Cause and Effect in Political Polarization: A Dynamic Analysis

Steven Callander

Stanford University

Juan Carlos Carbajal

University of New South Wales

Polarization is both a description of the current state of politics and a dynamic path that has rippled across the political domain over decades. We provide a simple model that explains why polarization appears incrementally and why it was elites who polarized first and more dramatically, whereas mass polarization came later and has been less pronounced. We incorporate an ostensibly unrelated finding about how voters form preferences into a dynamic model of elections. This change, when combined with the response of strategic candidates, creates a feedback loop that can replicate many features of the data. We explore the model's implications for other aspects of politics and trace what it predicts for the future of polarization.

We have benefited from the helpful comments of Avi Acharya, Dave Baron, Gabriel Carrol, Charlotte Cavaille, Dana Foarta, Jon Eguia, Gabriele Gratton, Andy Hall, Matt Jackson, Keith Krehbiel, Andrew Little, Andrea Mattozzi, Nolan McCarty, Kirill Pogorelskiy, Emir Kamenica, two anonymous referees, and seminar audiences at the Stanford Graduate School of Business, the University of Warwick, the University of Sydney, the London School of Economics, the Econometric Society Summer Meetings, the Australasian Economic Theory Workshop, and the Australian Political Economy Network. Jason Lambe provided excellent research assistance. Carbajal gratefully acknowledges financial support from the Australian Research Council under grant DP190101718. This paper was edited by Emir Kamenica.

Electronically published March 9, 2022

Journal of Political Economy, volume 130, number 4, April 2022.

https://doi.org/10.1086/718200

I. Introduction

Political polarization is an important and enduring puzzle. A big part of the challenge to explain polarization is that it is not just a single thing. Polarization is both a description of the current state of politics and a dynamic path that has rippled across the political domain over multiple decades. Polarization is, as the sociologists DiMaggio, Evans, and Bryson (1996, 693) put it, "both a state and a process."

Adding to the complexity is that polarization's dynamic path has not been simple. Polarization has affected different groups in different ways and to different degrees. In the United States, elite polarization has proceeded monotonically since the 1970s, accumulating to such a substantial degree that in the US Congress there remains no overlap ideologically between representatives of the two major parties.[1] In contrast, the mass public has not polarized to the same degree, and to the extent it has polarized, the process began later and has been far less pronounced (Gentzkow 2016).

The objective of this paper is to provide a simple model that accounts for the richness of polarization. Our model seeks to provide an explanation not only for why politics is polarized today but also for why it took so long to get to where it is and why different groups have polarized to different degrees, at different speeds, and at different times. Our model adopts a classic model of electoral competition and amends it with an intuitive and behaviorally justified change to the nature of voter preference. On its own, this change is innocuous and does not lead to polarization. However, when this change is interacted with strategic candidates and iterated, the impact on politics is dramatic. We show that it leads to a rich dynamic that can account for multiple moments in the data as well as other features of voting and political behavior, thereby providing an integrated explanation of polarization across time and across different groups.

The building block for our model is the behavioral finding that preferences and behavior coexist in a causal loop. The classic view of decision-making is that preferences are fixed and that they determine choice. Increasingly, evidence across many domains points to causality also running the other way: that behavior also affects preferences.[2] In politics this causal loop works through the voting decision. The act of voting not only reflects a citizen's preference but also causes those preferences to change. The evidence suggests that after voting for a candidate, a voter updates her preferences such that she likes that candidate a little more (Beasley and Joslyn 2001).

[1] Bartels (2016) provides similar evidence for US presidential candidates.

[2] Of course, one could consider fixed metapreferences such that past behavior influences marginal preferences over current choices, as in the Becker and Murphy (1988) model of addiction.

We incorporate this feedback loop into a model of preferences by endogenizing a voter's ideal point. Formally, a voter updates her ideal point by moving it toward the location of the party she voted for, even if by only a small amount. We make no other changes, retaining otherwise the classic conception of expressive voting with abstention.[3] This means that within each election voting behavior is standard—citizens vote for the nearest candidate or otherwise abstain—and remains consistent with the large body of evidence that has accumulated on how votes are cast.

The causal loop affects behavior only across periods. Even then it does not, on its own, lead to polarization. If party positions are fixed, the feedback loop leads to a congealing of voters around the parties. Voters on the flanks update inward, and voters toward the center update outward. This process produces homogenized but not necessarily polarized voting blocs. Indeed, if parties are located at moderate policies, this process produces an overall moderation of the electorate.

The key to our result is the addition of strategic parties and how they react to the evolving preferences of the citizenry. The congealing of preferences is impactful not because of the homogenization of preferences per se but because of the impact it has on the incentives of the parties. The updating process leads to gaps opening up in the distribution of ideal points, both between voters and abstainers and between voters for the two parties. The gap at the center of the distribution is important, as it is in the center that political competition plays out. A gap in the distribution implies that there are no voters to gain or lose, freeing parties to polarize toward their own, more extreme, preferences without fear of losing voters to the other party. This incentive to polarize depends only on the inside margin and exists whether the overall electorate is polarizing or moderating. This dynamic exposes how the preferences of voters and elites, despite their interdependence, can evolve at different speeds and, indeed, even in different directions.

The first step begins an iterative process of polarization. The more the parties polarize, the more voters update toward them, widening the gap in the center and allowing the parties to polarize farther, creating a feedback loop. Critically, this feedback loop is incremental, and polarization does not occur all in one step. The reason is abstention. Opening a gap in the voter distribution fundamentally changes the nature of electoral competition, as it relaxes competition for centrist voters. It does not, however, relax the competition that parties face with voter apathy. If the parties polarize too far too quickly, they risk alienating their own supporters. In equilibrium, the

[3] This conception of voter behavior was first promulgated by Hotelling (1929) and Smithies (1941). We briefly discuss other possible theories of voting when we present our model in sec. II.

parties adopt a "no-voter-left-behind" strategy, polarizing incrementally such that in each step, no voters are lost to abstention. This implies that the speed of polarization is tightly linked to the degree voters update their preferences after voting. The smaller, more incremental is the updating, the slower and more iterative is party polarization. This dynamic reflects a shift in the nature of electoral competition over time, from competition for the swing voter to competition for turnout, an evolution that is consistent with evidence from US election campaigns (Panagopoulos 2016).

That polarization in the model is progressive illustrates how the preferences of the elites and the masses coevolve. The equilibrium not only matches the polarization of both groups but also explains how the timing and degree of polarization that is observed in the data can emerge. In our model, it is the elites who polarize first and always lead the masses, regardless of the speed at which they polarize. The masses, in contrast, may even moderate at first before reversing course and polarizing. On the surface, this gives the appearance of polarization being an elite-driven phenomenon, but as our model demonstrates, a necessary part of the root cause is voter preferences, without which elite polarization would not occur.

The equilibrium in our model also rationalizes other features of political behavior related to polarization. The emergence endogenously of a gap in the distribution of preferences matches the often-lamented "missing middle" of the electorate, what Abramowitz (2010) refers to as the "disappearing center." In drawing a clear distinction between the preferences of voters and those of abstainers (abstainers do not update their preferences), we are able to show simultaneously how those engaged with politics can polarize while their fellow citizens become increasingly disenchanted with partisan politics, leading to a bimodal distribution of political preferences. This finding matches observation and goes some way to reconciling the competing findings about voter preferences that have riven the empirical literature.[4]

Behavior within our model also relates to some empirical puzzles that are ostensibly unrelated to polarization. One example is negative partisanship. Although voters update toward their favored party, the party itself is polarizing. This means that moderate voters do not, at first, get any closer to the party as they effectively chase it to the extremes during the polarization process. Consequently, despite their updating, these voters do not evaluate their favored party any higher. At the same time, these voters are moving away from the opposing party—at a rapid rate as that party moves in the opposite direction—and their evaluation of that party is declining. This creates a perplexing combination of preferences as voters seemingly become only more negative about the opposition and no less favorable about their

[4] We return to this controversy in detail later in the paper.

favored party, yet this is exactly the pattern of preferences that defines "negative partisanship" and has been extensively documented empirically (Abramowitz and Webster 2016).

The process of polarization will not stop at the present day, requiring us to look forward as well as back. To that end, we put our model to work to explore the future of polarization. Following the logic of our model through, it predicts that party elites will continue to polarize until they reach their own ideal points, where they will stabilize. That does not end voter polarization, however, as over time voters will increasingly converge on the positions of their favored party. That voters in practice are currently not as polarized as elites suggests that this process still has a ways to run. This future is surely not of comfort to those who lament the state of politics today, as it implies a future electorate that is as polarized as elites, with an ever larger "missing middle" and with partisan constituencies that approach homogeneity.[5] Moreover, at the limit of the model, the state of polarization is not only an extreme but also an increasingly stable outcome.

The power of our result is in its simplicity. With a single, empirically grounded change to voter behavior, the model is able to rationalize the rich dynamics of the past few decades in US politics and make predictions about the future. That said, politics in practice is complicated, and definitely more complicated than is our model. We make no claim that our explanation is the only force causing polarization, whether of elites or of the masses, and the explanatory power of our model has its limits, to be sure. Rather, we seek to illuminate a simple mechanism that organizes many key features of the data.

The focus of our paper is on the era of polarization in the United States that began in the second half of the twentieth century and has run through to the present day and, presumably, into the future. The evolution of political preferences did not begin with this era, of course, and before the era of polarization there was an extended period of convergence. We conclude the paper by suggesting a way to connect our model to this earlier era. By adding generational turnover to our model, we speculate on how polarization can sow the seeds of its own demise and initiate an era of moderation. Exploring more fully how our theory of polarization fits into the broader, longer-term landscape, as well as how it applies in political systems other than the United States, offers the promise of a more general understanding of the dynamics of political behavior.

Related literature.—Our model contributes to the literature on electoral competition in several ways. Our main point of departure is our focus on the updating of preferences and the dynamic path of policy. To be sure,

[5] This is consistent and a natural end point to the evidence on increasing within-party homogeneity presented in Levendusky (2009).

many theoretical models of political economy seek to explain the divergence of candidates and parties from the median, but we know of no model that provides a dynamic account of that movement or captures the simultaneous polarization of voters.

A second contribution is to the analysis of single elections, models of which have been the focus of much of the literature. An ongoing challenge in that literature is proving equilibrium existence when parties are policy motivated (Calvert 1985). Adding abstention complicates this challenge further. Nevertheless, we show how existence problems can be overcome. Moreover, the inclusion of abstention reveals a novel type of equilibrium in which the parties do not compete directly for the median voter, a structure that matches more closely electoral competition in practice.

By now, the question of what has caused polarization has produced a large literature. This predominantly empirical literature has eliminated many explanations (such as gerrymandering in the House of Representatives), yet a clear consensus on the underlying cause (or causes) has yet to emerge (McCarty 2019). We differ from the literature in developing a theory of the underlying mechanism that causes polarization. This enables us to not only provide an explanation of polarization per se but also explain the speed, timing, and differences in polarization across the different levels of politics and to produce testable predictions about other facets of voting behavior and political outcomes that can be used to verify the theory.

II. The Model

In each period, $t = 1, 2, \ldots$, two parties, D and R, compete in an election. The parties simultaneously announce policy positions, d_t and r_t, that they will implement if elected. Policies are points in the classic one-dimensional policy space such that $d_t, r_t \in \mathbb{R}$. The election is decided by plurality rule.

The parties are motivated by both winning office and policy outcomes—they have mixed motivations, in the classic parlance. To keep the notation simple, we denote the ideal policies for parties D and R by D and R, respectively, where $D < 0 < R$. The benefit to each party of winning office is $\beta \geq 0$. The period utility for D is

$$u_t^{\mathrm{D}} = \begin{cases} -|d_t - D| + \beta & \text{if D wins,} \\ -|r_t - D| & \text{if R wins.} \end{cases}$$

The intuition behind our results does not require forward-looking behavior by the parties, and, for simplicity, we focus on the extreme case in which parties discount the future completely. Also for simplicity, we focus on symmetric party preferences, assuming $D = -R$.[6]

[6] We discuss the case of forward-looking parties and asymmetric preferences in sec. VII.

A continuum of citizens possess ideal points distributed in $\mathbb{R}$. In each election, citizens either vote for one of the two parties or they abstain, and when they vote they do so on the basis of proximity in the policy space. The exact form of spatial voting is not important for the mechanism that drives our results. For concreteness, we adopt the expressive form of voting that originated with Hotelling (1929) and was extended by Smithies (1941) to allow for abstention. In this model, citizens evaluate parties relative to their own ideal point and vote for the closer one. If neither of the parties is sufficiently close, the citizen is alienated and abstains. This is known as abstention due to alienation. Formally, a voter with ideal point v_t votes for

$$\text{D if } |d_t - v_t| < |r_t - v_t| \text{ and } |d_t - v_t| \leq \lambda,$$
$$\text{R if } |d_t - v_t| > |r_t - v_t| \text{ and } |r_t - v_t| \leq \lambda;$$

otherwise, she abstains. If indifferent between the parties, she randomizes, although this tie-breaking rule will be unimportant. The constant $\lambda > 0$ is the *region of tolerance*, beyond which a citizen prefers to abstain rather than express a preference for either party. This voting rule is rationalized by a simple utility function. If party $J \in \{\text{D}, \text{R}\}$ has platform p, set the utility of voting for J to be $\pi(J; v_t) = \lambda - |p - v_t|$ and the utility of abstention to zero.[7]

The key novelty of the model is how voting leads to movement in a citizen's ideal point. The updating process is as follows. For a citizen with ideal point v_t who votes for a party with platform p_t, her ideal point at election $t + 1$ becomes

$$v_{t+1} = v_t + \tau(p_t - v_t), \tag{1}$$

where $0 < \tau < 1$ is the *dissonance* parameter that dictates the speed of updating. The ideal point of an abstainer does not change. We discuss the basis for this updating rule below.

The standard view of elections is that outcomes are to some degree random. The classic approach is to add some stochastic element into the electoral process, typically by adding an idiosyncratic noise term to voter utility. To avoid the distraction of excessive notation, and in the spirit of the original

[7] We adopt this perspective on voting because it is standard, simple, and empirically supported (Jessee 2009, 2010) and because expressive voting accords more naturally with behavioral voters who experience cognitive dissonance. It fits more closely with evidence from large elections than does the strategic view of voting, such as in Riker and Ordeshook (1968). The strategic view is plagued not only by the paradox of turnout but also by the fact that a purely relative evaluation of the parties induces abstention only when a citizen is close to indifferent between them—what is known as abstention due to indifference—as this also does not accord with empirical observation. Within the perspective of expressive voting, it is possible to incorporate a relative evaluation of the parties in addition to the direct evaluation that we focus on (such as in Callander and Wilson 2006). We expect that this extension would only reinforce the logic of our results, although we do not conduct a formal analysis.

reduced-form approach of Calvert (1985), we assume directly that a party's probability of winning an election is simply equal to its share of voters in that election.[8]

The distribution of citizen ideal points evolves from election to election as votes are cast and ideal points are updated. Initially, the distribution of citizen ideal points is given by a continuous distribution F, with zero mean and density function f. We assume that f is continuous, strictly increasing for all $v < 0$, symmetric around its mean of zero, and thus decreasing for $v > 0$, with full support on $\mathbb{R}$, and that it satisfies *midpoint ratio log concavity*. This final requirement is a new condition that we define formally in appendix A. It represents a strengthening of log concavity to allow for the fact that electoral competition in our model is relative rather than absolute. As with log concavity itself (Bagnoli and Bergstrom 2005), many familiar distributions satisfy this requirement, including the logistic and the uniform.[9]

Polarization of the masses reflects the spread of ideal points at a point in time. For concreteness, we define *voter polarization* and *citizen polarization* (including abstainers) as the average distance of voters' (citizens') ideal points from the mean of the distribution. Throughout our main model, the distribution will be symmetric in every election, and our measure of polarization is equivalent to the average distance of voter ideal points from zero.

Elite polarization is given by the gap between the expressed policy positions of the parties in each election. As will become clear, those positions will vary between the median citizen and the parties' ideal policies. To avoid corner solutions in the first election, we impose two conditions. To rule out the parties beginning at their ideal points, we impose a lower bound on the party ideal points, specifically, that $R = -D > \lambda^*$, where λ^* is the voter tolerance level that solves

$$1 + \left(\lambda^* + \frac{\beta}{2}\right)\frac{f(2\lambda^*) - f(0)}{F(2\lambda^*) - F(0)} = 0. \tag{2}$$

[8] This approach reduces the complexity of our analysis, both within each election and in keeping track of the evolving distribution of issue-voter ideal points over time. This approach can be microfounded by supposing that a subset of citizens are "noise" voters whose behavior is random, or at least conditioned only on features of the political landscape that are uncontrollable and even unidentifiable by the parties, and that the remaining citizens are issue voters and vote deterministically. Such a dichotomy is consistent with empirical evidence that some citizens pay attention to politics and vote spatially according to policy whereas others are essentially uninformed and seemingly cast their ballots on a whim or abstain altogether (Jessee 2009, 2010). A more direct foundation can be derived by supposing that a random fraction of votes are miscast or voided. We thank a referee for this interpretation.

[9] We prove this analytically in app. C (available online) for the logistic and uniform distributions. We provide numerical verification for the normal distribution in the relevant range for our model.

To rule out full convergence to the median in the first election, we suppose that the perks from winning office, β, are not too high, specifically, that $0 \leq \beta \leq M$, for some real number $M > 0$.[10]

Foundation of the voter updating rule.—The preference-updating rule we apply has many possible interpretations. The most natural, and the one we carry throughout, is that it is a generalized form of cognitive dissonance. In Festinger's (1962) classic formulation of cognitive dissonance, an individual who faces a tension between their preferences and their choice will respond by updating their preferences to remove the tension.

In our setting there is no tension per se, as the citizen always votes for the nearer of the two candidates. Yet, in the same manner as an individual in the classic theory of cognitive dissonance, a voter updates her preferences to make her choice seem more secure. Moreover, the magnitude of a voter's response increases the more insecure she is in the choice she made. This represents a smooth generalization of cognitive-dissonance theory, consistent with more recent evidence from psychology (Aronson, Fried, and Stone 1991). The classic theory, in contrast, imposes a sharp transition, such that updating preferences occurs only when the relative appeal of the alternatives crosses over.

Our formulation is motivated by evidence in political economy that shows that citizens update their preferences to rationalize their vote choice and make their choice seem more appealing. This was first identified by Beasley and Joslyn (2001) and developed further by Mullainathan and Washington (2009) and Dinas (2014) for the United States and by McGregor (2013) for Canadian voters. The evidence in Bølstad, Dinas, and Riera (2013) is particularly illuminating. By focusing on tactical voting in the United Kingdom, whereby citizens vote for a party other than their most preferred, they show that the act of voting causes voters to update their preference even toward a party for whom they voted for purely tactical reasons. The underlying psychological mechanism is also consistent with the considerable evidence that voting is habit forming, as noted by Fujiwara, Meng, and Vogl (2016).[11]

The specific functional form we adopt embeds several additional modeling choices. Updating is exclusively action driven. It does not depend on the identity of the party or even whether the party wins the election but depends only on the act of voting. (We discuss the latter two possibilities in sec. VII.) The specification also presumes that voters update toward the location of the party when votes are cast rather than where the party might

[10] The precise value of this upper bound M is provided in app. A; see lemma A5.

