum_{t=-T}^{0}(z_t\beta_L - y_t)^2$$

Notice that the history of outcomes considered by the voter includes the past record of T outcomes and $T+1$ policy choices that are publicly observed, plus the outcome y_0 which is privately observed by the voter.

The ideological contest stage is described visually in Figure 1. In one dimension, the history takes the form of a scatterplot, and ideologies take the form of scalars (representing the slope of the line connecting policy and outcome).

---

[7]Because the voter's utility is linear, this would also be the optimal behavior for a Bayesian voter with a diffuse prior over β_0.

The election. The second stage of the model is the election. The voter casts her vote for one of the two parties. The winner then implements a policy $z_1 \in \mathbb{R}$. We assume that parties have no credible commitment ability, and thus once in power will implement their bliss point.[8]

Payoffs. The parties are purely policy motivated, with quadratic loss utility: $U_i = -(z_1 - z^i)^2$. z^i denotes party i's preferred policy. Assume $z^R > z^L$.

The voter cares only about outcomes. In particular, we interpret outcomes y_t as a measure of the voter's welfare. For simplicity, her utility is simply equal to y_t. Given her ideology β_w, and the implemented policy z_1 the voter's expected policy utility is then $\beta_w z_1$.[9]

In addition, the voter's evaluation of the right-wing party is subject to an idiosyncratic shock $\xi \sim U[-\frac{1}{2\psi}, \frac{1}{2\psi}]$.

To sum up, the game proceeds as follows:

1. The players observe a public record of $T < \infty$ policy-outcome pairs, as well as the (exogenous) policy z_0

2. The two parties propose their models of the world $\beta_R, \beta_L \in Z \subset \mathbb{R}$

3. The voter privately observes outcome realization y_0

4. The voter adopts the model with best fit to observed history $\beta_w \in \{\beta_R, \beta_L\}$

5. The idiosyncratic preference shock ξ realizes

6. The voter elects party R or L

7. The winner implements policy $z_1 \in \mathbb{R}$

For simplicity, we will assume that the parties know the true value of the coefficient β_0, although this is not essential to our story nor our results. Since we do not allow the voter to draw inferences about β_0 from the parties' proposals, our solution concept is Subgame Perfect Nash Equilibrium.

---

[8]The assumption of no commitment simplifies the analysis and allows us to focus attention on the ideological stage, but is not crucial. We relax this assumption in the Online Appendix, and discuss the robustness of the results in this richer environment in a separate section below.

[9]Our results do not depend on the assumption that the voter's utility is defined over outcomes, nor that it is linear in z_t. For example, we could assume the voter has single-peaked utility over the policy space, with her ideal policy a function of her ideology β_w. As we show in Appendix E.1, this alternative version of the model is essentially isomorphic to the one presented here, and all of our results survive.

Analysis

We now describe the formal analysis of the game and the derivation of comparative statics. All proofs are in the Online Appendix.

As usual, we proceed by backwards induction, beginning with the election winner's policy choice. Our parties do not have credible commitment ability. Straightforwardly, the election winner will therefore always implement its own preferred policy.

The election

Moving one step backward, consider the voter's problem at the electoral stage. Having adopted ideology β_w, and anticipating the equilibrium policy choice of the two parties, the voter can compute the expected outcome under R and L. Thus, her expected policy utility from electing the right-wing party R is $\beta_w z^R$. Similarly, her expected utility from electing L is simply $\beta_w z^L$. Further, recall that the voter's evaluation of the right-wing party is subject to the idiosyncratic shock ξ. Thus, in equilibrium the voter chooses to elect the right-wing party if and only if

$$\beta_w z^R + \xi > \beta_w z^L \tag{2}$$

From the parties' perspective, the election is probabilistic: even after conditioning on the voter's ideology, the election outcome depends on the realization of the shock ξ. Nonetheless, winning the ideological battle gives an electoral advantage.

Proposition 1. *The probability that the right-wing party wins the election is linearly increasing in β_w. Specifically, the probability that R wins the election is given by:*

$$\mathbb{P}_R = \frac{1}{2} + \psi \beta_w (z^R - z^L) \tag{3}$$

Recall that $z^R > z^L$. A direct implication of Proposition 1 is that the right-wing (left-wing) party's expected utility is linearly increasing (decreasing) in β_w. The larger the (absolute) value of β_w, the larger the difference in the voter's expected utility from the two parties. The right-wing party thus wants the voter to adopt an extreme right-wing ideology, as this insures against

unfavorable realizations of the shock ξ. A symmetric reasoning holds for the left-wing party L. Thus, both parties have a preference for inducing extreme (and directionally opposed) ideologies.

The ideological contest

Let us now consider the strategic problem the parties face in the ideological contest. Each party wants the voter to adopt an extreme and favorable ideology. However, parties compete with each other over ideological influence, and must consider how proposing an extreme ideology influences their chances of persuading the voter, which depend on the expected fit of the proposed model to the history of outcomes. Crucially, while the parties observe the public record of past policies and outcomes, the outcome of the policy z_0 in place at the beginning of the game is privately observed by the voter. For any pair of proposed models, the parties are therefore unsure of which one will best explain the outcomes observed by the voter. They must trade off the probability of convincing the voter against the amount of extremism induced in the voter's beliefs.

Proposition 2 is a direct consequence of this uncertainty and the resulting tradeoff.

Proposition 2. *An equilibrium exists for any $T \geq 1$. The (unique) equilibrium can be characterized by:*

1. *The two parties never propose the same model of the world: ideological polarization always emerges in equilibrium. $\beta_R^* > \beta_L^*$.*

2. *The parties always win the ideological contest with equal probability, and propose models centered around the expected OLS estimate for $\hat{\beta}^z$:*

$$\frac{\beta_R^* + \beta_L^*}{2} = \beta_{OLS} = \frac{(\beta_0 z_0^2 + \sum_{t=-T}^{-1} z_t y_t)}{\sum_{t=-T}^{0} z_t^2} \tag{4}$$

3. *Ideological polarization between the parties is given by:*

$$\beta_R^* - \beta_L^* = \frac{\sigma}{\phi(0)} \frac{|z_0|}{\sum_{t=-T}^{0} z_t^2} \tag{5}$$

We discuss each part of the proposition in turn.

No convergence

Part 1 of the proposition is reminiscent of the classical divergence theorem from probabilistic voting models (Calvert, 1985). Indeed, the proof follows an identical logic. If the parties were to propose the same model of the world, a unilateral deviation in the preferred direction would always be profitable for each of them. Depending on the realization of y_0, this deviation will either be payoff-irrelevant or strictly increase the party's expected utility. It is payoff irrelevant if y_0 is such that the deviating party loses the ideological contest, since in this case the deviation has no impact on the ideology adopted by the voter. It improves expected utility if y_0 is such that the deviating party wins, since in this case the voter's ideology has moved in the party's preferred direction. Since parties do not know the realization of y_0, ex-ante the deviation is always strictly profitable, and ideological convergence cannot be sustained in equilibrium. Ideological polarization (here defined as the difference between β_R^* and β_L^*) always emerges.

Symmetry

Substantively, part 2 of Proposition 2 indicates that a favorable history of outcomes does not generate an advantage in the ideological contest for political parties. In our world, the two parties face symmetric strategic problems. As a consequence, the equilibrium will take a symmetric form, with the parties ex-ante equally likely to win the ideological contest and persuade the voter to adopt their model of the world.

However, the results also highlight that a favorable history of outcomes does translate into an *electoral* advantage for political parties. As the history becomes more favorable to the right-wing party (i.e., the OLS estimate for the coefficient β increases), *both* equilibrium ideologies move to the right. In turn, this improves the right-wing party's prospects of winning the election, thereby increasing its expected payoff. The converse holds if the history becomes more favorable to the left-wing party.

Finally, a corollary of the above is that some ideological manipulation always occurs in equilibrium:

Corollary 1. *For any $T < \infty$, $\beta_R^*, \beta_L^* \neq \beta_0$ with probability 1.*

In equilibrium, the ideologies proposed by the two parties are centered around the model that maximizes the expected fit to the history of outcomes observed by the voter. This is the expected OLS estimate for the coefficient β (which we denote as β_{OLS}), where the unobserved outcome realization y_0 is replaced by the expectation $\beta_0 z_0$. Thus, while both parties' equilibrium ideologies are a function of β_0, because the expression in (4) includes the (random) outcome realizations, the true model of the world emerges only in the limit where $T \to \infty$. For any finite history, therefore, neither party offers the true model in equilibrium.

Comparative Statics on Polarization

Part 3 of Proposition 2 describes the determinants of ideological polarization, i.e., the difference $\beta_R^* - \beta_L^*$. Substantively, this difference captures the extent to which the models of the world proposed by the parties disagree on the consequences of the policy dimension for the voter's welfare.

We show that ideological polarization is a function of the history of policies:

Corollary 2. *Equilibrium ideological polarization:*

- *Decreases as past policies become more extreme (i.e., move away from 0);*

- *Increases as today's policy (z_0) becomes more extreme.*

As past policies become more extreme, historical data become more informative for the voter. If policies in the past history are close to 0, proposing a more radical ideology (i.e., a model of the world farther away from the expected best fitting one) has a small impact on the expected model fit, since the predicted outcomes under any model will be close to 0 as well. In contrast, for more extreme past policies, moving the proposed model of the world away from β_{OLS} reduces the fit to the observed history of outcomes more and, thus, also reduces the likelihood of winning the ideological contest. Radical ideologies become more costly, and ideological polarization decreases.

Consider instead the impact of today's policy extremism. Recall that parties face uncertainty over the exact realization of the outcome of today's policy, which is privately observed by the voter. As z_0 becomes more extreme, favorable realizations of the outcome y_0 have a larger (positive)

impact on overall model fit. Thus, as today's policy moves away from 0, the parties can gamble on extreme outcomes and can afford to propose more radical ideologies. Ideological polarization, therefore, increases.

Finally, ideological polarization increases with the uncertainty in the true data-generating process.

Corollary 3. *Equilibrium ideological polarization increases in the variance of the distribution of the error (ε), σ^2.*

As the variance σ^2 increases, the realization of y_0 becomes more uncertain from the parties' perspective. Thus, the outcome of the ideological contest becomes more arbitrary, and parties are willing to propose ideologies that are farther away from the expected best fitting model.

An Illustrative Example: the Reconstruction Era

A key idea in our model is that political entrepreneurs strategically exploit voters' experiences and shared histories to shape the ideology voters adopt and this, in turn, shapes their policy preferences. An important example of this dynamic process comes from the Reconstruction era in the United States. Following the US civil war, formerly enslaved Black freedmen gained political rights and wielded substantial political power in the legislatures of Southern states for the first time. This change in policy coincided with a sharp rise in corruption by government officials.

Although the simultaneity of these two events was coincidental and the industrial-revolution-era rise in corruption is seen by modern scholars (e.g. Skowronek, 1982) as the result of other factors,[10] it was exploited by white Southern politicians to argue that Black empowerment had been the cause of the increase in corruption and insider dealing. This belief became deeply embedded in the minds of many voters due to the deliberate efforts of Southern Democrats.[11] According to DuBois (2007, pp. 511), "The Negro vote and graft were indissolubly linked in the public mind by incessant propaganda" launched by Democratic Redeemers. With this understanding of the world—this ideology—adopted by many Southerners, policy change easily followed, and Black political rights were reversed soon after (DuBois, 2007).

---

[10]An unfavorable realization of ϵ, in the terms of our model.

[11]In this interpretation, in the true model is no relationship between this policy dimension (Black political rights) and the outcome (corruption). The observed history indicated a positive relationship and this allowed Southern politicians to offer such an ideology and have it accepted by voters.

This association between Black empowerment and political corruption, which was literally written into the history books by Southern historians, provided the intellectual foundation for Jim Crow and one-party rule in the South for nearly a century thereafter. The Reconstruction era left a profound mark on American, and particularly Southern, politics, and research has established how it continues to affect modern politics (Acharya et al., 2018). It is natural to think it still runs through political ideologies today, demonstrating the importance of the ideological formation stage and the path dependence of policy that follows from this choice.

Ideology, Policies and Dynamic Incentives

The key intuition that our paper builds on is that political parties strategically manipulate voters' ideologies in order to obtain an advantage in the electoral arena. In the baseline model we consider a world were parties have no commitment ability. Therefore, their policy positions are exogenous and the voter's ideology influences the probability of one or the other party winning the election. However, another reason why political entrepreneurs may want to induce voters to adopt a favorable ideology is to change parties' policy stances. If the voter's ideal policy is a function of her ideology, and parties have commitment ability, moving the voter's ideology will shift the platforms that parties propose in equilibrium.

In the Online Appendix E.2, we analyze this version of the model. Here, the parties no longer face a symmetric problem in the ideological contest. Proposing a more extreme ideology decreases the probability the party can persuade the voter but, conditional on doing so, moves the voter's optimal policy *and* the equilibrium of the platform game in the party's preferred direction. The net gain (or cost) is a function of how far the party's own ideal point is from the voter's induced preferences. While this complicates the analysis, we nonetheless show that the nature of the parties' strategic incentives is analogous to those emerging in the baseline model. Thus, ideological polarization always emerges in equilibrium, and increases as the history of policies becomes more informative (i.e., extreme).

Further, we analyze a dynamic version of the game where the history of past policies is (partially) endogenized, to study how the prospect of engaging in an ideological battle influences parties' incentives to propose moderate or extreme platforms. We show that, in equilibrium, the party that is favored by the true model of the world proposes more extreme policies to facilitate

17

voter learning, even at the cost of decreasing the chances of winning the upcoming election. The party sacrifices immediate electoral success in order to move the equilibrium of the ideological contest in his preferred direction, and thus obtain better policies in the future (similarly to Izzo (2020)).

The Multidimensional Model

So far, we have assumed that parties are constrained to include only the policy dimension z in their models of the world. Real-world ideologies often weave in non-policy dimensions of the social world. We aim to capture this richness here, and to understand what happens to equilibrium polarization as models of the world become more complex and expand to encompass more features of the environment.

In this section we allow parties to include in their models, in addition to the policy variable z, an additional vector-valued set of covariates $\mathbf{x}$. The key distinction between policy and covariates is that the policy dimension z is the one over which parties have preferences and/or are differentiated, and on which they have control. Non-policy covariates may nonetheless be relevant because of the possibility that they affect outcomes. For example, we may think of some components of $\mathbf{x}$ as immutable ethnic, cultural or demographic characteristics of the population. Alternatively, covariates may represent exogenous features of the environment, such as the state of the global economy, or geopolitical factors. These elements cannot be altered by policy means (at least in the short run), but parties may nonetheless find it useful to include them in equilibrium ideologies if they help to explain variation in outcomes.

