[zoomed no-margin version for easier reading on computers]

Mutual Party Extremism and the Primary Trap

Ivo Welch

July 28, 2026

Abstract

It is well known that candidates may have to stake out more partisan positions to win their primaries. These then also determine their success in the general election. The equilibrium consequences of primaries are less well understood. There is a robust externality — the primary trap — that can sustain multiple equilibria (along a band). Candidate extremism (moderation) on one side facilitates candidate extremism (moderation) on the other. Electorate polarization and turnout are less important than other candidates' choices. The paper also discusses how other considerations can alter the set of equilibria.

Abstract Word Count: 92

Keywords: political polarization, party primaries, candidate extremism, spatial voting

JEL Codes: D72, D71, C72

I Introduction

The classic theorem of Downs (1957) predicts that candidates want to locate themselves near the median voter. This is not always the case in politics. A number of explanations have been proposed,[1] one of which is the presence of a primary. As Owen and Grofman (2006) emphasize, to win the party nomination, a competitive candidate must first appeal to a more partisan base. Yet, despite the intuitive economic appeal — and some evidence of recent polarization of both the American electorate and its candidates — the general empirical evidence has remained puzzling. Sides et al. (2020) found that those who vote in primaries are demographically and ideologically close to their party's general-election supporters. Changes in the electorate and changes in candidate positioning have also often appeared not strongly connected: Hare and Poole (2014) document that the congressional parties have diverged steadily since the 1970s even as the distribution of voters has shifted little. Finally, the type of primary a state runs — whether open or closed — has had little measurable effect on how extreme its legislators are, despite the reform hypothesis that more open nomination systems should produce more moderate candidates (McGhee et al., 2014).

In equilibrium, it is not only the marginal influence that primaries exert on individual candidates that should matter, but the strategic interplay of this influence on all candidates. The central mechanism of the model presented in the paper here is a robust and intuitive externality, termed the *primary trap*.

Consider the simplest canonical example. A candidate derives utility only from winning office. Voters are distributed uniformly from -1 to $+1$. There are two $\mathcal{D}$ and two $\mathcal{R}$ candidates who choose where to locate on the same interval. Each voter supports the nearest candidate, breaking indifference by a fair coin. Democrat ($\mathcal{D}$) voters lie on the left, in $[-1, 0]$, and Republican ($\mathcal{R}$) voters on the right, in $[0, 1]$.

Now consider the choice of an $\mathcal{R}$ candidate, faced with two $\mathcal{D}$ candidates who have both located at -0.3. If this $\mathcal{R}$ candidate locates at $+0.5$ (by Downsian logic, the location of her median $\mathcal{R}$ voter), she will win or tie the primary election (PE). However, she will lose the general election (GE), because all voters between $[-1, +0.1]$ will vote $\mathcal{D}$, and only voters from $[+0.1, +1]$ will vote for her. Thus, her utility will be zero. If instead she chooses $+0.2$, although this would

[1] Adams et al. (2005) describe alienation-based centrifugal and indifference-based centripetal forces. In Smithies (1941), Hinich and Ordeshook (1969), and Comanor (1976), voters far from candidates stay home. Jones et al. (2022) points out that skewed or bimodal electorates can do the same. Section IV discusses the empirical evidence in more detail.

win the GE, the other $\mathcal{R}$ candidate would locate at $+0.2 + \epsilon$ and win both the primary and the general election. Given the other side's position at -0.3, her only viable equilibrium choice (yielding her 1/4) is +0.3, as is that of her $\mathcal{R}$ co-partisan and primary opponent. The result is an equilibrium in which both parties are locked into mirrored candidate extremism, here at -0.3 for $\mathcal{D}$ and +0.3 for $\mathcal{R}$. Each party has a 50% chance of winning, and each candidate a 25% chance. No other pure equilibria exist when the two $\mathcal{D}$ candidates locate at -0.3, and mixed equilibria are inappropriate for the economic environment considered here.

The same logic and an equivalent equilibrium can exist if the two $\mathcal{D}$ candidates locate themselves anywhere in the band between -0.5 and 0. Both $\mathcal{R}$ candidates mirror the $\mathcal{D}$ choice. The central object of the model is not a unique single equilibrium point but a set of robust equilibrium points along a *band*. (Importantly, the band here does not mean that *individual* candidates have discretion to choose any value inside the wide band. Each equilibrium itself is a point, and it strongly constrains the four individual candidates.) The same voter configuration can sustain two moderate parties just as it can sustain two extreme parties. It follows that primaries can but *need not* radicalize. This reconciles the puzzle that primaries constrain candidates to more partisan choices and yet the rules governing primaries may leave little empirical trace. The primary constraint is ubiquitous and real but slack from a joint perspective. Moderate and extreme equilibria can be viable at the same parameters, and parameter changes (e.g., on voter polarization or turnout) need not shift the current equilibrium.

This paper develops this intuition into a more general model. Two candidates in each party compete first in a primary, and the two winners then compete in a general election. Candidates who seek to win the primary must cater to the median *party* voter, who is more extreme than the median *overall* voter. Catering to a more extreme base can harm electability in the general election. This creates the mutual externality the paper analyzes.

The key assumption of the paper here is that candidates optimize for electability. (The paper considers "principled candidates" later.) If and when they do, elections often end up competitive. Although a perfectly symmetric scenario is not always the most realistic case, it provides the clearest and most general expression of the primary trap. In equilibrium, under symmetric voter distribution and turnout, candidates interested in being elected will stake out mirror-identical positions to survive both stages. One party can stake out more extreme positions if and only if the other party also stakes out more extreme positions. The same configuration can often also sustain two moderate parties as readily as it can sustain two extreme ones. What the basic

model excludes *in equilibrium* is one party that is more moderate or more extreme than the other party.

The main insight of this model about the primary trap should not be its literal equal winning probabilities under certainty. Instead, it should be the insight that if candidates want to remain competitive, it is the endogenous positioning of candidates that matters more than exogenous electorate parameters. This endogeneity can facilitate the presence of more moderate equilibria and more extreme equilibria under the same parameters. Furthermore, if some candidates are principled and don't care about electability at all (Adams et al., 2005), they can doom the entire party to unelectability. They do so by choosing positions that win the primary but lose the general election (Section III.A).

The paper now proceeds as follows. Section II sets up the model — voters on a line, four candidates, a primary followed by a general election — derives the equilibrium band for the specialized case of exponentially decaying turnout, and then states the result in general form: under four assumptions on a symmetric electorate and an otherwise arbitrary turnout rule, the equilibria are a band of mirror-symmetric profiles, each candidate winning office with probability 1/4 (Theorems 1–2). Section III considers departures from the symmetric baseline: a principled co-partisan can leave the free candidate no winning position, differential voter polarization and participation can shut a whole party out of office, open primaries may admit cross-over voting, and more than two candidates per party can preclude a pure equilibrium existence altogether. Section IV places the model against the empirical evidence. For readability, many proofs and several formal results are relegated to the appendices.

II Symmetric Party Scenarios

This section maintains a *symmetric* electorate with four purely office-seeking candidates throughout. This yields the paper's central result: the presence of a band of simultaneously viable point equilibria, each with mirror-symmetric candidates, leaving every one of the four candidates expecting a 1/4 probability of winning (analogous to Priest and Klein (1984)). This section begins with a parameterized symmetric model before proceeding to the more general symmetric model. The next section considers non-symmetric cases. Table 1 collects the notation for convenient reference.

[Insert Table 1 here: **Glossary of Notation**]

A Assumptions of a Specific Parametric Model

In the specific parameterized model, voters are distributed on the support $[-1, +1]$. There are four candidates, with exogenous party labels: two $\mathcal{R}$ and two $\mathcal{D}$ candidates. Each candidate chooses her position, which then remains fixed across the PE and the GE.[2] The scale need not be about a policy (e.g., abortion) that is all-deciding. It can be a composite of many issues. It can also represent a general "left" (at -1) versus "right" (at $+1$) identity, where the center (0) is moderate and the endpoints (± 1) are extreme.

[Insert Table 2 here: **Primary Systems as Voter-to-Primary Assignment Rules**]

The PE is fought within party, the GE across the two party nominees. Table 2 shows the prevalence of the two types of primaries in the United States: *closed primaries*, in which voters can participate in the primary only on their side of zero; and *open primaries*, in which they can vote instead for the candidate of the other party in their primary if that candidate is closer to them. These two cases cover most of the U.S. electoral system. The analysis begins with closed primaries, which are slightly more tractable. The results transfer readily to open primaries, especially in the symmetric case.[3]

[2] A candidate who switched or randomized her degree of moderation versus extremism would lose credibility, and the electorates can observe position. In an extended model, one could allow mixing or changing position at a given cost. The model can also be extended to multiple attributes (dimensions), which would complicate the algebra but still result in similar symmetric band insights. For conditions under which a multi-attribute problem is equivalent to a one-dimensional one, see Gans and Smart (1996).

[3] One difference is a second mirror equilibrium of $(-q, +q, -q, +q)$ in the symmetric situation, because voters show zero party affiliation in the open primary. The full analysis — the double-cross proposition, the collapse lemmas, and their proofs — in an open primary are discussed in Appendix W4.

The game has two stages. Two $\mathcal{D}$ and two $\mathcal{R}$ candidates choose positions, fixed across both stages. In the closed primary stage, each voter joins her own party's contest: voters left of center vote in the $\mathcal{D}$ primary and voters right of center in the $\mathcal{R}$ primary, each supporting the nearer of her party's two candidates (ties split half-half). The two primary winners then compete in the general election where all voters participate.

The specific model uses two parameterizations of the voter distribution and turnout:

Voter Distribution: The distribution of voters has density

$$g(x;\,\alpha) \;=\; \frac{(1-\alpha) \;+\; 2{\cdot}\alpha\cdot|x|}{2}\,, \qquad \alpha\in[-1,1]\,. \tag{1}$$

The electorate has unit mass, every turnout-weighted vote total below is a fraction of it, and — by symmetry — each half carries mass $1/2$ for every α, so a change in α reshapes an electorate without resizing it. Because the normalizer does not depend on α, the parameter has a direct reading: $\alpha = g(1) - g(0)$, the edge-minus-center density gap, with $g(1) + g(0) = 1$ throughout. At $\alpha = 0$ the electorate is uniform. Positive αs give bimodal distributions while negative αs give single-peaked distributions, both centered at 0. The admissible range is not a convention but a consequence of $g \geq 0$, and it binds at both ends: at $\alpha = 1$ the density is the bimodal V-shape $|x|$, zero at the center, and at $\alpha = -1$ it is the single-peaked tent $1 - |x|$, zero at the two edges. (The solutions are similar when the absolute value is replaced by the square, in which case the functional forms appear like $\cup$ and $\cap$, respectively.)