[11] Akerlof and Dickens (1982) is the seminal introduction of cognitive dissonance into economics. Acharya, Blackwell, and Sen (2018) provide an application to politics, although that work does not consider the role of strategic candidates that is central to the equilibrium dynamic in our model. Penn (2017) provides an interesting application of these ideas to a formal model of political values.

subsequently move. This is consistent with the logic of cognitive dissonance and the feedback loop between decisions and preferences. It is also appropriate, given that the attention of most voters is turned on during elections and off subsequently, and resonates with the scant empirical evidence on this point (see again Beasley and Joslyn 2001).[12]

III. The First Election

With policy-motivated parties and uncertainty over the election outcome, the first election presents the parties with a classic trade-off between the probability of winning and the policy outcome. By inching toward the center, a party increases the chance it wins the election, but at the cost of a less attractive policy should it win. As has been known since the seminal contribution of Calvert (1985), this trade-off leads to an equilibrium in which the two parties do not fully converge to the center as long as the pure benefit of winning office, β, is not too large.

In the classic formulation, the competitive tension plays out exclusively at the center of the distribution, with the parties competing intensely for the median voter. The logic depends, however, on full turnout. With full turnout, every citizen votes and the only competitive margin is halfway between the parties, where the swing voters sit. Adding abstention changes this. If the parties' positions are far enough apart, the intervals of their support do not intersect. This leaves abstainers in the middle of the distribution and multiple margins at which citizens are indifferent between abstaining and voting for one or other party. There is no citizen, however, who votes and who is at the margin of deciding which party to support.

This formulation has not been analyzed previously in the literature, even for one-shot elections. We show that it is important, as it leads to a new type of equilibrium, one in which competition is between parties and abstention rather than between the parties directly. In this equilibrium, parties stop converging before their intervals of support meet and intense competition at the center of the distribution does not occur.

Proposition 1 establishes that the possibility of this new equilibrium coexists with the traditional equilibrium in which parties compete for the median voter at the center. The equilibria are distinguished by the level of voter tolerance λ. For high voter tolerance, the parties converge sufficiently that they compete in the center, whereas for low voter tolerance, centrist citizens abstain and the parties appeal to very distinct constituencies.

[12] An alternative dynamic linkage is to keep ideal points fixed and suppose that voters update the valence of the party they support. With only two parties, this formulation would lead to the same dynamic polarization of elites as generated in our model, but, clearly, it would not generate polarization of the masses. That said, because most surveys of spatial preferences do not include a valence dimension, the empirical connection between this formulation and polarization is unclear and in need of further research.

The inclusion of abstention—and the new type of equilibrium—complicates the analysis considerably, as the parties' objective functions now are only piecewise differentiable and not necessarily quasi-concave. Nevertheless, we are able to establish the uniqueness of a symmetric equilibrium for each set of parameter values and show that there is a unique cut point demarcating the two types of equilibrium.

PROPOSITION 1. In the first election, a unique symmetric equilibrium exists with $r_1^* = -d_1^* \in (0, R)$. The parties win the election with equal probability. Party R's equilibrium location r_1^* is implicitly defined by

$$1 + \left(r_1^* + \frac{\beta}{2}\right)\frac{f(r_1^* + \lambda) - f(r_1^* - \lambda)}{F(r_1^* + \lambda) - F(r_1^* - \lambda)} = 0 \quad \text{for } \lambda \leq \lambda^*, \tag{3}$$

$$1 + \left(r_1^* + \frac{\beta}{2}\right)\frac{f(r_1^* + \lambda) - f(0)}{F(r_1^* + \lambda) - F(0)} = 0 \quad \text{for } \lambda > \lambda^*, \tag{4}$$

where λ^* is the tolerance level implicitly defined in equation (2).

Figure 1 depicts the two cases that are possible. For lower levels of voter tolerance, $\lambda < \lambda^*$, the parties stop converging before their intervals of support intersect. As a result, citizens abstain on either flank as well as in the middle. This is the solution given by equation (3) and is depicted in the top panel of figure 1. For higher voter tolerance, $\lambda > \lambda^*$, the intervals of support do intersect, leaving abstainers only on the flanks. This solution is given by equation (4) and is depicted in the bottom panel of the figure.

Equations (3) and (4) differ only on whether the inside boundary of party R's support is at zero or λ to the left of R's equilibrium position. This difference appears in the final term of the numerator and denominator. Each equation represents the first-order condition for party R. The first term represents the direct policy gain from moving to the right conditional on winning the election. The second term represents, approximately, the decrease in R's probability of winning because of this movement multiplied by the cost of losing in both direct policy and the benefit of winning. In the proof of proposition 1 in appendix A, we establish that setting the sum of these effects to zero defines the unique symmetric equilibrium.[13] Our key observation is that the expected utility for the parties is a product function (see eq. [A3]) and that both components of this function are log concave, given our assumptions on *f*. This allows us to establish the existence and uniqueness of an optimal location for each party. In this way, we overcome the problem of second-order conditions that normally plagues models of this sort.

[13] For the logistic distribution, we prove in Callander and Carbajal (2020) the stronger result that this is the unique equilibrium.

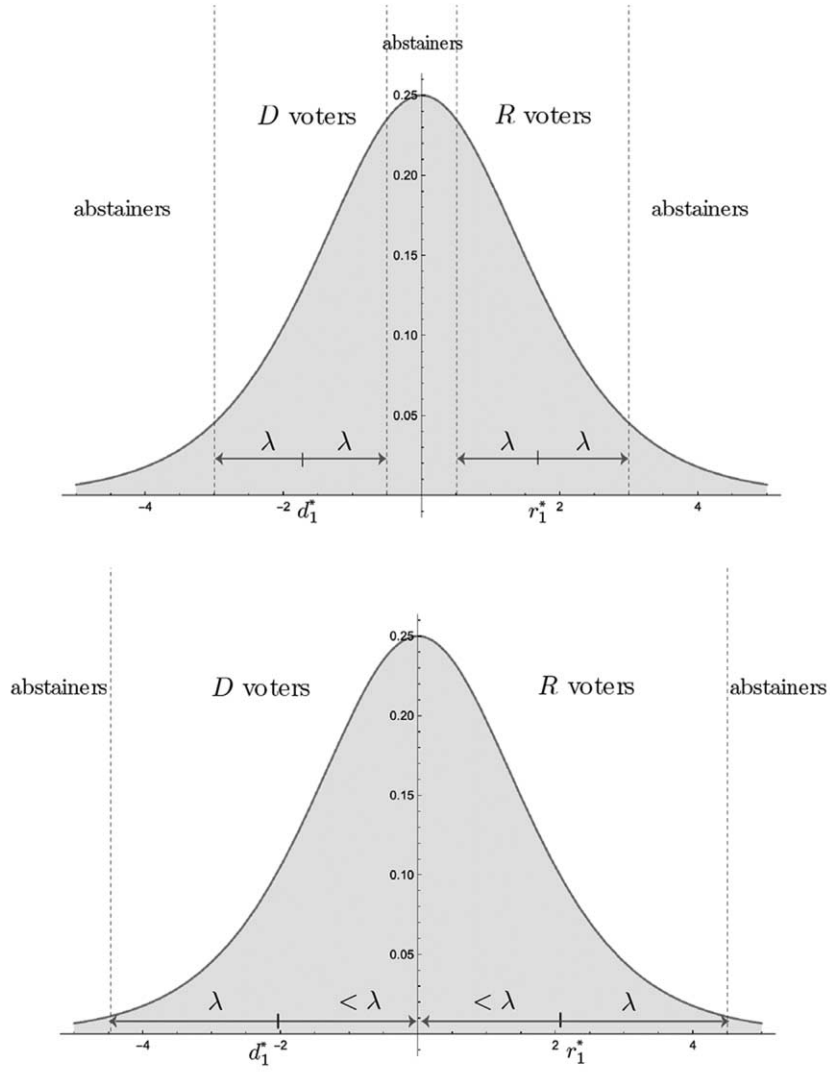

FIG. 1.—Equilibrium configurations of voters and abstainers. A color version of this figure is available online.

The two configurations that are possible in equilibrium resonate with the prominent debate in political science over whether it is better for parties to appeal to their base or to swing voters (Hall and Thompson 2018). For low voter tolerance, the parties seemingly abandon efforts to persuade voters to vote for them rather than the opposing party, concentrating

exclusively on voters on the flank who would otherwise abstain. For high voter tolerance, the parties do seek to persuade as well as mobilize voters, and they compete head to head for centrist voters. This result shows how these two strategic options, rather than being fundamentally in contrast, can in fact emerge from a single model of electoral competition, differentiated only by parameter values. In both types of equilibrium the competitive tension is the same: creep inward for more voters, at the expense of a worse policy. The novelty of the low-voter-tolerance equilibrium is simply that this competitive drive can exhaust itself well before the battle is met with the other party and instead resemble a mobilize-the-base strategy. We see in later sections that this type of equilibrium, rather than being a peculiarity, in fact emerges over time as the dominant style of electoral competition, matching the dominance of the mobilize-the-base strategy in practice (Panagopoulos 2016).

To better see the equilibrium, and the continuity between the two forms of competition, the left-hand panel of figure 2 depicts the equilibrium positions as a function of voter tolerance, λ, for three different values of β, the direct benefit of winning office, when the distribution of citizen ideal points is logistic with scale equal to one.

An intuitive comparative static is that the parties converge toward the center (and each other) as β increases. This is evident in figure 2 (*left*). In corollary A1 in the appendix, we prove that this holds generally in our model. As the direct benefit of winning office increases, the parties become more willing to sacrifice policy in pursuit of electoral victory and compete more intensively by converging toward the center.

The effect of voter tolerance on competition depends on the distribution of ideal points. What we do know is that for $\lambda > \lambda^*$, such that the intervals

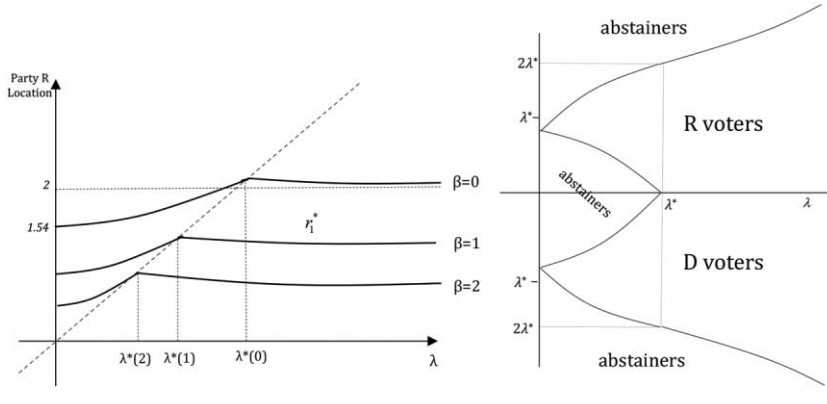

FIG. 2.—*Left*, equilibrium platforms in the first election as a function of λ. *Right*, regions of turnout as λ varies for $\beta = 0$. A color version of this figure is available online.

of support meet in the center, an increase in λ causes the parties to choose more moderate positions, as evident in figure 2. This convergence is bounded, however, and the limit equilibrium position as $\lambda \to \infty$ is bounded away from zero.[14]

For lower voter tolerance, a general comparative static is not available. For the class of logistic distributions, we can prove that the parties respond to an increase in voter tolerance by moving farther from the center, the opposite to that for higher tolerance. This is of interest, as it implies the possibility of nonmonotonicity in party positions as λ varies, as evident in figure 2. This suggests that efforts to increase turnout among citizens—to increase λ—may induce more party polarization rather than less.

To further understand how λ affects the structure of elections, it is helpful to visualize the regions of turnout as λ varies. In the right-hand panel of figure 2, we show this for a logistic distribution with β set to zero. For levels of voter tolerance below λ^*, there are abstainers in the center as well as on the flank. Although the parties diverge as λ increases, they do so at a slow rate, such that the interval of abstainers contracts, eventually disappearing at λ^*. Similarly, although the parties converge for higher λ, they do so sufficiently slowly that the outside boundary of their support expands. Thus, for the logistic distribution at least, turnout strictly increases in λ. Regardless of the distribution of voter ideal points, however, as $\lambda \to \infty$ turnout approaches one.

IV. Fixed Party Locations: A Benchmark

After the first election, the winning party is installed in office and voters update their preferences. This changes the distribution of ideal points in two ways: the ideal points of voters compress toward the party positions, and gaps open up. The gaps appear because voters update whereas abstainers do not, such that at the margin between them a discontinuity is created. If there are no abstainers in the middle, then the gap is between the voters themselves as D voters shift left and R voters shift right. The two possible configurations are depicted in figure 3. In both cases, the compression in ideal points of voters leads to higher density in those regions. The top (bottom) panel in figure 3 shows the density of voters' ideal points in the second election for a low (high) tolerance level λ. Compare this to figure 1, which represents the density of voters' ideal points in the first election.

This process then iterates over time. In the full model, the evolution of voter preferences interacts with the strategic response of parties. To understand the forces at play, we begin by disentangling these effects. We

[14] We prove each of these claims in app. A.

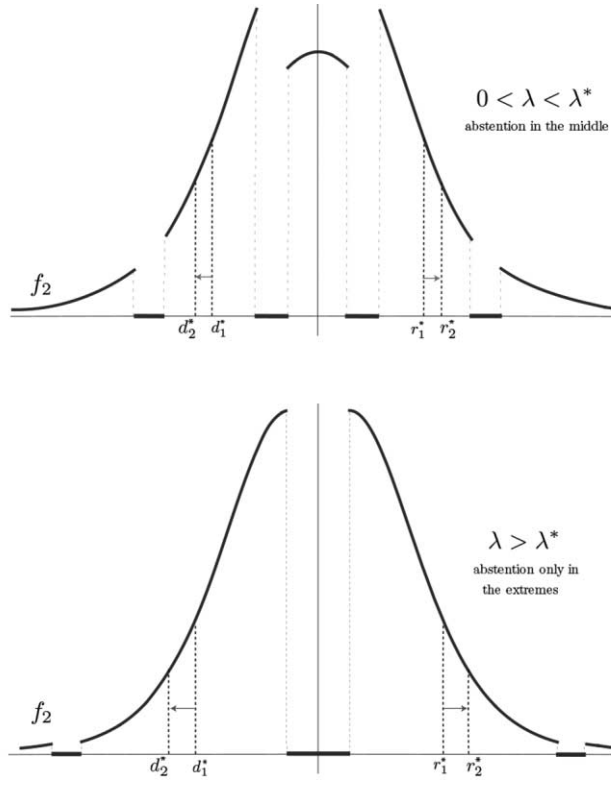

FIG. 3.—Ideal-point updating for low λ (*left*) and high λ (*right*). A color version of this figure is available online.

fix the party positions at $\hat{r}$ and $\hat{d}$ for all elections and focus exclusively on voters.

With fixed party positions, the evolution of voter preferences is straightforward. Voters compress around the position of their preferred party, becoming increasingly homogeneous. At the same time, abstainers remain unmoved and are never tempted in off the sidelines to vote, leaving the turnout rate constant. The combination of these two facts implies that, over time, the gap between voters and abstainers grows ever larger. We collect these properties in the following proposition, whose proof is immediate.

PROPOSITION 2. Fix the party positions at $\hat{r} = -\hat{d} > 0$. At each election $t \geq 2$,

1. the ideal point of a voter evolves monotonically, converging on $\hat{r}$ or $\hat{d}$ as $t \to \infty$;

2. the average distance between copartisans (who vote for the same party) decreases, approaching zero as $t \to \infty$;
3. the minimum distance between a voter and an abstainer increases from zero, approaching λ as $t \to \infty$; and
4. turnout is constant throughout elections.

The evolution of individual ideal points changes the degree of polarization at the electorate level. The aggregate effect is ambiguous, however, as it depends on the relative position of the parties and the degree of voter tolerance, λ. The key measure is how the initial average ideal point of a party's voters compares to the position of the party. Proposition 3 establishes that the average is all that matters. If the party is more moderate initially than this average, then voters will moderate over time; otherwise, they will polarize.

PROPOSITION 3. Fixing party positions $\hat{r} = -\hat{d} > 0$, voter polarization decreases monotonically over time if

$$\hat{r} < \int_0^{\hat{r}+\lambda} vf(v)\,dv;$$

otherwise it increases monotonically.

To understand this result, begin with the case in which voter tolerance is low and abstainers exist in the middle of the electorate. Because the parties' intervals of support do not intersect, each interval is symmetric around a party's position. It follows from the single-peakedness of f, the distribution of voter ideal points, that there are more voters to the inside of a party position than to the outside. Thus, as voters update, more voters are updating outward than are updating inward, and the electorate is progressively polarizing. Intuitively, as voters collapse in on the party position, so too must their average.

The electorate will moderate instead if the average voter begins more extreme than the party. This is possible for high λ, as then the parties' intervals of support intersect at the center and each party's support is compressed. Party R's support will be narrower to its inside than it is on its outside, and the broader support on the flank may be enough to overcome the lower density of voters there.

For this to occur, the compression toward the center has to be sufficient. The condition in proposition 3 simply represents this requirement. The integral is the average ideal point of a voter for party R.[15] Another way of stating

[15] Note that the condition cannot be satisfied if the intervals of support do not intersect.

this is to say that, for a given λ, voters will moderate if and only if $\hat{r}$ is sufficiently close to zero. A necessary condition for this to hold is that $\hat{r} < \lambda$ such that the intervals of support intersect. It is also necessary that party R not be too polarized in an absolute sense, specifically, that $\hat{r}$ is more moderate than the average of all citizens on the right-hand side of the distribution, as otherwise the compression at the center is insufficient to overcome the lower density on the flank even as $\lambda \to \infty$.

An interesting case emerges when party R is located between the average citizen and the median citizen on the right-hand side of the distribution (that the median is more moderate than the average follows from the single-peakedness of f). In this case, more voters are on the inside of the party position than on the outside, and thus more voters are shifting outward in their ideal point than are shifting inward. Nevertheless, in aggregate, the electorate is moderating because the voters on the outside are more extreme than the centrist voters are moderate, and the convergence of extreme voters outweighs the polarization of moderate voters.[16]

Voter updating with fixed party positions leads to rich dynamics, but, ultimately, it can explain only so much. On their own, voters may polarize, but they may also moderate. Even if they polarize, the effect is bounded by the locations of the parties' initial policies. Moreover, the preference profiles that do emerge are inconsistent with other known properties of voting, such as negative partisanship.[17] On top of this, there is, of course, no polarization of the parties. In the next section, we reintroduce strategic parties and show how their reactions to voter updating creates an interdependence and coevolution of elite and mass positions that do resonate with the data.