In this section, we ask if and when parties choose to include non-policy variables in their models, and whether conflict on these dimensions emerges in equilibrium, i.e. whether the proposed models differ on non-policy dimensions. Furthermore, we analyze whether there is an association between model complexity and ideological polarization in equilibrium.

The generalization from the one-dimensional setting is that the data-generating process is now taken to be $y_t = \beta_0' w_t + \varepsilon_t$, where

$$w_t \equiv \begin{bmatrix} z_t \\ \mathbf{x}_t \end{bmatrix},$$

and $\mathbf{x}_t$ is a k-dimensional vector of attributes. Without loss, we assume that the mean of each dimension of $\mathbf{x}$ is zero. It will be useful in what follows to define the stacked $(T \times (k+1))$ matrix of history W, where each row of W is an observation of w_t for $t \in \{-T, \ldots, -1\}$, and similarly the "extended" history W_+ equal to W with an additional row $w_0 = (z_0, \mathbf{x}_0)$.

Each party i proposes a vector of coefficients β_i of dimension $k+1$. We use the superscript j to denote the j-component of β_i. Thus, β_R^z is the z-coefficient proposed by the R party. Note that the model need not encompass all available dimensions, as proposing a coefficient $\beta_i^j = 0$ is equivalent to excluding dimension j from the model of the world. As in the baseline setup, the voter selects the ideology that best explains, in the likelihood sense, the history of outcomes she observes. Importantly, we do not assume that the true coefficients $\beta_0^j \neq 0$, i.e. we allow for the possibility that some dimensions of $\mathbf{x}$'s true correlation with outcomes y are 0. Initially we will take the set of covariates included in the common-knowledge history to be exogenous; later we will endogenize the choice of what to make available.

Analysis

Proposition 3 establishes the multidimensional equivalents of each part of Proposition 2. As a result, the main comparative static results from the baseline single-dimensional setup (Corollaries 1-3) will continue to hold in the multidimensional setting.[12]

Proposition 3. *If W has full rank and $w_0 \neq 0$, then a Nash equilibrium exists. In any such equilibrium:*

1. *The two parties' models of the world differ on all dimensions. $\beta_R^{j*} \neq \beta_L^{j*} \; \forall \, j$.*

---

[12]Corollaries 1-3 hold locally, close to the mean value of the $\mathbf{x}$'s. The implicit expression in (6) reduces to an explicit one in one dimension, such that the corollaries can be stated unconditionally. In the general case, we rely on the implicit function theorem to establish local comparative statics close to the mean of the data.

2. *The parties always win the ideological contest with equal probability, and propose models whose z-component centers around the expected OLS estimate for $\hat{\beta}^z$:*

$$\frac{\beta_R^z + \beta_L^z}{2} = \left((W_+' W_+)^{-1} W_+' y_+\right)^z,$$

$$y_+ \equiv \begin{bmatrix} y \\ w_0' \beta_0 \end{bmatrix}$$

3. *Ideological polarization on all dimensions between the parties is given (implicitly) by the solution to the vector equation:*

$$\beta_R - \beta_L = \frac{\sigma}{\phi(0)} \frac{|w_0'(\beta_R - \beta_L)|}{\beta_R^z - \beta_L^z} \left(W_+' W_+\right)^{-1} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{6}$$

In addition to preserving the results established in one dimension, the multidimensional environment delivers some new comparative static results on which we now expand.

Complete politicization

Part 1 of Proposition 3 shows that equilibrium ideologies are always total: *all* available non-policy dimensions are always included (in the sense of having nonzero coefficient in β_i) by at least one of the parties, and the models offered by the parties differ on all dimensions. Because the size of the history is finite, the within-sample correlation between each x^j and z is always nonzero.[13] Therefore, the (expected) OLS estimate for β^z is a function of whether x^j is included in the model, and because equilibrium proposals move with the expected OLS, at least one party will have an incentive to include it.

The intuition for divergence is that parties can use the x dimensions to improve the overall fit of their model, and take advantage of this better fit to push for more extremism on the policy

---

[13]Outside of a set of outcomes with probability zero.

dimension. In other words, parties increase the dimensionality of their model so as to blame policy failures on unrelated factors. Consider the incentives facing the right-wing party. The party trades-off the desire to induce the voter to adopt an extreme right-wing ideology on the z dimension, with the need to win the ideological contest. Suppose that x^j and z are positively correlated, with an analogous argument holding for the mirroring case. Then, the cost of ideological extremism on the z dimension can be partially offset by proposing a low coefficient on x^j.[14] Given the positive correlation with z, this tends to move predicted outcomes closer to the observed history, thereby improving the overall fit of the model. Directionally opposite incentives emerge for the left-wing party. As a consequence, the parties' models of the world always differ on the x^j dimension. Specifically, the sign of the element of the covariance matrix corresponding to the pair (z, x^j) determines the role that x^j plays in the parties' narratives:

Corollary 4. *Suppose the element of the within-sample inverse covariance matrix $(W'_+ W_+)^{-1}$ corresponding to the pair (z, x^j) is positive. Then, in an open neighborhood around the line $w_0 = [z_0, 0, 0, \ldots, 0]$, $\beta_R^{j^*} > \beta_L^{j^*}$. The converse also holds.*

In the simple bivariate setting (with scalar x), the (z, x^j) element of the inverse covariance matrix has the opposite sign as the correlation between the two variables. So, an x that is positively correlated with z will be emphasized more by the left-wing party than the right ($\beta_R^{j^*} < \beta_L^{j^*}$) and vice versa.

Additionally, although all available dimensions are politicized in equilibrium, some are more politicized than others. The larger is the correlation between policy and covariate in the history, the larger will be the parties' polarization on that dimension. This statement is Corollary 5.

Corollary 5. *The stronger the (within-sample) correlation between x^j and z, the larger is the magnitude of the difference in equilibrium coefficients on the j dimension, $|\beta_R^{j^*} - \beta_L^{j^*}|$.*

Model Complexity and Ideological Polarization

Following the logic above, parties use additional dimensions to compensate for more extremism on the policy dimension. A consequence is that adding additional dimensions to the history allows the possibility of greater observed polarization on the policy dimension. It turns out that

---

[14]That is, low relative to the OLS estimate on this dimension.

in equilibrium, there is an association between model complexity and ideological polarization, which is conditional on the realization of the last-period policy and covariates:

Corollary 6. *Let* $\tilde{W}_+ \equiv \begin{bmatrix} W_+ & x^{k+1} \end{bmatrix}$, *and let* $w_0 = [z_0, 0, 0, \ldots, x_0^{k+1}]$. *The ratio of equilibrium ideological polarization on the z dimension under $\tilde{W}_+$ to that under W_+ is:*

$$
\frac{\tilde{\beta}_R^z - \tilde{\beta}_L^z}{\beta_R^z - \beta_L^z} = \begin{cases} r \geq 1, & x_0^{k+1} = 0 \\ r > 1, & sign(z_0) = sign(x_0^{k+1}(W_+'W_+)_{z,k+1}^{-1}) \\ r > 0, & sign(z_0) \neq sign(x_0^{k+1}(W_+'W_+)_{z,k+1}^{-1}) \end{cases}
$$

Substantively, Corollary 6 says that when today's realization of policy and covariates aligns with the pattern observed in the historical data, observed polarization is higher than it would be without the additional covariate. When there is a "surprise," in the sense that the correlation today is in the opposite direction of the historical pattern, polarization may be lower. If the realization of the covariate today is exactly at its mean (0), there is a weakly positive effect on polarization of adding the additional covariate.[15] If the period 0 realization is drawn from the same conditional distribution that generated the history, a surprise is unlikely, and thus the additional covariate on average increases polarization. We thus expect the addition of more complexity in parties' models to generally increase observed polarization.

Endogenous Salience

In the model analyzed in the previous section, it is costless for parties to propose more complex models. Consequently, even when the correlation with the policy dimension z is arbitrarily small, parties will always chose to include all available variables in their narratives, and ideologies are always total.

Here, we instead consider that parties must pay a cost to render non-policy dimensions salient in the voters' views, and that additional dimensions can play a role in ideologies only when they are salient.

---

[15]The effect is strictly positive except in the case that x^{k+1} is orthogonal to z in the observed history.

We thus add an additional stage to the game, occurring prior to the ideological contest. In this stage, either party may pay a cost C to make non-policy dimension x available in the history. If at least one party pays the cost, then the parties play the multidimensional version of the game; otherwise they play the baseline unidimensional game. We interpret C as the cost of informing the voter of the past history of x, through a campaign of public education (or propaganda, depending on one's point of view).[16] Absent such costly efforts, the voter is unaware of the history of x, and will not be able to use the x dimension when computing the predicted outcomes under the parties' models of the world. Therefore, parties cannot gain from including this additional dimension in their ideologies unless and until the voter is made aware of its past history.

We show that, in equilibrium, x will become politically salient if and only if the within-sample correlation with z (and y) is sufficiently large. Denote β^z_{OLS} the (expected) OLS estimate for β^z if x is included in the model, and $\beta^z_{Omitted}$ the estimate from the single-dimensional model. The difference $\beta^z_{Omitted} - \beta^z_{OLS}$ is the omitted variable bias, which we denote Δ.

Proposition 4. *There exists a unique threshold $\widehat{|\Delta|}$ such that the non-policy dimension x becomes endogenously salient if and only if $|\Delta| \geq \widehat{|\Delta|}$.*

As discussed in relation to Corollary 4, parties include non-policy variables in their models in order to exploit the within-sample correlations and move the equilibrium ideology in their preferred direction. Suppose $\beta^z_{Omitted} - \beta^z_{OLS} < 0$, so that the omitted variable bias moves the estimate to the left. Then, the right-wing party benefits from making x politically salient. However, for the party to be willing to pay the cost C, the magnitude of the bias (and thus the correlation between variables) must be sufficiently large. An analogous reason applies to the left-wing party when $\beta^z_{Omitted} - \beta^z_{OLS} > 0$.

The results of this section highlight that random events may have a crucial importance in the genesis of political ideologies. Recall that parties have no control, or preferences, over the non-policy dimension. Therefore, if a strong enough historical correlation with the policy variable (and thus a large enough omitted variable bias) emerges, it is purely by chance. Political parties, acting as ideological entrepreneurs, will exploit such correlations between variables to their favor, constructing complex models of the world that support their favored policy position.

---

[16]Qualitatively equivalent results would obtain if, prior to the ideological contest, each party decides whether to invest a cost in order to increase the dimensionality of its own model.

Furthermore, Proposition 4 indicates that the sign of the historical correlation with the policy dimension will determine how the parties position themselves on other issues. Thus, random historical events will determine how social issues, religion, race, and other issues become integrated into political ideologies.

An Illustrative Example: The Politics of Nostalgia

The multidimensional model shows that the ideological debate is in part a contest over which dimensions voters should pay attention to. An example from modern American politics illustrates this feature in practice.

Pat Buchanan, the Republican political operative and avatar of the "paleoconservative" wing of the party, wrote in *Suicide of a Superpower* (2012, pp. 44):

> How, in a generation, did we reach a point where. . . our children will not know the good life their parents had, [where] the American Dream may never become reality for scores of millions of our countrymen?
>
> The answer: the failure of our system is rooted in a societal failure. We are not ruled by the same ideas nor do we possess the same moral character as our parents did . . . our intellectual and cultural elites reject the God our parents believed in and the moral code they lived by.

Buchanan's book, like much American social-conservative rhetoric, explains the country's supposed decline from the prosperous and optimistic years of the 1950s as the result of the country's deviation from traditional family and moral strictures. In Buchanan's telling, the American golden age is in the past, lost due to the increasing degeneracy of a godless cultural elite and its corrupting influence on the public's moral character. Importantly, in this understanding the lamented decline is not to be attributed to economic policy changes since the 1950s, such as reductions in the top marginal income tax rates, budget cuts to public institutions, or the relaxation of antitrust enforcement.

In the logic of our model, two features of this rhetoric stand out. The first notable feature is that Buchanan's statement contains no reference to any particular policy. It instead invites the

audience to adopt a model of the causal forces that underlie the variation in outcomes they observe in the world. This is consistent with how we represent the ideological contest in our model. The second notable feature is that, in presenting an ideology, Buchanan is arguing for a particular weighting of dimensions in how voters see the world that allows him to attribute blame away from conservative economic policies. This aligns with the logic underlying our multidimensional narratives.

Conclusion and Discussion

In this paper we have developed a model of hierarchical policy preference formation, where parties provide causal explanations (ideologies) for patterns observed in voters' past experience. Voters evaluate these ideologies according to their fit to experience, and then use the best-fitting one when deciding which policies to support. Winning the ideological battle thus confers an advantage in subsequent electoral competition. The advantage of political parties over voters in our model is a technological rather than an informational one: their exclusive control over the creation and dissemination of ideologies grants parties some ability to persuade even in the absence of any private information.

This persuasive ability is constrained by the features of the common-knowledge history of outcomes. A lower degree of policy extremism in the history means the data is less informative about the effects of policy, which in turn allows parties to offer more divergent ideologies. The persuasive ability of elites is also determined by the history that is considered politically relevant. We showed how ideology can be shaped by variables that are beyond the control of politicians, and even issues that are not inherently political, as long as they are correlated and, to an uninformed voter's mind, plausibly related.

In concluding this paper, we briefly discuss how our model may be enriched and how our conceptualization of ideology further exploited to improve our understanding of the formation of policy preferences, polarization, and electoral competition.

History, Ideology, and Policy Preferences. One valuable contribution of the model is to help understand why the construction of consensus history and its dissemination in schools and

universities is so often hotly politicized. Even in a world of perfect information about parties' policy agendas and no commitment to platforms — where what parties will do if given power is perfectly predictable ex ante — what citizens know about the past can change how they respond to party rhetoric, and ultimately which party they are willing to support. Strategic politicians and political movements understand this dependence well, and intervene when they can to shape the "dataset" that future voters draw upon.

Intriguingly, this suggests that the relationship between voter knowledge and political outcomes is conditional on the type of information held. On one hand, a longer memory of policy history reduces polarization as voters draw upon a larger time series of data to tighten their beliefs about the underlying world. On the other hand, a broader knowledge of other variables, whether policy related or not, may increase polarization as parties are better able to fit the policy variable to their preferred ideology. Proposals to improve policy making by expanding voter knowledge depend critically, therefore, on the direction of knowledge expansion.

Ideological Persuasion, Ideological Revision.   To streamline the presentation and focus on the key logic underlying parties' ideological competition, we have focused on a one-period game. Altering the model so that the game is repeated more than once, with a fresh ideological contest in every period, does not impact the results. As in the one-shot model, parties have symmetric yet directionally opposed incentives to pull the voter to the extreme. This then drives results qualitatively identical to those presented above (see Appendix E.4.1).