Voter Turnout: Following Aldrich (1983), voters at distance d from their preferred candidate turn out to vote with probability weight

$$w(d;\,\tau) \;=\; e^{-\tau\cdot d}, \qquad \tau\geq 0\,, \tag{2}$$

where τ measures turnout decay per unit distance of position. At $\tau = 0$, all voters participate regardless of preferred candidate distance. The w function can be viewed as a normalization to the maximum turnout. The paper considers either a baseline of $\tau_0 = 0$, in which voters always turn out, or a baseline of $\tau_0 = 1$, at which a voter one unit from her candidate turns out at $e^{-\tau_0} \approx 0.37$ of the peak rate. A higher τ corresponds to a more discriminating electorate.[4]

[4]The appendix shows that the primary trap remains the same if turnout is an arbitrary function of all four candidates' positions *as long as* certain symmetry conditions are met. For example, voters may tend to turn out more if the distance between candidates is larger — but this then has to be the case both for $\mathcal{D}$ and for $\mathcal{R}$ voters.

[Insert Figure 1 here: **Voter Polarization, Turnout Decay, and the Two Elections**]

Figure 1 illustrates these two functional forms. Panel A shows the three possible forms of voter location distributions. Panel B shows a general election. The left case has everyone vote. The right case shows two candidates in the GE, one at −0.7 the other at +0.3, when voters without a candidate nearby are less inclined to vote. The voter turnout effectively creates two endogenous peak distributions around the two candidates. Note that in this specification, it is not just the candidates that locate where the voters are, but also the converse. However, voters are *not* strategic but atomistic.[5] Panel C shows the primary election turnout for four candidates. The left panel shows a closed primary, while the right panel shows an open primary, where some $\mathcal{D}$ voters prefer to vote in the $\mathcal{R}$ primary.

B Primary Trap Equilibria in the Specific Parametric Model

This subsection derives the set of equilibria. The graphical illustration in Section II.B.6 may be helpful before following the algebraic derivation.

The two functional forms (1) and (2) were chosen to deliver closed-form integrals. Consider an $\mathcal{R}$ candidate at $+q$, writing $q \geq 0$ for her distance from the center. Define her **outer mass** as the turnout-weighted electorate between her and the right edge,

$$m_{\text{out}}(q) = \int_q^1 g(x)\, e^{-\tau\cdot(x-q)}\, dx = \frac{(1-\alpha)\cdot\tau + 2\cdot\alpha + 2\cdot\alpha\cdot\tau\cdot q - \left[\,(1-\alpha)\cdot\tau + 2\cdot\alpha\cdot(1+\tau)\,\right]\cdot e^{-\tau\cdot(1-q)}}{2\cdot\tau^2}, \tag{3}$$

and her **inner mass** as the turnout-weighted electorate between her and the center:

$$m_{\text{in}}(q) = \int_0^q g(x)\, e^{-\tau\cdot(q-x)}\, dx = \frac{(1-\alpha)\cdot\tau - 2\cdot\alpha + 2\cdot\alpha\cdot\tau\cdot q - \left[\,(1-\alpha)\cdot\tau - 2\cdot\alpha\,\right]\cdot e^{-\tau\cdot q}}{2\cdot\tau^2}. \tag{4}$$

At $\tau = 0$ the turnout weight is $e^0 = 1$, so each mass is just the voter density integrated over its interval — the electorate share of $[\,q, 1\,]$ (outer) and $[\,0, q\,]$ (inner).

Now presume that the four candidates locate at $(-q, -q, +q, +q)$, the two $\mathcal{D}$ at $-q$, the two $\mathcal{R}$ at $+q$. This is an equilibrium if and only if no candidate can profitably deviate. By mirror

[5] Sincere voting is an assumption here, not a finding. It sits comfortably with the evidence of Dal Bó et al. (2018) that voters weigh the direct effect of a change and underappreciate the equilibrium response that it induces. A primary voter who supports the co-partisan nearest her own position is not reasoning through the general election that her nomination brings about.

symmetry it is enough to test the off-equilibrium choices of one $\mathcal{R}$ candidate. Her potential deviations will be related to comparing two masses, so it is useful to define the difference as

$$\Delta m(q) \equiv m_{\text{out}}(q) - m_{\text{in}}(q) = \frac{4\cdot\alpha + \left[\tau\cdot(1-\alpha) - 2\cdot\alpha\right]\cdot e^{-\tau\cdot q} - \left[\tau\cdot(1-\alpha) + 2\cdot\alpha\cdot(1+\tau)\right]\cdot e^{-\tau\cdot(1-q)}}{2\cdot\tau^2}. \tag{5}$$

B.1 Primary Election

In the closed primary, the two $\mathcal{R}$ candidates compete for electorates on the right half $[0, 1]$. A co-located pair at $+q$ splits it evenly. A co-partisan who undercuts *inward*, toward the center, takes the inner mass $m_{\text{in}}(q)$ and leaves the other the outer mass $m_{\text{out}}(q)$. Leaning *outward* does the reverse. The inward undercut is the dangerous one, because a more moderate nominee would go on to carry the general election. It fails — and the co-located pair holds the primary — when the candidate who stays at $+q$ keeps the larger flank, $\Delta m(q) \geq 0$. Under the functional forms this simplifies to

$$\Delta m(q) \geq 0 \quad \Leftrightarrow \quad q \leq -\left(\frac{1}{\tau}\right)\cdot \ln z_{\text{hi}}(\alpha, \tau). \tag{6}$$

The root $z_{\text{hi}} \in (e^{-\tau}, 1)$ solves $\left[\tau\cdot(1-\alpha) - 2\cdot\alpha\right]\cdot z^2 + 4\cdot\alpha\cdot z - \left[\tau\cdot(1-\alpha) + 2\cdot\alpha\cdot(1+\tau)\right]\cdot e^{-\tau} = 0$. For uniform voters ($\alpha = 0$) it gives $z_{\text{hi}} = e^{-\tau/2}$, so $q \leq 1/2$. This limit is the upper edge q_{hi}.

B.2 General Election

In the general election, consider the $\mathcal{R}$ candidate leaning out to $+p$ and facing the $\mathcal{D}$ nominee fixed at $-q$. Turnout is unaffected by voters' party, so the two nominees split the line at their midpoint $(p-q)/2$. The $\mathcal{R}$ candidate's total is the sum of her outer flank, her inner flank, and a signed *central sliver* reaching across the center (to the midpoint between the two candidates):[6]

$$ge_{\mathcal{R}}(p,q) = \underbrace{\int_{(p-q)/2}^{0} g(x)\, e^{-\tau\cdot(p-x)}\, dx}_{\text{getm}(p;+)} + \underbrace{\int_{0}^{p} g(x)\, e^{-\tau\cdot(p-x)}\, dx}_{m_{\text{in}}(p)} + \underbrace{\int_{p}^{1} g(x)\, e^{-\tau\cdot(x-p)}\, dx}_{m_{\text{out}}(p)} . \tag{8}$$

The $\mathcal{D}$ nominee at $-q$ collects the mirror-image total $ge_{\mathcal{D}}(p,q) = m_{\text{out}}(q) + m_{\text{in}}(q) + \text{getm}(q;-)$ from $[-1,(p-q)/2]$. The first term is the sliver from the center $x=0$ out to the midpoint, won by whichever candidate is the more moderate. The other two terms are the candidate's flank masses (3) and (4), measured from her position.

Among choices that would have also survived the primary, the deviation pays iff it wins the general election, $ge_{\mathcal{R}}(p,q) > ge_{\mathcal{D}}(p,q)$. The two general-election totals tie at the mirror, $ge_{\mathcal{R}}(q,q) = ge_{\mathcal{D}}(q,q)$, because the two nominees then split a symmetric turnout-weighted electorate at the center. The only deviation that also survives the primary is the outward lean: by (6), an inward undercut forfeits the primary wherever $\Delta m(q) \geq 0$, so only the outward lean — which carries the larger flank — reaches the general election. The outward lean pays — so the symmetric profile fails — iff it raises the $\mathcal{R}$ margin at first order,

$$\frac{\partial}{\partial p}\Big[\, ge_{\mathcal{R}}(p,q) - ge_{\mathcal{D}}(p,q) \,\Big]\Big|_{p=q} = \tau\cdot\Delta m(q) - \frac{(1-\alpha)\cdot e^{-\tau\cdot q}}{2} > 0 . \tag{9}$$

The pair therefore survives the general-election deviation iff the reverse (weak) inequality holds,

$$\Delta m(q) \leq \frac{(1-\alpha)\cdot e^{-\tau\cdot q}}{2\cdot\tau} . \tag{10}$$

[6] The sliver is the central segment from the midpoint $(p-q)/2$ to the center, weighted by the candidate's own turnout. It does not reduce further, but up to the sign of α and the candidate's own exponential it is a single function,

$$\text{getm}(x;\pm) \equiv \pm\int_{(p-q)/2}^{0} \frac{(1-\alpha) - 2\cdot\alpha\cdot t}{2}\, e^{-\tau\cdot(x\mp t)}\, dt =$$
$$\frac{1}{2}\cdot\left[\frac{\big[(1-\alpha)\cdot\tau \pm 2\cdot\alpha\big]\cdot\big(e^{-\tau\cdot x} - e^{-\tau\cdot(p+q)/2}\big)}{\tau^2} - \frac{2\cdot\alpha\cdot(q-p)}{2\cdot\tau}\cdot e^{-\tau\cdot(p+q)/2}\right], \tag{7}$$

written for the more-moderate candidate ($p \leq q$, so the sliver lies center-left of $x=0$ and both nominees read the same density branch there, $g = [\,(1-\alpha) - 2\cdot\alpha\cdot t\,]/2$). The upper sign is the more-moderate nominee's own sliver, *added* to her total at $x=p$; the lower sign is the same stretch *subtracted* from the other nominee's total at $x=q$, whose turnout is measured from $-q$ rather than from $+p$ — which is why the exponent's sign flips with the branch. The reverse ordering mirrors it, swapping p and q.