V. Adding Strategic Parties Back In

Strategic parties respond to the changing distribution of ideal points, which, in turn, changes the evolution of voter preferences. In this section, we characterize the dynamic process that results, beginning with the second

[16] The divide that forms between voters and abstainers and the growing homogenization of voters resonate with the axiomatic measures of polarization of Esteban and Ray (1994, 2012). A key difference is that the compression of voters here is around the location of each party rather than the mean of the group distribution, as it is in Esteban and Ray's notion of a "squeeze." Differences aside, the growing divide between voters and abstainers in US politics increasingly resonates with the in-group and out-group measures in Esteban and Ray's work, suggesting that their theory of polarization, which was developed in the context of conflict and ethnic and tribal allegiance, could also find profitable application in the domain of US politics. (Clark 2009 offers an application of their ideas to the United States in the context of Supreme Court justices.)

[17] All voters experience increasing preference for their favored party, whereas negative partisanship finds that this is relatively stable.

election. Throughout, we presume that the policy positions in the first election are those given by the unique symmetric equilibrium described in proposition 1.

A. *The Second Election*

The gaps in the distribution of ideal points fundamentally change the incentives of the parties. The equilibrium positions in the first election balance the incentive to converge to gain more votes against the incentive to diverge to a better policy position. Starting from the same positions, that trade-off now collapses. The gap(s) in the center of the distribution imply that a party can diverge slightly without losing any votes. No votes are lost because there are no voters there to lose. Those who had been on the inside margin and who did turn out to vote updated toward the party, leaving behind an empty space. This changes the calculus of the parties, and they respond by moving their positions toward the extremes.

This logic provides the foundation for polarization. It is, however, only half of it. As one party shifts outward, so does the other party, and, as a result, the midpoint between them remains unchanged. This implies that the incentive to shift outward is recreated anew, leading to more polarization and potentially a substantial unwinding of party positions. This unwinding replicates but turns on its head the classic logic of convergence due to Hotelling (1929). In the classic intuition, parties inch toward the center to win the median voter, and, as the opposition does the same, this creates an iterative process that leads to full convergence. In our model, in contrast, a party inches outward without losing votes, the opposition party responds, and the iterative process leads instead to polarization.

Unlike in Hotelling, however, the iterative process need not lead to complete unraveling instantaneously. The parties in our model may no longer be constrained by each other, but that does not allow them to escape from competition altogether. Instead of competing against each other, the parties compete against voter apathy. If they polarize too much, the parties will lose voters to abstention.

The exact nature of that divergence depends on the type of the first-election equilibrium and, thus, on the level of voter tolerance. For low levels of voter tolerance, that is, $\lambda \leq \lambda^*$, the parties do not compete directly in the first election, and their ability to polarize is limited only by the extent of their own voters' updating. The largest shift in preferences is by the voters who were on the margin between voting and abstaining in the first election. Proposition 4 shows that this amount, $\lambda\tau$, is exactly the amount that the parties polarize at the second election.

PROPOSITION 4. At the second election, the unique equilibrium for $\lambda \leq \lambda^*$ is

$$r_2^* = \min\{r_1^* + \lambda\tau, R\},$$
$$d_2^* = \max\{d_1^* - \lambda\tau, D\}.$$

The equilibrium represents a sort of "no-voter-left-behind" strategy.[18] The parties polarize only as much as they can without losing any voters to abstention. Any larger polarization and the marginal voter on the inside would roll off and abstain, even allowing for that voter's own outward shift.

While the parties are leaving no voters behind on the inside, they are also gaining voters on the flank. Abstainers on the flank who were just outside the margin of voting in the first election are now $\lambda\tau$ closer to the new party position, and as a result, an interval of abstainers of that length switch to voting and turnout goes up.

The updating of voter preferences allows the parties to, in a sense, secure their core supporters, and this, in turn, gives the parties freedom to move. Rather than move to the center to appeal to abstaining moderates, however, the parties use the opportunity to polarize outward, drawing more extreme citizens into the voting pool. This implies that as the voting pool grows, it is the newer voters who are the most extreme.

The situation when voter tolerance is high ($\lambda > \lambda^*$) leads to even more polarization in the second election, although it can also lead to the symmetric equilibrium failing to exist. For high voter tolerance, the intervals of support in the first election intersect, which implies that the length of party support on its inside is less than the full length of λ. The shortened length means less updating by voters, with the marginal supporter for party R at zero moving her ideal point outward by only $r_1^*\tau < \lambda\tau$. This might suggest that the freedom of the parties to shift outward is also compressed, but, in fact, the opposite is true, and the parties polarize to a greater extent.

The increased freedom to polarize comes from the fact that the compressed intervals of support represent slack in the parties' ability to win voters. In the first election, the parties win only an interval of support of length r_1^* to their inside, whereas voter tolerance is λ, meaning that there is $\lambda - r_1^*$ in slack that can be exploited. To put it another way, the inside boundary of party support does not so quickly hit its limit as the parties shift outward. Combining this with the fact from above that competition is against voter apathy rather than the other party directly, slackness allows the parties to polarize faster. To implement the "no-voter-left-behind" strategy, therefore, party R can at most leave the voter located at $r_1^*\tau$ indifferent, which translates to a location for the party at $r_1^*\tau + \lambda$. This can represent a substantial jump from the first election position when λ is high, potentially all the way to the

[18] The parties are myopic and, thus, are not taking τ into consideration directly. Rather, τ appears only because it affects the distribution of ideal points in the second election.

parties' ideal points. Proposition 5 confirms that this is, indeed, the equilibrium for $\lambda \geq \lambda^*$, with polarization bounded by the parties' own ideal points.

PROPOSITION 5. At the second election, there exist $\bar{\lambda} > \lambda^*$ and $\bar{\tau} \in (0, 1)$ such that for all $\lambda^* < \lambda \leq \bar{\lambda}$ and $\bar{\tau} \leq \tau < 1$, the unique equilibrium is

$$r_2^* = \min\{r_1^* \tau + \lambda, R\},$$
$$d_2^* = \max\{-|d_1^*|\tau - \lambda, D\}.$$

For $\lambda > \bar{\lambda}$ and $\tau < \bar{\tau}$, there exists no symmetric equilibrium.

The equilibrium implies that even small changes in the distribution of voter ideal points can lead to substantial and immediate polarization of the parties. This is because voter updating of any size causes a gap to open up in the distribution, and it is this gap that induces tit-for-tat divergence, such that the parties unwind their positions to the point where they are competing no longer against each other but against voter apathy and abstention.

The logic of the result does have a limit, as the equilibrium fails for sufficiently high λ and low τ. This failure derives from failure of the second-order condition: eventually, as polarization increases, a point is reached at which the parties find it profitable to deviate and jump to the center. Failure occurs because, in effect, the logic of divergence is too powerful and the parties otherwise get too far apart too quickly. To see this, note that the distribution of ideal points in the second election when τ is low is very close to that in the first election. The "no-voter-left-behind" strategy, however, can generate substantial polarization. This leaves a large block of voters in the middle who can be exploited, and, eventually, one of the parties prefers to do so. This does not mean, though, that an equilibrium exists with centrist positions, as then the same unwinding logic would again apply.[19]

The most striking feature of the equilibrium in proposition 5 is that the type of electoral competition changes. Electoral competition in the first election for high λ is of the classic win-the-median-voter form, yet by the second election, this style of competition has given way to a mobilize-the-base strategy. Therefore, by the second election, electoral competition is such that the parties do not compete head to head for the median voter but instead compete only indirectly, focusing on the margin of turnout rather than persuasion. This pattern continues through later elections, suggesting that rather than being the unusual case, this style of competition is the norm.

The "no-voter-left-behind" strategy is intuitive, yet seeing exactly why it is optimal requires some digging. It is clear that if the parties polarize, they should polarize no less than they do with this strategy. Polarizing less would strictly decrease their vote share and implement a less appealing policy. What

[19] For an equilibrium to exist in this situation, it would have to be asymmetric or in mixed strategies. We leave the nature of this equilibrium (or whether one exists) as an open question.

is less clear is why the parties do not polarize farther, in fact leaving some voters behind, or why they do not instead exploit the opportunities created by the gaps in the distribution to converge. The answer to both questions comes from the logic of the first-period equilibrium.

The first-election equilibrium tells us that it is not profitable for a party to deviate outward, as the loss of centrist voters outweighs the gain in voters on the flank and the more appealing policy position. That the parties do not wish to polarize more than with the "no-voter-left-behind" strategy follows from this by a simple dominance argument. At its new position, the inside flank consists of exactly the same voters as in the first election (as they updated by $\lambda\tau$ in proposition 4 and by $r_1^*\tau$ in proposition 5), although now these voters are packed more densely and the rate of loss from further divergence is higher. At the same time, the marginal voters who would be gained on the flank are fewer in number (lower density farther out), and the policy cost of losing is now higher, as the opposition party has shifted farther away. Consequently, if deviating outward from r_1^* in the first period is not profitable, deviating outward from r_2^* in the second election is also not profitable.

Typically, this strict dominance argument would imply that the party must then find it optimal to instead shift inward. This would be true if expected utility were continuous. However, because of the gaps in the distribution of ideal points, expected utility is discontinuous in location (in contrast to the first election) and inward deviations are not profitable. Putting the two pieces together ensures that the strategies in propositions 4 and 5 constitute local optima.

Surprisingly, the same logic can also be used to support a second local optimum, in which the parties converge from the first-period locations (proposition 4). When voter tolerance is low and centrist citizens abstain in the first election, the logic of the "no-voter-left-behind" strategy works in reverse. In the same way as when it diverges, by converging a party does not lose voters on the flank (who have updated inward), whereas it gains voters in the center. Even though converging involves a less appealing policy should the party win, we could use the fact that the parties are indifferent about converging from the equilibrium location in the first election to show that they strictly prefer to converge from the same position at the second election.

This possibility complicates the analysis, as expected utility is no longer quasi-concave. This makes it difficult to establish not only that the strategy in proposition 4 is a global as well as a local optimum—and, therefore, an equilibrium—but also that a second, more convergent equilibrium does not exist. In the proof in appendix B, we construct a dominance argument that shows that convergence is dominated by polarization, thereby accounting for both of these concerns and establishing our equilibrium result. The main intuition boils down to the relative weights of the voters who update outward and those who update inward. Because more voters

are on the inside than the outside of the party position, the relative gain of no-voter-left-behind is greater for polarization than for moderation.[20] On top of this, the opposition party's voters, who have also updated their ideal points, are farther away and harder to capture, and the relative cost of losing is higher when the opponent polarizes rather than moderates. By putting these pieces together, combined with the fact that jumping to the center is not profitable in the first election, we establish that "no voter left behind" with polarization is the unique equilibrium.

B. The Third and Subsequent Elections

Voters update their ideal points again after the second election, the parties respond, and the process iterates. As this process continues, the parties progressively polarize, continuing until they reach their ideal points, at which they stabilize. This can take a few elections, or it can take many, depending on the party ideal points themselves as well as the speed at which voters update their ideal points and follow the parties' positions.

A difference at the third election and thereafter is that the nature of the equilibrium no longer depends on the size of λ. In the second election, the parties polarize such that the marginal voter is λ from the party position. The "no-voter-left-behind" logic then implies that the parties can polarize exactly $\lambda\tau$ farther in each election. The recursive process of polarization that this sets in motion is described in proposition 6.

PROPOSITION 6. For election $t \geq 3$, the unique equilibrium is

$$r_t^* = \min\{r_{t-1}^* + \lambda\tau, R\},$$
$$d_t^* = \max\{d_{t-1}^* - \lambda\tau, D\}.$$

Figure 4 shows the polarization process for different values of $\lambda\tau$. The rate of polarization is constant in each case, with the exception of panel *C* for high voter tolerance. In this case, a kink appears at the second election as the parties exploit their latent appeal to voters before settling down at rate $\lambda\tau$. The faster start in case *C* does not necessarily imply faster polarization overall. Should τ be low such that $\lambda\tau$ is also low, then, as depicted in panel *C*, polarization is thereafter slow and drawn out.[21]

The polarization of the parties leads to the polarization of voters, although the pattern of voter polarization is more varied and the timing different. We emphasize four features of voter behavior that stand out.

[20] In the first election, convergence from the equilibrium may lose x voters on the flank and gain $y > x$ voters at the center. From this same location, the gains in either direction in the second election are the same, whereas diverging now avoids losing y voters while converging avoids losing only x voters. This implies that divergence is relatively more attractive.

[21] The starting party positions, d_1^* and r_1^*, can be calibrated by varying the parameters of the model and the density f.

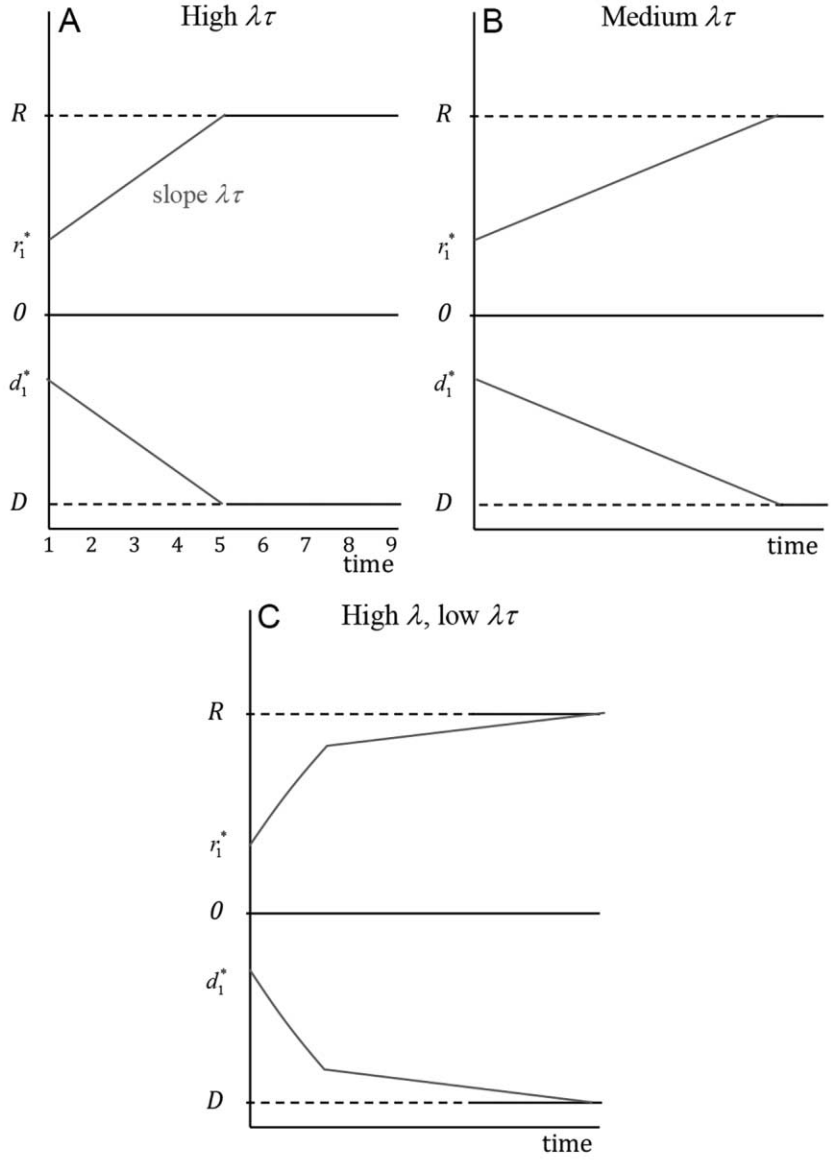

FIG. 4.—Party polarization over time. A color version of this figure is available online.

The first notable feature of voter polarization is that in the long run it is unambiguous. If the parties polarize, then so too do voters. This differs from the case of fixed party positions in which voter polarization is ambiguous (proposition 2), confirming that it is the interaction of party positions and voter updating that drives voter polarization.

The second notable feature of voter polarization is that it need not be monotonic, both at the aggregate level and for individual voters. This also differs from the situation with fixed party positions and is also different from the polarization trajectory of the parties. Figure 5 depicts the possible paths of voter ideal points for supporters of party R.

The trajectory of ideal points is monotonic for two sets of voters, the most moderate and the most extreme, although the direction of movement is the opposite for the two groups. Voters who are initially more moderate than the parties polarize monotonically, chasing, in effect, their favored party as it moves outward. In contrast, the voters who are most extreme—with ideal points beyond the party ideal points—begin by abstaining but are eventually drawn in off the sidelines to vote, and when they are, they traverse a monotonic path inward. Both sets of voters ultimately converge on their preferred party's ideal points.

The path of polarization is nonmonotonic for voters between these two extremes. A citizen with an ideal point between a party's ideal point and the first-election equilibrium position will begin to moderate her position once drawn in to vote. Eventually, however, the party will cross over her ideal

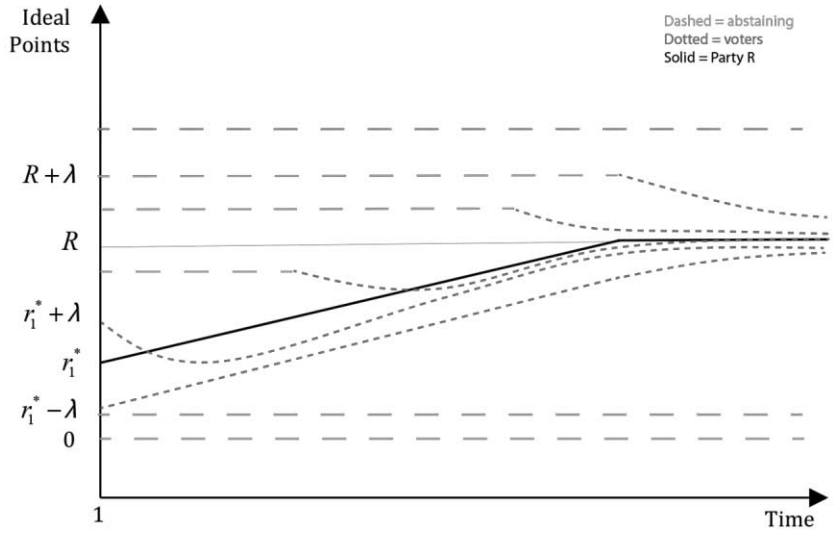

FIG. 5.—Path of citizen ideal points. A color version of this figure is available online.

point, causing that voter to reverse course and begin to polarize outward. Ultimately, this voter will polarize past her own initial ideal point, continuing outward until the party stops at its ideal point and the voter begins to catch up.[22]

The third notable feature of voter polarization is that it lags that of the parties. Interestingly, the parties not only polarize ahead of most voters but also polarize faster. At each election, each party polarizes by an amount $\lambda\tau$ (and farther in the second election when λ is high), whereas the voters polarize strictly less than this amount. Only precisely at the boundary of a party's support—the voter exactly λ from the party's position—is updating by the full amount of $\lambda\tau$. All voters with ideal points closer to the party update by less. Consequently, the parties actually move away from their more moderate supporters during the polarization process. Only when the parties stabilize at their ideal points do the more moderate voters begin to catch up.