A perhaps more meaningful way to think about dynamics in our setting is to consider a world where, once it is formed, ideology is sticky, but can be called into question if it does not resonate with the voter's future experiences. Here, when engaging in the ideological contest, parties think about persuading the voter today *and* ensuring their proposed model of the world is not susceptible to being revised in the future. This may alter their incentives in two ways. First, both parties may be pushed towards the true model. Second, parties may have asymmetric incentives to avoid ideological revision, depending on whether they believe the outcome of a new ideological contest would be more or less favorable to them than the current situation. In turn, this possibility would influence parties' preferences over policies, creating an additional set of dynamic incentives. Analyzing this richer setting is certain to prove technically challenging, but is particularly promising

for future research. In the Online Appendix (E.4.2) we take a first step in this direction. We show that, as expected, the parties no longer face a symmetric problem in the original ideological contest. The party that is disfavored by the history in the first period would like to encourage ideological reassessment in the future, and it moves its ideology towards the truth, albeit at the cost of some reduction in the probability of winning the first-period contest. Despite this asymmetry we are able to show that, under some conditions, the comparative statics from the baseline model continue to hold.

Groups and Identity. We have operationalized ideology as a theory of how the world works, albeit with potentially rich understandings that incorporate many variables and that may extend beyond politics. It is possible that one's ideology is chosen in conjunction with one's identity and even membership of a group (Mason, 2018). If identity and group membership is chosen first, then this will shape the beliefs voters bring to politics and how they experience and perceive the history that, in turn, affect how they respond to ideologies proposed by the parties.[17]

One simple extension of our model is to suppose that society consists of two different identities—say, conservative and liberal—that manifest as two distinct groups. The true data-generating model is different for the different groups, e.g., the same policy may in expectation help one group but hurt the other. Thus, all members of the same group share a common history, which is distinct from the one observed by the other group. The parties must now weigh how the different groups will evaluate the ideology a party offers and in which group will the median voter reside at the election.[18] We conjecture that the existence of heterogeneous groups in society has the effect of increasing ideological polarization in equilibrium. The uncertainty about how the median voter will interpret ideologies adds to the other uncertainty in the model, and this empowers the parties to diverge further in their ideological offerings. The increase in polarization is larger the more evenly divided are the groups.

In the equilibrium of this expanded model, the two groups will, with high likelihood, adopt different ideologies, with each adopting the ideology of a different party. That the divide between the groups extends beyond policy preferences to their underlying views of the world resonates

---

[17]An alternative possibility is that identity and membership of a group is chosen simultaneously with ideology.

[18]The model we present in the paper can be interpreted as the case in which one group dominates the other and holds the median voter with certainty.

with the current state of American politics. This extension of the model may go some way to explaining the hollowing of the middle in the US electorate that emerged in recent decades and why the two sides of politics so often seem to talk past each other, unable to communicate. Formally analyzing this richer setting is thus an important avenue for future research.

Values vs. Ideology. An important assumption in our model is that a citizen's objective is taken as given. Thus, she knows what outcome she wants to achieve and her uncertainty is only over the policy tool that achieves it. Ideology in our model is, therefore, a practical ideology, concerned with the pragmatics of achieving particular outcomes. It does not concern itself with the values of the citizenry, which are taken as fixed. As mentioned in the introduction, this represents the narrow view of ideology, and is a subset of broader conceptions of ideology in which values are also at stake (Bonilla-Silva, 1997).

To expand our framework to endogenize values as well, we can think of an ideological hierarchy in which different principles and beliefs are laid down sequentially. People first formulate their values from experience and some innate metric of happiness of meaning. The values so constructed can then inform or constrain the beliefs about the world that are the focus here and, in turn, shape policy preferences and political behavior. It is an open question how this hierarchy is best formalized, whether it is a purely sequential set of decisions or whether there is some overlap or feedback between levels, as Bonilla-Silva (1997) seems to suggest. Applying our formal structure to this question, and connecting to the myriad accounts of ideology in the literature, is a promising direction for future work.

References

Acharya, Avi, Matthew Blackwell, and Maya Sen, *Deep Roots: How Slavery Still Shapes Southern Politics*, Princeton University Press, 2018.

Benabou, Roland and Jean Tirole, "Belief in a Just World and Redistributive Politics," *The Quarterly Journal of Economics*, 2006, *121* (2), 699–746.

Bonilla-Silva, Eduardo, "Rethinking Racism: Toward a Structural Interpretation," *American Sociological Review*, 1997, *62* (03), 465–480.

Brady, H.E. and P.M. Sniderman, "Attitude attribution: A group basis for political reasoning," *The American Political Science Review*, 1985, pp. 1061–1078.

Bullock, John G., "Elite Influence on Public Opinion in an Informed Electorate," *The American political science review*, 2011, *105* (3), 496–515.

Callander, Steven, "Searching for good policies," *American Political Science Review*, 2011, *105* (04), 643–662.

Calvert, Randall L., "Robustness of the Multidimensional Voting Model: Candidate Motivations, Uncertainty, and Convergence," *American Journal of Political Science*, 1985, *29* (1), 69–95.

Converse, Philip E., "The Nature of Belief Systems in Mass Publics," in D.E. Apter, ed., *Ideology and Discontent*, Free Press, 1964.

Downs, Anthony, "An economic theory of democracy," 1957.

DuBois, W. E. B., *Black Reconstruction in America.* The Oxford W.E.B. DuBois, Oxford: Oxford University Press, 2007.

Eliaz, Kfir and Ran Spiegler, "A Model of Competing Narratives," *The American Economic Review*, 2020, *110* (12), 3786–3816.

Freeden, Michael, "Editorial: What is special about ideologies," *Journal of Political Ideologies*, 2001, *6* (1), 5–12.

Gerring, John, "Ideology: A definitional analysis," *Political Research Quarterly*, 1997, *50* (4), 957–994.

Gramsci, Antonio, *I Quaderni dal Carcere*, Einaudi, 1929.

Hafer, Catherine and Dimitri Landa, "Deliberation as self-discovery and institutions for political speech," *Journal of Theoretical Politics*, 2007, *19* (03), 329–360.

Izzo, Federica, "Ideology For the Future," 2020.

Kinder, Donald R. and Nathan P. Kalmoe, *Neither liberal nor conservative: Ideological innocence in the American public*, University of Chicago Press, 2017.

Klar, Samara and Yanna Krupnikov, *Independent politics*, Cambridge University Press, 2016.

Knight, Kathleen, "Transformations of the Concept of Ideology in the Twentieth Century," *American Political Science Review*, 2006, *100* (4), 619–626.

Lenz, Gabriel S., *Follow the leader?: how voters respond to politicians' policies and performance*, University of Chicago Press, 2013.

Levy, Gilat, Ronny Razin, and Alwyn Young, "Misspecified Politics and the Recurrence of Populism," March 2020. Working paper.

Mason, Lilliana, *Uncivil agreement: How politics became our identity*, University of Chicago Press, 2018.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer, "Coarse Thinking and Persuasion," *The Quarterly Journal of Economics*, 2008, *123* (2), 577–619.

Ortoleva, Pietro, "Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News," *The American Economic Review*, 2012, *102* (6), 2410–2436.

Sartori, Giovanni, "Politics, ideology, and belief systems," *The American political science review*, 1969, *63* (2), 398–411.

Schwartzstein, Joshua and Adi Sunderam, "Using Models to Persuade," *American Economic Review*, 2021, *111* (1), 276–323.

Simon, Herbert A., "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics*, 1955, *69* (1), 99–118.

Skowronek, Stephen, *Building a new American state: The expansion of national administrative capacities, 1877-1920*, Cambridge University Press, 1982.

Zaller, John R., *The nature and origins of mass opinion*, Cambridge university press, 1992.

Online Supplemental Information:

Ideological Competition

Federica Izzo

University of California, San Diego

Gregory J. Martin

Stanford Graduate School of Business

Steven Callander

Stanford Graduate School of Business

August 17, 2022

Contents

A   Proof of Proposition 1

Fixing the winning ideology β_w^z, the voter's expected utility from electing the right-wing party is $\beta_w^z z^R$. Similarly, her expected utility from electing the left-wing party is $\beta_w^z z^L$. Further, recall that her evaluation of the right-wing party is subject to the shock ξ. Thus, the voter will elect R if and only if $\beta_w^z z^R + \xi \geq \beta_w^z z^L$ (the indifference breaking assumption is without loss of generality since ξ is drawn from a continuous distribution). Thus, the probability that R wins the election is:

$$\mathbb{P}(R\ elected) = \mathbb{P}(\xi \geq \beta_w^z(z^L - z^R)).$$

Exploiting the fact that $\xi \sim U[-\frac{1}{2\psi}, \frac{1}{2\psi}]$, the above reduces to:

$$\frac{1}{2} + \psi\beta_w^z(z^R - z^L).$$

In what follows, we will assume that the bounds of the ideological space are such that this probability is always interior.

B   Proof of Propositions 2 and 3

We start from the general (multidimensional) version of the model. The result in the one-dimensional case presented in the main text (Proposition 2) follows as an immediate corollary of the general proof (Proposition 3). The existence statement is Property B1, part 1 is Property B2, part 2 is Property B4, and part 3 is Property B3.

B.1   Preliminaries

We begin with the common-knowledge history h consisting of triples $(y_t, \mathbf{x}_t, z_t)$ and the final period covariates and policy $(\mathbf{x}_0, z_0)$. The extended history h^1 observed only by the citizen contains an additional observation of the outcome y_0 corresponding to the final period policy. Each $\mathbf{x_i} \in \mathcal{R}^k$.

The parties' maximization problems are:

$$\max_{\beta_R} (\beta_R^z - \beta_L^z) \, \mathbb{P}\left(\log \mathcal{L}(y|h^1, \beta_R) - \log \mathcal{L}(y|h^1, \beta_L) \geq 0\right) + \beta_L^z$$

$$\max_{\beta_L} -(\beta_R^z - \beta_L^z) \, \mathbb{P}\left(\log \mathcal{L}(y|h^1, \beta_R) - \log \mathcal{L}(y|h^1, \beta_L) \geq 0\right) - \beta_L^z$$

The probability appearing in the maximand is the probability that the citizen adopts R's model, $\mathbb{P}(R\ win)$. This can be written as:

$$\mathbb{P}(R\ win) \equiv \mathbb{P}(\log \mathcal{L}(y|h^1, \beta_R) - \log \mathcal{L}(y|h^1, \beta_L) \geq 0) =$$

$$\mathbb{P}\left(\sum_{t=-T}^{-1} \log\left(\frac{f(y_t|x_t, z_t, \beta_R)}{f(y_t|x_t, z_t, \beta_L)}\right) + \log\left(\frac{f(y|x_0, z_0, \beta_R)}{f(y|x_0, z_0, \beta_L)}\right) \geq 0\right)$$

Plugging in the distributional assumptions, and letting $w_t \equiv \begin{bmatrix} z_t \\ x_t \end{bmatrix}$, this reduces to:

$$\mathbb{P}\left(w_0'(\beta_R - \beta_L)\left((w_0'\beta_0 + \epsilon) - \frac{w_0'(\beta_R + \beta_L)}{2}\right) \geq -\sum_{i=1}^{n} \sigma^2 \log r_t\right),$$

$$\sigma^2 \log r_t = w_t'(\beta_R - \beta_L)\left(y_t - \frac{w_t'(\beta_R + \beta_L)}{2}\right)$$

The right hand side of the inequality above can be more compactly written in matrix notation as

$$-\frac{1}{\sigma}(\beta_R - \beta_L)'W'(y - W\frac{\beta_R + \beta_L}{2})$$

Where y is the vector of T outcome observations, and W is the $(T \times k)$ matrix of observed data.

Rearranging, we get

$$\mathbb{P}\left(\epsilon/\sigma \geq \frac{1}{\sigma w_0'(\beta_R - \beta_L)}\left(-(\beta_R - \beta_L)'W'(y - W\frac{\beta_R + \beta_L}{2})\right) - w_0'\frac{(\beta_0 - (\beta_R + \beta_L)/2)}{\sigma}\right)$$

for $w_0'(\beta_R - \beta_L)$ positive or the reverse condition if $w_0'(\beta_R - \beta_L)$ is negative. So:

$$\mathbb{P}(R \ wins) = \begin{cases} (1 - \Phi(\xi(\beta_R, \beta_L))), & w_0'(\beta_R - \beta_L) > 0 \\ \Phi(\xi(\beta_R, \beta_L)), & w_0'(\beta_R - \beta_L) < 0 \end{cases} \tag{1}$$

where

$$\xi(\beta_R, \beta_L) = \frac{1}{\sigma w_0'(\beta_R - \beta_L)}\left(-(\beta_R - \beta_L)'W'(y - W\frac{\beta_R + \beta_L}{2})\right) - w_0'\frac{(\beta_0 - (\beta_R + \beta_L)/2)}{\sigma} \tag{2}$$

B.2   Existence

We now show that a Nash equilibrium exists in the first (ideological) stage of the contest.

Lemma 1. *The payoff functions are continuous in β_R and β_L.*

Proof. Both parties' utilities conditional on the model adopted by the citizen are linear in the β's (Proposition 1), so the only question concerns the probability $\mathbb{P}(R \ wins)$. $\xi(\beta_R, \beta_L)$ is continuous everywhere except the hyperplane with $w_0'(\beta_R - \beta_L) = 0$. However, the switch from $1 - \Phi(\cdot)$ to $\Phi(\cdot)$ that occurs here counteracts this and preserves the continuity of the payoff function. Taking limits from below and above respectively:

$$\lim_{w_0'(\beta_R - \beta_L) \uparrow 0} \mathbb{P}(win) = \lim_{w_0'(\beta_R - \beta_L) \uparrow 0} \Phi(\xi(\beta_R, \beta_L)) = \begin{cases} 1, & (\beta_R - \beta_L)'W'(y - W\frac{\beta_R + \beta_L}{2}) > 0 \\ 0, & (\beta_R - \beta_L)'W'(y - W\frac{\beta_R + \beta_L}{2}) < 0 \end{cases}$$

$$\lim_{w_0'(\beta_R - \beta_L) \downarrow 0} \mathbb{P}(win) = \lim_{w_0'(\beta_R - \beta_L) \downarrow 0} 1 - \Phi(\xi(\beta_R, \beta_L)) = \begin{cases} 1, & (\beta_R - \beta_L)'W'(y - W\frac{\beta_R + \beta_L}{2}) > 0 \\ 0, & (\beta_R - \beta_L)'W'(y - W\frac{\beta_R + \beta_L}{2}) < 0 \end{cases}$$

□

Property B1. *If W has full rank and $w_0 \neq 0$, then a Nash equilibrium (possibly in mixed strategies) exists.*

Proof. We apply Theorem 1.3 of Fudenberg and Tirole (1991). Lemma 1 established continuity of the payoff functions, the first condition of the theorem. The second condition, compactness of the choice space, reduces to closedness and boundedness in our setting. We show now that the equilibria where strategies are restricted to a closed and bounded subset of $\mathbb{R}^k$ always coincide with the unbounded equilibria for some arbitrary large but finite bound.