The right-hand side is the marginal turnout at the center, scaled by $1/\tau$: the density there is $g(0) = (1-\alpha)/2$, weighted by the participation $e^{-\tau \cdot q}$ of voters there, at distance q from either nominee.

This lower edge q_{lo} is a manifestation of the classic result that candidates abandon the median once alienation-driven abstention meets a sufficiently bimodal electorate (Smithies, 1941; Hinich and Ordeshook, 1969; Comanor, 1976). However, here the abandonment of the median arises from endogenous voter participation, and it occurs only for sufficiently high voter polarization.

B.3 Surviving both the PE and the GE

The primary condition (6) and the general-election condition (10) bracket the deviation-proof positions,

$$0 \;\le\; \Delta m(q) \;\le\; \frac{(1-\alpha)\cdot e^{-\tau \cdot q}}{2 \cdot \tau}\,, \tag{11}$$

which are the primary binding from below and the general election from above. Because Δm is *decreasing* in q, the same two conditions read directly as a band on the candidate's own position,

$$\underbrace{q_{\text{lo}}(\alpha,\tau)}_{\text{general election}} \;\le\; q \;\le\; \underbrace{q_{\text{hi}}(\alpha,\tau)}_{\text{primary}}\,, \tag{12}$$

where the primary sets the ceiling q_{hi} (from $\Delta m \ge 0$, the largest position that still holds the primary) and the general election the floor $q_{\text{lo}} = \max(0, \cdot)$ (from $\Delta m \le (1-\alpha)\cdot e^{-\tau \cdot q}/(2\cdot\tau)$). The floor is 0 below the centrist threshold α_c — with the general election tying, nothing binds from below — and lifts above it. Both edges can be written out:

$$q_{\text{lo}} \;=\; \max\left\{0,\; -\left(\frac{1}{\tau}\right)\cdot \ln\!\left(1 - \sqrt{1-\kappa(\alpha,\tau)\cdot e^{-\tau}}\right)\right\}, \qquad \kappa(\alpha,\tau) \;\equiv\; 1 + \tau + \frac{\tau\cdot(1-\alpha)}{2\cdot\alpha}\,, \tag{13}$$

with the general-election radical real — and the floor lifting off 0 — precisely when $\alpha > \alpha_c \equiv \tau/(2e^{\tau} - 2 - \tau)$.[7] Writing out the primary-election root $z_{\text{hi}} \in (e^{-\tau}, 1)$ of the quadratic in (6),

$$q_{\text{hi}} = -\left(\frac{1}{\tau}\right) \cdot \ln z_{\text{hi}}\,, \quad z_{\text{hi}} = \frac{-2{\cdot}\alpha + \sqrt{4{\cdot}\alpha^2 + \left[\tau{\cdot}(1-\alpha) - 2{\cdot}\alpha\right]{\cdot}\left[\tau{\cdot}(1-\alpha) + 2{\cdot}\alpha{\cdot}(1+\tau)\right]{\cdot}e^{-\tau}}}{\tau{\cdot}(1-\alpha) - 2{\cdot}\alpha}\,. \tag{14}$$

These two edges coincide only at full bimodality: at $\alpha = 1$, where the density vanishes at the center, the general-election floor rises to meet the primary ceiling and the band pinches to the single point $q^* = q_{\text{hi}} = q_{\text{lo}} \approx 0.722$ (at $\tau_0 = 1$), the two parties locked at their common primary median for every turnout decay $\tau > 0$.

One clarification is warranted. Mutual candidate extremism is not a coordination failure across the aisle, with each party extreme because the other is. The binding constraint is a collective-action problem internal to each party: the party would do better with a more central nominee, every one of its candidates may recognize this, and yet none can be the one to move, because the first to step toward the center loses the nomination to a co-partisan who moves less. The rival party enters only through the general election, which determines *whether* being pinned is costly, not *where* the q_{hi} pin sits. Positions are therefore determined within a party and penalized across parties. This is also why the two levers that move the q_{hi} pin (Section III.B) are properties of a party's own electorate, and why a party can be made worse off by its own voters — not through how they vote in the general election, but through which nominee they select (Section III.B.1). The band is an interval rather than a point because it is cut out by two deviation constraints rather than by any agent's optimization: no candidate has a preference over the common position, and the party that does cannot act on it.

[7] The linear density is not a "lucky" functional form. For the family $g(x) \propto (1-\alpha) + 2{\cdot}\alpha\,|x|^n$ the same calculation gives

$$\alpha_c(n, \tau) = \frac{\tau^n/n!}{\tau^n/n! \;+\; 2\sum_{k>n} \tau^k/k!}\,,$$

the Poisson(τ) probability at n against itself plus twice its upper tail. (The normalizer cancels, so the threshold is well defined for every n even though only $n = 1$ has a normalizer free of α.) The threshold used here is the case $n = 1$, which reduces to $\tau/(2e^{\tau} - 2 - \tau)$. The quadratic $n = 2$ gives ≈ 0.53 at $\tau_0 = 1$. The threshold is a property of how the density thins toward the center, not of the parameterization chosen for it. Little else moves with n: the proof of Lemma W1.1 uses the density only through its value at the center, which is proportional to $1 - \alpha$ for every n, so the gap between the two edge functions and the pinch at $\alpha = 1$ carry over unchanged. A larger n costs closed forms elsewhere — at $\tau = 0$ the turnout-weighted half-median solves a cubic rather than a quadratic — and α is not comparable across the family, the unnormalized half-masses being 1 at $n = 1$ but $1 - \alpha/3$ at $n = 2$. That the half-mass is free of α only at $n = 1$ is what singles the linear case out.

B.4 Equilibrium Selection and Dynamics

As in all other static game theories with non-fragile multiple equilibria, the model is not rich enough to provide guidance about the dynamics of how equilibria are selected.[8] (If it did, there would not be multiple equilibria.)

A plausible but not necessary hypothesis is that past prevailing equilibria are selected again. For example, parties that have nominated moderate candidates may tend to do so again. (Unmodelled latent aspects that do not change or change only slowly could play a role in selecting among multiple equilibria, in effect maintaining them over time.) However, all multiple-equilibrium models are also consistent with sudden equilibrium switches instead.[9]

For analogous reason, equilibria are also often presumed to move modestly if underlying parameters move modestly. A plausible hypothesis is that the comparative statics provide guidance. A small shift in the underlying parameters that render a current equilibrium no longer viable but a nearby equilibrium viable can result in a "more plausible" or "more likely" shift of participants to this nearby equilibrium.

This paper follows this common interpretation, assuming that the comparative statics on the feasibility of equilibria also provide guidance to changes in equilibria. When an equilibrium $(-q, -q, +q, +q)$ was previously feasible but after a parameter shift only $(-q-e, -q-e, +q+e, +q+e)$ ($e > 0$) remains feasible, the parameter change is taken to facilitate a potential shift of candidates outwards. This interpretation is informal and does not follow from the equilibrium analysis itself.

B.5 Comparative Statics

The band's two edges respond to more voter polarization (α) and to steeper turnout decay (τ) as

$$\frac{\partial q_{\text{lo}}}{\partial \alpha} \geq 0\,, \qquad \frac{\partial q_{\text{lo}}}{\partial \tau} \geq 0\,, \qquad \frac{\partial q_{\text{hi}}}{\partial \alpha} \geq 0\,, \qquad \operatorname{sign} \frac{\partial q_{\text{hi}}}{\partial \tau} = \operatorname{sign} \alpha\,. \tag{15}$$

Three of the four point outward: more voter polarization lifts both edges, and steeper decay lifts the floor. Only the primary ceiling's response to decay changes sign with α — decay pushes q_{hi}

[8] One dynamic to unsettle an equilibrium is to assign an order to candidates' moves. For example, if the first two candidates assigned to move are $\mathcal{D}$, the first candidate will choose q_{hi} in order not to be undercut to lose the primary. If the first candidate is $\mathcal{D}_1$ followed by $\mathcal{R}_1$, she may choose q_{lo} instead. (The in-text example had both candidates of the $\mathcal{D}$ party having chosen a position already, followed by the $\mathcal{R}$ candidates.)

[9] For example, the multiple equilibria of Diamond and Dybvig (1983) are often read as allowing the better (no-run) equilibrium to persist over many periods. It is unknown what may trigger bank runs one day but not the other.

outward for bimodal voters, holds it at 1/2 for uniform voters, and pulls it inward for single-peaked voters. The displacement is small at plausible decay: the ceiling stands at 1.027 times the party's median voter at $\alpha = 0.6$ and $\tau_0 = 1$, and at 0.952 times it at $\alpha = -0.6$, reaching 1.15 and 0.74 only by $\tau = 5$.

z_{hi} decreases in α and in τ. Writing $z_{\text{hi}} = e^{-\tau q_{\text{hi}}}$, more voter polarization pushes the primary ceiling q_{hi} outward and steeper turnout decay discounts distant voters more, so the boundary turnout weight falls under either force (for uniform voters $z_{\text{hi}} = e^{-\tau/2}$). For the uniform distribution, when voter turnout is unrelated to candidate position ($\tau = 0$), the band ranges from 0 to 0.5, as noted in the introduction. As $\tau \to \infty$, the only equilibrium is the extreme $(-1,-1,+1,+1)$ when $\alpha > 0$, and the moderate $(0,0,0,0)$ when $\alpha < 0$.

B.6 Graphical Illustration

[Insert Figure 2 here: **The Band's Two Edges on the Candidate Line** ($\alpha = 2/3$, $\tau_0 = 1$)]

Figure 2 illustrates the two binding deviations on the candidate line, drawn for a bimodal electorate ($\alpha = 2/3$) where *both* edges bind at once — a position too moderate loses the primary, one too extreme loses the general election. (The uniform electorate of this section has only the upper, primary edge — the lower edge opens once voters polarize, as Figure 3 shows.)

[Insert Figure 3 here: **The Continuum Equilibrium Band as a Function of Voter Polarization**]

Figure 3 traces those same two edges across the whole range of voter polarization, in two panels: Panel A at full participation ($\tau = 0$) and Panel B at the baseline decay ($\tau_0 = 1$). The upper (primary) edge rises with voter polarization in both: it sits at 1/2 for uniform voters at every τ, climbs above 1/2 once $\alpha > 0$, and for single-peaked voters ($\alpha < 0$) more decay lowers it. This upper edge is also relatively insensitive to voter turnout. The lower (general-election) edge differs more sharply between the two panels. At full voter participation, it is 0 for every $\alpha < 1$ — with no abstention the general election ties and nothing binds from below.