The slower polarization of voters manifests itself at the aggregate level in the distribution of ideal points that develops. As the parties polarize, they pass many of their own voters, which leads to a large majority of voters being located on the inside of each party's position. Moreover, the faster polarization of the parties leads to much of this mass accumulating at the inside fringe of each party's support. Thus, the great bulk of a party's support ends up being more moderate than the party itself, reinforcing the impression that voters lag the parties as they polarize.

Figure 6 depicts the distribution of citizen ideal points to the right of zero after the election at which party R first locates at its ideal point. The coarseness of the figure obscures much of the richness of the distribution. At a fine micro level, the distribution has both lumps and discontinuities. On the right side of party R, each new interval of citizens drawn in to vote create their own voting block disconnected from the other voters, with the gaps contracting over time. On the left side of R, the distribution has no gaps, but it has lumps as these cohorts are drawn in to vote and become embedded into the overall distribution. Of course, this distribution is itself only transitory, because, even though the parties no longer change positions, the voters continue to converge on the parties and the distribution of ideal points increasingly collapses around these points.[23]

The fourth and final feature of voter behavior we emphasize is turnout. As the parties polarize, the "no-voter-left-behind" strategy means that no

[22] Citizens with ideal points only just beyond the party ideal point will exhibit a mixture of these properties. When drawn in to vote, they may cross over *R*. Thus, when the party position crosses their ideal point, they will reverse course and polarize, but they will converge only on *R* and not again reach their own original ideal point.

[23] The lumpiness and gaps in the distribution are not necessary for our equilibrium result. The party's calculus depends on the density of voters at the inside and outside margins (at $\pm\lambda$ from the party position). Therefore, adding noise, or heterogeneity, in voter updating that smooths out the distribution of ideal points will not upset the intuition for the result, although it will complicate the analysis.

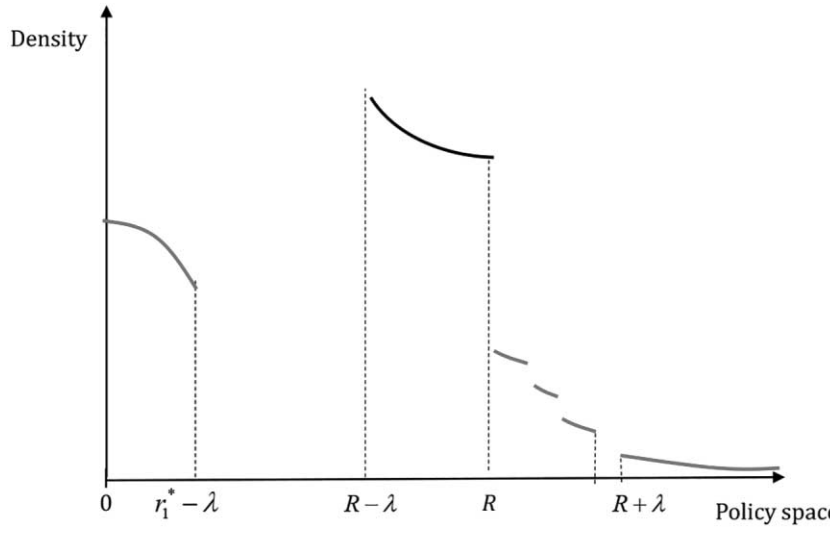

FIG. 6.—Distribution of citizen ideal points. A color version of this figure is available online.

voters are ever lost to abstention, whereas new voters are gathered in on the flanks (as is evident in fig. 5). Thus, aggregate turnout strictly increases from election to election until the parties reach their ideal points, after which turnout stabilizes. Turnout will remain incomplete, as abstainers out on the far flanks will never be drawn in to vote and, for low levels of voter tolerance, centrist abstainers will only grow ever more alienated. Figure 7 depicts the limit levels of turnout overlaid on the turnout rate at the first election for $\beta = 0$. An interesting effect is that the turnout gap between low and high values of λ contracts over time, both in an absolute sense and more dramatically in a relative sense. This is because, regardless of the level of voter tolerance, the parties end up at the same policy position and sweep up all of the voters in their path as they traverse their path outward.

VI. Theoretical and Empirical Implications of the Model

The features of polarization described in the previous section, with the exception of turnout, resonate with the data. Polarization of elites has been significant, and it occurred earlier, faster, and to a greater extent than polarization of voters. The model also rationalizes the ostensibly distinct phenomenon of negative partisanship. All voters are moving away from the opposing party as the parties polarize. More strikingly, at the early stages of the

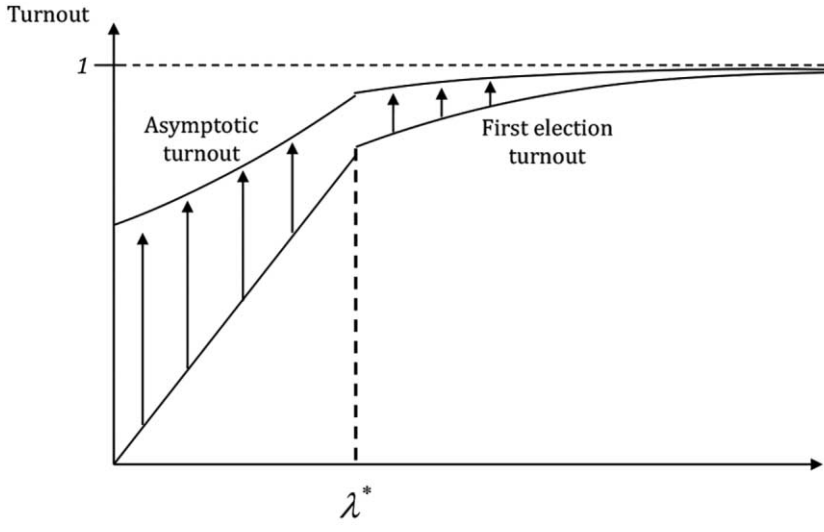

FIG. 7.—Dynamic turnout.

polarization process most voters are not getting closer to their favored party, and many are actually falling farther away. It is the distinct combination of candidate evaluations that defines negative partisanship.

The dynamic path of turnout does not fit the data as well. Although it is common to lament a decline in turnout in US national elections, the evidence suggests that turnout has been relatively constant throughout the era of polarization (McDonald and Popkin 2001). Either way, it does not match the prediction of increasing turnout in the model. In the next section, we offer an extension to the model to allow for generational turnover, and we discuss how this realistic enrichment can reconcile the model's prediction with empirical observation and do so in a way consistent with the differential patterns of turnout across generations.

In addition to explaining the past, we can also put the model to the work of prediction. What is the future of polarization? Will elites and masses continue to polarize? Will elites always be more polarized than the masses? Our model suggests answers to these questions. According to our model, the parties continue to polarize until they reach their own ideal points. Only then will polarization of elites stop. The polarization of the masses will continue beyond this time, albeit at a slower rate as voters converge upon the party location. The end point of the model is for voters of each party to form a homogeneous block at the location of their favored party, far removed from nonvoters. That voters in practice are currently less polarized than the parties suggests that this process has not yet run its course, which surely

is discomforting news for those who already lament the polarized state of politics today.[24]

This dynamic in voter ideal points implies that negative partisanship will ultimately weaken in practice. In the final convergent phase, most voters will continue to increase in their dislike of the opposition, whereas all voters will begin to rate their favored party more highly. This weakening will not imply a positive awakening in voters; rather, it will simply reflect the fact that the parties have stopped moving and the voters can catch up.

The model also speaks to an ongoing debate in the political science literature on the extent to which the masses have polarized, if at all. At the heart of this debate are two conflicting pieces of evidence. On one hand, support for moderate positions remains high, even after decades of elite polarization (Fiorina, Abrams, and Pope 2010). On the other hand, voters, and particularly those most engaged in politics, have polarized considerably (Abramowitz 2010). The patterns of behavior in our model offer one way to adjudicate this dispute. Although the mechanism in our model is simple, the dynamic shows how rich and varied voter behavior of this sort can emerge from the simple process of elite polarization. In our model, low voter tolerance implies the existence of centrist abstainers, and over time, these citizens remain exactly where they are. This rationalizes the evidence that moderate policies remain supported. At the same time, those who choose to vote polarize, and, consistent with Abramowitz (2010), the polarizers are those most engaged in politics, which creates a bimodal distribution of ideal points among voters.

More broadly, our model connects to the broader debates in political economy about the origin and nature of political preferences. On one side is the classic spatial theory of voting familiar from political economy. This theory places ideology and policy at its center, with a measure of distance that determines vote choice and political outcomes. In this theory, what Hall and Thompson (2018) label the "institutional literature," voters can swing from one party to the other if it moves closer. Opposing this view is what Hall and Thompson (2018) label the "behavioral literature." In this literature, dating back to Campbell et al. (1960) and Converse (1964), ideology plays little role and swing voters do not exist. Instead, voters are rigid partisans who are rallied to their team at each election. According to this view, voting is a purely partisan endeavor and ideological preferences are nothing more than ex post rationalizations of behavior. The model we introduce demonstrates how these contrasting perspectives, and the evidence in support of each, can be unified. Our theory is very much in Hall and Thompson's

[24] One possible prediction is that extreme polarization will induce entry by a third party. Although we do not consider this in our model, the evolution in our model points to a simple explanation for why entry does not occur despite the large gap between the parties—there may not be any voters left in the middle for an entrant to exploit.

"institutional literature." Yet by endogenizing the ideal points of citizens, we demonstrate how patterns are generated that resonate with the findings of the behaviorists. We show how voting is spatial within elections, as found by Jessee (2009, 2010), yet at the same time preferences can be responsive to party cues.

This duality can also inform the puzzle of why parties do not choose more moderate positions to appeal to the median voter. The behaviorists interpret this absence as evidence that voting is not spatial and ideological. Our theory suggests the potential error in this inference. In our model, voting is purely ideological, but because of the updating process and the polarization of the parties, a party relocating to the middle is not profitable, as the middle has been hollowed out, with the unattractiveness of this strategy increasing the longer the polarization process goes on. The "disappearing center," as Abramowitz (2010) calls it, implies that even in a purely ideological world, the distribution of citizen ideal points can become bimodal, and political competition can resemble a contest of voter mobilization rather than one of persuading the median voter, as evidence suggests has become the dominant strategy amongst parties in US elections.

These broader debates have previously found their way into the literature on polarization. The behaviorist literature, in dismissing ideology and voter agency, naturally converged on the conclusion that polarization is elite led (Lenz 2012; Achen and Bartels 2016). As neatly as this behaviorist view rationalizes polarization, it fails to explain why polarization has been slow and iterative rather than dramatic, and indeed, why there had to be a process at all—if elites are unconstrained ideologically, why were they ever not polarized? These gaps demonstrate the importance of explaining both the state and the process of polarization. We match the data that the behaviorist literature emphasizes—that it is the elites that polarize first—but we are also able to explain both the progressivity of polarization and the sequencing of polarization by the elites and the masses. Critically, the rationale for polarization that we offer is entirely at odds with the behaviorist view of politics. We show that what might appear to be an elite-driven polarization process is actually a process driven by voters and the manner in which their ideological preferences are constructed.

Our model opens a new perspective on political representation. If the preferences of the citizenry are evolving, what does it mean for them to be represented in politics? Should representation be measured by where citizens begin, by where they end, or by a series of snapshots at each point along the dynamic process? Interestingly, the updating of preferences by voters embeds a positive force for representation into the system, as, at least measured naively, representation inexorably increases as voters and parties converge. In contrast to standard notions of representation, however, the convergence is not because parties move to where voters are, but because voters move to the parties.

VII. Extending the Model: Applications and Conjectures

The power of our model is in its simplicity, that a simple amendment to voter behavior can generate a feedback loop between elite and mass behavior and drive rich dynamics. To focus on the feedback loop, we chose to keep the rest of the model as simple and transparent as possible. This raises two questions: To what extent are the predictions of the model robust to alternative specifications? and What additional features of the data can be explained by extensions of the model? In this section, we consider both of these questions informally. We sketch out the intuition for how the model and its results might be broadened, although we do not provide formal results.

A. Party Preferences

In the model, we assume that the parties are short-sighted and that their ideal points are fixed. We see as a strength that our result emerges even with parties that are not forward-looking and, therefore, that it does not rely on any sort of intertemporal trade-offs. Progressive polarization follows from the feedback loop between voters and parties. If the parties were forward-looking, the feedback loop would continue and the polarization should only be more dramatic. A straightforward intuition is that the intensity of competition in the first election would increase as the parties foresee that a higher vote share initially will carry over to future elections. This is akin to the impact of switching costs in markets. Klemperer (1987) shows how price competition is more intense early and weakens later once consumers are attached to one supplier.[25] The same dynamic is likely to play out here: more convergence in the first period as parties compete intensely for votes but with the same end point (full polarization), creating an even broader sweep of polarization by the parties over time.[26]

The assumption that party ideal points are fixed implies that parties have always been extreme and that polarization has empowered them to express those preferences in policy. In practice, the preferences of political parties are a complicated amalgamation of many forces, and it is plausible,

[25] An even more explicit strategy was laid out by the Chinese philosopher Laozi, on how a leader can exploit the attachment of the people to serve his own ends: "The ruler is thus able to accomplish everything, but it will seem to the people as though everything is simply occurring naturally, without any directing will: 'When his achievements are completed and tasks finished, the commoners say that "We are like this naturally (*zi ran*)"'" (Puett and Gross-Loh 2016).

[26] Allowing for forward-looking behavior suggests some interesting possibilities. In particular, it may be that the parties polarize beyond their own ideal point to gather more voters, then reverse course to converge back to their ideal points. (This possibility may even occur with myopic parties and convex utility as the competitive margin of winning more votes actually pushes parties outward rather than inward.)

indeed likely, that the political parties have themselves developed more extreme preferences over time. A natural formulation is that party preferences are an aggregation of the preferences of their members or, more narrowly, their elected representatives. As long as the preferences of those within the party evolve in a way that is more polarized than those among the electorate as a whole, our results should carry over directly to such an environment. Indeed, this relative ordering between party insiders and rank-and-file supporters is evident in the data and has held throughout the period of polarization (Abramowitz 2010; Bartels 2016). In a richer model such as this, the feedback loop would move from voter preferences to the preferences and membership of the parties and then into the party platform, before closing the loop to voter preferences. Documenting and understanding this mechanism rigorously is a promising direction for further work.

B. Voter Updating

For simplicity, we posited a simple voter updating rule that is linear and, more importantly, operates exclusively through the act of voting. This channel of feedback is empirically supported, as we noted above, although it is not the only way that behavior can feed back into preferences. Indeed, Beasley and Joslyn (2001) provide evidence that the outcome of the election also matters, with voters who support the winner updating more than those who voted for the loser. In our model, this would create an asymmetry in the distribution of ideal points and a path dependence in the trajectory of polarization. (We sketch below some of the possibilities this gives rise to.) It is possible that other actions or observations of politics can affect preferences, such as attending a political rally, witnessing a powerful speech, or door knocking for a campaign. To induce cognitive dissonance, such an event must be important enough to create a conflict with preferences. More empirical work is needed to determine which acts and observations, beyond the act of voting, are sufficient to create this effect.[27]

Another feature of updating in our model is that it distinguishes only between voters and abstainers. A generalization that may more closely match the data is to calibrate updating to the degree of a citizen's engagement in politics. Another possibility is to incorporate noise, or randomness, in which voters update and by how much. This would seem to be particularly pertinent, given the above discussion, as which citizens participate in rallies or hear speeches will be unevenly distributed across the population. One upside of this randomness is that it could rationalize the existence of party

[27] A difficulty in identifying such a causal link empirically is separating any effect from the flow of information about political positions. Dinas, Hartman, and van Spanje (2016) provide suggestive evidence that emotion—namely, sympathy—can feed back into political preferences, although whether the mechanism is cognitive dissonance is unclear.

switchers, those who support a Democrat, say, in one election and then switch to a Republican in the next. We do include randomness in the model, although, for tractability, we do it in a reduced form. Incorporating these effects directly into voter utility would enrich the model and also offers the promise of deeper insight into voter behavior and party strategy, although a formal analysis is beyond the scope of this paper.

C. *Presidential versus Congressional Elections*

The model is of a single election, yet in the US context, citizens vote for president, the House, and typically the Senate and a host of lesser offices. This raises the question, Which vote cast feeds back into preferences? This is particular relevant if the voter splits the ticket, casting votes for candidates from different parties. The preceding discussion suggests that the most salient office is the most likely to induce cognitive dissonance if a vote conflicts with preferences. In the United States, this would most likely be the vote for president; in countries with parliamentary systems, it would be votes for legislative representatives. We know of no work that considers the impact of cognitive dissonance on voter preferences in different types of elections (presidential, congressional, etc.), although it would be of clear interest.

Much of the emphasis on polarization in the United States is on members of Congress. Data on the polarization of presidential candidates is noisier (not having legislative roll call votes) and more idiosyncratic, given the smaller number, yet it too displays the same trends as evident in Congress. Bartels (2016, 155) documents the polarization of presidential candidates over time, concluding that "[t]he positions of both parties' presidential candidates have become more extreme over the past three decades."[28] Consistent with our model, Bartels finds that the positions of the presidential candidates are more extreme than those of core partisan voters.[29]

D. *Asymmetric Polarization*

Throughout the baseline model, we work with symmetric parties and voters. In practice, polarization in the United States has been asymmetric. For Congress, the ordering is clear, with one side (the Republicans) polarizing faster and to a greater extent than the other. The evidence on presidential evidence is less clear, although also less robust, yet an asymmetry is evident, particularly across issues (Bartels 2016).

[28] To overcome what he describes as the "rough and often idiosyncratic summary" of ideological self-identification, Bartels (2016, 149) employs positions on four salient policy issues as well as the overall ideological positioning.

[29] Bartels (2016) also finds that core partisans have polarized more than the presidential elites, closing the gap between them. This finding resonates with the idea that those more engaged in politics update more than those who have only a fleeting interest.

In the context of the feedback loop in our model, one can see how small asymmetries can reverberate over time into larger asymmetries in aggregate polarization. For instance, if, say, Republican voters update their ideal points slightly more than Democratic voters, then the Republican Party will be free to polarize faster. Another possibility is that the Republican Party simply has more extreme preferences than does the Democratic Party. In this case, the Republican Party will polarize no faster than Democrats but will polarize farther than them. The presence of both factors will generate both faster and larger polarization. As neither of these extensions changes behavior in the first election, the existence of equilibria of the form just described follows immediately from our results as long as the asymmetry in updating or in preferences is not too large.

A third possibility is that asymmetries in polarization follow from the random luck of winning elections. As mentioned above, voters may update their ideal points more if the party they support wins the election. Thus, an early stroke of good fortune with an election win can lead to more voter updating and, along the lines of the previous paragraph, more polarization.

This possibility resonates with the path dependence and momentum effects that are ubiquitous features of politics (Pierson 2000, 2004; Page 2006). Momentum emerges because greater updating of ideal points after voting for a winner creates its own feedback effect. If, say, the good fortune of Ronald Reagan emerging and winning the 1980 presidential election early in the era of polarization induced more updating by Republican voters, then the party would have been able to polarize more without losing voters toward the center while gaining more voters on the flank. This means that the Reagan win would have mechanically increased the Republican vote share in the subsequent election and would, therefore, increase the party's probability of winning, creating even more voter updating in a reinforcing cycle. It is significant that the electoral advantage this creates is not despite greater Republican polarization but precisely because of it.