We first show that β_R^z times the probability of winning goes to zero as $\beta_R^z \to \pm\infty$ for fixed β_L. An identical argument applies in the opposite case. Hence the best response to any finite strategy is also finite.

Given the structure of payoffs, we need to ensure that $\lim_{\beta_R^z \to \infty} \xi(\beta_R, \cdot)$ is $+\infty$ when $w_0'(\beta_R - \beta_L) > 0$ and $-\infty$ when $w_0'(\beta_R - \beta_L) < 0$. This is guaranteed if $W'W$ is positive definite, e.g. if W has full rank, as in this case $-(\beta_R - \beta_L)'W'(y - W\frac{\beta_R + \beta_L}{2})$ is strictly convex. $\xi(\beta_R, \cdot)$ is $O(\beta_R^z)$ and hence $\beta_R^z \phi(\xi(\beta_R, \cdot))$ converges to 0 in the limit as required.

To show that there can be no equilibrium where both strategies are infinite, note that any such equilibrium requires that $\xi(\beta_R, \beta_L)$ remains finite in the limit. Otherwise, one player wins the ideological contest with probability zero, which cannot be an equilibrium since she can guarantee a strictly positive probability of winning by proposing any finite β.

□

Property B1 requires that the matrix of common-knowledge history W must have full rank. Substantively, this requires that there is sufficiently independent variation in all dimensions j of

the history. If there is not, parties can find a direction of change in β_i which holds the likelihood exactly constant, and exploit this to push the coefficient on z they offer to $\pm\infty$ without loss in the chance of winning the ideological contest; the contest thus exhibits infinite ideological polarization and no equilibrium in finite strategies. A necessary condition for full rank to obtain is that $T > k$, and thus the complexity of the environment relative to the length of voters' memory will be a key determinant of equilibrium polarization. This observation connects to the result in Corollary 6.

B.3 Characterization

The first-order conditions generate a system of equations:

$$\nabla_R \mathbb{P}(R \ wins) = \begin{bmatrix} -\frac{\mathbb{P}(R \ wins)}{\beta_R^z - \beta_L^z} \\ 0 \end{bmatrix} \tag{3}$$

$$\nabla_L \mathbb{P}(R \ wins) = \begin{bmatrix} -\frac{1-\mathbb{P}(R \ wins)}{\beta_R^z - \beta_L^z} \\ 0 \end{bmatrix} \tag{4}$$

where ∇_i is the gradient with respect to β_i. By inspection, we have immediately the non-convergence result:

Property B2. *For $T < \infty$, $\beta_R^{z^*} > \beta_L^{z^*}$.*

Proof. From the first row of (3) and (4), $\beta_R^z = \beta_L^z$ implies that the derivative of $\mathbb{P}(R \ wins)$ with respect to β_R^z and β_L^z is infinite. This is possible only if the likelihood surface degenerates to a step function (e.g. in the limit $T \to \infty$). For any finite T the slope is finite and hence $\beta_R^z \neq \beta_L^z$. Given the utility functions it must be that $\beta_R^{z^*} > \beta_L^{z^*}$ in equilibrium. $\square$

We next show that equilibrium proposals are symmetric and center around the expected OLS estimate on the z dimension. It will be useful first to establish the property that divergence appears on all dimensions:

Property B3. *Ideological polarization on all dimensions between the parties is given (implicitly) by the solution to the vector equation:*

$$\beta_R - \beta_L = \frac{\sigma}{\phi(0)} \frac{|w_0'(\beta_R - \beta_L)|}{\beta_R^z - \beta_L^z} \left(W_+' W_+ \right)^{-1} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{5}$$

Equilibrium platforms diverge on all dimensions with probability 1.

Proof. Add the two FOC's (3) and (4), plug in the definition in (1) and rearrange to get (5). The ith entry of the vector on the right hand side of (5) can be zero only if $(W_+' W_+^{-1})_{(z,i)}$ is zero. This is a zero probability event if the w's are iid from a distribution with full support on $\mathcal{R}^k$. $\qquad\square$

Property B4. *The parties always win the ideological contest with equal probability, and propose models whose z-component centers around the expected OLS estimate for $\hat{\beta}^z$:*

$$\frac{\beta_R^z + \beta_L^z}{2} = \left((W_+' W_+)^{-1} W_+' y_+ \right)^z,$$

$$y_+ \equiv \begin{bmatrix} y \\ w_0' \beta_0 \end{bmatrix}$$

Proof. From the first row of the FOCs, it is evident that any symmetric equilibrium (with $\mathbb{P}(R \ wins)$ = $\frac{1}{2}$) requires $\frac{\partial}{\partial \beta_R^z} \mathbb{P}(R \ wins) = \frac{\partial}{\partial \beta_L^z} \mathbb{P}(R \ wins)$. We now show that this is guaranteed to hold in equilibrium.

From the FOC's we have that $\mathbb{P}(A \ wins) = \frac{1}{2}$ iff $\nabla_A \mathbb{P}(R \ wins) = \nabla_B \mathbb{P}(R \ wins)$. From the expression in (1) we see that $\nabla_i \mathbb{P}(R \ wins) = \alpha \nabla_i \xi(\beta_R, \beta_L)$ for some strictly nonzero scalar α, hence checking equality reduces to checking $\nabla_A \xi(\beta_R, \beta_L) = \nabla_B \xi(\beta_R, \beta_L)$. Plugging in the definition from (2) and taking the differences between the gradients with respect to the choice of each party, we get that the condition $\nabla_A \xi(\beta_R, \beta_L) - \nabla_B \xi(\beta_R, \beta_L) = 0$ reduces to:

$$w_0'(\beta_R - \beta_L)'W'(y - W\frac{\beta_R + \beta_L}{2}) - w_0'(\beta_R - \beta_L)W'(y - W\frac{\beta_R + \beta_L}{2}) = 0 \qquad (6)$$

Which is true for any β_R, β_L, hence any equilibrium must be symmetric. This result implies that in equilibrium, $\xi(\beta_R, \beta_L) = 0$. Plugging in the expression in (2), and multiplying through by $\sigma w_0'(\beta_R - \beta_L)$ we get:

$$
\begin{aligned}
0 =& ((\beta_R - \beta_L)'w_0)'w_0'(\beta_0 - \frac{\beta_R + \beta_L}{2}) + (\beta_R - \beta_L)'W'(y - W\frac{\beta_R + \beta_L}{2}) \\
=& (\beta_R - \beta_L)'w_0 w_0'(\beta_0 - \frac{\beta_R + \beta_L}{2}) + (\beta_R - \beta_L)'W'(y - W\frac{\beta_R + \beta_L}{2}) \\
=& (\beta_R - \beta_L)'\left(w_0 w_0'(\beta_0 - \frac{\beta_R + \beta_L}{2}) + W'y - W'W(\frac{\beta_R + \beta_L}{2})\right) \\
=& (\beta_R - \beta_L)'\left(W_+'y_+ - W_+'W_+\frac{\beta_R + \beta_L}{2}\right)
\end{aligned}
$$

where the first step uses the fact that $w_0'(\beta_R - \beta_L)$ is a scalar.

We know from Equation (5) that $\beta_R - \beta_L$ is generically nonzero in all components. Plugging in the expression from (5) and canceling the scalar terms gives:

$$\begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}(W_+'W_+)^{-1}W_+'y_+ = \frac{\beta_A^z + \beta_B^z}{2} \qquad (7)$$

In other words, the z components of the vectors β_R and β_L must average to the z component of the OLS estimate on the observed data $\hat{\beta}_{OLS} = (W_+'W_+)^{-1}W_+'y_+$. The average on the other dimensions is unconstrained. Hence while (5) pins down the difference between A and B's positions on all dimensions, only the z component of the sum is restricted.

Payoffs are thus unique, but equilibrium strategies are unique only on the z dimension. We note that a natural focal point is the choice of the OLS estimate on all dimensions:

$$\frac{\beta_R + \beta_L}{2} = (W_+'W_+)^{-1}W_+'y_+ \qquad (8)$$

This choice solves for $\left(W'_+ y_+ - W'_+ W_+ \frac{\beta_R + \beta_L}{2}\right) = 0$ and thus satisfies (6) for any choice of $\beta_R - \beta_L$.

$\square$

C Proofs of Corollaries 1-5

Corollary 1. *For any* $T < \infty$, $\beta_R^{z^*}, \beta_L^{z^*} \neq \beta_0^z$ *with probability 1.*

Proof. The inequality is immediate from Proposition 3. The proposals center around the expected OLS estimate, which can be written as:

$$
\begin{aligned}
\hat{\beta}_{OLS} &= (W'_+ W_+)^{-1} W'_+ y_+ \\
&= (W'_+ W_+)^{-1} W'_+ (W_+ \beta_0 + \varepsilon) \\
&= \beta_0 + (W'_+ W_+)^{-1} W'_+ \varepsilon
\end{aligned}
$$

The last term converges in probability to zero but is nonzero in all components with probability 1 for any finite T and full-rank W_+.

$\square$

Corollary 2. *Equilibrium ideological polarization is not a function of the history of outcomes.*

Proof. By inspection of (5), in which y does not appear. $\square$

To get comparative statics on w_0, W_+, σ we apply the implicit function theorem around the point where $w_0 = [z_0, 0, 0, \ldots, 0]$, i.e. close to the mean of $\mathbf{x}$. We prove some comparative static statements only in a neighborhood of this line. In one dimension the condition is satisfied trivially and hence the comparative statics hold without qualification.

We define $\Delta \equiv \beta_R - \beta_L$ and using D_Δ to denote the Jacobian matrix of (5) with respect to Δ, we note that

$$D_\Delta = I - \frac{\sigma}{\phi(0)}(W_+'W_+)^{-1}\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}\begin{bmatrix}\frac{\partial}{\partial\Delta^z}\frac{|w_0'\Delta|}{\Delta^z} & \frac{\partial}{\partial\Delta^{x_1}}\frac{|w_0'\Delta|}{\Delta^z} & \cdots & \frac{\partial}{\partial\Delta^{x_k}}\frac{|w_0'\Delta|}{\Delta^z}\end{bmatrix}$$

In a neighborhood around $w_0 = [z_0, 0, 0, \ldots, 0]$, the gradient vector on the right hand side of the above is the zero vector. Hence D_Δ is just I, and by the implicit function theorem the Jacobian matrix of Δ with respect to the parameters q is just $-D_q$. It therefore suffices to differentiate (5) with respect to the parameter of interest to get local comparative statics.

Corollary 3. *In an open neighborhood around the line $w_0 = [z_0, 0, 0, \ldots, 0]$, equilibrium ideological polarization:*

- *Decreases as past policies become more extreme (i.e., move away from 0);*

- *Increases as today's policy (z_0) becomes more extreme.*

Proof. At $w_0 = [z_0, 0, 0, \ldots, 0]$, $|w_0'(\beta_R - \beta_L)| = |z_0|(\beta_A^z - \beta_B^z)$.

Hence, (5) reduces at this point to:

$$\beta_R^z - \beta_L^z = \frac{\sigma}{\phi(0)}(W_+'W_+)_{zz}^{-1}|z_0| \tag{9}$$

$$\beta_R^j - \beta_L^j = \frac{\sigma}{\phi(0)}(W_+'W_+)_{zj}^{-1}|z_0|, \ \forall j \tag{10}$$

The statement follows by inspection of (9).

$\square$

Corollary 4. *In an open neighborhood around the line $w_0 = [z_0, 0, 0, \ldots, 0]$, equilibrium ideological polarization increases in the variance of the distribution of the error (ε), σ^2.*

Proof. By inspection of (9), as $(W_+'W_+)_{zz}^{-1}$, $|z_0|$, and the normal density $\phi(0)$ are all positive. $\square$

Corollary 5. *Suppose the element of the within-sample inverse covariance matrix $(W_+'W_+)^{-1}$ corresponding to the pair (z, x^j) is positive. Then, in an open neighborhood around the line $w_0 = [z_0, 0, 0, \ldots, 0]$, $\beta_R^{j^*} > \beta_L^{j^*}$. The converse also holds.*

Proof. By inspection of (5), as σ, $|w_0'(\beta_R - \beta_L)|$, $\beta_R^z - \beta_L^z$, and the normal density $\phi(0)$ are all positive. $\qquad\square$

Corollary 6. *Let $\tilde{W}_+ \equiv \begin{bmatrix} W_+ & x^{k+1} \end{bmatrix}$, and let $w_0 = [z_0, 0, 0, \ldots, x_0^{k+1}]$. The ratio of equilibrium ideological polarization on the z dimension under $\tilde{W}_+$ to that under W_+ is:*

$$\frac{\tilde{\beta}_R^z - \tilde{\beta}_L^z}{\beta_R^z - \beta_L^z} = \begin{cases} r \geq 1, & x_0^{k+1} = 0 \\ r > 1, & sign(z_0) = sign(x_0^{k+1}(W_+'W_+)_{z,k+1}^{-1}) \\ r > 0, & sign(z_0) \neq sign(x_0^{k+1}(W_+'W_+)_{z,k+1}^{-1}) \end{cases}$$

Proof. (9) shows that at $w_0 = [z_0, 0, 0, \ldots, 0]$, $\beta_R^z - \beta_L^z$ is a function only of the variance on the z dimension. We consider adding an additional element with $x_0^{k+1} \neq 0$. The first row of (5) now becomes:

$$\tilde{\beta}_R^z - \tilde{\beta}_L^z = \frac{\sigma}{\phi(0)} \frac{\left| z_0(\tilde{\beta}_R^z - \tilde{\beta}_L^z) + x_0^{k+1}(\tilde{\beta}_R^{k+1} - \tilde{\beta}_L^{k+1}) \right|}{\tilde{\beta}_R^z - \tilde{\beta}_L^z} \left(\tilde{W}_+'\tilde{W}_+ \right)_{z,z}^{-1}$$

While under the lower dimension the corresponding expression is:

$$\beta_R^z - \beta_L^z = \frac{\sigma}{\phi(0)} |z_0| \left(W_+'W_+ \right)_{z,z}^{-1}$$

Dividing the two, we get that the ratio $r = \frac{\tilde{\beta}_R^z - \tilde{\beta}_L^z}{\beta_R^z - \beta_L^z}$ satisfies:

$$r = \frac{\left| z_0(\tilde{\beta}_R^z - \tilde{\beta}_L^z) + x_0^{k+1}(\tilde{\beta}_R^{k+1} - \tilde{\beta}_L^{k+1}) \right|}{|z_0| \left(\tilde{\beta}_R^z - \tilde{\beta}_L^z \right)} \frac{\left(\tilde{W}_+'\tilde{W}_+ \right)_{z,z}^{-1}}{(W_+'W_+)_{z,z}^{-1}}$$

Obviously, r must be positive. Using the partitioned matrix inverse formula we have that $\left(\tilde{W}_+'\tilde{W}_+ \right)_{z,z}^{-1} = (W'M_{x^{k+1}}W)^{-1}$, where $M_{x^{k+1}}$ is the "residual maker" matrix $I - x^{k+1}(x^{k+1'}x^{k+1})^{-1}x^{k+1'}$.