The multiplicity of equilibria reflects the primary trap itself. Voter turnout decay lifts the lower edge off 0, but only once voter polarization crosses the threshold $\alpha_c = \tau/(\,2e^{\tau} - 2 - \tau\,) \approx 0.41$. Left of this critical α, the band always begins at 0.

Note that candidate polarization is not assumed. Even when the voter distribution is uniform, it is not necessarily the case that the equilibrium has candidates that sit far off zero. An

equilibrium in which both parties are located extremely near zero is just as viable. The result is structural — it is the primary trap itself that can facilitate moderate just like extremist equilibria — and it is not an artifact of a bimodal electorate or of alienated voters abstaining.

[Insert Figure 4 here: **The Band Edges and Their Width over the** (α, τ) **Plane**]

Figure 4 maps the two edges over the whole (α, τ) plane, with turnout decay τ on a logarithmic axis. **Panel A** is the general-election floor q_{lo}: it is identically 0 — the moderate equilibria survive — to the left of the single curve $\alpha_c(\tau) = \tau/(2e^{\tau} - 2 - \tau)$, and lifts off to its right, where at some point only non-centrist equilibria remain. However, for *all* unimodal distributions, the most moderate equilibrium is always feasible. Bimodal distributions are intrinsically different from unimodal distributions.

Panel B is the primary ceiling q_{hi}: its $1/2$ contour is the uniform line $\alpha = 0$ (invariant in τ), and the ceiling rises toward the extremes as voters polarize. Turnout decay pivots on that uniform line: as τ grows it pushes the ceiling farther out for bimodal electorates ($\alpha > 0$) and pulls it back toward the center for single-peaked ones ($\alpha < 0$). **Panel C** shows that the widest equilibrium band occurs for high voter polarization with small turnout decay.

C Primary Trap Equilibria in a More General Model

The specialized model above used specific parameterizations for voter distribution and participation. The primary trap is, however, far more general. It rests on one substantive restriction on the electorate — symmetry — and otherwise leaves turnout almost free. Symmetry is what makes candidates interchangeable, and from it alone the primary election ties, the general election ties, payoffs equalize, and the equilibrium set collapses to the band — the content of Theorems 1–2 below. Turnout may be any **participation weight** $P(x;\, c_1, c_2, c_3, c_4) \geq 0$ depending on all four candidate positions. (This can include, e.g., higher turnout where a voter may turn out more readily when her nearest candidate is close *and* the second-nearest is far.) The model requires of these objects four assumptions:

(A1) **Symmetric electorate:** each voter's concerns collapse to a single position x on a one-dimensional left–right axis, and the density g of these positions is symmetric, $g(-x) = g(x)$.

(A2) **Anonymity:** votes depend only on the candidate positions — on the *set* of positions on the ballot (all four in a primary, the two nominees in the general election), not on which candidate holds which.

(A3) **Turnout reflection symmetry:** $P(-x;\, -c_1, -c_2, -c_3, -c_4) = P(x;\, c_1, c_2, c_3, c_4)$.

(A4) **Regularity and Continuity:** Regularity: Each candidate's turnout-weighted vote is single-peaked in her *own* position (holding the others fixed) — a restriction on the *shape* of a candidate's payoff, not on what turnout may read. (It may still depend on the whole slate, e.g. on the gap to the second-nearest candidate. See the covered example below.) Two further clauses of the same kind — single-crossing along the mirror family, and the *global* form of the primary comparison — are stated in Appendix $\mathcal{A}$.[10] Continuity: the density g and the turnout weight P are continuous, g is everywhere positive on its support, and total turnout is positive.

(A1) is the one restriction on the electorate.[11] (A2)–(A3) say only that turnout carries no built-in candidate-label or left–right bias. (A4) gathers the standard spatial-competition regularity

[10] With whole-slate turnout, a candidate's vote total — as she moves her own position, the others fixed — could in principle have two local maxima: the global one, and a second, lower peak elsewhere. Single-peakedness rules this out. Without it, the marginal-deviation analysis the proof relies on can miss a profitable jump to the far peak, and Theorem 1 may no longer apply.

[11] The axis x may itself be a fixed one-dimensional index of several underlying concerns, and the results then apply verbatim to that index — but the reduction is a restriction, not a normalization. In U.S. congressional voting the reduction has become more defensible over time: Hare and Poole (2014) document race and social issues being absorbed into the dominant economic liberal–conservative dimension, with the spatial fit of the first dimension

and continuity conditions: for the single-peakedness of (A4) a clean sufficient primitive is that participation is a *log-concave* function of the voter's distance to her nearest candidate — the exponential rule $P = e^{-\tau d}$ of Section II is the leading, log-linear case. The remaining, technical clauses of (A4) — single-crossing along the mirror family, which makes the band a single interval, and the global form of the primary comparison, which says that of two co-partisans standing at or inside their half's turnout-weighted median the more extreme is nominated — are stated in Appendix $\mathcal{A}$, with the proof of Theorems 1–2. These two clauses buy the *shape* of the equilibrium set, not its size. The multiplicity comes from the two boundaries bracketing q rather than solving for it, so without single-crossing the surviving set may be disconnected — but it is still a set, and only the interval structure of clause (iii) is lost. All four assumptions are silent at any mirror-symmetric profile. They do their work off equilibrium, where deviations break the symmetry. The band is then cut out by two boundaries: the *primary* boundary, the turnout-weighted half-electorate median, and the *general-election* boundary, where a more extreme deviation's first-order win margin changes sign.

Theorem 1 (Symmetric Equilibria Bands) *Under (A1)–(A4), the symmetric co-located profile* $(-q, -q, +q, +q)$ *is a Nash equilibrium* if and only if q *lies in a closed interval* $[\, q_{lo}, q_{hi} \,]$ *(possibly empty), where:*

(i) ***Primary boundary.*** q_{hi} *is the turnout-weighted half-electorate median: beyond it a centrist undercut captures more than half the party's primary electorate and wins both stages.*

(ii) ***General-election boundary.*** q_{lo} *is the smallest* q *at which a more extreme deviation no longer wins the general election — the first-order outward margin turns nonpositive. Below a centrist threshold it is* 0.

(iii) ***Contiguity.*** *The Nash-equilibrium set is exactly this interval — it never splits.*

(iv) ***Moderation is blocked, not unwanted.*** *At every equilibrium with* $q > 0$*, the party would win the general election by moving* both *of its candidates toward the center — weakly, and strictly for* $q > q_{lo}$*. No candidate can make that move alone, because it costs her the primary.*

rising across the period. What it takes is an order restriction on the positions themselves together with a second that orders voters independently of the position chosen, at which point majority comparisons behave as they would in a one-dimensional model (Gans and Smart, 1996). A monotone aggregate of the underlying concerns does not suffice on its own, because it orders the voters rather than their comparisons. Genuine multidimensional competition, in which candidates choose points in $\mathbb{R}^k$ with $k \geq 2$, is not covered: central symmetry of the density is not enough, because the general-election tie survives but the primary's half-median pivot and the single-crossing that makes the band a single interval have no multidimensional counterpart.

The proof is in Appendix $\mathcal{A}$.

Clause (iv) carries economic content, and it follows from the mirror tie alone. The mirror profile ties the general election. If in addition no outward move can win it — which is what $q \geq q_{\text{lo}}$ says — then the mirror already maximizes the party's outward general-election margin, so an inward step must lift that margin above the tie. This fixes only the direction of a marginal step, not the center as its destination. What the primary blocks is a move inward as far as q_{lo}, the position at which the party's first-order general-election incentive vanishes. Where that stopping point lies depends on the electorate. When it is centrist enough — turnout nearly flat, or bimodality below α_c — the incentive points inward all the way to the middle, $q_{\text{lo}} = 0$, and the primary blocks a genuine pull toward the center. When the electorate is polarized and turnout decays sharply, so that a concentrated base rewards candidate extremity, the incentive reverses before the center, $q_{\text{lo}} > 0$: the party then seeks only a bounded moderation down to q_{lo}, and a nominee already at q_{lo} wants none. A candidate who moderates *unilaterally* off the equilibrium always improves her party's general-election position, yet never her own primary standing — which is precisely why no candidate is the one to move.

Read the other way, q_{lo} is the general-election counterpart of the primary ceiling. q_{hi} is the turnout-weighted half-electorate median, past which a marginally more moderate co-partisan takes the primary. q_{lo} is the position at which a marginally more polarizing nominee — voters counted at their turnout weights — is just indifferent in the general election, gaining below it and losing above it. The band closes from below as this general-election reward for polarizing reaches farther out. The argument never reads the shape of g or of P: the band is the set of positions from which a party's collective interest in moderation, however far that interest reaches, is blocked by its own primary — for every symmetric electorate, every degree of bimodality, and every turnout rule admitted by (A1)–(A4).

The primary boundary has a reading that needs no parameterization. When everyone votes the turnout weights are constant, so q_{hi} is the plain median of the party's own half of the electorate — which, for a symmetric density, is the electorate's upper quartile. The most extreme sustainable nominee is then her own party's median voter, whatever the shape of the electorate. At full participation the general-election floor sits at the center, because a step outward loses at first order when turnout is flat, so the band is exactly $[\,\text{median}, \text{quartile}\,]$. Turnout decay is the only force that moves the ceiling off the quartile.

The theorem covers broader than the above parameterized form. More general g and P functions move the two boundaries q_{lo} and q_{hi}, but they change neither the interval structure

nor contiguity. Only the boundaries' *locations* are distribution- and participation-specific. For the specialized parameterized model, the boundary functions and their closed-form edges were derived above. As a further example of an admissible participation function, let $d_1 \le d_2$ be a voter's distances to her nearest and second-nearest candidate, and set

$$P = e^{-\tau d_1} \cdot \left[\, 1 + \gamma\,(d_2 - d_1) \,\right]$$

with $\gamma \ge 0$ small. Turnout then decays with the distance to the favorite but is boosted when that favorite is *clear*, the next-best option being farther away. This participation function depends on all four positions, yet it reads them only through the voter's unordered set of distances to the four candidates, not through their identities — so it satisfies anonymity (A2); it is unchanged under reflection, since reflecting every position leaves those distances the same (A3); and each candidate's vote stays single-peaked in her own position for small γ (A4). Theorem 1 therefore applies verbatim, the γ term merely shifting the two boundaries.