E. Convergence and Polarization in US History

The focus of our paper is on the era of polarization that began in the second half of the twentieth century and has run through to the present day and, presumably, into the future. The evolution of political preferences did not begin with this era, of course, and before the era of polarization there was an extended period of convergence. In fact, the major parties in the United States had reached such a point of convergence in the 1950s that the American Political Science Association famously issued a report lamenting that the parties were too close together and that they offered voters an insufficiently differentiated choice. This juxtaposition leads to the question of why, then. Why did polarization begin when it did? And what caused the moderation of parties in the earlier era? An explanation of the broader sweep of

American political history is, unfortunately, beyond the scope of this paper, although obtaining such an explanation is of clear importance. Toward this end, and to show how our theory of polarization can be used as a piece of this broader understanding, we sketch here how our model could be extended to capture an even longer and richer period of preference dynamics.

One restriction of our model is that the population of citizens is fixed. This is particularly important because we seek to explain a dynamic phenomenon that stretches over more than half a century, and during that time there has been substantial turnover in the citizens of voting age. Consider, then, our model with the addition of births, deaths, and generational turnover. To be more specific, suppose that after each election a new generation is born of mass ρ and that each generation lives for some finite number of periods, T.[30] To fix ideas, suppose that each generation arrives according to the same original distribution F.

The population at the first election will be the same as in the baseline model, and the equilibrium will go through unchanged. For the second election, the equilibrium logic continues to go through as long as the mass of the new generations born is not too large, and the same holds for the third election, and so on.[31] The parties begin the same polarization process, and, as in the baseline model, the same patterns of elite and mass polarization begin to emerge.

Fast-forward, then, to the point where the parties reach their own ideal points and stabilize. In the baseline model, no new voters are won over and the set of existing voters remains stable, converging in preference around the parties. With births and deaths, however, this set of voters is not stable but is dying off approximately at rate ρ. At the same time, the new generations are arriving with distribution F, and if the parties are polarized, not many of these new voters are falling into each party's interval of support. This leads to a different trajectory for turnout than in the baseline model. Rather than increasing throughout the polarizing phase and remaining stable thereafter, turnout is tempered by the newly born who are not captured by one of the parties, and it declines once the parties stabilize at their ideal points. This pattern matches more accurately the evidence from US elections (McDonald and Popkin 2001), reconciling a discordant prediction from the baseline model.

The changing pattern of turnout also matters for polarization and the location of the parties. As a large mass of unattached citizens accumulates in the center of the distribution, the opportunity emerges for a party to move its position and appeal to them. It is intuitive to see that this generational

[30] To allow a seeding process, we can think of the original generation dying off at rate ρ from the second election onward.

[31] As this is only a sketch, we do not dwell on the precise range of ρ that satisfies this argument, although it is clear that it is not empty.

turnover will lead inevitably to the end of polarization. Eventually, one of the parties will find it optimal to move to the center to win over abstainers. An open question is whether the movement inward will come suddenly or progressively. Will a party make a sudden jump to the center, alienating its core supporters, while awakening a new generations of voters? Or will the parties inch inward progressively, attracting new voters and dragging their old voters with them toward the center? Whichever is the answer, the inevitable lure of the center will surely comfort those concerned with today's state of polarization, yet each possibility implies very different timing for the end of polarization and suggests different types of politics.

The inclusion of generational turnover provides a natural and simple explanation for cycles of moderation and convergence. Appealingly, it also creates patterns at the micro level consistent with observation. The assumption that each generation arrives according to the same distribution, F, presumes that citizens come of age unmolded by politics. Although somewhat extreme, this assumption is nevertheless consistent with the core assumption of our model that the act of voting itself shapes political preferences. Empirically, it is supported by evidence that a citizen's political preferences are shaped significantly by the first presidential election in why they are eligible to vote. Indeed, Ghitza and Gelman (2014) show how a citizen's lifetime of presidential election shapes their preferences in a sort of running-tally way, with by far the most weight on the first election.[32]

Regardless of how polarization breaks down and party moderation obtains, it is clear that the parties, having captured the mass of young centrist voters, will once again discover the incentive to polarize and that the polarization process will begin anew. Of course, throughout the period of polarization, new generations will continue to be born, and with these new generations arriving uncaptured at the center, the process of polarization will, eventually, break down again. The cycle that this creates demonstrates how the moderation of the first half of the twentieth century can fit naturally with the polarization of the second half.[33] It also suggests why the alarm of the

[32] The combination of this assumption and the dynamics of party polarization also creates cross-generational patterns that match long-standing empirical findings. The most striking implication is that the propensity to vote strictly increases as citizens age. This is because older generations are, in a sense, captured by the party and pulled with it toward the extreme. A newly born citizen may land at the same moderate location as her mother did, yet the daughter will abstain, as the party has polarized and alienated her, whereas the mother updates toward the party position and continues to vote. Notably, this prediction is not dependent on party polarization. All that is required is movement by the parties, and so the same pattern would emerge during a period of party moderation. It is significant, therefore, that higher turnout among older generations is a prominent and persistent feature of the data, dating back to the seminal work by Wolfinger and Rosenstone (1980) and continued since then. See http://www.electproject.org/home/voter-turnout/demographics.

[33] Such a cycle of polarization through births and deaths is also consistent with the scattered evidence and popular conception that today it is largely the old who are the radicals, whereas in the 1960s it was the young. This turnover matches exactly the beginning of the

American Political Science Association was misplaced. With the inevitability of generational turnover, cycles of moderation and polarization are likely to be the norm rather than the exception of political dynamics.

F. Polarization around the World

The United States is not alone in experiencing movements of positions among the elites and masses. The United States does stand out, however, in the consistency and degree of polarization that has occurred since the mid-to-late twentieth century (Rehm and Reilly 2010). Indeed, the experience in the United Kingdom was a steady decline in polarization through the final decades of the century, followed by a change in direction in the past 20 years leading up to and including the vote over Brexit (Boxell, Gentzkow, and Shapiro 2021).[34]

The previous section made clear that depolarization is not inconsistent with our theory. At heart, our theory predicts a comovement of party and electorate positions. Given this, the logic of party positions then depends on the nature of electoral competition, in particular the number of parties competing, the electoral rules in place, and the distribution of ideal points that the parties face. For instance, in the United Kingdom the decline of the union movement may have led to new generations being more centrist than in the past, inducing the parties to converge to the center to capture these votes. Another difference in the United Kingdom is the entry of a credible third party in the Liberal Democrats. With a center-left location, the Liberal Democrats changed the calculus for the Labour Party, perhaps necessitating a move to the center for competitive reasons.[35] The strategy of political position is even more complicated in countries with multiparty systems and a proportional representation electoral rule.

What is important for the purposes of our theory is that this elite depolarization induced depolarization of the masses, as voters update their ideal points after each election, and that this depolarization is more modest and later than for elites. This pattern is evident in the data, and as Adams, Green, and Milazzo (2012, 507) conclude, the trends in the United Kingdom in the last two decades of the twentieth century are a "mirror image"

---

modern era of polarization in the 1970s. (To the extent that young people are radical today, they are more likely to fall on the left. This asymmetry may follow from the slower polarization of the Democratic Party or may suggest that the arrival of new generations of voters is skewed left.)

[34] Evidence across countries is scant and difficult to compare. Boxell, Gentzkow, and Shapiro (2021) provides the best evidence of comparative polarization, although it reports only measures of affective polarization at the mass level.

[35] Indeed, as we point out in the discussion following proposition 5, moderate convergence also strictly increases vote share as the party brings supporters with them. With two parties and a distribution of ideal points f, we show that divergence is nevertheless strictly preferred. If a third sits at the center, or if the distribution of ideal points takes a different shape, then that logic may change and one or both of the parties prefer to converge.

of those in the United States. Since that time, and in line with our theory, once the parties have become very centrist, their incentive is to polarize, and this has been evident in the most recent decade.

These are only speculations, to be sure. And much more detailed work, both empirical and theoretical, is necessary to trace through the implications of the mechanism we uncover here for the United Kingdom and other countries. The core takeaway is that polarization is not just something that happens exogenously. It is the consequence of strategic calculation by political parties and the process by which voters react to the changing circumstance and update their political preferences. Applying this idea to settings beyond the United States is a promising direction of work and offers the prospect of unifying understanding of polarization (and depolarization) trends around the world.

VIII. Conclusion

To return to the motivating question of what causes polarization, our answer is that it is complicated. We show that the necessary ingredient for polarization is the interaction of voters and elites. Voter updating is necessary, yet on its own does not guarantee polarization. It is only when updating is combined with the strategic maneuverings of party elites that a feedback loop is created and polarization occurs. However one interprets responsibility within this relationship, the depth and subtlety of the codetermination in this process point to why researchers have had such difficulty in isolating an individual cause of polarization. Our results establish that a focus on the process of polarization—and not just on the state of polarization—is essential to an understanding of polarization's origins, its impact, and its future.

Appendix A

Equilibrium in the First-Period Electoral Competition

For simplicity of notation, we ignore the time subscript to analyze the first-period election. Throughout, we assume that policy positions satisfy $d \leq r$. The opposite case never occurs in equilibrium, given parties' preferences. These preferences also imply that, in any equilibrium, $D \leq d$ and $r \leq R$. Thus, we treat the first-period policy space for both parties as the compact, nonempty interval $\mathcal{P} = [D, R]$. Throughout, we assume that the tolerance parameter λ is bounded away from zero by some $\underline{\lambda} > 0$, although this last can be arbitrarily close to zero.

Given $r, d \in \mathcal{P}$, and $\lambda \geq \underline{\lambda}$, the *vote total functions* for each party are

$$\mathrm{VR}(r, d; \lambda) = \int_{x \in \eta_R} f(x)\, dx \quad \text{and}$$

$$\mathrm{VD}(r, d; \lambda) = \int_{x \in \eta_D} f(x)\, dx,$$

where η_R and η_D are the intervals of voters' support for the parties and f is a log-concave density function, symmetric around zero, strictly positive, and continuous on $(-\infty, \infty)$. Party R's *vote share function* (winning probability) is therefore

$$S(r, d; \lambda) = \frac{\mathrm{VR}(r, d; \lambda)}{\mathrm{VR}(r, d; \lambda) + \mathrm{VD}(r, d; \lambda)}.$$

Party D's vote share function is, of course, $1 - S(r, d; \lambda)$.

Voters' support for each party depends on the value of the tolerance region with respect to the distance $r - d$. There are two distinct cases to consider. In regime A, λ is sufficiently large relative to the distance $r - d$ that the parties' intervals of support have a common boundary: only extreme voters abstain. In regime B, the tolerance region is sufficiently small that parties' intervals of support do not touch: there is abstention in the middle as well as in the extremes. To be precise:

Regime A. abstention only in the extremes; when $\lambda \geq (r - d)/2$, we have

$$\eta_D^A = \left[d - \lambda, \frac{r + d}{2}\right] \quad \text{and} \quad \eta_R^A = \left[\frac{r + d}{2}, r + \lambda\right];$$

Regime B. abstention in the middle and in the extremes; when $\lambda \leq (r - d)/2$, we have

$$\eta_D^B = [d - \lambda, d + \lambda] \quad \text{and} \quad \eta_R^B = [r - \lambda, r + \lambda].$$

To distinguish between these cases, we denote the vote total functions VR^k and VD^k, for $k = \mathrm{A}, \mathrm{B}$. Likewise, we let $S^k(r, d; \lambda)$ denote party R's vote share function for $k = \mathrm{A}, \mathrm{B}$. To be explicit, the vote total functions for regime A and regime B are, respectively,

$$\begin{aligned} \mathrm{VR}^A(r, d; \lambda) &\equiv F(r + \lambda) - F\left(\frac{r + d}{2}\right), \\ \mathrm{VD}^A(r, d; \lambda) &\equiv F\left(\frac{r + d}{2}\right) - F(d - \lambda), \end{aligned} \tag{A1}$$

and

$$\begin{aligned} \mathrm{VR}^B(r, d; \lambda) &\equiv F(r + \lambda) - F(r - \lambda), \\ \mathrm{VD}^B(r, d; \lambda) &\equiv F(d + \lambda) - F(d - \lambda). \end{aligned} \tag{A2}$$

A1. Expected Utility as a Product Function

With linear preferences, R's expected utility, given $d \leq r$, can be written as

$$\mathbb{E}U(r, d; \lambda, \beta) = S(r, d; \lambda)(r - d + \beta) - (R - d). \tag{A3}$$

Thus, R's expected utility is proportional to the difference between policy positions. Increasing r yields a higher payoff, conditional on winning, but can negatively affect

R's winning probability. Since the expected utility is a product function, in our analysis we employ the following result —for a proof, see Kantrowitz and Neumann (2007).

THEOREM A1. Consider log-concave continuous functions g_1 and g_2 defined on a closed, bounded interval $[a, b]$, and suppose that $g_i(x) > 0$ for all $x \in (a, b)$, for $i = 1, 2$. If one of these functions is strictly log concave, then there exists a point $x^* \in [a, b]$ such that the product function $h = g_1 g_2$ is strictly increasing on $[a, x^*]$ and strictly decreasing on $[x^*, b]$.

An affine function is strictly log concave. Thus, to employ theorem A1, it suffices to show that R's vote share function $S^k(r, d; \lambda)$ is log concave on r, for $k = \text{A, B}$. Since f is log concave, the associated distribution function F is log concave (see, e.g., Bagnoli and Bergstrom 2005). Unfortunately, this fact does not translate immediately into the log concavity of the vote share function, because this is a quotient function for which both the numerator and the denominator are affected by the choice of party R. We start by showing that the individual components of S^k are log concave on r.

LEMMA A1. Fix $d \in \mathcal{P}$ and $\lambda \geq \underline{\lambda}$. For $k = \text{A, B}$, the vote total functions $\text{VR}^k(\cdot, d; \lambda)$ and $\text{VD}^k(\cdot, d; \lambda)$ are log concave on r for all $r \geq d$.

Proof. To deal with both regimes simultaneously, we show that if $p(\cdot)$ and $q(\cdot)$ are increasing, affine functions of r, with $p(r) > q(r)$ for all $r \geq d$, then $F(p(r)) - F(q(r))$ is log concave with respect to r. To do so, we adapt the proof technique of Pratt (1981).

For any two real numbers $p > q$, write the difference between $F(p)$ and $F(q)$ as

$$F(p) - F(q) = \int I(x, p, q) f(x)\, dx,$$

where $I(x, p, q) = 1$ for all $q < x < p$ and $I(x, p, q) = 0$ otherwise. Note that the integrand of the above expression is the product of two functions. Since $I(x, p, q)$ is log concave in (x, p, q) and $f(x)$ is log concave in x, and hence in (x, p, q), the integrand is also log concave in (x, p, q). By the Prékopa theorem, log concavity is preserved by marginalization, and thus $F(p) - F(q)$ is log concave in (p, q). Thus, we have that the function

$$g(p, q) = \log(F(p) - F(q))$$

is concave on (p, q) for all $p > q$. Now, a standard argument on concave functions shows that if $p(r)$ and $q(r)$ are increasing, affine functions, with $p(r) > q(r)$ for all $r \geq d$, then the function

$$h(r) = g(p(r), q(r)) = \log(F(p(r)) - F(q(r)))$$

is concave in r for all $r \geq d$, as desired. QED

To simplify our notation, we use subscripts to refer to the partial derivatives of multivariate functions; for instance, we write $S^k_r(r, d; \lambda)$ to denote the partial derivative of S^k with respect to r in regime $k = \text{A, B}$, and so on. Recall that the log concavity of $S^k(r, d; \lambda)$ on r is equivalent to the quotient S^k_r / S^k being everywhere decreasing with respect to r for fixed d and λ. The next lemma allows us to express this quotient in a convenient way.

LEMMA A2. Given the vote share function $S^k(r, d; \lambda)$ $(k = \mathrm{A}, \mathrm{B})$, one has

$$\frac{S^k_x(r, d; \lambda)}{S^k(r, d; \lambda)} = (1 - S^k(r, d; \lambda))\left(\frac{\mathrm{VR}^k_x(r, d; \lambda)}{\mathrm{VR}^k(r, d; \lambda)} - \frac{\mathrm{VD}^k_x(r, d; \lambda)}{\mathrm{VD}^k(r, d; \lambda)}\right)$$

for $x = r, d$.

Proof. The proof is immediate from taking the partial derivative of $S^k(r, d; \lambda)$ with respect to $x = r, d$ and simplifying the resulting expression. QED

Employing lemma A2 for $x = r$, we obtain

$$\begin{aligned}\frac{\partial}{\partial r}\left(\frac{S^k_r(r, d; \lambda)}{S^k(r, d; \lambda)}\right) &= -S^k_r(r, d; \lambda)\left(\frac{\mathrm{VR}^k_r(r, d; \lambda)}{\mathrm{VR}^k(r, d; \lambda)} - \frac{\mathrm{VD}^k_r(r, d; \lambda)}{\mathrm{VD}^k(r, d; \lambda)}\right)\\ &\quad + (1 - S^k(r, d; \lambda))\frac{\partial}{\partial r}\left(\frac{\mathrm{VR}^k_r(r, d; \lambda)}{\mathrm{VR}^k(r, d; \lambda)} - \frac{\mathrm{VD}^k_r(r, d; \lambda)}{\mathrm{VD}^k(r, d; \lambda)}\right)\\ &= -S^k(r, d; \lambda)(1 - S^k(r, d; \lambda))\left(\frac{\mathrm{VR}^k_r(r, d; \lambda)}{\mathrm{VR}^k(r, d; \lambda)} - \frac{\mathrm{VD}^k_r(r, d; \lambda)}{\mathrm{VD}^k(r, d; \lambda)}\right)^2\\ &\quad + (1 - S^k(r, d; \lambda))\frac{\partial}{\partial r}\left(\frac{\mathrm{VR}^k_r(r, d; \lambda)}{\mathrm{VR}^k(r, d; \lambda)} - \frac{\mathrm{VD}^k_r(r, d; \lambda)}{\mathrm{VD}^k(r, d; \lambda)}\right).\end{aligned} \tag{A4}$$

As $0 < S^k(r, d; \lambda) < 1$, the log concavity of the vote share function hinges on the sign of the last expression in equation (A4). In regime B, the vote total function for D is constant on r, so immediately we obtain the following lemma.

LEMMA A3. The vote share function $S^{\mathrm{B}}(\cdot, d; \lambda)$ is log concave on r for all $r \geq d$, consistent with regime B.