This implies $\left(W'_+W_+\right)^{-1}_{z,z} \leq \left(\tilde{W}'_+\tilde{W}_+\right)^{-1}_{z,z}$ and hence if $x^{k+1}_0 = 0$ then $r \geq 1$, with equality achieved only when x^{k+1} is orthogonal to z.

By Corollary 4, the sign of $\beta^{k+1}_R - \tilde{\beta}^{k+1}_L$ is the same as the sign of $\left(W'_+W_+\right)^{-1}_{z,j}$. If $x^j_0 \left(W'_+W_+\right)^{-1}_{z,j}$ has the same sign as z_0 then the absolute value in the numerator can be separated and the ratio becomes $r = 1 + R$ where R is strictly positive. Otherwise, it cannot, and we can only show $r \geq 0$. $\qquad \square$

Corollary 7. *The stronger the (within-sample) correlation between x^j and z, the larger is the magnitude of the difference in equilibrium coefficients on the j dimension, $|\beta^{j^*}_R - \beta^{j^*}_L|$.*

Proof. By inspection of (10), as σ, T, $|z_0|$, and the normal density $\phi(0)$ are all positive. $\qquad \square$

D Proof of Proposition 4

Proceeding by backwards induction, we can compute the parties' expected utility as a function of the set of salient dimensions. Given property B4, we know that parties always win the ideological contest with equal probability. Given properties B3 and B4, the z-component of the equilibrium ideologies is a function of the set of covariates. Furthermore, we know from Proposition 1 that each party's probability of winning the election is a function of β^z_w. Thus, the set of salient dimension influences the parties' expected utility only via the z-component of their equilibrium models.

In particular, let $\mathbb{I}_x$ a binary indicator taking value 1 if x is salient, and 0 otherwise. Then, we can denote $\beta^{z^*}_R(\mathbb{I}_x)$ and $\beta^{z^*}_L(\mathbb{I}_x)$ the parties' equilibrium proposed coefficients on the z dimension as a function of the set of salient dimensions.

With this, we can write the parties' expected utility as:

$$E[U_R] = -\frac{1}{2}(z^R - z^L)^2(\frac{1}{2} - \beta^{z^*}_R(\mathbb{I}_x)\psi(z^R - z^L)) - \frac{1}{2}(z^R - z^L)^2(\frac{1}{2} - \beta^{z^*}_L(\mathbb{I}_x)\psi(z^R - z^L)) \quad (11)$$

and

$$E[U_L] = -\frac{1}{2}(z^R - z^L)^2(\frac{1}{2} + \beta_R^{z^*}(\mathbb{I}_x)\psi(z^R - z^L)) - \frac{1}{2}(z^R - z^L)^2(\frac{1}{2} + \beta_L^{z^*}(\mathbb{I}_x)\psi(z^R - z^L)) \quad (12)$$

Thus, the right-wing party's expected utility is higher when x is salient if and only if

$$-\frac{1}{2}(z^R - z^L)^2(\frac{1}{2} - \beta_R^{z^*}(1)\psi(z^R - z^L)) - \frac{1}{2}(z^R - z^L)^2(\frac{1}{2} - \beta_L^{z^*}(1)\psi(z^R - z^L)) > \quad (13)$$
$$-\frac{1}{2}(z^R - z^L)^2(\frac{1}{2} - \beta_R^{z^*}(0)\psi(z^R - z^L)) - \frac{1}{2}(z^R - z^L)^2(\frac{1}{2} - \beta_L^{z^*}(0)\psi(z^R - z^L))$$

Which reduces to

$$\beta_R^{z^*}(1) + \beta_L^{z^*}(1) > \beta_R^{z^*}(0) + \beta_L^{z^*}(0) \quad (14)$$

From Property B4, we know that in equilibrium the parties always propose models whose z-component is centered around the expected OLS estimate for β^z. Denote β_{OLS}^Z the expected OLS estimate for β^z if x is salient, and $\beta_{Omitted}^Z$ the estimate if x is excluded by both models.

The above can be rewritten as

$$\beta_{OLS}^z > \beta_{Omitted}^z \quad (15)$$

Thus, the right-wing party's expected equilibrium utility increases when x is salient if and only if the omitted variable bias is negative. Similarly, we can verify that the left-wing party's expected equilibrium utility increases when x is salient if and only if the omitted variable bias is positive. Therefore, only one party at a time may be willing to pay the cost of making x salient.

Let us focus on the right-wing party, with symmetric results holding for the left-wing one. In equilibrium, R chooses to pay the cost C if and only if

$$-\frac{1}{2}(z^R - z^L)^2(\frac{1}{2} - \beta_R^{z^*}(1)\psi(z^R - z^L)) - \frac{1}{2}(z^R - z^L)^2(\frac{1}{2} - \beta_L^{z^*}(1)\psi(z^R - z^L)) - C > \quad (16)$$
$$-\frac{1}{2}(z^R - z^L)^2(\frac{1}{2} - \beta_R^{z^*}(0)\psi(z^R - z^L)) - \frac{1}{2}(z^R - z^L)^2(\frac{1}{2} - \beta_L^{z^*}(0)\psi(z^R - z^L))$$

Which reduces to

$$C < (z^R - z^L)^3\psi\frac{\beta_R^{z^*}(1) + \beta_L^{z^*}(1) - \beta_R^{z^*}(0) - \beta_L^{z^*}(0)}{2} \quad (17)$$

Which, noticing that the z-components of the equilibrium models are always centered around the expected OLS estimate for β^z can be rewritten as

$$C \quad < -\psi(z^R - z^L)^3\Delta, \quad (18)$$
$$\Delta \quad \equiv \beta_{Omitted}^z - \beta_{OLS}^z \quad (19)$$

E Extensions and Robustness

E.1 Alternative Assumptions on Voter's Utility

In the baseline model, the voter's utility is defined over outcomes, and is linear in the policy z. Here, we show that neither of these assumptions is essential for our results.

Consider an amended version of the baseline model where the voter's utility is defined over the policy space, and takes the familiar quadratic loss form:

$$u^v = -(z^v(\beta_w^z) - z)^2. \quad (20)$$

14

Here, $z^v(\beta_w^z)$ denotes the voter's ideal point, which is a linearly increasing function of her adopted ideology (i.e., $\frac{\partial z^v(\beta_w^z)}{\partial \beta_w^z} > 0$ and $\frac{\partial^2 z^v(\beta_w^z)}{\partial^2 \beta_w^z} = 0$). For simplicity, assume that $z^v = \gamma \beta_w^z$, where $\gamma \geq 0$.

As in the baseline model, increasing (decreasing) β_w^z then translates into an electoral advantage for the right-wing (left-wing) party. This preserves the directional incentives that drive the equilibrium of the ideological contest. Indeed, we can show that this version of the model is essentially isomorphic to the baseline.

To see this, it is enough to consider how the voter's ideology influences the electoral stage. Given winning ideology β_w^z, the voter elects the right-wing party iff

$$- (z^R)^2 + 2z^R z^v(\beta_w^z) + \xi > -(z^L)^2 + 2z^L z^v(\beta_w^z). \tag{21}$$

The probability that R wins the election is then

$$P_w = \frac{1}{2} - \psi(z^R - z^L)(z^R + z^L - 2z^v(\beta_w^z)).$$

Recall that $z^R - z^L > 0$, therefore P_w is linearly increasing in z^v and thus linearly increasing in β_w^z. As a consequence, the parties' expected utility is also linear in β_w^z, just as in the baseline model. In fact, the marginal effect of the winning ideology on the probability of winning the election stage is the same as in the baseline model, simply scaled by a factor 2γ: $2\gamma\psi(z^R - z^L)$. Thus, the parties' strategic problem is identical to the baseline, and all our results follow.[1]

E.2 Parties' Commitment Ability

We analyze an extension of the baseline model that considers endogenous policy platforms. We thus endow our parties with credible commitment ability, and assume they strategically choose their policy positions before the election but after the ideological contest. The outcome of the ideological contest will therefore determine the equilibrium of the platform game, (potentially)

---

[1] We also note that another version of the model isomorphic to the baseline is one where the voter's utility is $u_t^v = y_t - \gamma z_t^2$. This could, for example, capture a risk-averse voter: extreme policies generate more uncertain outcomes, and therefore the voter pays a cost that is increasing as the implemented policy moves away from 0. This preference structure can also be interpreted as the voter having a quadratic model of how policy affects outcomes.

impacting parties' expected utility via their own and their opponent's platform choices *and* via their probability of being elected.

As in the previous section **E.1**, the voter's utility is

$$u^v = -(z_v(\beta_w^z) - z)^2, \tag{22}$$

where $\frac{\partial z^v(\beta_w^z)}{\partial \beta_w^z} > 0$ and $\frac{\partial^2 z^v(\beta_w^z)}{\partial^2 \beta_w^z} = 0$. More generally, any symmetric and single-peaked utility function would generate the results below.

As in the baseline model, parties also have quadratic loss utility, $-(z^j - z)^2$, for $j \in \{L, R\}$. We will be assuming that the parties are sufficiently extreme that, for any β_w^z, the voter's ideal policy is in between their bliss points. Further, as in the baseline, the support of the valence shock ξ is sufficiently large that the probability of winning is always interior. For simplicity, we will focus on the version of the model where parties' models of the world are unidimensional.

First, we analyse the electoral stage:

Lemma E.2. *In equilibrium of the election stage, the parties' platforms are located on opposite sides of $z^v(\beta_w^z)$. Further, R's expected utility is increasing in β_w^z, L's is decreasing.*

Proof. Suppose that both parties locate to the left of the voter's induced ideal point $z^v(\beta_w^z)$. By the assumption on the voter's utility, there is a (symmetric) point to the right of $z^v(\beta_w^z)$ which delivers the voter an equivalent utility and thus generates an equivalent probability that R wins the election. R strictly prefers this deviation. An identical argument applies if both locate to the left of $z^v(\beta_w^z)$. The parties cannot converge to $z^v(\beta_w^z)$ because this would allow for one party to deviate to a strictly preferred policy with strictly positive probability of winning, which would be a strict improvement over getting $z^v(\beta_w^z)$ with certainty. $z^v(\beta_w^z)$ acts as a constraint on the parties' maximization problems and hence their expected utility must (weakly) increase as this constraint is relaxed. □

Let $v_R(\beta_w)$ and $v_L(\beta_w)$ be the equilibrium expected utility of R and L in the election stage, given that the voter has adopted β_w in the ideological stage. Lemma E.2 implies that for any pair of possible adopted ideologies β_w, β_w' with $\beta_w' > \beta_w$, $v_R(\beta_w') > v_R(\beta_w)$ and $v_L(\beta_w') < v_L(\beta_w)$. Moving backwards, we analyze the ideological contest. Unlike the baseline model, the equilibrium

of the ideological contest is no longer symmetric. This is because, while the voter's ideal point is linear in β_w^z, the parties' expected utility in equilibrium is not. The marginal gain from moving the voter's bliss point depends on the location of this policy relative to the party's own ideal point, which will in turn determine the equilibrium platforms. As a consequence, the parties no longer face a symmetric problem. Nonetheless, we can show that the key results from the baseline model (the emergence of ideological polarization and its determinants) survive, at least under some conditions.

Proposition E.1. *An equilibrium exists for any $T \geq 1$. In equilibrium, the two parties never propose the same model of the world: ideological polarization always emerges in equilibrium, with $\beta_R^{z^*} > \beta_L^{z^*}$.*

Proof. The proof is exactly as in the baseline model, and is therefore omitted. In particular, the proof of the divergence results follows the usual logic of the standard divergence theorems in probabilistic voting models. $\square$

Next, we examine the determinants of ideological polarization. In particular, we consider how polarization responds to past policies becoming more extreme. Due to the lack of symmetry in the ideological (as well as electoral) contests, it is not easy to show that polarization always decreases as past policies become more extreme. This is because changing past policies has two potential effects. First, it changes the informativeness of the history of outcomes. Second, if policies change but the outcomes remains fixed, this also affects the OLS estimate β_{OLS}^z. In the baseline model, this second effect is washed out by the symmetric nature of the equilibrium, that ensures that both parties respond to changes in the OLS in identical ways to maintain the probability of winning at $\frac{1}{2}$. Absent such symmetry, it is hard to characterize in general the net effect of a change in policy through both channels. However, we can show that if we perturb the history of policies *and* outcomes together in a way that shuts down the second channel, we recover the results from the baseline. To this aim, we make the following assumption:

Assumption E.1. *Suppose the history of policies is perturbed. Then, we will assume that the history of outcomes is also perturbed, in a way that preserves the equilibrium probability of each party winning the ideological contest. Formally, consider a history (y_t, z_t) with associated equilibrium ideologies β_R, β_L. We define the α-perturbed history $(y_t', z_t'), t \in \{-T, \ldots, -1\}$ with $z_t' \equiv \alpha z_t$ for some $\alpha > 1$, and $y_t' \equiv \frac{y_t + \frac{\beta_R + \beta_L}{2} z_t (\alpha^2 - 1)}{\alpha}$.*

As discussed above, this assumption allows us to capture the effect of perturbations that make the history more informative, without changing the relative favorability of the history to the two parties. The parameter $\alpha > 1$ controls the degree of increase to the extremity (and, therefore, informativeness) of the policies in the history.