Theorem 1 describes which symmetric co-located profiles are equilibria. A companion result establishes that there are no others and fixes the payoffs.

Theorem 2 (Only Symmetric Equilibria) *Under (A1)–(A4) and closed primaries, every pure-strategy Nash equilibrium is a mirror-symmetric co-located profile* $(-q, -q, +q, +q)$ *of Theorem 1 — no asymmetric profile is an equilibrium. In it each candidate is nominated with probability* $1/2$ *and her party wins the general election with probability* $1/2$*, so each of the four holds office with probability* $1/4$*. (Under open primaries the only further equilibria are the interleaved profiles* $(-q, +q, -q, +q)$*, Section III.B.1.)*

This proof is also in Appendix $\mathcal{A}$.

The equal payoff follows from symmetry alone — no feature of the turnout rule enters. It breaks only when the two sides are *asymmetric*: the $1/4$-each tie then splits into $(1/2, 1/2)$ for one party and $(0, 0)$ for the other, whose two candidates are shut out though the office is still filled (Appendix W2 and Section III.B). (A stated payoff of zero always marks an equilibrium that is Nash but not strict — a party that loses for certain is indifferent among all of its positions, so its pair is held where it stands only as the limit of a general election decided with noise.) (A measure-zero set of knife-edge ties — a primary tied at distinct positions, or a non-mirror general-election tie — is set aside. By (A4) none is an equilibrium.)

So far a candidate values only holding office, earning 1. Suppose she also has a small preference (prize ϵ) when winning her party's nomination and losing, rather than losing already

at the primary.[12] The band of Theorem 1 and the payoffs of Theorem 2 are unchanged: at a symmetric equilibrium the prize raises every candidate's payoff by the same amount and it changes who wins no election. Appendix $\mathcal{A}$ carries the prize through the proof.

Corollary 1 (Collapse to a Single Point) *The band* $[\, q_{lo}, q_{hi} \,]$ *pinches to a single point exactly when a marginally more polarizing nominee can no longer improve the general election even at the primary ceiling — that is, when the general-election reward for polarizing reaches* q_{hi}*, so that* $q_{lo} = q_{hi}$*. In the parameterized model this happens only at full bimodality (*$\alpha = 1$*, the density vanishing at the center). Both parties are then locked at their common primary median* $q^* = q_{hi}$*, and no moderate co-located equilibrium survives.*

Uniqueness here should not be read as stability. The band collapses only when voter polarization is severe enough to carry the general-election reward out to the ceiling, and the single surviving point lies far toward the extreme rather than at the center. Uniqueness in this model therefore indicates that candidate extremism is already forced.[13]

III Model Extensions

The analyses in these sections return to the specific parameterized density of $g(x;\, \alpha) = \big[\, (1 - \alpha) + 2{\cdot}\alpha \cdot |x| \,\big]/2$, and $w(d;\, \tau) = e^{-\tau \cdot d}$. These functional forms suffice to exposit some other economic forces in a frugal manner.

A The Principled-Candidate Curse

This section analyzes the effect of a principled candidate, i.e., a candidate who chooses a position regardless of electability.

[12] The function of epsilon is to exclude some off-equilibrium choices when the best equilibrium choice still cannot win. The prize need not be infinitesimal. Measured against the value of holding office, the band and the equal-1/4 payoffs survive intact for every ϵ up to 1/3. A larger prize breaks them, since each candidate then prefers to step outward and secure the certain nomination over sharing the tied general election.

[13] The pinch is exact only at $\alpha = 1$, but the band is already narrow well short of it, because turnout decay closes it from both sides. At $\alpha = 0.6$ the band spans $[\, 0.472, 0.656 \,]$ at the baseline $\tau_0 = 1$, narrows to a width of 0.004 by $\tau = 5$, and to $5{\cdot}10^{-5}$ by $\tau = 10$. Multiplicity is therefore a property of moderate turnout decay. Where voters abstain sharply, the band survives as a formal interval but leaves candidates no practical room, and the model's prediction becomes a point.

[Insert Figure 5 here: **The Principled-Candidate Curse (Uniform $\alpha = 0$, No Decay $\tau = 0$)**]

Figure 5 returns to the simplest case, the uniform distribution $\alpha = 0$ with perfect turnout $\tau = 0$. Both $\mathcal{D}$ candidates have located at -0.3. The only rational position that allows a $\mathcal{R}$ candidate to tie both the PE and the GE is the mirror $+0.3$. However, $\mathcal{R}_1$ is principled rather than interested in winning the election. The figure shows her choosing $+0.6$. This means that the only PE-winning positions for $\mathcal{R}_2$ are between $+0.4$ and $+0.6$, but no value in this range is moderate enough to win the general election. The principled candidate has thus cost not only herself but also her co-partisan the election. Such mechanisms work in general regardless of voter turnout or participation.

[Insert Figure 6 here: **The Principled-Candidate Curse $\mathcal{R}_1$: Winnable and Unwinnable Regions**]

Figure 6 maps the range of possible outcomes: the horizontal axis is the common $\mathcal{D}$ position $-q$, the vertical axis the pinned $\mathcal{R}_1 = r_1$. The region color marks the free candidate's outcome — blue where $\mathcal{R}_2$ can win, red the cursed region where she cannot. The left middle rectangle shows the just-described scenario, where $\alpha = 0$ (uniform) and $\tau = 0$ (full participation). The principled curse appears when the $\mathcal{D}$ candidates have coordinated on a moderate (GE-winning) position, and the $\mathcal{R}_1$ candidate has chosen a modestly more extreme value. (Too extreme a value would allow $\mathcal{R}_2$ with a modest choice to eliminate $\mathcal{R}_1$ in the PE *and* still win the GE.) The reason why all other position pairs lose against $\mathcal{R}_2$ is that $\mathcal{R}_1$ no longer "traps" $\mathcal{R}_2$. $\mathcal{R}_2$ can tailor her response optimally to $\mathcal{D}$'s choices.

The figure illustrates the effects of voter density and turnout more generally. Under unimodality, the principled-curse area shrinks. Voter turnout does not matter. Under bimodality, the principled curse region is larger, but shrinks substantially with more discriminating voter turnout.

In sum, a principled candidate costs her party the election only by standing *somewhat* more extreme than the opposition's mirror — far enough out to be unelectable in the general election, yet not so far that a moderate co-partisan could win both her primary and the office. (It can be shown that a cursed co-partisan cannot escape by moving to the other side of the spectrum in an open primary.) The opposite can also be the case. A principled candidate who instead *moderates* costs only herself, never her party (Theorems W3.1 and W3.2).

B Differential Voter Polarization and Participation

Adding voters to one side enlarges a party and always helps the side where the voters are added. (This can be proven in the specific parameterized model.) Thus, this case is conceptually less interesting than the case where the mass of $\mathcal{R}$ and $\mathcal{D}$ voters are held constant, but only either their polarization or their turnout changes.

B.1 Voter Polarization

Polarizing the electorate can have complex effects. As Smithies (1941) showed in the absence of a primary, candidates may not always want to locate near the median voter and thus moderate towards the central location. If voter turnout is sufficiently discriminative and voter polarization is extreme enough, then more moderation hurts candidates because the standard Downsian effect reverses. The presence of primary elections and variations in both voter polarization and voter turnout also complicate not only the single-election point equilibrium but also the dual-election band equilibrium (i.e., q_{lo} and q_{hi}). The mechanism and result are instructive, although the effects of one side's electorate density on the strategic choices of both candidates are not easily signable (except under certain strong restrictions).

[Insert Figure 7 here: **Increasing Polarization Moving Probability Mass**]

Point Changes: Figure 7 illustrates some of the effects of changing voter preferences on the viable equilibrium band. Start the analysis by considering a symmetric voter situation with a requisite symmetric candidate equilibrium in Panel A (showing only the $\mathcal{R}$ voter density side of the symmetric density and the $\mathcal{R}$ candidate).[14] Now consider the effects of polarization of voters on q_{hi} maintainable in equilibrium. Hold $\mathcal{D}$'s half of the electorate fixed.[15] The thought experiment is to reshape the voter density in $\mathcal{R}$'s half to become more polarized and consider only the $\mathcal{R}$ candidate. (The $\mathcal{D}$ candidate's primary in the closed primary remains unaffected for now.) Own voter base polarization has two effects: (1) it may push the candidates' q_{hi} outward because this is where the (raw) voters have moved to; but (2) with high enough voter

[14] An inner threshold q_{lo} exists because the primary rewards a step outward, nominating the more extreme of two co-partisans standing inside the ceiling q_{hi}. Where q_{lo} sits is then fixed by the general election, which decides how far out such a step still pays.

[15] The *marginal* addition of voters to one side always weakly helps that side, wherever the voters are placed, at every τ — full turnout ($\tau = 0$) included: it is a free vote, and even at the extreme it outweighs the slight outward drift it induces in the nominee.

discriminating tastes (i.e., a high τ), these now more extreme (effective) voters may actually vote less now. This may push q_{hi} inward.

Panels B and C both polarize voters, sending the same mass of voters to the same extreme destination, differing only in where the mass is drawn from. In Panel B, the mass comes from *well inside* the nominee, where it was already too far from her to turn out heavily. Moving that thinly-turning-out inner mass outward costs the nominee little and pushes q_{hi} outward. In Panel C, it comes from *just outside* the nominee (where voter turnout was high). It lands farther out still, where at $\tau = 3$, it turns out too thinly to hold the position. Moving that heavily-turning-out outer mass outward now pushes q_{hi} inward. What sets the sign is where the mass is removed from, not the direction it travels.[16]

Integral Changes: Now let voter polarization differ across the two halves ($\alpha_{\mathcal{D}} \neq \alpha_{\mathcal{R}}$), with masses held equal. Each party remains pinned inside its own band, with the primary changing the upper limit. The solution now depends on how eager candidates are to win a primary that later loses the general election.

If there is a small outward step by the more moderate party that helps a candidate win the primary but is not far enough to lose her the general election against the more polarized party, then the equilibrium can unravel from a band to a point.[17] In such cases, one or the other party is at its own ceiling, the general-election loser keeping only the nomination prize $\epsilon/2$. Either the moderate or the polarized side can be the loser, as the example below shows.