Proof. The vote total function $\mathrm{VD}^{\mathrm{B}}(r, d; \lambda)$ is constant on r (see eq. [A2]). This allows us to simplify equation (A4) to obtain

$$\begin{aligned}\frac{\partial}{\partial r}\left(\frac{S^{\mathrm{B}}_r(r, d; \lambda)}{S^{\mathrm{B}}(r, d; \lambda)}\right) &= -S^{\mathrm{B}}(r, d; \lambda)(1 - S^{\mathrm{B}}(r, d; \lambda))\left(\frac{\mathrm{VR}^{\mathrm{B}}_r(r, d; \lambda)}{\mathrm{VR}^{\mathrm{B}}(r, d; \lambda)}\right)^2\\ &\quad + (1 - S^{\mathrm{B}}(r, d; \lambda))\frac{\partial}{\partial r}\left(\frac{\mathrm{VR}^{\mathrm{B}}_r(r, d; \lambda)}{\mathrm{VR}^{\mathrm{B}}(r, d; \lambda)}\right).\end{aligned}$$

Since $\mathrm{VR}^{\mathrm{B}}(\cdot, d; \lambda)$ is log concave with respect to r, the term $\mathrm{VR}^{\mathrm{B}}_r(r, d; \lambda)/\mathrm{VR}^{\mathrm{B}}(r, d; \lambda)$ is decreasing in r for all $r \geq d$. It follows that the above expression is weakly negative, and thus $S^{\mathrm{B}}(\cdot, d; \lambda)$ is log concave on r, as desired. QED

Obtaining the log concavity of the vote share function for regime A is more complicated, as both $\mathrm{VR}^{\mathrm{A}}(\cdot, d; \lambda)$ and $\mathrm{VD}^{\mathrm{A}}(\cdot, d; \lambda)$ depend on r. Intuitively, we require that the first vote total function be "more log concave" than the second one, that is,

$$\frac{\partial}{\partial r}\left(\frac{\mathrm{VR}^{\mathrm{A}}_r(r, d; \lambda)}{\mathrm{VR}^{\mathrm{A}}(r, d; \lambda)}\right) \leq \frac{\partial}{\partial r}\left(\frac{\mathrm{VD}^{\mathrm{A}}_r(r, d; \lambda)}{\mathrm{VD}^{\mathrm{A}}(r, d; \lambda)}\right).$$

We introduce the following condition.

DEFINITION A1. Given a log concave density function f defined on $\mathbb{R}$, for all $x > y$ let

$$P(x, y) = F(x) - F\left(\frac{x + y}{2}\right) \quad \text{and}$$

$$Q(x, y) = F\left(\frac{x + y}{2}\right) - F(y).$$

The density f is said to be midpoint ratio log concave (MRLC) if the quotient function $P(x, y)/Q(x, y)$ is log concave with respect to x on $(y, +\infty)$ for all $y \in \mathbb{R}$.

In appendix C, we show that the class of MRLC densities is nonempty. In particular, we prove analytically that MRLC is satisfied if the density function f is that of the logistic or the uniform distributions. Employing numerical methods, we also show that MRLC is also satisfied for the normal distribution for the relevant parameter values. The importance of MRLC is that it is a sufficient condition for the log concavity of the vote share function in regime A.

LEMMA A4. Suppose that f is MRLC. Then the vote share function $S^{A}(\cdot, d; \lambda)$ is log concave on r for all $r \geq d$, consistent with regime A.

Proof. Note that we can write

$$\frac{\partial}{\partial x}\left(\ln \frac{P(x, y)}{Q(x, y)}\right) = \frac{\partial}{\partial x}(\ln P(x, y) - \ln Q(x, y))$$

$$= \frac{P_x(x, y)}{P(x, y)} - \frac{Q_x(x, y)}{Q(x, y)}.$$

Next, observe that in regime A we can express the vote total as

$$\mathrm{VR}^{A}(r, d; \lambda) = F(r + \lambda) - F\left(\frac{(r + \lambda) + (d - \lambda)}{2}\right) = P(r + \lambda, d - \lambda),$$

$$\mathrm{VD}^{A}(r, d; \lambda) = F\left(\frac{(r + \lambda) + (d - \lambda)}{2}\right) - F(d - \lambda) = Q(r + \lambda, d - \lambda).$$

Thus, since f satisfies MRLC, we have that

$$\frac{\partial}{\partial r}\left(\frac{\mathrm{VR}^{A}_r(r, d; \lambda)}{\mathrm{VR}^{A}(r, d; \lambda)} - \frac{\mathrm{VD}^{A}_r(r, d; \lambda)}{\mathrm{VD}^{A}(r, d; \lambda)}\right) \leq 0.$$

This last expression, combined with equation (A4), obtains the result. QED

A2. Equilibrium

Obtaining Nash equilibria in the first-period electoral competition game is complicated by the fact that there can be an endogenous switch between regimes, depending on what policy positions the parties announce. We proceed by treating these regimes as two independent games. That is, we fix λ and β and restrict the policy space for each party to correspond to either regime A or regime B. Focusing on party R, we obtain R's best-response function in each regime and show that this leads to a unique symmetric equilibrium for each case. Finally, we put both regimes

together and construct the unique symmetric equilibrium for the electoral competition game between R and D in period 1.

Ignoring the constant $-(R - d)$, party R chooses r in regime $k = \mathrm{A}, \mathrm{B}$ to maximize

$$\mathbb{E}U^k(\cdot, d; \lambda, \beta) = S^k(r, d; \lambda)(r - d + \beta)$$

on its respective domain.

PROPOSITION A1. Let $\lambda > \underline{\lambda}$, $\beta \geq 0$, and $d \in \mathcal{P}$ be given.

a. Suppose that f is MRLC. The expected utility function $\mathbb{E}U^{\mathrm{A}}(\cdot, d; \lambda, \beta)$ defined for r on the interval $[d, d + 2\lambda] \cap \mathcal{P}$ has a unique maximum

$$r^{\mathrm{A}}(d; \lambda, \beta).$$

Moreover, $\mathbb{E}U^{\mathrm{A}}(\cdot, d; \lambda, \beta)$ is strictly increasing to the left of $r^{\mathrm{A}}(d; \lambda, \beta)$ and strictly decreasing to the right of $r^{\mathrm{A}}(d; \lambda, \beta)$.

b. The expected-utility function $\mathbb{E}U^{\mathrm{B}}(\cdot, d; \lambda, \beta)$ defined for r on the interval $[d + 2\lambda, +\infty) \cap \mathcal{P}$ has a unique maximum

$$r^{\mathrm{B}}(d; \lambda, \beta).$$

Moreover, $\mathbb{E}U^{\mathrm{B}}(\cdot, d; \lambda, \beta)$ is strictly increasing to the left of $r^{\mathrm{B}}(d; \lambda, \beta)$ and strictly decreasing to the right of $r^{\mathrm{B}}(d; \lambda, \beta)$.

Proof. Both parts a and b follow immediately from theorem A1, in conjunction with lemma A3 and lemma A4. QED

An important implication of proposition A1 is that first-order conditions are necessary, as well as sufficient, to obtain an interior maximizer of $\mathbb{E}U^k(\cdot, d; \lambda, \beta)$. Allowing for directional derivatives at the boundaries of the respective domains, the derivative of party R's objective function with respect to r in regime $k = \mathrm{A}, \mathrm{B}$ is expressed as

$$\begin{aligned}\frac{\partial \mathbb{E}U^k}{\partial r} &= S^k(r, d; \lambda) + S^k_r(r, d; \lambda)(r - d + \beta) \\ &= S^k(r, d; \lambda)\left[1 + (r - d + \beta)(1 - S^k(r, d; \lambda))\left(\frac{\mathrm{VR}^k_r(r, d; \lambda)}{\mathrm{VR}^k(r, d; \lambda)} - \frac{\mathrm{VD}^k_r(r, d; \lambda)}{\mathrm{VD}^k(r, d; \lambda)}\right)\right],\end{aligned}$$

where the second line of the equation uses lemma A2.

A3. Finding Nash Equilibria in Regime A

Assume throughout this section that the density function f satisfies MRLC. We start the analysis in regime A by showing that party R's optimal policy position $r^{\mathrm{A}}(d; \lambda, \beta)$ is never at the lower boundary of its domain, provided that the value of holding office is not too high. Recall that the policy space $\mathcal{P}$ is compact, and thus

$$M(\lambda) = \min_{x \in \mathcal{P}} \frac{F(x + \lambda) - F(x)}{f(x)} > 0.$$

Moreover, it is clear that $M(\lambda)$ is an increasing function, and thus we have $0 < M(\underline{\lambda}) \leq M(\lambda)$ for all $\lambda \geq \underline{\lambda}$.

LEMMA A5. Fix $\lambda \geq \underline{\lambda}$ and $d \in \mathcal{P}$. If the value of holding office satisfies $0 \leq \beta \leq 2M(\underline{\lambda})$, then $r^{\mathrm{A}}(d; \lambda, \beta) > d$.

Proof. Evaluating the partial derivative of $\mathbb{E}U^{\mathrm{A}}$ with respect to r at $r = d$ obtains

$$\frac{\partial \mathbb{E}U^{\mathrm{A}}}{\partial r}\bigg|_{r=d} = S^{\mathrm{A}}(d, d; \lambda)\left[1 + \beta(1 - S^{\mathrm{A}}(d, d; \lambda))\left(\frac{\mathrm{VR}^{\mathrm{A}}_r(r, d; \lambda)}{\mathrm{VR}^{\mathrm{A}}(r, d; \lambda)} - \frac{\mathrm{VD}^{\mathrm{A}}_r(r, d; \lambda)}{\mathrm{VD}^{\mathrm{A}}(r, d; \lambda)}\right)\bigg|_{r=d}\right].$$

Since $S^{\mathrm{A}}(d, d; \lambda) > 0$, it suffices to show that the expression in square brackets is strictly positive. To do this, note that

$$\begin{aligned} -\left(\frac{\mathrm{VR}^{\mathrm{A}}_r(r, d; \lambda)}{\mathrm{VR}^{\mathrm{A}}(r, d; \lambda)} - \frac{\mathrm{VD}^{\mathrm{A}}_r(r, d; \lambda)}{\mathrm{VD}^{\mathrm{A}}(r, d; \lambda)}\right)\bigg|_{r=d} &= -\left[\frac{f(d + \lambda) - (1/2)f(d)}{\mathrm{VR}^{\mathrm{A}}(d, d; \lambda)} - \frac{(1/2)f(d)}{\mathrm{VD}^{\mathrm{A}}(d, d; \lambda)}\right] \\ &< \frac{f(d)}{2}\frac{\mathrm{VR}^{\mathrm{A}}(d, d; \lambda) + \mathrm{VD}^{\mathrm{A}}(d, d; \lambda)}{\mathrm{VR}^{\mathrm{A}}(d, d; \lambda)\mathrm{VD}^{\mathrm{A}}(d, d; \lambda)}. \end{aligned}$$

As $1 - S^{\mathrm{A}} = \mathrm{VD}^{\mathrm{A}}/(\mathrm{VR}^{\mathrm{A}} + \mathrm{VD}^{\mathrm{A}})$, multiplying the above expression by $\beta(1 - S^{\mathrm{A}}(d, d; \lambda))$ and substituting the resulting equation obtains

$$\begin{aligned} -\beta(1 - S^{\mathrm{A}}(d, d; \lambda))\left(\frac{\mathrm{VR}^{\mathrm{A}}_r(d, d; \lambda)}{\mathrm{VR}^{\mathrm{A}}(d, d; \lambda)} - \frac{\mathrm{VD}^{\mathrm{A}}_r(d, d; \lambda)}{\mathrm{VD}^{\mathrm{A}}(d, d; \lambda)}\right) &< \frac{\beta}{2}\frac{f(d)}{\mathrm{VR}^{\mathrm{A}}(d, d; \lambda)} \\ &\leq \frac{\beta}{2M(\lambda)} \leq \frac{M(\underline{\lambda})}{M(\lambda)} \leq 1. \end{aligned}$$

The marginal utility of party R in regime A, evaluated at $r = d$, is strictly positive. Hence, d cannot be a solution to its maximization problem. QED

The bound on the value of holding office is related to the shape of the density function f. Intuitively, one expects that a higher dispersion permits a larger bound on β, but verifying this claim depends on the details of the density function in a nontrivial way. In the working-paper version of this work (Callander and Carbajal 2020), we showed that for the logistic distribution with mean zero and scale $\alpha > 0$, a higher scale (larger variance) indeed allows for a larger bound on β.

Given lemma A5, we can focus on the case where $r \in (d, d + 2\lambda] \cap \mathcal{P}$, as long as β is not too large. Since the vote share function $S^{\mathrm{A}}(r, d; \lambda)$ is positive, in an interior solution the first-order conditions are expressed as

$$1 + (r - d + \beta)(1 - S^{\mathrm{A}}(r, d; \lambda))\left(\frac{\mathrm{VR}^{\mathrm{A}}_r(r, d; \lambda)}{\mathrm{VR}^{\mathrm{A}}(r, d; \lambda)} - \frac{\mathrm{VD}^{\mathrm{A}}_r(r, d; \lambda)}{\mathrm{VD}^{\mathrm{A}}(r, d; \lambda)}\right) = 0. \tag{A5}$$

Let $r^{\mathrm{A}}_{\mathrm{int}}(d; \lambda, \beta)$ denote the interior maximizer, that is, the implicit solution to equation (A5). We now express R's best-response function in regime A as

$$r^{\mathrm{A}}(d; \lambda, \beta) = \min\left\{r^{\mathrm{A}}_{\mathrm{int}}(d; \lambda, \beta), d + 2\lambda, R\right\}. \tag{A6}$$

We focus on symmetric equilibria. Assume that the symmetric equilibrium is a corner solution, that is, $\min\{d + 2\lambda, R\}$. One has either $r = d + 2\lambda = -r + 2\lambda$ or $r = R = -d$. Thus, R's policy position in a corner equilibrium is $r^{A}(\lambda, \beta) = \lambda$ or $r^{A}(\lambda, \beta) = R$. Assume now that the symmetric equilibrium is interior. We substitute $d = -r$ in equation (A5) to obtain

$$0 = 1 + \frac{2r + \beta}{2}\left(\frac{\mathrm{VR}_r^{A}(r, -r; \lambda)}{\mathrm{VR}^{A}(r, -r; \lambda)} - \frac{\mathrm{VD}_r^{A}(r, -r; \lambda)}{\mathrm{VD}^{A}(r, -r; \lambda)}\right).$$

Since f is symmetric around zero, $\mathrm{VR}^{A}(r, -r; \lambda) = \mathrm{VD}^{A}(r, -r; \lambda)$. Using this and simplifying obtains, from the previous expression, the following equation, which characterizes the interior symmetric equilibrium strategy:

$$1 + \left(r + \frac{\beta}{2}\right)\frac{f(r + \lambda) - f(0)}{F(r + \lambda) - F(0)} = 0. \tag{A7}$$

This leads to the following proposition.

PROPOSITION A2. For all $\lambda \geq \underline{\lambda} > 0$ and $0 \leq \beta \leq 2M(\underline{\lambda})$, the unique symmetric equilibrium in regime A is given by

$$r^{A}(\lambda, \beta) = \min\left\{r_{\mathrm{int}}^{A}(\lambda, \beta), \lambda, R\right\} \quad \text{and}$$
$$d^{A}(\lambda, \beta) = -r^{A}(\lambda, \beta),$$

where $r_{\mathrm{int}}^{A}(\lambda, \beta)$ is the implicit solution to equation (A7).

Proof. Our previous analysis establishes the existence of the symmetric equilibrium stated in the proposition. It remains to show that this is, indeed, the unique symmetric equilibrium. In any corner equilibrium, one has either $r^{A}(\lambda, \beta) = \lambda = -d^{A}(\lambda, \beta)$ or otherwise $r^{A}(\lambda, \beta) = R = -d^{A}(\lambda, \beta)$.

We now show that there exists a unique solution to equation (A7). For that, we rely on two observations related to the function

$$\psi^{A}(r; \lambda) = 1 + \left(r + \frac{\beta}{2}\right)\frac{f(r + \lambda) - f(0)}{F(r + \lambda) - F(0)}.$$

First, the arguments in the proof of lemma A5 show that $\psi^{A}(0; \lambda) > 0$ as long as $0 \leq \beta \leq 2M(\lambda)$. Second, lemma A1 allows us to conclude that for all $r > 0$,

$$\frac{\partial \psi^{A}(r; \lambda)}{\partial r} = \frac{f(r + \lambda) - f(0)}{F(r + \lambda) - F(0)} + \left(r + \frac{\beta}{2}\right)\frac{\partial}{\partial r}\left(\frac{f(r + \lambda) - f(0)}{F(r + \lambda) - F(0)}\right) < 0.$$

This follows from the log concavity of $F(r + \lambda) - F(0)$. Together, these two observations show that the function $\psi^{A}(r)$ attains a value of zero at most once on the nonnegative real line. Hence, the solution to equation (A7) is unique. QED

We can now fully describe the symmetric equilibrium in regime A, using some comparative statics. Applying the implicit function theorem obtains

$$\frac{\partial r_{\mathrm{int}}^{A}(\lambda, \beta)}{\partial \lambda} = -\left(r + \frac{\beta}{2}\right)\frac{\partial}{\partial \lambda}\left(\frac{f(r + \lambda) - f(0)}{F(r + \lambda) - F(0)}\right)\left(\frac{\partial \psi^{A}(r; \lambda)}{\partial r}\right)^{-1}.$$

Using once more the fact that $F(r + \lambda) - F(0)$ is log concave for all $r \geq 0$, one concludes that the signs of the second and third factors in the right-hand side above coincide. Thus, we have verified that

$$\frac{\partial r^{A}_{int}(\lambda, \beta)}{\partial \lambda} < 0.$$

Moreover, for values of λ sufficiently close to zero, equation (A7) shows that $r^{A}_{int}(\lambda, \beta)$ is bounded away from zero.[36]

The preceding analysis leads to the following: the symmetric equilibrium policy in regime A starts at $r^{A}(\underline{\lambda}, \beta) = \underline{\lambda}$ for sufficiently low $\underline{\lambda} > 0$ and is strictly increasing. With party R's ideal point large, the equilibrium policy changes from the corner equilibrium to the interior symmetric equilibrium at a tolerance level λ^{A} satisfying

$$\lambda^{A} = r^{A}_{int}(\lambda^{A}, \beta).$$

For values of λ above λ^{A}, the symmetric equilibrium in regime A is $r^{A}(\lambda, \beta) = r^{A}_{int}(\lambda, \beta)$ and is strictly decreasing, although asymptotically.[37] We illustrate this equilibrium for the case of the logistic distribution with scale $\alpha = 1$ and office value $\beta = 0$ in figure A1, which plots the tolerance parameter λ in the horizontal axis and the equilibrium strategy $r^{A}(\lambda, \beta)$ in the vertical axis.

To understand the nature of this equilibrium, we note that for small values of the tolerance region the parties diverge as much as possible. Intuitively, near the center of the distribution a move slightly to the left improves the likelihood of winning the election as much as a move slightly to the right. Thus, R will position itself as far right as possible, that is, at $\lambda \leq \lambda^{A}$. For larger values of the tolerance region, a move to its flank is marginally less advantageous. Now party R trades off improving its electoral chances to hold office and implementing a policy closer to its policy ideal. Indeed, note that $f(0) > f(r + \lambda)$, and thus we can write the condition for the interior equilibrium as

$$1 = \left(r + \frac{\beta}{2}\right) \frac{|f(r + \lambda) - f(0)|}{F(r + \lambda) - F(0)}.$$

This expression makes clear that the value of holding office and the value of implementing a desired policy act as substitutes. A higher β forces party R to move its policy position r to the center—this trade-off is captured by the first term in the right-hand side of the expression. How much the party shifts position depends on how many net new votes it acquires, which is expressed in the second term of the right-hand side. This, in turn, depends on the distribution of voters' ideal points and the value of the tolerance region. A larger tolerance region compels parties to fight more aggressively for votes.