Corollary E.1. *In an α-perturbed history, equilibrium ideological polarization decreases as the history of policies becomes more extreme, i.e. as α increases. Formally, $\frac{\partial |\beta_R^* - \beta_L^*|}{\partial \alpha} < 0$.*

Proof. Let β_L, β_R be the equilibrium ideologies under the history (y_t, z_t), and consider the perturbation to (y_t', z_t'). From R's first order condition:

$$\mathbb{P}(R \text{ wins}) = -\frac{\partial \mathbb{P}(R \text{ wins})}{\partial \beta_R} \frac{v_R(\beta_R) - v_R(\beta_L)}{\frac{\partial v_R(\beta_R)}{\partial \beta_R}}$$

A similar expression obtains for L:

$$1 - \mathbb{P}(R \text{ wins}) = -\frac{\partial \mathbb{P}(R \text{ wins})}{\partial \beta_L} \frac{v_L(\beta_R) - v_L(\beta_L)}{\frac{\partial v_L(\beta_L)}{\partial \beta_L}}$$

From the expressions in the main text:

$$\mathbb{P}(R \text{ wins}) = 1 - \Phi(\xi(y_t, z_t)),$$

$$\frac{\partial \mathbb{P}(R \text{ wins})}{\partial \beta_L} = \frac{\partial \mathbb{P}(R \text{ wins})}{\partial \beta_R} = -\phi(\xi(y_t, z_t)) \left(\frac{\sum_{t=-T}^{-1} z_t^2}{2\sigma |z_0|} \right),$$

$$\xi(y_t, z_t) = (\beta_R + \beta_L) \sum_{t=-T}^{0} z_t^2 - 2\beta_0 z_0^2 - 2 \sum_{t=-T}^{-1} z_t y_t$$

We note that our assumption guarantees $\xi(y_t, z_t) = \xi(y_t', z_t')$, which implies $\mathbb{P}(R \text{ wins})$ is unchanged in the perturbation from (y_t, z_t) to (y_t', z_t'). However, crucial for our result is that the derivatives of the winning probability with respect to the parties' strategies, $\frac{\partial \mathbb{P}(R \text{ wins})}{\partial \beta_L}$ or $\frac{\partial \mathbb{P}(R \text{ wins})}{\partial \beta_R}$ (the two are equal due to the symmetry of the expression for the winning probability) strictly increase by a factor α^2.

From R's FOC, this implies either $v_R(\beta'_R) - v_R(\beta'_L) < v_R(\beta_R) - v_R(\beta_L)$ or $\left.\frac{\partial v_R(\beta_R)}{\partial \beta_R}\right|_{\beta'_R} > \left.\frac{\partial v_R(\beta_R)}{\partial \beta_R}\right|_{\beta_R}$. Monotonicity and concavity of v in the region interior to the parties' ideal policies means these conditions are equivalent and imply that R's proposed ideology must move towards L's. A symmetric argument holds for L's strategy, implying $\beta'_R - \beta'_L < \beta_R - \beta_L$. $\qquad\square$

As an aside, we note that, even without perturbing the history of outcomes, making the history of policies sufficiently more extreme always decreases ideological polarization: in the limit as the history becomes arbitrarily informative, both parties converge to β_{OLS}.[2]

E.3   Dynamic Policymaking

A crucial feature of our model is that the voter's lived experiences constrain parties' ideological competition, and thus determine the ideology she adopts in equilibrium. Therefore, parties' policy choices today influence the constraints they will face when trying to shape voters' ideology tomorrow. It is then natural to ask how such dynamic considerations may influence parties' strategic platform positioning in elections. To address this question, we analyze an amended version of the model presented above, where a first platform game precedes the ideological contest. Thus, the timing of the game is as follows:

1. Parties simultaneously propose platforms z_1^R and z_1^L.

2. The voter chooses whom to elect, on the basis of a default ideology β_d.

3. The winner implements the announced policy, and the outcome is privately observed by the voter.

4. The second period begins with the ideological contest: the parties simultaneously propose models β_R and β_L.

5. The voter adopts one of the proposed ideologies, as in the baseline.

6. A new election takes place, where parties simultaneously propose z_2^R and z_2^L.

---

[2]This holds because each party's probability of winning must be bounded away from 0 in equilibrium. This implies $\frac{\partial \mathbb{P}(R \text{ wins})}{\partial \beta_R} \to -\infty$ as $\sum z_t^2 \to \infty$. Therefore, from the first order condition, $v_R(\beta_R) - v_R(\beta_L)$ must go to zero as $\sum z_t^2 \to \infty$, implying $\beta_R \to \beta_L$.

7. The voter decides whom to elect.

8. The winner implements the announced platform, payoffs are realized, and the game ends.

For tractability, we assume that players have absolute-value loss utility over the implemented policy. Thus, party j's per-period payoff is $-|z^j - z_t|$, where z_t is the policy implemented in period t. The voter's utility is also linear, but as above her bliss point is a function of her adopted ideology in period t: $-|z_t^v(\beta_w^z) - z_t|$, with $z_t^v = \gamma\beta_t^z$. Recall that β_1^z is fixed exogenously at the default β_d while β_2^z is the winner of the ideological contest β_w^z. As above, we will assume that the parties' bliss points are sufficiently extreme that the probability of winning, as well as the equilibrium of the platform game, is always interior.

To simplify the analysis we will also impose:

Assumption E.2. $z_1^v = 0$ *(i.e., $\beta_d = 0$).*

This assumption reduces the number of cases to consider, but is not essential for the strategic logic to hold. Parties must consider how their policy choice today influences their probability of winning the upcoming election (with this probability increasing as the policy moves closer to $z_1^v = 0$), *and* what it will do to the history of policies and outcomes that voters will take into account when forming ideologies in the next period.

As usual, we proceed by backwards induction characterizing the equilibrium of the second period platform game:

Lemma E.3. *In equilibrium the two parties win the second-period election with equal probability, and propose policies equidistant from $z_2^v(\beta_w^z)$.*

Proof. Denote $\widehat{z^j}$ party j's policy proposal. Then, the voter elects the right-wing party if and only if

$$u_2^v(\widehat{z^R}) + \xi > u_2^v(\widehat{z^L}).$$

Thus, the right-wing party chooses $\widehat{z^R}$ to maximize

$$-\left(\frac{1}{2} + \psi[u_2^v(\widehat{z^R}) - u_2^v(\widehat{z^L})]\right)|\widehat{z^R} - z^R| - \left(\frac{1}{2} - \psi[u_2^v(\widehat{z^R}) - u_2^v(\widehat{z^L})]\right)|\widehat{z^L} - z^R|,$$

20

and the left-wing party maximizes

$$-\left(\frac{1}{2} - \psi[u_2^v(\widehat{z^R}) - u_2^v(\widehat{z^L})]\right)|\widehat{z^L} - z^L| - \left(\frac{1}{2} + \psi[u_2^v(\widehat{z^R}) - u_2^v(\widehat{z^L})]\right)|\widehat{z^R} - z^L|.$$

As usual in probabilistic models we know that, in equilibrium, $\widehat{z^R} \leq z^R$ and $\widehat{z^L} \geq z^L$. In what follows, we will assume that z^R and z^L are sufficiently extreme to avoid corner solutions. Then, we have that the parties' FOCs are, respectively

$$+\frac{\partial u_2^v(\widehat{z^R})}{\partial \widehat{z^R}}\psi(\widehat{z^R} - \widehat{z^L}) + \left(\frac{1}{2} + \psi[u^v(\widehat{z^R}) - u_2^v(\widehat{z^L})]\right) = 0$$

and

$$+\frac{\partial u_2^v(\widehat{z^L})}{\partial \widehat{z^L}}\psi(\widehat{z^R} - \widehat{z^L}) - \left(\frac{1}{2} - \psi[u_2^v(\widehat{z^R}) - u_2^v(\widehat{z^L})]\right) = 0$$

It is easy to verify that the equilibrium must be symmetric, due to the linearity of the voter's utility function. $\qquad\square$

Next we show that, due to the linearity of the players' utility functions, the equilibrium of the ideological contest is identical to the baseline model with exogenous policies. Denote the equilibrium second-period policy proposals as $z_2^v(\beta_w^z) + \lambda$ and $z_2^v(\beta_w^z) - \lambda$ for the right-wing and left-wing party respectively. Then, the right-wing party chooses β_R^z that maximizes

$$\mathbb{P}_R(\beta_R^z, \beta_L^z)[-\frac{1}{2}(z^R - z_2^v(\beta_R^z) - \lambda) - \frac{1}{2}(z^R - z_2^v(\beta_R^z) + \lambda)]$$
$$+(1 - \mathbb{P}_R(\beta_R^z, \beta_L^z))[-\frac{1}{2}(z^R - z_2^v(\beta_L^z) - \lambda) - \frac{1}{2}(z^R - z_2^v(\beta_L^z) + \lambda)],$$

which reduces to

$$-(z^R - z_2^v(\beta_L^z)) + \mathbb{P}_R(\beta_R^z, \beta_L^z)(z_2^v(\beta_R^z) - z_2^v(\beta_L^z)),$$

where $\mathbb{P}_R$ is as in the baseline model. We can similarly compute the objective function for the left-wing party.

Notice that, since $z_2^v(\beta_w^z)$ is linearly increasing in β_w^z, the parties' strategic problem is exactly as in the baseline, and all the results describing the equilibrium of the ideological contest continue to hold. Thus, the equilibrium is symmetric, and ideological polarization increases in the extremism of the policies in the history.

Finally, we analyze the parties' choice in the first election. Denote $\widehat{\beta}_t$ the OLS estimate for β_0 given the history observed up to period t (so that $\widehat{\beta}_1$ is the OLS estimate at the beginning of the game, and $\widehat{\beta}_2$ is the estimate following the realization of the first-period policy outcome). The party that wins the election in period 1 gets to add an additional observation of policy and outcome to the history. The larger is the implemented z, the more voters will learn, in expectation, about the true β_0. These dynamic considerations shape the equilibrium of the first-period platform game:

Proposition E.2.

- *Suppose that $\beta_0 = \widehat{\beta}_1$. Then, in equilibrium $z_1^R = -z_1^L$ and R wins the first-period election with probability $P_1 = \frac{1}{2}$.*

- *Suppose that $\widehat{\beta}_1 > \beta_0$, then in equilibrium $z_1^R < -z_1^L$ and R wins the first-period election with probability $P_1 > \frac{1}{2}$.*

- *Suppose instead that $\widehat{\beta}_1 < \beta_0$. Then, in equilibrium $z_1^R > -z_1^L$ and R wins the first-period election with probability $P_1 < \frac{1}{2}$.*

Proof. We begin by describing how the first-period policy influences the parties' expected continuation value. Recall that $\widehat{\beta}_2$ is the OLS estimate as a function of the history observed prior to the second period. We have:

Lemma E.4. $E[\widehat{\beta}_2] = (1 - w(|z_1|))\widehat{\beta}_1 + w(|z_1|)\beta_0$, where $w(|z|)$ is everywhere positive, increasing in $|z|$, $w(0) = 0$, and $\lim_{|z| \to \infty} w(|z|) = 1$.

Proof.

$$\widehat{\beta}_2 = \frac{\sum_{-T}^{1} z_t y_t}{\sum_{-T}^{1} z_t^2}$$

$$= \frac{\sum_{-T}^{0} z_t y_t + z_1 y_1}{\sum_{-T}^{0} z_t^2 + z_1^2}$$

$$= \frac{\sum_{-T}^{0} z_t^2}{\sum_{-T}^{0} z_t^2 + z_1^2} \widehat{\beta}_1 + \frac{z_1 y_1}{\sum_{-T}^{0} z_t^2 + z_1^2}$$

$$E[\widehat{\beta}_2] = \frac{\sum_{-T}^{0} z_t^2}{\sum_{-T}^{0} z_t^2 + z_1^2} \widehat{\beta}_1 + \frac{z_1^2}{\sum_{-T}^{0} z_t^2 + z_1^2} \beta_0$$

The postulated properties follow immediately by inspection, with $w(|z_1|) = \frac{z_1^2}{\sum_{-T}^{0} z_t^2 + z_1^2}$. $\qquad\square$

Recall that, as established in the baseline model, the right-wing party's expected continuation value is linearly increasing in $\widehat{\beta}_2$. The opposite is true for the left-wing party. Thus, a party which is advantaged by the truth, in the sense that β_0 is more favorable to the party than the current OLS, will have an incentive to facilitate learning by choosing more extreme policy in period 1, in expectation moving the next period OLS closer to β_0. A party which is disadvantaged will have an incentive to suppress learning by choosing more moderate policy to keep the next period OLS closer to the current one. Importantly, only the magnitude of the implemented policy (and not its direction) matters for what voters expect to learn following period 1.

We note first that a direct consequence of Lemma E.4 is that if $\widehat{\beta}_1 = \beta_0$, then the dynamic considerations are irrelevant (regardless of z_1, the expected second period OLS $\widehat{\beta}_2$ is the same). The game is thus identical to the static platform game, which we previously showed to have a unique symmetric equilibrium.

We next consider the case where $\widehat{\beta}_1 > \beta_0$. Party R is advantaged in this case, as the first-period history is more favorable to R than the truth. We let $v_i(z_1)$ denote player i's continuation value from the equilibrium of the subsequent ideology and platform game as a function of the first-period implemented policy z_1, noting that z_1 affects v_i only through its influence on the expected OLS $\widehat{\beta}_2$ of the dataset that voters in period 2 will observe when evaluating ideologies. The parties' maximization problems are, respectively:

$$\max_{z_1^R} -P_1|z_1^R - z^R| - (1 - P_1)|z_1^L - z^R| + P_1 E[v_R(\widehat{\beta}_2(z_1^R))] + (1 - P_1)E[v_R(\widehat{\beta}_2(z_1^L))]$$

$$\max_{z_1^L} -P_1|z_1^R - z^L| - (1 - P_1)|z_1^L - z^L| + P_1 E[v_L(\widehat{\beta}_2(z_1^R))] + (1 - P_1)E[v_L(\widehat{\beta}_2(z_1^L))]$$

Taking first-order conditions[3] and rearranging, we have that:

$$P_1\left[\frac{\partial v_R}{\partial z_1^R} + 1\right] + \frac{\partial P_1}{\partial z_1^R}\left[z_1^R - z_1^L + E[v_R(\widehat{\beta}_2(z_1^R))] - E[v_R(\widehat{\beta}_2(z_1^L))]\right] = 0 \qquad (23)$$

$$(1 - P_1)\left[\frac{\partial v_L}{\partial z_1^L} - 1\right] + \frac{\partial P_1}{\partial z_1^L}\left[z_1^L - z_1^R + E[v_L(\widehat{\beta}_2(z_1^R))] - E[v_L(\widehat{\beta}_2(z_1^L))]\right] = 0 \qquad (24)$$

Conjecture a symmetric equilibrium in period 1 strategies, i.e. $z_1^L = -z_1^R$, and $P_1 = \frac{1}{2}$. Then, using the symmetry property of the weights in Lemma (E.4), $\widehat{\beta}_2(z_1^R) = \widehat{\beta}_2(z_1^L)$. Further, linearity implies $\frac{\partial P_1}{\partial z_1^L} = \frac{\partial P_1}{\partial z_1^R}$, and thus adding (23) and (24) reduces to:

$$P_1\left[\frac{\partial v_R}{\partial z_1^R} - \frac{\partial v_L}{\partial z_1^L}\right] + \frac{\partial v_L}{\partial z_1^L} + 2P_1 - 1 = 0$$

$$\frac{\partial v_R}{\partial z_1^R} + \frac{\partial v_L}{\partial z_1^L} = 0$$

But this is a contradiction because increasing z_1^R away from zero reduces $\widehat{\beta}_2$ by Lemma E.4 and hence hurts R (i.e., $\frac{\partial v_R}{\partial z_1^R} < 0$), while increasing z_1^L towards zero increases $\widehat{\beta}_2$ and hence hurts L (i.e., $\frac{\partial v_L}{\partial z_1^L} < 0$). Thus, the equilibrium cannot be symmetric.