For example, consider an asymmetric scenario in which only one side is polarized for intuition. Assume that the left $\mathcal{D}$ side is bimodal ($\alpha_{\mathcal{D}} = 1$), while the right side is uniform ($\alpha_{\mathcal{R}} = 0$). If there is a tiny amount of voter turnout discrimination (to exclude an edge case), the only viable equilibrium is the asymmetric $(-0.707, -0.707, +0.500, +0.500)$. In this case, the bimodal $\mathcal{D}$ side always loses in the GE. $\mathcal{D}$ candidates keep only the consolation prize $\epsilon/2$. One of the two moderate $\mathcal{R}$ candidates of the moderate $\mathcal{R}$ electorate wins.

[16] Relocating mass inward (no farther from the nominee than before) always weakly helps $\mathcal{R}$ win the general election, guaranteed off the symmetric profile and at every $\tau > 0$. The effect vanishes only in the limit $\tau \to 0$.

[17] Not the collapse but the point location itself depends on the relative preference of candidates to win the primary election while losing the general election. The unraveling in a winner-take-all situation is taking the model too seriously. Although it is logically correct that a candidate should prefer even the tiniest derived utility more, the perfect certainty with extremeness of the payoff (and no candidate preferences) loses realism here. For example, at $\tau = 0.01$ a candidate takes the nomination outright with an outward step of 0.0002, which costs her about 0.0001 of her party's general-election margin *but* doubles her probability of winning office. The economic intuition of the primary trap that candidates are constrained by a need to remain near their competitors is unaffected; and that this can have multiple equilibrium solutions in a band remains reasonable.

However, raising voter turnout decay τ (while holding mass constant on the left and the right side) changes the outcome. Consider $\tau = 1$.[18] The only equilibrium still has the $\mathcal{D}$ candidate more extreme than the $\mathcal{R}$ candidate, because her $\mathcal{D}$ electorate is more extreme. The equilibrium is $(-0.722, -0.722, +0.500, +0.500)$. However, the more polarized $\mathcal{D}$ candidates now beat the more moderate $\mathcal{R}$ candidate in the GE (an effect noted at the outset of this subsection).

[Insert Table 3 here: **Moderating One Side's Voters: Help or Harm — and Even the Direction — Depends on the Electorate**]

Table 3 quantifies a milder version of the same thought experiment. Starting from a symmetric exact-tie equilibrium at the ceiling, only $\mathcal{D}$'s voters are moderated — flattened toward uniform, with mass preserved — while $\mathcal{R}$'s electorate and candidates remain untouched. From a bimodal baseline (Panel A), the flattening lowers $\mathcal{D}$'s ceiling and forces its candidates inward, which helps $\mathcal{D}$ when turnout decay is low but hurts it once τ exceeds roughly 0.88. From a centrist baseline (Panel B), the same flattening raises the ceiling, so the primary re-pins $\mathcal{D}$ farther outward, where it loses at every τ. Moderating one side's voters thus need not moderate — or help — its candidates.

Closed vs. Open Primaries: In an open primary, a candidate can jump across to the other side of the spectrum. This creates two concerns: (1) a new interleaved equilibrium arises at the same locations, but each location now carries one candidate from each party. With a modest penalty for lack of party loyalty, this equilibrium vanishes. (2) In an asymmetric equilibrium (which can happen with asymmetric parameters), one party's candidates can obtain $1/2$ and the other party's candidates $\epsilon/2$. The concern is that a losing candidate would want to jump across to the other side of the spectrum. It can be proven that this turns out never to be the case *in equilibrium* under assumptions (A1)–(A4) (Appendix W5).

B.2 Voter Turnout

In Theorem 2, symmetry gave every candidate $1/4$. This subsection relaxes symmetry in the electorate's participation rather than in a candidate, and in one respect only. The voter distribution $g(\cdot; \alpha)$ stays *symmetric*: the two halves carry equal mass and have the same shape,

[18] The decay rate $\tau = 1$ exceeds the critical rate $\tau = 0.3603$ at which the general-election winner flips from the moderate $\mathcal{R}$ side to the polarized $\mathcal{D}$ side. At the critical rate itself the equilibrium is $(-0.712, -0.712, +0.500, +0.500)$ and the general election ties.

so neither party is bigger than the other and neither party's voters sit any differently from the other's. What differs is only how readily each side turns out — $\mathcal{D}$'s voters at decay rate $\tau_{\mathcal{D}}$ and $\mathcal{R}$'s at $\tau_{\mathcal{R}}$.

The result is a conditional: the side with the more discriminating voters — the steeper own decay rate, hence the more concentrated turnout — always wins the general election if (i) each side's *total* participation is held fixed, by normalizing the kernel to $w(d;\tau) = (\tau/2)\cdot e^{-\tau\cdot d}$ so that $\int_{-\infty}^{\infty} w(d;\tau)\,dd = 1$ for every τ;[19] (ii) the electorate itself is symmetric — the same density and equal mass on both halves, so only the decay rates differ; and (iii) both parties stand at their primary anchors $q_{\text{hi}}(\alpha, \tau_P)$, which the closed primary delivers endogenously. Under the raw kernel of equation (2) the lever reverses, because raising τ then removes participation rather than relocating it, and the side with the more loyal base wins instead. The two kernels answer different questions — how a fixed turnout budget is allocated across voters, and how much turnout there is — and the conditional above is a statement about the first. The mechanism is leakage: concentration keeps a fixed turnout budget from spilling past the general-election divide.

Each condition binds. Condition (i) is the reversal just described, proved in Appendix W6.

[19] The normalized kernel is a participation *density*, not a per-voter probability: $(\tau/2)\cdot e^{-\tau\cdot d}$ exceeds 1 near the nominee once $\tau > 2$. The reading is that a fixed aggregate turnout budget is allocated across voters by proximity — not that each voter's turnout probability stays below one, as in the raw kernel of equation (2).

C More Than Two Candidates

The base model assumed two viable candidates per party (though each of which could have emerged previously from a larger pool of potential entrants). Having two leading contenders with a realistic chance of winning the primary is a common configuration in practice, possibly justified by the large fixed costs of mounting a serious political campaign.

Nevertheless, fields of three or four candidates can also occur. Under the same assumptions (A1)–(A4), all equilibria still have candidates choose symmetric positions and live within continuous bands. Only the primary ceiling changes: with N co-located co-partisans, an inward undercutter must beat each of the other $N-1$, so q_{hi} is the root of $m_{\text{out}}(q) = (N-1)\, m_{\text{in}}(q)$, which tightens as the field grows. The general-election floor q_{lo} is untouched, because the general election remains a contest between two nominees (Appendix W7).

Under full participation ($\tau = 0$) the viable band is $[0, 1/N]$ for the uniform distribution and $[0, 1/\sqrt{N}]$ for the fully bimodal one. Both closed forms are exact only at $\tau = 0$: with turnout decay the ceiling falls below these values for $N \geq 3$ (the uniform $N = 2$ ceiling of $1/2$ is the one that holds at every τ). There is no closed-form solution for arbitrary τ and N.

[Insert Figure 8 here: **Symmetric Equilibria With More Than Two Candidates ($\tau_0 = 1$)**]

Figure 8 draws the bands for $N = 3$ and $N = 4$ at the baseline $\tau_0 = 1$. The new feature is that some parameter regions admit no pure equilibrium: the primary forces a co-located party so moderate that an outside challenger beats it in the general election, while any position extreme enough to survive the general election is undercut in the primary.[20]

Three reinforcing conditions push a party system into the non-equilibrium region. **(i) A crowded primary** — the effect needs $N \geq 3$. With two serious candidates per party the ceiling always clears the floor. It is the third and later entrants, splitting the protective flank, that pull q_{hi} down toward $1/N$. **(ii) A polarized electorate** — the floor q_{lo} is positive only above the centrist threshold ($\alpha > \alpha_c$). An electorate massed at the center keeps $q_{\text{lo}} = 0$ and the band never empties, whatever the field size. **(iii) Alienation, i.e. non-voting** — larger turnout decay τ lowers $\alpha_c = \tau/(2e^{\tau} - 2 - \tau)$ and lifts q_{lo}, so the more sharply distant voters abstain, the less

[20] Discontinuous location games of this Hotelling type generically admit a *mixed*-strategy equilibrium even when no pure one exists. That is not an economically meaningful resolution here, because a candidate runs on a single platform and cannot randomize her position. The empty region is therefore read as genuine nonexistence. Nor does a non-co-located pure profile fill the gap: asymmetric profiles are non-generic (Theorem 2), and the interleaved "double-cross" is excluded in Section III.B.1.

bimodality it takes to open the gap. At the baseline decay the region is remote: it opens only above $\alpha \approx 0.59$ for $N = 3$ and $\alpha \approx 0.49$ for $N = 4$, beyond the modest bimodality that pooled self-identification shows, though not beyond what a party-conditioned primary electorate might reach. Because α_c falls with τ, the region is approached through abstention rather than through voter polarization alone. The equilibrium therefore breaks down where a fractured primary field meets a polarized electorate with low turnout: the crowded primary drives each survivor toward the center to win the nomination, the general election rewards a concentrated extreme base, and when both pressures are strong enough no single platform answers both.

D Further Model Generalizations

D.1 Previous Nomination Literature

A nomination process before the general election was modelled early and then largely set aside.[21] In Owen and Grofman (2006), the primary stage enters as a preference shaping the party's one position. The unitary maximizing agent is the pivotal primary voter, who trades proximity to her own position against the probability that her party's nominee prevails in the general election. Each party selects a single platform — the population median if it is office-seeking, an interior position if it is policy-motivated. The equilibrium is a point, and the comparative statics move this point (rather than trace a set). As a two-player (rather than four-player) game, the model may apply better to earlier (elite-driven) coordinated party choice than the competitive primary elections of today.