[36] In the case of the logistic distribution, one can solve for r^{A}_{int} numerically. For example, when the scale of the logistic distribution is $\alpha = 1$ and $\beta = 0$, one obtains $r^{A}_{int}(0, 0) \approx 2.3994$.

[37] This last follows from the fact that $f'(x) \to 0$ as $x \to \infty$, and thus $\partial r^{A}_{int}(\lambda, \beta)/\partial \lambda$ goes to zero as λ increases without bound.

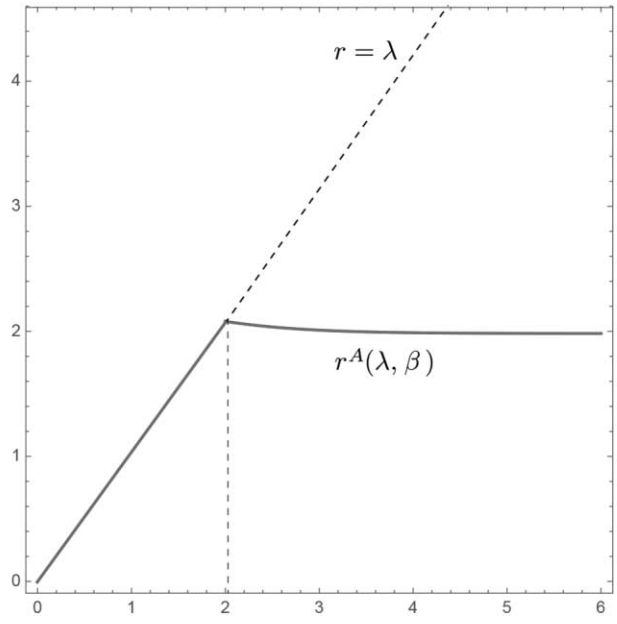

FIG. A1.—Symmetric equilibrium in regime A with the logistic distribution. The horizontal axis represents λ; the vertical axis represents *r*. A color version of this figure is available online.

A4. Finding Nash Equilibria in Regime B

Fix $\lambda \geq \underline{\lambda}$, $\beta \geq 0$, and $d \in \mathcal{P}$. Recall that R's policy domain in regime B is the interval $[d + 2\lambda, +\infty) \cap \mathcal{P}$. Assume that $d + 2\lambda < R$, else the domain of regime B is empty. Focusing first on the interior solution, we differentiate party R's expected utility with respect to r (allowing for directional derivatives on the boundaries). Using lemma A2 and recalling that $\mathrm{VD}^{\mathrm{B}}_r(r, d; \lambda) = 0$ obtains

$$\begin{aligned}\frac{\partial \mathbb{E}\, U^{\mathrm{B}}}{\partial r} &= S^{\mathrm{B}}(r, d; \lambda) + S^{\mathrm{B}}_r(r, d; \lambda)(r - d + \beta) \\ &= S^{\mathrm{B}}(r, d; \lambda)\left[1 + (r - d + \beta)(1 - S^{\mathrm{B}}(r, d; \lambda))\frac{\mathrm{VR}^{\mathrm{B}}_r(r, d; \lambda)}{\mathrm{VR}^{\mathrm{B}}(r, d; \lambda)}\right].\end{aligned}$$

Because vote share function $S^{\mathrm{B}}(r, d; \lambda)$ is strictly positive, the first-order conditions of the interior solution for this regime are expressed simply as

$$1 + (r - d + \beta)(1 - S^{\mathrm{B}}(r, d; \lambda))\frac{\mathrm{VR}^{\mathrm{B}}_r(r, d; \lambda)}{\mathrm{VR}^{\mathrm{B}}(r, d; \lambda)} = 0. \tag{A8}$$

As before, let $r^{B}_{int}(d;\lambda,\beta)$ denote the interior maximizer, that is, the solution to equation (A8). In regime B, we cannot rule out a priori either corner solution. To spare on notation, let $\text{mid}\{a, b, c\}$ denote the middle value of three real numbers a, b, and c. Party R's best-response function can be written as

$$r^{B}(d;\lambda,\beta) = \text{mid}\{r^{B}_{int}(d;\lambda,\beta),\ d + 2\lambda,\ R\}. \tag{A9}$$

We look again for the symmetric Nash equilibria. It is immediate to see that a corner solution yields $r = d + 2\lambda = -r + 2\lambda$, or $r = R$. Thus, in any corner symmetric equilibrium, R's position is either $r^{B}(\lambda,\beta) = \lambda$ or $r^{B}(\lambda,\beta) = R$. Assume instead that the symmetric equilibrium is interior. Substituting $d = -r$ in equation (A8), we obtain

$$1 + \left(r + \frac{\beta}{2}\right)\frac{f(r+\lambda) - f(r-\lambda)}{F(r+\lambda) - F(r-\lambda)} = 0. \tag{A10}$$

This equation characterizes the interior symmetric equilibrium strategy in regime B. We have arrived at the following result.

PROPOSITION A3. Given $\lambda \geq \underline{\lambda}$ and $\beta \geq 0$, the unique symmetric equilibrium in regime B is

$$\begin{aligned} r^{B}(\lambda,\beta) &= \text{mid}\{r^{B}_{int}(\lambda,\beta), \lambda, R\} \quad \text{and} \\ d^{B}(\lambda,\beta) &= -r^{B}(\lambda,\beta), \end{aligned}$$

where $r^{B}_{int}(\lambda,\beta)$ is the implicit solution to equation (A10).

Proof. We have already established the existence of the symmetric equilibrium stated in our proposition. It remains to show that this is the unique symmetric equilibrium. In any corner equilibrium, one has $r^{B}(\lambda,\beta) = \lambda = -d^{B}(\lambda,\beta)$, or alternatively, $r^{B}(\lambda,\beta) = R = -d^{B}(\lambda,\beta)$. Thus, it only remains to show that there exists a unique solution to equation (A10).

We employ two observations related to the function

$$\psi^{B}(r;\lambda) = 1 + \left(r + \frac{\beta}{2}\right)\frac{f(r+\lambda) - f(r-\lambda)}{F(r+\lambda) - F(r-\lambda)}.$$

First, the symmetric of f immediately implies $\psi^{B}(0;\lambda) = 1$ for any $\lambda > 0$. Second, from lemma A1 we know that $VR^{B}(r, d;\lambda) = F(r+\lambda) - F(r-\lambda)$ is log concave on r, and thus for all $r > 0$,

$$\frac{\partial \psi^{B}(r;\lambda)}{\partial r} = \frac{f(r+\lambda) - f(r-\lambda)}{F(r+\lambda) - F(r-\lambda)} + \left(r + \frac{\beta}{2}\right)\frac{\partial}{\partial r}\left(\frac{f(r+\lambda) - f(r-\lambda)}{F(r+\lambda) - F(r-\lambda)}\right) < 0.$$

Together, these two observations show that the function $\psi^{B}(r;\lambda)$ attains a value of zero at most once on the positive real line. Hence, the solution to equation (A10) is unique. QED

As in the previous case, we use some comparative statics to describe the nature of the symmetric equilibrium in regime B. Equation (A10) shows that, for any $\lambda \geq \underline{\lambda}$, $r = 0$ is not a solution to the interior equilibrium condition. Thus, $r^{B}_{int}(\lambda, \beta)$ is bounded away from zero.[38] Applying the implicit function theorem in regime B obtains

$$\frac{\partial r^{B}_{int}(\lambda, \beta)}{\partial \lambda} = -\left(r + \frac{\beta}{2}\right) \frac{\partial}{\partial \lambda} \left(\frac{f(r + \lambda) - f(r - \lambda)}{F(r + \lambda) - F(r - \lambda)}\right) \left(\frac{\partial \psi^{B}(r; \lambda)}{\partial r}\right)^{-1}.$$

Here it is not possible the assert the monotonicity of $r^{B}_{int}(\lambda, \beta)$ with respect to λ. This is because the vote total function $VR^{B}(r, d; \lambda) = F(r + \lambda) - F(r - \lambda)$, while log concave with respect to r, need not be log concave with respect to the tolerance level λ.[39]

On the other hand, using equation (A10) once more, we argue that there exists a unique $\lambda^{B} > 0$ such that

$$\lambda^{B} = r^{B}_{int}(\lambda^{B}, \beta).$$

Indeed, if there were $0 < \lambda_1 < \lambda_2 < R$, satisfying this condition, we would have

$$\left(\lambda_1 + \frac{\beta}{2}\right) \frac{f(2\lambda_1) - f(0)}{F(2\lambda_1) - F(0)} = \left(\lambda_2 + \frac{\beta}{2}\right) \frac{f(2\lambda_2) - f(0)}{F(2\lambda_2) - F(0)}.$$

But this is impossible, since $0 > f(2\lambda_1) - f(0) > f(2\lambda_2) - f(0)$ and the function $F(x) - F(0)$ is log concave.

We can now describe the unique symmetric equilibrium for regime B. For large enough values of the ideal policy position R, the equilibrium policy starts at $r^{B}(\underline{\lambda}, \beta) = r^{B}_{int}(\underline{\lambda}, \beta)$. It changes according to $\partial r^{B}_{int}(\lambda, \beta)/\partial \lambda$ for all $\underline{\lambda} \leq \lambda < \lambda^{B}$. At this last tolerance level, the equilibrium switches from the interior solution to the corner solution $r^{B}(\lambda, \beta) = \lambda$ and remains there for all $\lambda^{B} \leq \lambda \leq R$.[40] This equilibrium is illustrated in figure A2, for the case of the logistic distribution with scale $\alpha = 1$ and value of office $\beta = 0$. As before, the figure plots the tolerance parameter λ in the horizontal axis and the equilibrium strategy $r^{B}(\lambda, \beta)$ in the vertical axis.

[38] Solving for r^{B}_{int} numerically in the case of the logistic distribution with mean zero and scale $\alpha = 1$ obtains $r^{B}_{int}(0, 0) \approx 1.5434$.

[39] In the working-paper version of this work (Callander and Carbajal 2020), we show that for the case of the logistic distribution $r^{B}_{int}(\lambda, \beta)$ is indeed monotone increasing in λ.

[40] Note that regime B restricts the admissible values of λ to be less than R. For $\lambda > R$, the corner solution would imply a switch to regime A. We analyze this type of switch shortly.

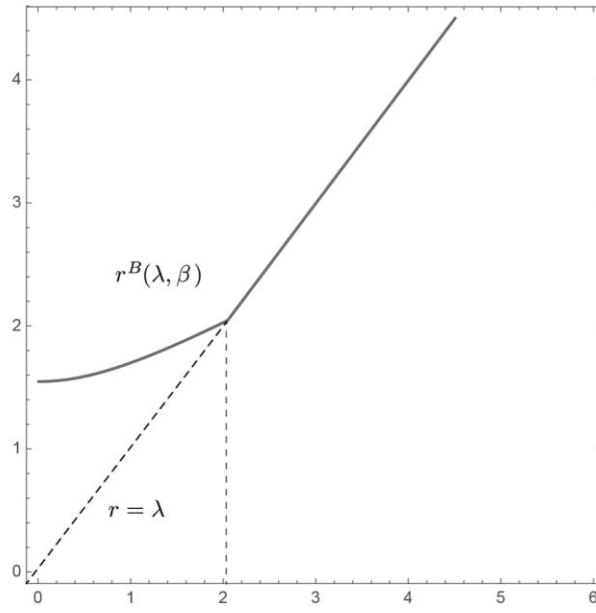

FIG. A2.—Symmetric equilibrium in regime B with the logistic distribution. The horizontal axis represents λ; the vertical axis represents *r*. A color version of this figure is available online.

A5. Nash Equilibrium of the First-Period Election

Fix a tolerance level $\lambda \geq \underline{\lambda}$ and a value of office $\beta \geq 0$. To find the Nash equilibria of the first-period electoral competition game, we consider candidate R's actual expected utility, which is given by

$$\mathbb{E}\, U(\cdot, d; \lambda, \beta) = \begin{cases} \mathbb{E}\, U^{\mathrm{A}}(\cdot, d; \lambda, \beta) \text{ if } d \leq r \leq d + 2\lambda, \\ \mathbb{E}\, U^{\mathrm{B}}(\cdot, d; \lambda, \beta) \text{ if } d + 2\lambda \leq r \leq R. \end{cases}$$

If there were no regime switch at all, the Nash equilibrium would be given by the symmetric solution for either regime A or for regime B. Yet there may be some regime switch, so that for some specific value of λ, R's best response switches from $r^{\mathrm{A}}(d; \lambda, \beta)$ to $r^{\mathrm{B}}(d; \lambda, \beta)$ at some given *d*, or vice versa. In that case, the best-response function for the actual electoral competition game may be discontinuous, so one would have to consider mixed strategies to ensure the existence of Nash equilibria. Despite this possibility, we show below that a unique symmetric equilibrium (in pure strategies) does exist. To rule out uninteresting corner solutions, that is, where $r = 0$ or $r = R$, we impose the following assumption.

ASSUMPTION A1. The value of holding office is $0 \leq \beta \leq 2M(\underline{\lambda})$, and the symmetric ideal policy positions of the parties are $-D = R > \lambda^*$, where λ^* is the tolerance level that solves

$$1 + \left(\lambda^* + \frac{\beta}{2}\right)\frac{f(2\lambda^*) - f(0)}{F(2\lambda^*) - F(0)} = 0.$$

Note that λ^* in assumption A1 satisfies

$$\lambda^* = r^{\mathrm{A}}_{\mathrm{int}}(\lambda^*, \beta) = r^{\mathrm{B}}_{\mathrm{int}}(\lambda^*, \beta).$$

That is, $\lambda^* = \lambda^{\mathrm{A}} = \lambda^{\mathrm{B}}$—see equations (A10) and (A7). This is also the value of the tolerance parameter given in equation (2).[41] The following proposition, which corresponds to proposition 1 in the main text, fully describes the unique symmetric equilibrium of the electoral competition game in the first period.

PROPOSITION A4. Suppose that f is MRLC. Under assumption A1, the unique symmetric equilibrium of the first-period electoral competition game between the parties is the strategy profile $(r^*(\lambda, \beta), d^*(\lambda, \beta))$ such that

$$r^*(\lambda, \beta) = \begin{cases} r^{\mathrm{B}}_{\mathrm{int}}(\lambda, \beta) & : 0 < \lambda < \lambda^*, \\ \lambda^* & : \lambda = \lambda^*, \\ r^{\mathrm{A}}_{\mathrm{int}}(\lambda, \beta) & : \lambda > \lambda^*, \end{cases} \quad \text{and}$$

$$d^*(\lambda, \beta) = -r^*(\lambda, \beta),$$

where $r^{\mathrm{A}}_{\mathrm{int}}(\lambda, \beta)$ and $r^{\mathrm{B}}_{\mathrm{int}}(\lambda, \beta)$ are implicitly defined by equations (A7) and (A10), respectively.

Proof. As a result of assumption A1, we ignore $r = 0$ or $r = R$ in the proof of the proposition. Building on propositions A2 and A3, we divide the analysis into four exhaustive cases and exploit the fact that $\mathbb{E}U^k(\cdot, d, ; \lambda, \beta)$ is strictly increasing to the left of $r^k(d; \lambda, \beta)$ and strictly decreasing to the right of $r^k(d; \lambda, \beta)$.

Fix $d \in \mathcal{P} = [D, R]$, $\lambda \geq \underline{\lambda}$, and $0 \leq \beta \leq 2M(\underline{\lambda})$. We consider four exhaustive cases.

CASE 1. Both regimes have a corner solution:

$$r^{\mathrm{B}}_{\mathrm{int}}(d; \lambda, \beta) < r^{\mathrm{B}}(d; \lambda, \beta) = d + 2\lambda = r^{\mathrm{A}}(d; \lambda, \beta) < r^{\mathrm{A}}_{\mathrm{int}}(d; \lambda, \beta).$$

It follows that $\mathbb{E}U^{\mathrm{A}}$ is strictly increasing on its domain and $\mathbb{E}U^{\mathrm{B}}$ is strictly decreasing on its domain. Therefore, the actual expected utility $\mathbb{E}U$ has a single peak at $r = d + 2\lambda$. Party R's unique best response is therefore $r^*(d; \lambda, \beta) = d + 2\lambda$.

CASE 2. Regime A has a corner solution and regime B an interior solution:

$$r^{\mathrm{A}}(d; \lambda, \beta) = d + 2\lambda < r^{\mathrm{B}}_{\mathrm{int}}(d; \lambda, \beta) = r^{\mathrm{B}}(d; \lambda, \beta).$$

Note also that $d + 2\lambda < r^{\mathrm{A}}_{\mathrm{int}}(d; \lambda, \beta)$. It follows that $\mathbb{E}U^{\mathrm{A}}$ is strictly increasing on its domain and $\mathbb{E}U^{\mathrm{B}}$ is strictly increasing around $d + 2\lambda$ and has a single peak at $r^{\mathrm{B}}_{\mathrm{int}}(d; \lambda, \beta)$. Therefore, the actual expected utility $\mathbb{E}U$ has a single peak at $r^{\mathrm{B}}_{\mathrm{int}}(d; \lambda, \beta)$. Party R's unique best response is therefore $r^*(d; \lambda, \beta) = r^{\mathrm{B}}_{\mathrm{int}}(d; \lambda, \beta)$.

CASE 3. Regime A has an interior solution and regime B a corner solution:

$$r^{\mathrm{A}}(d; \lambda, \beta) = r^{\mathrm{A}}_{\mathrm{int}}(d; \lambda, \beta) < d + 2\lambda = r^{\mathrm{B}}(d; \lambda, \beta).$$

We must also have $r^{\mathrm{B}}_{\mathrm{int}}(d; \lambda, \beta) < d + 2\lambda$. It follows that $\mathbb{E}U^{\mathrm{A}}$ has a single peak at $r^{\mathrm{A}}_{\mathrm{int}}(d; \lambda, \beta)$ but $\mathbb{E}U^{\mathrm{B}}$ is strictly decreasing on its domain. The actual expected utility $\mathbb{E}U^{\mathrm{R}}$ has a single peak at $r^{\mathrm{A}}_{\mathrm{int}}(d; \lambda, \beta)$. Therefore, R's unique best response is $r^*(d; \lambda, \beta) = r^{\mathrm{A}}_{\mathrm{int}}(d; \lambda, \beta)$.

[41] Again for the logistic distribution, one can numerically estimate the value of λ^*. With $\alpha = 1$ and $\beta = 0$, we obtain $\lambda^* \approx 2.06534$.