Abandoning the symmetry assumption and examining L's first order condition (24), it is clear that L must be better off in period 2 if z_1^L is implemented than if z_1^R is implemented (else the left hand side is strictly negative, since $\frac{\partial P_1}{\partial z_1^L} < 0$ and $\frac{\partial v_L}{\partial z_1^L} < 0$). Given Lemma E.4 and the assumption that $\widehat{\beta}_1 > \beta_0$, this implies z_1^L must be more extreme than z_1^R, and therefore that L wins with probability less than $\frac{1}{2}$ in period 1, i.e. $P_1 > \frac{1}{2}$.

Exactly symmetric reasoning holds for the case where $\widehat{\beta}_1 < \beta_0$.

---

[3]Under the assumption that the parties' ideal points are extreme enough to guarantee an interior solution, i.e. $z^R > z_1^R \geq z_1^L > z^L$.

E.4   Dynamic Ideological Persuasion

While our previous extension analyzes how dynamic considerations may influence parties' policy stances, a distinct although related argument is that parties may look beyond the upcoming election when designing the ideologies to present to the voter. We examine extensions to the model of this form in this section.

E.4.1   Repeated Game

First, we consider an extended version of the model where the game is repeated over two periods. Thus, in each period the parties engage in an ideological contest, the voter adopts the model that best fits the history, and then chooses whether to elect R or L. As in the baseline model, the parties' policies are exogenously fixed at their respective bliss points, z^R and z^L.

Here, the dataset that the voter observes prior to the second period ideological contest is a function of the policy implemented in the first period, and thus the outcome of the first election. This generates a dynamic link across periods. When engaging in the ideological battle in the first period, parties consider how the voter's adopted ideology will influence the outcome of the upcoming election and how that, in turn, impacts the equilibrium of the second-period ideological contest.

Nonetheless, we show that such dynamic considerations do not alter our qualitative results from the baseline model. The reason is that even in this dynamic model, the game is zero-sum. Parties have identical, although directionally opposite, incentives to pull the voter's ideology to the extreme. Thus, they face a symmetric problem and the equilibrium of the ideological contest is symmetric as well, in the first as in the second period. This then drives comparative static results identical to those established in the baseline model.

To see this, consider the parties' maximization problem. In the second period, the game is identical to the baseline analyzed in the main body. Moving back to the first period, denote as $v_j(\widehat{\beta}_2 | z_1)$ party j's expected second-period continuation value, as a function of the expected second-period OLS estimate for the coefficient of z on y, $\widehat{\beta}_2$. As mentioned above, this estimate

depends on the policy that is implemented in the first period, z_1. Recall that $\mathbb{P}_R$ is the probability that R wins the first-period ideological contest. Further, denote $P_t(\beta_t)$ the probability that R wins the period-t election, as a function of the voter's ideology β_t.

Then, we have that in the first-period R chooses β_R to maximize

$$\mathbb{P}_R\Big(-(z^R-z^L)^2\big(1-P_1(\beta_R)\big)+P_1(\beta_R)v_R(\widehat{\beta}_2|z^R)+\big(1-P_1(\beta_R)\big)v_R(\widehat{\beta}_2|z^L)\Big) \tag{25}$$

$$\tag{26}$$

$$nonumber +(1-\mathbb{P}_R)\Big(-(z^R-z^L)^2 P_1(\beta_L)+P_1(\beta_L)v_R(\widehat{\beta}_2|z^R)+\big(1-P_1(\beta_L)\big)v_R(\widehat{\beta}_2|z^L)\Big). \tag{27}$$

Analogously, β_L maximizes

$$\mathbb{P}_R\Big(-(z^R-z^L)^2 P_1(\beta_R)+P_1(\beta_R)v_L(\widehat{\beta}_2|z^R)+\big(1-P_1(\beta_R)\big)v_L(\widehat{\beta}_2|z^L)\Big) \tag{28}$$

$$\tag{29}$$

$$nonumber +(1-\mathbb{P}_R)\Big(-(z^R-z^L)^2 P_1(\beta_L)+P_1(\beta_L)v_L(\widehat{\beta}_2|z^R)+\big(1-P_1(\beta_L)\big)v_L(\widehat{\beta}_2|z^L)\Big). \tag{30}$$

From this, the FOCs are, respectively

$$\left[\frac{\partial\mathbb{P}_R}{\partial\beta_R}\Big(P_1(\beta_R)-P_1(\beta_L)\Big)+\mathbb{P}_R\frac{\partial P_1}{\partial\beta_R}\right]\Big(v_R\left(\widehat{\beta}_2|z^R\right)-v_R(\widehat{\beta}_2|z^L)+(z^R-z^L)^2\Big)=0, \tag{31}$$

and

$$\left[\frac{\partial\mathbb{P}_R}{\partial\beta_L}\Big(P_1(\beta_R)-P_1(\beta_L)\Big)+(1-\mathbb{P}_R)\frac{\partial P_1}{\partial\beta_L}\right]\Big(v_L\left(\widehat{\beta}_2|z^R\right)-v_L(\widehat{\beta}_2|z^L)-(z^R-z^L)^2\Big)=0. \tag{32}$$

Recall that $\frac{\partial P_1}{\partial\beta_R}=\frac{\partial P_1}{\partial\beta_L}$ and $\frac{\partial\mathbb{P}_R}{\partial\beta_R}=\frac{\partial\mathbb{P}_R}{\partial\beta_L}$. Further, $v_R=-(z^R-z^L)^2\big(1-P_2(\widehat{\beta}_2|z_1)\big)$ and $v_L=-(z^R-z^L)^2 P_2(\widehat{\beta}_2|z_1)$, which implies $v_L(\widehat{\beta}_2|z^R)-v_L(\widehat{\beta}_2|z^L)=-[v_R(\widehat{\beta}_2|z^R)-v_R(\widehat{\beta}_2|z^L)]$. Plugging this in, and adding the FOCs, we get that in equilibrium

26

$$(2\mathbb{P}_R - 1)\frac{\partial P_1}{\partial \beta_R}\left(v_R\left(\widehat{\beta}_2|z^R\right) - v_R(\widehat{\beta}_2|z^L) + (z^R - z^L)^2\right) = 0, \tag{33}$$

which reduces to

$$(2\mathbb{P}_R - 1)\frac{\partial P_1}{\partial \beta_R}(z^R - z^L)^2\left(P_2(\widehat{\beta}_2|z^R) - P_2(\widehat{\beta}_2|z^L) + 1\right) = 0. \tag{34}$$

This can only be satisfied if $\mathbb{P}_R = \frac{1}{2}$. Thus, in equilibrium the two parties must win the first-period ideological contest with equal probability, locating equidistant from the first-period expected OLS. The proof of Proposition 2, and the corresponding corollaries, then proceed exactly as in the baseline model.

E.4.2 Ideological Reassessment

Second, we consider a version of the game where there is some potential for durability in ideology: that ideologies, if successful, may survive for multiple elections. In particular, parties may want to propose a model of the world that not only persuades the voter today, but is also likely to survive as the voter's experiences evolve over time.

Here, we allow the voter to potentially reassess her ideology as she experiences more policies and outcomes. We analyze how this influences parties' strategies in the ideological contest and thus the ideology initially adopted by the voter. Formally, we analyze a game with the following sequence of play:

1. Ideological contest: parties simultaneously propose models of the world β_R and β_L

2. The voter privately observes outcome realization y_0, and chooses which ideology to adopt.

3. First-period election with exogenous platforms $z_R > z_L$.

4. The voter privately observes the outcome of the implemented policy y_1

5. The second period begins with the reassessment stage: the voter retains her ideology from the first stage if the likelihood ratio $\mathcal{L}(y_1 \mid z, \beta_w)/\mathcal{L}(y_1 \mid z, \beta_{-w}) > 1$. Otherwise, reassessment occurs and there is a new ideological contest, which proceeds as in stage 1.

6. Second-period election with exogenous platforms $z_R > z_L$.

7. Payoffs realize and the game ends.

For simplicity, we will assume that the parties' exogenous platforms are symmetric around 0: $z_R = -z_L = z$. The players' utility is as in the baseline model in the main body (although identical results would obtain if the voter's utility was defined over policies as in section E.2).

We begin by characterizing how the outcome of the first-period ideological contest and election stages influence the second period game. The symmetry of the parties' policy commitments around 0 simplifies the analysis, as the following two lemmas show.

Lemma E.5. $\mathbb{P}(R \text{ is reassessed} \mid R \text{ wins ideological contest}) = 1 - \mathbb{P}(L \text{ is reassessed} \mid L \text{ wins ideological contest})$. *Denoting the probability that R is reassessed by $\rho \equiv \mathbb{P}(R \text{ is reassessed} \mid R \text{ wins ideological contest})$, ρ does not depend on the winner of the election stage. Furthermore, $\frac{\partial \rho}{\partial \beta_R} = \frac{\partial \rho}{\partial \beta_L}$.*

Proof. First consider the case where R wins the first period ideological contest. Denote by z_1 the policy implemented in the first period. The probability of reassessment is

$$\mathbb{P}(\mathcal{L}(y_1 \mid z_1, \beta_R)/\mathcal{L}(y_1 \mid z_1, \beta_L) > 1)$$
$$=\mathbb{P}(\log \mathcal{L}(y_1 \mid z, \beta_R) - \log \mathcal{L}(y_1 \mid z, \beta_L) > 0)$$
$$=\mathbb{P}((y_1 - \beta_R z_1)^2 - (y_1 - \beta_L z_1)^2 > 0)$$
$$=\mathbb{P}(z_1^2(\beta_R + \beta_L - 2\beta_0) - 2z_1\epsilon > 0)$$
$$=\begin{cases} \mathbb{P}\left(\left(\epsilon < z_1\left(\frac{\beta_R + \beta_L}{2} - \beta_0\right)\right)\right) & z_1 > 0 \\ \mathbb{P}\left(\left(\epsilon > z_1\left(\frac{\beta_R + \beta_L}{2} - \beta_0\right)\right)\right) & z_1 < 0 \end{cases}$$

By symmetry of the normal distribution, and given our assumption that $z^L = -z^R = z$, these two probabilities are the same. We can thus write the reassessment probability as:

$$\rho \equiv \Phi \left(\frac{|z| \left(\frac{\beta_R + \beta_L}{2} - \beta_0 \right)}{\sigma} \right)$$

The statement that the derivatives with respect to β_R and β_L are equal follows by inspection. A symmetric argument holds for the case where L wins the first period ideological contest. □

Lemma E.6. *The outcome of the first-period election stage has no effect on the parties' second stage continuation values. The parties' expected continuation values in the second period are given by:*

$$v_R = \begin{cases} 4z^2(-\frac{1}{2} + 2\psi z \tilde{\beta}_2^R) & \text{R wins the ideological contest and is reassessed} \\ 4z^2(-\frac{1}{2} + 2\psi z \beta_R) & \text{R wins the ideological contest and is not reassessed} \\ 4z^2(-\frac{1}{2} + 2\psi z \tilde{\beta}_2^L) & \text{L wins the ideological contest and is reassessed} \\ 4z^2(-\frac{1}{2} + 2\psi z \beta_L) & \text{L wins the ideological contest and is not reassessed} \end{cases} \tag{35}$$

$$v_L = \begin{cases} 4z^2(-\frac{1}{2} - 2\psi z \tilde{\beta}_2^R) & \text{R wins the ideological contest and is reassessed} \\ 4z^2(-\frac{1}{2} - 2\psi z \beta_R) & \text{R wins the ideological contest and is not reassessed} \\ 4z^2(-\frac{1}{2} - 2\psi z \tilde{\beta}_2^L) & \text{L wins the ideological contest and is reassessed} \\ 4z^2(-\frac{1}{2} - 2\psi z \beta_L) & \text{L wins the ideological contest and is not reassessed} \end{cases} \tag{36}$$

Where β_R and β_L are R and L's proposed ideologies in the first period, respectively. $\tilde{\beta}_2^R$ and $\tilde{\beta}_2^L$ are the expected values of the second period OLS estimate of the effect of z on y, conditional on R or L being reassessed, respectively.

Proof. First, note that the first period affects the second period only through the expected OLS estimate of the effect of z_t on y_t in the history at the beginning of the second period. This expectation is the same regardless of who wins the first period election, by Lemma E.4 applied to the situation here with symmetric policies. Combined with the previous lemma showing that the election-stage winner has no effect on the probability of reassessment, this shows that the election stage has no effect on second-period continuation values.

The cases we need to consider are therefore the combinations of first-period ideological stage winner, and the occurrence of reassessment. Recall that given the voter's adopted ideology β^z, R's probability of winning the second-period election is $\frac{1}{2} + \psi(z^R - z^L)\beta^z$, or $\frac{1}{2} + 2\psi z \beta^z$ in the case of symmetric policies. R's expected utility in the election stage conditional on β^z is thus:

$$-4z^2 \left(1 - \left(\frac{1}{2} + 2\psi z \beta^z \right) \right)$$
$$= 4z^2 \left(-\frac{1}{2} + 2\psi z \beta^z \right)$$

Similarly, L's expected utility in the election stage conditional on β^z is:

$$-4z^2 \left(\left(\frac{1}{2} + 2\psi z \beta^z \right) \right)$$
$$= 4z^2 \left(-\frac{1}{2} - 2\psi z \beta^z \right)$$

In the cases where there is no reassessment, the voter simply retains the first-period ideological stage winner's proposed ideology, i.e. β_R or β_L. This yields utility of $4z^2 \left(-\frac{1}{2} \pm 2\psi z \beta_R \right)$ or $4z^2 \left(-\frac{1}{2} \pm 2\psi z \beta_L \right)$, respectively.

In the cases where there is reassessment, the continuation game is exactly the one analyzed in the main text. In this game, each party wins with probability one-half, and proposed ideologies are centered around the OLS estimate of the effect of z on y in the history at the beginning of period 2. Because both parties' utilities are linear in the voter's adopted ideology, in expectation this yields payoffs of $4z^2 \left(-\frac{1}{2} \pm 2\psi z \tilde{\beta}_2^R \right)$ and $4z^2 \left(-\frac{1}{2} \pm 2\psi z \tilde{\beta}_2^L \right)$ in cases when R or L is reassessed, respectively. $\qquad \square$

As a final preliminary result, we unpack the conditional expectation of the second-period OLS in cases where there is reassessment. Note that the reassessment event depends on the realization of the outcome y in period 1: R is reassessed when the realization under-shoots the voter's expectation, and L is reassessed when the outcome over-shoots it. Therefore, the conditional expectation of the second period OLS is *below* its unconditional mean when R is reassessed,

and *above* its unconditional mean when L is reassessed. Furthermore, the choice of proposed ideologies affects not only the probability of reassessment but also the parties' continuation values in the event that reassessment occurs. The following lemma shows that nonetheless, the marginal influence of each party's proposed first-period ideology on the conditional expectations $\tilde{\beta}_2^R$ and $\tilde{\beta}_2^L$ is the same, which simplifies the analysis of equilibrium.