D.2 Valence

Work by Gilles Serra and others has focused on valence. There are two threads. The first asks what a quality advantage does: it can win a candidate either election within a region, and surviving a contested primary may itself carry valence into the general. The second asks why

[21] The two earliest treatments are recorded here rather than discussed in the text, because both circulated poorly and are now seldom read. Coleman (1971) first made the point that a nominee must satisfy a more partisan primary electorate before facing the general election, and Owen and Grofman (2006) describe their own model as differing from his in no more than technical details. His article appeared in an early issue of *Public Choice*, shortly after that journal was renamed and before many libraries subscribed to it, and it drew almost no citations for three decades. Aranson and Ordeshook (1972) is the other, and is not covered here at first hand: in the account given by Owen and Grofman (2006), its candidates form expectations about their probability of winning the primary and their probability of winning the general election, each as a function of the position they take, and choose a position so as to maximize the product of the two. That chapter appeared in a volume whose publisher folded shortly afterwards, so the description here is secondhand.

parties hold primaries at all. Serra (2011) frames the nomination as a tradeoff between a policy cost — the primary pulls the nominee toward the primary electorate — and a valence gain, because the primary selects a higher-quality nominee, so a party holds a primary when the quality dividend is large enough. The mechanism is *selection* rather than the building of quality into a fixed candidate. The stronger claim that primaries *make* candidates better is not established, and the divisive-primary evidence (Bernstein, 1977) points the other way.

The first thread is a literature of its own, and in it divergence has a different source. Ansolabehere and Snyder (2000) (two-party spatial competition in more than one dimension) showed that a valence dimension can restore the existence of pure-strategy equilibria, with the advantaged party taking the moderate position. Groseclose (2001) (policy-motivated candidates uncertain about the median, in the Calvert-Wittman tradition) found that a small valence advantage moves the disadvantaged candidate away from the center and the advantaged candidate toward it. Aragonès and Palfrey (2002), Hummel (2010), and Xefteris (2013) (office-seeking candidates, an uncertain median, and an advantage large enough that pure-strategy equilibria generally fail) characterized the mixed equilibria that replace them, again with the advantaged candidate at the moderate positions. Gul and Pesendorfer (2009) obtained divergence from a different primitive: some voters cannot observe policy and decide on personality instead, so that a candidate can win on a partisan platform that every voter would rather not have.

Three differences separate all of these from the present model. Each is a single-stage contest between two candidates, with no nomination to survive first. Each derives divergence from an exogenous asymmetry — a quality gap, a policy preference, or an information failure — whereas here four symmetric office-seekers, identical in quality and indifferent about policy, are held off-center by their own primary. And where these models deliver a *range* of positions, as in Hummel's ranges without gaps, that range is the support of a mixed strategy inside a single equilibrium. The band here is a set of distinct pure-strategy equilibria, any one of which can be played and then held.

These forces map onto the trap rather than replacing it. Appendix W8 carries valence through the present model, where Serra's tradeoff reappears as a threshold: a party holds a primary when the quality dividend clears ν_{crit}. Crowding and quality act on opposite edges of the trap, and the general-election floor q_{lo} answers to neither. More candidates tighten the ceiling q_{hi} downward (Section III.C), whereas a quality advantage loosens the trap *below* it, since $\nu^*_{\text{pri}}(q) = m_{\text{out}}(q)/m_{\text{in}}(q) - 1$ is the quality price of moving toward the center and still winning the primary (W8.4). A dividend shared by all four candidates cancels and leaves the band where

it was (Lemma W8.1), so a quality advantage buys a *particular* candidate room on the central side, not a wider band.

Serra (2015) reaches the opposite conclusion, that adding competitive nominations to Downs would still yield full convergence on a moderate candidate: a policy-motivated selectorate nominates not the candidate closest to its own position but its most electable one, and each candidate therefore moves toward the center to be that one. The difference here is the selectorate. The nominee is chosen by the party's own voters on proximity to themselves, so a candidate who moderates loses her primary to a more aligned co-partisan — candidate polarization within the band survives even at full, uniform turnout, where electability plays no role at all. Serra's convergence is the opposite pole, the pure-electability selectorate. Decaying turnout widens the gap between the poles, because it severs the link his argument relies on: with abstention the more moderate candidate need not win the general, so even an electability-minded voter would not necessarily favor her. Real selectorates plausibly mix the three types — voters who choose the nearest candidate, voters who choose the most electable, and voters who abstain — and the two frameworks could be integrated to weight these motives and to measure their distribution across contexts.

D.3 Other Potential Generalizations

Like all models, this model assumed away other aspects that can play important roles in practice. In exchange for gaining insights about the important strategic interplay when there are both a PE and GE, the model neglected, for example:

1. The electorate and party structure:
   (a) Unequal party sizes, relaxing the equal-mass symmetry of the two halves so that one party's electorate outnumbers the other's.
   (b) More than two parties.
   (c) Multidimensional attributes that do not simply collapse into a single line (see also Gans and Smart, 1996).
2. Who the candidates are and what they want:
   (a) Endogenous entry: who chooses to run, with entry costs and strategic deterrence determining the number of candidates N rather than fixing it (see also Callander and Wilson, 2007).

(b) An intrinsic preference for being more partisan — or, more generally, policy-motivated candidates who value the winning platform itself, not only holding office, so that a nominee trades her own ideal point off against the probability of winning (but see Calvert, 1985).[22]

(c) Differential payoff utility (potentially leading to a game of chicken where one candidate hopes to induce the other to drop out).

(d) The importance of activists and elites in selecting candidates (see also Aldrich, 1983; Amitai, 2024).

3. A richer candidate strategy space:

(a) The ability of candidates to move positions between PE and GE at a cost (see also Agranov, 2016).

(b) The ability of money to influence outcome.

(c) Side payments between candidates. This could allow for some coordination (rather than competition) among candidates, or allow one party's candidates to support a principled candidate on the other side.

4. Information, stochastics, and bounded optimization:

(a) A stochastic rather than deterministic election outcome, in which a more extreme candidate wins with lower probability rather than not at all, so that an outcome-motivated voter trades the chance of victory off against a platform closer to her own ideal — and more extreme voters may then favor a more extreme candidate (see also Hinich et al., 1973).

(b) Voter uncertainty about candidate positions, so that platforms are perceived with noise rather than perfectly known.

(c) Uncertainty about parameters (and thus outcomes).

(d) Learning about unknown parameters.

(e) Disagreement about the uncertainty.[23]

(f) Smarter strategic behavior by atomistic voters (but see Dal Bó et al., 2018).

[22] Some of the literature assumes that candidates want to be off-center. Here, office-seekers are instead pushed off-center by their own primary. In the policy-motivated account (Adams et al., 2005), divergence comes from the candidates' own (principled) preferences or from those of the activists who select them — so divergence is, in a precise sense, assumed: a principled candidate can remain viable, but the principle itself is exogenous. This paper's candidates have no such preference. They are pure office-seekers who would occupy the median at once, and they still cannot reach it. Candidate positioning is an outcome, not a primitive. (In the model above, candidates could only be either principled or office-seeking. The principled candidate of Section III.A was only the limiting illustration.)

[23] The basic model contains no private information, which helps facilitate both tractability and empirical applicability. This is a model advantage. However, this is not necessarily a good description of reality.

(g) Optimization only within bounds — candidates who act on first-order incentives and neglect second-order ones, so that the thin deviations pinning the edges of the band need not be found in practice.

5. Equilibrium selection: latent variables that reduce the multiplicity of available equilibria, especially across election cycles. Alternatively, there could be historical path dependence, focal points, elite coordination.

Appendix W10 records, for each of these neglected forces, a conjecture about how they would interact with this paper's sequential elections and resulting equilibrium bands.

Despite these omissions, the strategic choices constrained by both the PE and the GE — and the equilibrium externalities they generate — are likely to play a first-order role in practice, and they have not been analyzed in models of the factors just listed. The primary trap creates a robust externality, which facilitates a band of equilibria, in which the same voter configuration can sustain equilibria in which candidates on both sides are similarly moderate or polarized.

IV Empirical Background

The intuition that primary elections pull candidates to become more partisan is an old one (Section III.D.1). The contribution of the paper here is the equilibrium analysis of this intuition in a strategic setting of competitive candidates, which reconciles the fundamental importance with the lack of necessary strong links in the equilibrium.

In all models, the assumptions may fit real-world applications better in some situations than others. In the model of this paper, voter preferences and candidate positioning occur on an abstract single-dimensional line. This spectrum could be policy- or identity-based. In the latter case, candidates would choose ideological self-identification or out-party antipathy. (However, they may not be able to choose racial or ethnic tribal affiliations, which constitutes a spectrum where the model does not apply.) Moreover, over time, the relevant attributes may change, presenting a natural challenge to empirical measures of polarization. For example, in the US, it appears that the relative importance of identity over policy was lower in the past than it is today. The subsections below therefore note, for each study, the years covered and whether polarization measures are topical or identity-based.

This section discusses some literature about past polarization of the electorate, the polarization of candidates, the connection (or lack of one) between the two, the footprint of having primaries at all, and the footprint of open versus closed primary rules.

A Electorate Polarization

Policy On the distribution of issue positions, the electorate changed only modestly, and through the early 1990s it was unimodal and centrist. DiMaggio et al. (1996) (General Social Survey and National Election Study attitude items, mid-1970s to mid-1990s) found little electorate polarization, with abortion and the gap between partisan identifiers the exceptions. Whether the movement since then is polarization at all is contested. Fiorina and Abrams (2008) (issue-position evidence through 2008) and Fiorina et al. (2008) (raw National Election Study seven-point issue scales, 1984–2004) interpreted it as *sorting* — a tightening correlation between a citizen's views and her party label, which Levendusky (2009) attributed to elites redefining what the parties stand for — rather than as a widening of positions, while Abramowitz and Saunders (2008) (American National Election Studies multi-issue scale with ideological self-identification, through 2004) found divergence among the public, deepest among the most engaged, which Fiorina et al.

(2008) in turn argued conflated elite behavior with mass opinion. One measurement caveat spans the debate: Ansolabehere et al. (2008) showed that individual survey items carry enough measurement error to make voters appear to hold no attitudes at all, and that averaging many items within an issue area restores preferences that are stable, ideologically constrained, and about as predictive of the vote as party identification — so these distributions are reading genuine positions rather than noise. For the model, the pooled topical α is small while each party's own half is distinctly one-sided, which is the primary electorate the q_{hi} pin reads. Abramowitz (2010)'s sharpest division, between the engaged and the disengaged, is a claim about *who turns out* — the participation lever τ — not a rival account of the density.

Self-Identification On ideological self-identification, the electorate seems to have shifted more. The American National Election Studies seven-point self-placement[24] has peaked at "moderate" in every year since 1972. The Gallup self-identified moderate share fell from roughly 43% in 1992 to about 33% by 2024,[25] a thinning but not an emptying of the center. In the model's terms, the pooled α read off voter self-identification is small and has risen only slightly.