Case 4. Both regimes have an interior solution:

$$r^{A}(d;\lambda,\beta) = r^{A}_{\text{int}}(d;\lambda,\beta) < d + 2\lambda < r^{B}_{\text{int}}(d;\lambda,\beta) = r^{B}(d;\lambda,\beta).$$

Both $\mathbb{E}U^{A}$ and $\mathbb{E}U^{B}$ have a single peak in the interior of their respective domains, and the actual expected utility $\mathbb{E}U$ has two (local) maxima. Note, however, that this case is incompatible with a symmetric equilibrium. Indeed, because in each of the $\mathbb{E}U^{k}$ functions the maximum is interior, it must be the case that in the symmetric equilibrium $-d = r^{A}_{\text{int}}(d;\lambda,\beta) < r^{B}_{\text{int}}(d;\lambda,\beta) = -d$. Thus, this case will never occur in equilibrium.

We can now describe the unique symmetric equilibrium of the first-period electoral competition game as a function of λ, for $0 \le \beta \le 2M(\underline{\lambda})$ and sufficiently large R. For $\lambda < \lambda^{*}$, case 2 is relevant, and we have $r^{*}(d;\lambda,\beta) = r^{B}_{\text{int}}(d;\lambda,\beta)$. Thus, the symmetric equilibrium strategy $r^{*}(\lambda,\beta)$ is characterized by equation (A10). For $\lambda > \lambda^{*}$, case 3 is relevant, and we have $r^{*}(d;\lambda,\beta) = r^{A}_{\text{int}}(d;\lambda,\beta)$. Thus, the symmetric equilibrium strategy $r^{*}(\lambda,\beta)$ is characterized by equation (A7). For $\lambda = \lambda^{*}$, case 1 is relevant, and thus the symmetric equilibrium yields $r^{*}(\lambda^{*},\beta) = -d^{*}(\lambda^{*},\beta) = \lambda^{*}$, where λ^{*} is given by either equation (A7) or equation (A10) evaluated at $r = \lambda$. QED

The unique equilibrium of the first-period electoral game for the case of the logistic distribution with zero mean and scale $\alpha = 1$ is illustrated in figure A3, where we plot $r^{*}(\lambda,\beta)$ as a function of λ, for value of office $\beta = 0$.

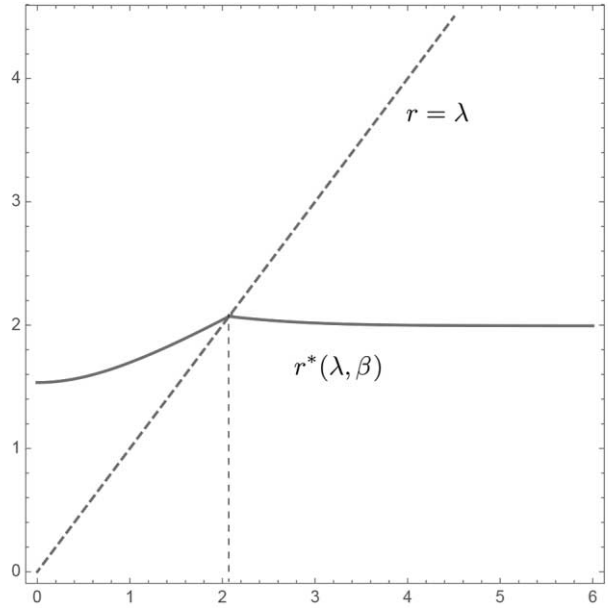

Fig. A3.—Equilibrium for the first electoral competition game with the logistic distribution. The horizontal axis represents λ; the vertical axis represents r. A color version of this figure is available online.

A6. Comparative Statics

Recall that β, the value derived from being in office, is assumed to be $0 \leq \beta \leq 2M(\underline{\lambda})$. A straightforward observation is that the effect of a higher β is a downward shift of R's best-response function in both regimes A and B. A higher value for office thus implies more convergence to the center and also, albeit less importantly, the fact that the switch between regimes occurs closer to the median. In other words, the cutoff value of λ^* where the equilibrium changes from regime B to regime A is also decreasing in β.

COROLLARY A1. The symmetric equilibrium position $r^*(\lambda, \beta)$ in the first-period election is decreasing in β, for all $0 \leq \beta \leq 2M(\underline{\lambda})$.

Proof. Directly from equations (A7) and (A10), one sees that the solution to each of these expressions is decreasing in β. Note also that the switch from regime B to regime A occurs at the λ^* stated in assumption A1. An immediate application of the implicit function theorem shows that the value of λ^* is decreasing in β. QED

What about different values of β? Suppose, for instance, that $\beta^R > 0$ and $\beta^D = 0$. Then the shift in the first-order conditions occurs only for party R. This would lead to asymmetric equilibria. Even though one has to be careful in considering the parameter values for which asymmetric equilibria would exist, the nature of our results for the first-period electoral competition should not change.

A different comparative statics exercise considers changes in the spread of the density function f. Unfortunately, given the dependence of the equilibrium position on the ratios

$$\frac{f(r+\lambda) - f(0)}{F(r+\lambda) - F(0)} \quad \text{and} \quad \frac{f(r+\lambda) - f(r-\lambda)}{F(r+\lambda) - F(r-\lambda)},$$

it is hard to draw any general conclusion about the potential changes in the symmetric equilibrium with an increase in the variance of the distribution. In the working-paper version (Callander and Carbajal 2020), we show that a higher value of the scale of the logistic distribution pushes the symmetric equilibrium to the extremes. Intuitively, a larger scale means that a larger proportion of voters can be found in the flanks, and thus parties have stronger incentives to move out of the center.

Appendix B

Equilibrium in the Second and Subsequent Elections

Here we gather the proofs of the results for the second and subsequent elections. For reasons of space, we omit proofs of results that are immediate.

B1. Proof of Proposition 4

Suppose that $\lambda \leq \lambda^*$. Recall that in this case the equilibrium in the first-period election consists of policy positions $r_1^* \in (0, R)$ and $d_1^* = -r_1^*$, such that there is abstention in the middle and in the extremes (see proposition A4). We assume that party R's ideal point is sufficiently large and show that its equilibrium position in the second election is $r_2^* = r_1^* + \lambda\tau$. This follows from three key observations.

The first observation is that, because the end points of the interval of support in the first election are $r_1^* - \lambda$ and $r_1^* + \lambda$ and the updating rule takes the form given in equation (1), any position that is more than $\lambda\tau$ to the left or the right of r_1^* leaves party R with the same vote total it would have had with said position in the first election. Thus, there is no equilibrium in which both parties choose positions farther than $\lambda\tau$ to the left or the right of their first-period equilibrium positions. Suppose, then, that D locates at d_2 in the second election, a position that is at most $\lambda\tau$ left or right of d_1^*.

The second key observation is that R's first period best-response function, which in this case corresponds to what we have referred to as regime B, attains a unique minimum at r_1^*. Thus, for any location $d \neq d_1^*$ chosen by D in period 1, party R's best response is some $\hat{r} > r_1^*$. We formally state and prove this observation as lemma C.1 in appendix C.

The final observation is that choosing a policy position less than $\lambda\tau$ away from r_1^* in the second period generates some gains in vote total compared to what R would have gotten with said position in period 1. Indeed, because of the compression of issue voters' ideal points, a small move to the left (right) of r_1^* does not lose votes on the outside (inside). However, because the initial density is singled peaked, the density of the compressed second-period distribution in the relevant interval will also be single peaked. Thus, moving slightly to the right of r_1^* generates higher gains (compared to the vote total in period 1) than moving slightly to the left of r_1^*.

Put together, these observations imply that for any position D chooses in period 2 that is at most $\lambda\tau$ to the left or right of d_1^*, R's best response will be slightly to the right of r_1^*. Thus, both parties have incentives to move slightly to the outside flank of their first-period equilibrium positions. These incentives stop at $r_2^* = r_1^* + \lambda\tau$ and $d_2^* = -r_2^*$, for any additional move to the outside flank loses more votes on the inside to alienation than the additional votes gained on the outside. QED

B2. Proof of Proposition 5

Suppose that $\lambda > \lambda^*$. Recall that in this case the equilibrium in the first-period election consists of policy platforms $r_1^* \in (0, R)$ and $d_1^* = -r_1^*$, such that there is abstention only in the extremes; that is, $\lambda > r_1^*$ (see proposition A4). As in the previous case, we implicitly assume that party R's ideal point is sufficiently large and argue that its equilibrium position in the second-period election is $r_2^* = r_1^*\tau + \lambda$.

We note that the observations made in the proof of proposition 4 are still valid, subject to the obvious accommodations due to the fact that R's marginal voters in the second election are compressed differently: the left-marginal voters situated at zero compressed by $r_1^*\tau$; the right-marginal voters situated at $r_1^* + \lambda$ compressed by $\lambda\tau > r_1^*\tau$.

Thus, it follows that in the second election both parties have incentives to move slightly to the outside flank of their first-period equilibrium positions. These incentives stop at $r_2^* = r_1^*\tau + \lambda$ and $d_2^* = -r_2^*$. At these positions, both parties win with half probability. Any additional move to the outside flank loses more votes on the inside to alienation than the votes gained on the outside, again because the density of the second-period election is single peaked in the relevant intervals.

It remains to show that neither party has an incentive to jump back to the middle as long as $\lambda^* < \lambda \leq \bar{\lambda}$ and $\tau \geq \bar{\tau}$ for some $\bar{\lambda} > \lambda^*$ and $\bar{\tau} \in (0, 1)$. To do so, let $\bar{\lambda}$ and $\bar{\tau}$ be parameter values such that party R's best response to $d = d_1^* \bar{\tau} - \bar{\lambda}$ in the first-period election is $\bar{r} > 0$ satisfying $\bar{r} - \bar{\lambda} = d_1^* \bar{\tau}$. From proposition A4 and lemma C.1 in appendix C, we know that $r_1^* < \bar{r} < r_1^* \bar{\tau} + \bar{\lambda}$. In this case, R has no incentive to jump to the middle. Indeed, any change from the second-period equilibrium position $r_1^* \bar{\tau} + \bar{\lambda}$ would have to be a move to the left in order to gain vote share. However, because of the gap around zero in the second-period distribution of voters, such a move would have to be to a position at or left of $\bar{r}$. At this point, the vote share for R in the second election is the same as the vote share in the first selection, and thus party R does not move left of $\bar{r}$ (which is by definition the first-election best response to $d_1^* \bar{r} - \bar{\lambda}$). But R has the same vote share at $\bar{r}$ and at $r_1^* \bar{\tau} + \bar{\lambda}$, the equilibrium position in the second-period election. Thus, R has no incentive to deviate.

The above argument remains valid as long as $\bar{\tau} \leq \tau < 1$ and $\lambda^* < \lambda \leq \bar{\lambda}$, because the first-election best-response function of R has a unique minimum at r_1^* (see again lemma C.1 in app. C). When $\lambda > \bar{\lambda}$ and $0 < \tau < \bar{\tau}$, party R has an incentive to jump from the position $r_1^* \tau + \lambda$ to what is the best response to position $d_1^* \tau - \lambda$ in the first-period election. In such a case, the symmetric profile given by $r_2^* = r_1^* \tau + \lambda$ and $d_2^* = -r_2^*$ is no longer an equilibrium. In this case, no symmetric equilibrium exists. If $-d_2 = r_2 > r_2^*$, then both parties have incentive to get closer to the middle. If $-d_2 = r_2 > r_2^*$, then both parties have incentive to move slightly to the outer flank. QED

B3. Proof of Proposition 6

This follows readily from propositions 4 and 5. QED

References

Abramowitz, Alan I. 2010. *The Disappearing Center: Engaged Citizens, Polarization, and American Democracy.* New Haven, CT: Yale Univ. Press.

Abramowitz, Alan I., and Steven Webster. 2016. "The Rise of Negative Partisanship and the Nationalization of U.S. Elections in the 21st Century." *Electoral Studies* 41:12–22.

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2018. "Explaining Preferences from Behavior: A Cognitive Dissonance Approach." *J. Polit.* 80 (2): 400–411.

Achen, Christopher H., and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government.* Princeton, NJ: Princeton Univ. Press.

Adams, James, Jane Green, and Caitlin Milazzo. 2012. "Has the British Public Depolarized along with Political Elites? An American Perspective on British Public Opinion." *Comparative Polit. Studies* 45 (4): 507–30.

Akerlof, George A., and William T. Dickens. 1982. "The Economic Consequences of Cognitive Dissonance." *A.E.R.* 72 (3): 307–19.

Aronson, Elliot, Carrie Fried, and Jeff Stone. 1991. "Overcoming Denial and Increasing the Intention to Use Condoms through the Induction of Hypocrisy." *American J. Public Health* 81 (12): 1636–38.

Bagnoli, Mark, and Ted Bergstrom. 2005. "Log-Concave Probability and Its Applications." *Econ. Theory* 26 (2): 445–69.

Bartels, Larry M. 2016. "Failure to Converge: Presidential Candidates, Core Partisans, and the Missing Middle in American Electoral Politics." *Ann. American Acad. Polit. and Soc. Sci.* 667:143–65.

Beasley, Ryan K., and Mark R. Joslyn. 2001. "Cognitive Dissonance and Post-Decision Attitude Change in Six Presidential Elections." *Polit. Psychology* 22 (3): 521–40.

Becker, Gary S., and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *J.P.E.* 96 (4): 675–700.

Bølstad, Jørgen, Elias Dinas, and Pedro Riera. 2013. "Tactical Voting and Party Preferences: A Test of Cognitive Dissonance Theory." *Polit. Behavior* 35 (3): 429–52.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. 2021. "Cross-Country Trends in Affective Polarization." Working Paper no. 26669 (November), NBER, Cambridge, MA.

Callander, Steven, and Juan Carlos Carbajal. 2020. "Cause and Effect in Political Polarization: A Dynamic Analysis." Working paper, Stanford Univ. and Univ. New South Wales Sydney.

Callander, Steven, and Catherine H. Wilson. 2006. "Context-Dependent Voting." *Q. J. Polit. Sci.* 1 (3): 227–54.

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

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. Chicago: Univ. Chicago Press.

Clark, Tom S. 2009. "Measuring Ideological Polarization on the United States Supreme Court." *Polit. Res. Q.* 62 (1): 146–57.

Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, edited by David E. Apter, 206–61. New York: Free Press.

DiMaggio, Paul, John Evans, and Bethany Bryson. 1996. "Have Americans' Social Attitudes Become More Polarized?" *American J. Sociology* 102 (3): 690–755.

Dinas, Elias. 2014. "Does Choice Bring Loyalty? Electoral Participation and the Development of Party Identification." *American J. Polit. Sci.* 58 (2): 449–65.

Dinas, Elias, Erin Hartman, and Joost van Spanje. 2016. "Dead Man Walking: The Affective Roots of Issue Proximity between Voters and Parties." *Polit. Behavior* 38 (3): 659–87.

Esteban, Joan-María, and Debraj Ray. 1994. "On the Measurement of Polarization." *Econometrica* 62 (4): 819–51.

———. 2012. "Comparing Polarization Measures." In *The Oxford Handbook of the Economics of Peace and Conflict*, edited by Michelle R. Garfinkel and Stergios Skaperdas, 127–51. Oxford: Oxford Univ. Press.

Festinger, Leon. 1962. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford Univ. Press.

Fiorina, Morris P., Samuel J. Abrams, and Jeremy C. Pope. 2010. *Culture War? The Myth of a Polarized America*. New York: Longman.

Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. "Habit Formation in Voting: Evidence from Rainy Elections." *American Econ. J. Appl. Econ.* 8 (4): 160–88.

Gentzkow, Matthew. 2016. "Polarization in 2016." White paper, Toulouse Network for Information Technology.

Ghitza, Yair, and Andrew Gelman. 2014. "The Great Society, Reagan's Revolution, and Generations of Presidential Voting." Working paper.

Hall, Andrew B., and Daniel M. Thompson. 2018. "Who Punishes Extremist Nominees? Candidate Ideology and Turning Out the Base in US Elections." *American Polit. Sci. Rev.* 112 (3): 509–24.

Hotelling, Harold. 1929. "Stability in Competition." *Econ. J.* 39 (153): 41–57.
Jessee, Stephen A. 2009. "Spatial Voting in the 2004 Presidential Election." *American Polit. Sci. Rev.* 103 (1): 59–81.
———. 2010. "Partisan Bias, Political Information and Spatial Voting in the 2008 Presidential Election." *J. Polit.* 72 (2): 327–40.
Kantrowitz, Robert, and Michael M. Neumann. 2007. "Is the Optimal Rectangle a Square?" *Pi Mu Epsilon J.* 12 (7): 405–12.
Klemperer, Paul. 1987. "Markets with Consumer Switching Costs." *Q.J.E.* 102 (2): 375–94.
Lenz, Gabriel S. 2012. *Follow the Leader? How Voters Respond to Politicians' Policies and Performance.* Chicago: Univ. Chicago Press.
Levendusky, Matthew. 2009. *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans.* Chicago: Univ. Chicago Press.
McCarty, Nolan. 2019. *Polarization: What Everyone Needs to Know.* New York: Oxford Univ. Press.
McDonald, Michael P., and Samuel L. Popkin. 2001. "The Myth of the Vanishing Voter." *American Polit. Sci. Rev.* 95 (4): 963–74.
McGregor, R. Michael. 2013. "Cognitive Dissonance and Political Attitudes: The Case of Canada." *Soc. Sci. J.* 50 (2): 168–76.
Mullainathan, Sendhil, and Ebonya Washington. 2009. "Sticking with Your Vote: Cognitive Dissonance and Political Attitudes." *American Econ. J. Appl. Econ.* 1 (1): 86–111.
Page, Scott E. 2006. "Path Dependence." *Q. J. Polit. Sci.* 1 (1): 87–115.
Panagopoulos, Costas. 2016. "All about That Base: Changing Campaign Strategies in U.S. Presidential Elections." *Party Polit.* 22 (2): 179–90.
Penn, Elizabeth Maggie. 2017. "Inequality, Social Context, and Value Divergence." *J. Polit.* 79 (1): 153–65.
Pierson, Paul. 2000. "Increasing Returns, Path Dependence, and the Study of Politics." *American Polit. Sci. Rev.* 94 (2): 251–67.
———. 2004. *Politics in Time: History, Institutions, and Social Analysis.* Princeton, NJ: Princeton Univ. Press.
Pratt, John W. 1981. "Concavity of the Log Likelihood." *J. American Statis. Assoc.* 76 (373): 103–6.
Puett, Michael, and Chrisine Gross-Loh. 2016. *The Path: What Chinese Philosophers Can Teach Us about the Good Life.* New York: Simon & Schuster.
Rehm, Philipp, and Timothy Reilly. 2010. "United We Stand: Constituency Homogeneity and Comparative Party Polarization." *Electoral Studies* 29 (1): 40–53.
Riker, William H., and Peter C. Ordeshook. 1968. "A Theory of the Calculus of Voting." *American Polit. Sci. Rev.* 62 (1): 25–42.
Smithies, Arthur. 1941. "Optimum Location in Spatial Competition." *J.P.E.* 49 (3): 423–39.
Wolfinger, Raymond E., and Steven J. Rosenstone. 1980. *Who Votes?* New Haven, CT: Yale Univ. Press.