Lemma E.7. $\frac{\partial \tilde{\beta}_2^R}{\partial \beta_R} = \frac{\partial \tilde{\beta}_2^R}{\partial \beta_L}$, *and* $\frac{\partial \tilde{\beta}_2^L}{\partial \beta_R} = \frac{\partial \tilde{\beta}_2^L}{\partial \beta_L}$.

Proof. By Lemma E.4, the expected value $E[\tilde{\beta}_2^R] = \frac{\sum_{-T}^0 z_t^2}{\sum_{-T}^0 z_t^2 + z_1^2} \widehat{\beta}_1 + \frac{z_1}{\sum_{-T}^0 z_t^2 + z_1^2} E[y_1 \mid \text{R reassessed}]$. R is reassessed if and only if the log-likelihood ratio of y_1 under R's model to that under L's model falls below zero, i.e.:

$$-(y_1 - \beta_R z_1)^2 + (y_1 - \beta_L z_1)^2 < 0$$

Which, after rearrangement, yields:

$$y_1 < \frac{\beta_R + \beta_L}{2} z_1, z_1 > 0$$
$$y_1 > \frac{\beta_R + \beta_L}{2} z_1, z_1 < 0$$

Plugging in the model $y_1 = \beta_0 z_1 + \epsilon$, we see that in the positive z_1 case, this corresponds to the event $\epsilon < \left(\frac{\beta_L + \beta_R}{2} - \beta_0\right) z_1$, and in the negative z_1 case it corresponds to the event $\epsilon > \left(\frac{\beta_L + \beta_R}{2} - \beta_0\right) z_1$. Again, given the symmetry assumptions on the parties' policy commitments and the symmetry of the normal distribution, these probabilities are the same. For the positive z_1 case we can write:

$$E[y_1 \mid \text{R reassessed}] = z_1 \beta_0 + \frac{\int_{-\infty}^{\left(\frac{\beta_L + \beta_R}{2} - \beta_0\right) z_1} \epsilon f_\epsilon(\epsilon) d\epsilon}{1 - F_\epsilon\left(\left(\frac{\beta_L + \beta_R}{2} - \beta_0\right) z_1\right)}$$

$$\Rightarrow E[\tilde{\beta}_2^R] = \frac{\sum_{-T}^0 z_t^2}{\sum_{-T}^0 z_t^2 + z_1^2} \widehat{\beta}_1 + \frac{\beta_0 z_1^2}{\sum_{-T}^0 z_t^2 + z_1^2} + \frac{z_1}{\sum_{-T}^0 z_t^2 + z_1^2} \frac{\int_{-\infty}^{\left(\frac{\beta_L + \beta_R}{2} - \beta_0\right) z_1} \epsilon f_\epsilon(\epsilon) d\epsilon}{1 - F_\epsilon\left(\left(\frac{\beta_L + \beta_R}{2} - \beta_0\right) z_1\right)}$$

The negative z_1 case is the same because the sign change of z_1 cancels the reflection of the conditional expectation around zero. Both parties' proposed ideologies enter only in the form $\beta_R + \beta_L$, and therefore the derivatives with respect to each party's proposed ideology are the same. An identical argument applies in the case when L is reassessed. $\qquad\square$

Proposition E.3. *In equilibrium, R wins the first stage ideological contest with probability*

$$\mathbb{P}(R \text{ wins}) = \frac{1+\rho}{3}$$

Where ρ is the probability that R is reassessed conditional on winning the first period contest defined above. The average of the two ideologies proposed in the first period satisfies:

$$\frac{\beta_R + \beta_L}{2} = \begin{cases} \tilde{\beta}_1 + \frac{\Phi^{-1}\left(\frac{2-\rho}{3}\right)}{\sum_{t=-T}^{0} z_t^2} & z_0 > 0 \\ \tilde{\beta}_1 + \frac{\Phi^{-1}\left(\frac{1+\rho}{3}\right)}{\sum_{t=-T}^{0} z_t^2} & z_0 < 0 \end{cases}$$

Where $\tilde{\beta}_1$ is the expected OLS in the first period.

Proof. By Lemma E.6, the probability of reassessment does not depend on the election-stage winner. The lemma also gives the continuation values in each combination of payoff-relevant outcomes in the first period. For notational convenience, denote these values by $v_i^{j,r}$, with $i \in \{L, R\}$ denoting the party, $j \in \{L, R\}$ denoting the winner of the first-period ideological contest, and $r \in \{0, 1\}$ denoting the occurrence of reassessment. Letting $P_R = \mathbb{P}(R \text{ wins})$, the two parties' objectives are, respectively:

$$\max_{\beta_R} - 2z^2 + 8P_R\psi z^3(\beta_R - \beta_L) + 8\psi z^3 \beta_L +$$
$$P_R(1-\rho)v_R^{R,0} + P_R\rho v_R^{R,1} + (1-P_R)\rho v_R^{L,0} + (1-P_R)(1-\rho)v_R^{L,1}$$
$$\max_{\beta_L} - 2z^2 - 8P_R\psi z^3(\beta_R - \beta_L) - 8\psi z^3 \beta_L +$$
$$P_R(1-\rho)v_L^{R,0} + P_R\rho v_L^{R,1} + (1-P_R)\rho v_L^{L,0} + (1-P_R)(1-\rho)v_L^{L,1}$$

And, taking first order conditions, we get:

$$8\psi z^3 \left(P_R + \frac{\partial P_R}{\partial \beta_R}(\beta_R - \beta_L) \right)$$

$$+ \frac{\partial P_R}{\partial \beta_R}(1-\rho)v_R^{R,0} - P_R\frac{\partial \rho}{\partial \beta_R}v_R^{R,0} + P_R(1-\rho)\frac{\partial v_R^{R,0}}{\partial \beta_R}$$

$$+ \frac{\partial P_R}{\partial \beta_R}\rho v_R^{R,1} + P_R\frac{\partial \rho}{\partial \beta_R}v_R^{R,1} + P_R\rho\frac{\partial v_R^{R,1}}{\partial \beta_R}$$

$$- \frac{\partial P_R}{\partial \beta_R}\rho v_R^{L,0} + (1-P_R)\frac{\partial \rho}{\partial \beta_R}v_R^{L,0} + (1-P_R)\rho\frac{\partial v_R^{L,0}}{\partial \beta_R}$$

$$- \frac{\partial P_R}{\partial \beta_R}(1-\rho)v_R^{L,1} - (1-P_R)\frac{\partial \rho}{\partial \beta_R}v_R^{L,1} + (1-P_R)(1-\rho)\frac{\partial v_R^{L,1}}{\partial \beta_R} = 0$$

For R, and similarly

$$-8\psi z^3 \left(1 - P_R + \frac{\partial P_R}{\partial \beta_L}(\beta_R - \beta_L) \right)$$

$$+ \frac{\partial P_R}{\partial \beta_L}(1-\rho)v_L^{R,0} - P_R\frac{\partial \rho}{\partial \beta_L}v_L^{R,0} + P_R(1-\rho)\frac{\partial v_L^{R,0}}{\partial \beta_L}$$

$$+ \frac{\partial P_R}{\partial \beta_L}\rho v_L^{R,1} + P_R\frac{\partial \rho}{\partial \beta_L}v_L^{R,1} + P_R\rho\frac{\partial v_L^{R,1}}{\partial \beta_L}$$

$$- \frac{\partial P_R}{\partial \beta_L}\rho v_L^{L,0} + (1-P_R)\frac{\partial \rho}{\partial \beta_L}v_L^{L,0} + (1-P_R)\rho\frac{\partial v_L^{L,0}}{\partial \beta_L}$$

$$- \frac{\partial P_R}{\partial \beta_L}(1-\rho)v_L^{L,1} - (1-P_R)\frac{\partial \rho}{\partial \beta_L}v_L^{L,1} + (1-P_R)(1-\rho)\frac{\partial v_L^{L,1}}{\partial \beta_L} = 0$$

For L. We note first that $\frac{\partial v_i^{-i,0}}{\partial \beta_i} = 0$ for both i. In addition, we take advantage of the facts that $\frac{\partial \rho}{\partial \beta_R} = \frac{\partial \rho}{\partial \beta_L}$ (Lemma E.5), that $\frac{\partial P_R}{\partial \beta_R} = \frac{\partial P_R}{\partial \beta_L}$ (Proposition 2 in the main text), that $v_R^{i,r} + v_L^{i,r} = -4z^2$ for all i, r (Lemma E.6) and that $\frac{\partial v_L^{i,1}}{\partial \beta_L} = -\frac{\partial v_R^{i,1}}{\partial \beta_R}$ (Lemma E.6). These simplify the sum of the two equations, which reduces to:

$$8\psi z^3 \left(2P_R - 1 + P_R(1-\rho) - \rho(1-P_R) \right) = 0$$

And rearranging yields:

34

$$P_R = \frac{1 + \rho}{3}. \tag{37}$$

Noting that $P_R = 1 - \Phi(\xi(y_t, z_t))$ when $z_0 > 0$ and $P_R = \Phi(\xi(y_t, z_t))$ when $z_0 < 0$, this relationship implies that $\xi(y_t, z_t) = \Phi^{-1}\left(\frac{2-\rho}{3}\right)$ or $\xi(y_t, z_t) = \Phi^{-1}\left(\frac{1+\rho}{3}\right)$), depending on the sign of z_0. Plugging in for $\xi(y_t, z_t)$ in the first case gives:

$$\Phi^{-1}\left(\frac{2-\rho}{3}\right) = (\beta_R + \beta_L)\sum_{t=-T}^{0} z_t^2 - 2\beta_0 z_0^2 - 2\sum_{t=-T}^{-1} z_t y_t$$

$$\frac{\beta_R + \beta_L}{2} = \frac{\beta_0 z_0^2 + \sum_{t=-T}^{-1} z_t y_t}{\sum_{t=-T}^{0} z_t^2} + \frac{\Phi^{-1}\left(\frac{2-\rho}{3}\right)}{\sum_{t=-T}^{0} z_t^2}$$

$$= \tilde{\beta}_1 + \frac{\Phi^{-1}\left(\frac{2-\rho}{3}\right)}{\sum_{t=-T}^{0} z_t^2}$$

An analogous result holds for the $z_0 < 0$ case. $\qquad\square$

The possibility of reassessment thus generates an asymmetry in the first stage ideological contest. In particular, the candidate who is favored by the history (i.e. for whom $\tilde{\beta}_1$ is preferred to β_0) has an incentive to avoid reassessment, and therefore to propose more moderate ideology than she would in the static game. The incentive for the party that is disfavored by the history is exactly the opposite, and this party wants to encourage reassessment by proposing a more extreme ideology.

Finally, we show that the results on the determinants of ideological polarization are analogous to those emerging in the static model. As for the case of our extension in **E.2**, lack of symmetry in the ideological stage implies that making the history of policies more extreme may influence the parties' strategic incentives via two channels. First, the history becomes more informative. Second, the OLS estimate for β_0^z may change as well. The dynamic game adds additional complexity because the behavior depends on the OLS estimate in the past as well as its expectation in the future (period 2). Unlike in the **E.2** case, there is no perturbation that preserves both the probability of winning and the second-period expected OLS, which makes the comparative statics for small

changes hard to characterize. Nonetheless, the behavior for large changes in the extremism of past policies is the same: as the past history becomes informative enough, polarization decreases.

Proposition E.4. *Consider an initial equilibrium pair of first-period ideologies (β_R, β_L) given a history (y_t, z_t) $t \in \{-T, \ldots, -1\}$. Suppose we increase the extremism of the past history by multiplying all past policies z_t by a factor γ, i.e. $z'_t \equiv \gamma z_t$, where $\gamma > 1$. For sufficiently large γ, equilibrium ideological polarization is smaller under (y_t, z'_t) than under (y_t, z_t).*

Proof. We begin by examining party R's first order condition defined in Proposition E.3, which after some rearrangement of terms is:

$$
\begin{aligned}
8\psi z^3 & \left(P_R + \frac{\partial P_R}{\partial \beta_R}(\beta_R - \beta_L) \right) \\
& + \frac{\partial P_R}{\partial \beta_R} \left[(1-\rho)(v_R^{R,0} - v_R^{L,1}) + \rho(v_R^{R,1} - v_R^{L,0}) \right] \\
& + \frac{\partial \rho}{\partial \beta_R} \left[P_R(v_R^{R,1} - v_R^{R,0}) + (1-P_R)(v_R^{L,0} - v_R^{L,1}) \right] \\
& + P_R(1-\rho)\frac{\partial v_R^{R,0}}{\partial \beta_R} + P_R\rho\frac{\partial v_R^{R,1}}{\partial \beta_R} + (1-\rho)(1-P_R)\frac{\partial v_R^{L,1}}{\partial \beta_R} = 0
\end{aligned}
$$

We note that the second and third lines in the above are both negative, and the third line is bounded in the interval $[\frac{8}{3}\psi z^3, \frac{16}{3}\psi z^3]$ as γ becomes large. To see the first statement, note that $v_R^{R,0}$ must be at least as large as $v_R^{L,1}$ because R can always guarantee L's reassessment by proposing the same ideology as L. Because this is a constant-sum game, the same logic applied to L implies $v_R^{R,1}$ must be at least as large as $v_R^{L,0}$. Similarly $v_R^{R,1} \geq v_R^{R,0}$ and $v_R^{L,0} \geq v_R^{L,1}$ because otherwise one player would have a profitable deviation to move to a more extreme position, increasing the likelihood of reassessment. $\frac{\partial \rho}{\partial \beta_R} > 0$ and $\frac{\partial P_R}{\partial \beta_R} < 0$ by inspection.

To see the second statement, note that $\frac{\partial v_R^{R,0}}{\partial \beta_R} = 8\psi z^3$. Proposition E.3 bounds the product $P_R(1-\rho)$ in the range $[\frac{1}{3}, \frac{2}{3}]$. The last and second to last terms go to zero as γ gets large, because the second period OLS converges to the first period OLS as the variance in the history increases (Lemma E.7).

Because $\frac{\partial P_R}{\partial \beta_R}$ is always negative and gets unboundedly large in magnitude as γ increases, the difference between β_R and β_L must shrink at some point to preserve the equality. $\square$

36