Identity On pure identity, voters have perhaps moved the most. Since the 1980s Americans have divided increasingly by partisan identity rather than by issue position (Iyengar et al., 2012; Iyengar et al., 2019), and party has come to weigh more heavily than policy in how voters place themselves (Mason, 2018; Achen and Bartels, 2016). By 2024 the party-conditioned distribution was bimodal: 77% of Republicans called themselves conservative and 55% of Democrats liberal, both record highs (Gallup).

From the model's perspective, there seems to be more bimodality in each party's own electorate than in the pooled density. The voter polarization the mechanism requires is therefore found where the data most clearly show it: at the primary, not in the electorate as a whole.

[24] American National Election Studies, Time Series Cumulative Data File (https://electionstudies.org).

[25] Gallup, Party Affiliation and Ideology historical trends (https://news.gallup.com/topic/ideology.aspx). The party-conditioned shares reported below are from Gallup's 2024 release on party ideology (https://news.gallup.com/poll/655190/political-parties-historically-polarized-ideologically.aspx).

B Voter Turnout

The model's second electorate parameter is turnout decay τ, and the evidence bearing on it is thinner than the evidence on voter polarization. Hirano and Snyder (2019) found ideologically extreme House members as likely to be elected in high-turnout primaries as in low-turnout ones. This does not bear against the mechanism here, because τ is the rate at which participation falls with a voter's distance to her nearest candidate rather than the level of primary turnout, and the two need not move together. The two parameters also do not act separately: the general-election floor q_{lo} lifts off zero only above the centrist threshold $\alpha_c(\tau)$, which falls as τ rises, so sharper decay lowers the voter bimodality needed to foreclose the moderate equilibria, and decay pushes the ceiling q_{hi} outward only where the electorate is already bimodal. Direct evidence on the decay rate would require participation measured against that distance, which the available series do not report.

C Candidate Polarization

Empirical measures of candidate polarization have largely been topical, as positions revealed in roll-call votes and campaign behavior. Hare and Poole (2014) (DW-NOMINATE scaling of congressional roll calls, 1879–2013) showed the parties pulling apart steadily since the mid-1970s, the ideological center hollowing out, and moderates in Congress virtually disappearing — candidate polarization now stands higher than at any time since the end of the Civil War.

The effects seem to have worked through *who runs and who wins* rather than through a pull on a fixed candidate. Hall (2015) (U.S. House, 1980–2010, primaries nearly tied between an extremist and a moderate) found that nominating the extremist carried a large general-election penalty, much as Bernstein (1977) (U.S. Senate, 1956–1972) found that a close nomination fight itself depressed general-election performance — and yet extremists kept winning nominations. Hall (2019) (House candidates, winners and losers alike, scaled by their campaign donors from 1980 onward) attributed the divergence to candidate supply: moderates increasingly declined to run. Thomsen (2017) (state legislators, 2000–2010, deciding whether to seek a congressional seat) documented the same self-selection, and found sitting House moderates retiring faster. The primary trap of this paper is a constraint of this kind: it is about which position can *survive* a primary — a selection restriction on the viable set — not a behavioral pull on a candidate who would otherwise moderate.

D Candidate-and-Electorate Co-polarization

The two preceding subsections show a stable electorate and a diverging set of candidates. Bafumi and Herron (2010) (members of the 109th and 110th Congresses and their constituents, from a large 2006 survey, placed on one common topical scale) gave the best evidence on the gap between them: members were more extreme than the voters who elected them, and a seat changing hands replaced one relative extremist with an opposing one — "leapfrog" representation, in which the moderate middle is less represented even as the district's own median barely moves. This is perhaps the best empirical evidence of "mutual extremism." The same gap appears outside the United States: Iversen (1994) (elite and mass surveys covering 37 parties in seven West European countries) documented parties standing more extreme than their own electorates, and read the divergence as a centrifugal pull held in check by the need to remain representative — a bounded divergence rather than an unbounded one.

The model suggests that as the electorate drifts, candidate extremism could but need not have drifted with it. Electorate polarization does not fix candidate polarization at a point. It fixes a *band* of candidate positions, any of which can be observed with the same electorate, though the band is bounded at both ends. The comparison the model pins is to the selectorate rather than to the population. At full turnout the nominee coincides with the median voter of the primary electorate that nominates her, so the model does not predict that candidates are more polarized than the voters who select them. Measured against the electorate as a whole she stands at its outer quartile, above the overall median. Turnout decay moves that ratio in either direction, and within the band the realized position is not determined at all.

E Primaries vs. Other Systems

The model helps resolve the tension between the intuition that candidates need to polarize if they want to win (closed) partisan primaries and the lack of a strong connection between partisan primaries and candidate extremism. Strategically, candidates need to be about as extreme as their peers on either side. The prior question is not how open a primary is but whether there is one at all. Hirano et al. (2010) (U.S. congressional roll calls, testing the introduction of primaries, primary turnout, and the threat of primary competition) found nulls throughout, and Hirano and Snyder (2019) (U.S. House, comparing primaries against the conventions, caucuses, and committee meetings that would replace them) found the alternatives no more moderate — pre-primary conventions endorsed extreme candidates more often than the primaries that

followed them — and noted that primaries were adopted decades before polarization began rising around 1980. Cintolesi (2022) (staggered introduction of primary elections for U.S. senators and representatives in Indiana and New York, difference-in-differences) found that the switch away from elite selection narrowed the pre-reform ideological gap between the parties by about a fifth. Fernandes et al. (2025) (matched difference-in-differences across thirty-four democracies, plus a survey experiment in three) found that democratizing candidate selection did not pay off at the polls either, although their outcome was the party's vote share rather than the nominee's position.

Introducing a nomination stage therefore need not polarize — a puzzle under any theory in which the primary pulls positions outward, but not under one in which it fixes only the range of positions that can survive.

F Open vs. Closed Primaries

The non-strategic interpretation of open primaries is that they should reduce polarization, because candidates should have to locate on the Downsian central median. Open vs. closed primary reforms have likewise left little empirical footprint. McGhee et al. (2014) (U.S. state legislators over two decades, five system types ordered by openness) and Rogowski and Langella (2015) (donor-based ideology scores for more than 85,000 U.S. House and state-legislative candidates, 1980–2012) found no systematic association between the openness of the primary and elected extremism.

Where an effect has appeared, it was neither large nor monotone in openness. Gerber and Morton (1998) (interest-group ratings of U.S. House winners, 1982–1990) found the most moderate representatives under *semi*-closed rather than open rules. Bullock and Clinton (2011) (California's 1996–2000 blanket primary) found moderation only in the least partisan districts, which are also the rarest. Ahler et al. (2016) (survey experiment on California voters just before the 2012 top-two reform) found that voters often could not tell moderate from extreme candidates, so the reform did not reliably advantage the moderates it was meant to help. Grose (2020) (U.S. House roll calls, 2003–2018) is the dissent, and it is about top-two rather than open rules: legislators elected under top-two primaries were substantially less extreme than those elected under closed primaries, with open primaries also associated with moderation but at smaller magnitude.

Even the mechanism is softer than assumed. Sides et al. (2020) (validated primary turnout matched to five surveys, 2008–2014) found primary voters close to their own party's rank and file in demographics and policy attitudes, and found the similarity not to vary with the openness of the primary. Hill (2015) (turnout-validated surveys and an item-response model of the primary and general electorates of every U.S. House district, 2010 and 2012) found that closed rules did not produce more polarized primary electorates. Who turned out mattered, the formal rule little.

The band is a resolution. First, the constraint is real but slack: the same voter configuration sustains moderate and extreme equilibria alike, so cross-sectional variation in the rule need not translate into cross-sectional variation in positions. Second, the open and closed rules coincide on the entire symmetric band — a profile inside the band survives both rules and a profile outside fails both (Appendix W4) — and even under asymmetry a shut-out party's candidates never profitably cross to the other side in equilibrium (Appendix W5). The two rules part company only off the symmetric configurations, in scenarios the reduced-form literature rarely isolates — asymmetric electorates, where the open rule admits cross-over relocations (Appendix W4). Even differently polarized electorates can tolerate moderate and extremist equilibria over a wide range of underlying voter parameters. A literature that regresses candidate extremism on the primary rule is therefore looking for the trap in the one place the model says it leaves little trace.

V Conclusion

It has long been known that primaries push candidates towards more partisan positions. This paper has explored the equilibrium consequences of the strategic positioning of candidates who first have to survive such a primary before the general election. If moderation helps in the general election, then candidates more extreme than their peers cannot survive the general election. Candidates more moderate than their peers cannot survive their primary election. This paper called this "the primary trap." It is not possible for a competitive candidate to escape this trap.

The key assumption of the model is that candidates want to be competitive. This means that their positions *vis-à-vis* their electorate do not matter as much as their positions *vis-à-vis* their competitors. The all-important positioning of candidates is not in absolute but in relative terms. Under reasonable assumptions, the candidates can find themselves in one of several simultaneously viable equilibria — some in which all candidates are more moderate, others in which all candidates are more extreme. The model's central implication is that when parties are

reasonably mirror similar, it rejects competitive situations in which one party's candidates are moderate while the other's are extreme. Viable equilibria are determined by the interplay of voter polarization, voter turnout, candidate preferences, and other considerations. These are all empirically measurable quantities in principle. Furthermore, the model allows principled candidates with little interest in electability to make their own parties' other candidates also unelectable.

The model can thus explain simultaneously the first-order importance of strategic positioning in the primary election, and the weak measured association between electorate polarization and candidate polarization. Candidates can be locked to the equilibrium, not to the underlying parameters.

References

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

Abramowitz, A. I. and K. L. Saunders (2008). "Is Polarization a Myth?" In: *Journal of Politics* 70.2, pp. 542–555.

Achen, C. H. and L. M. Bartels (2016). *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton University Press.

Adams, J. F., S. Merrill III, and B. Grofman (2005). *A Unified Theory of Party Competition*. Cambridge University Press.

Agranov, M. (2016). "Flip-Flopping, Primary Visibility, and the Selection of Candidates". In: *American Economic Journal: Microeconomics* 8.2, pp. 61–85.

Ahler, D. J., J. Citrin, and G. S. Lenz (2016). "Do Open Primaries Improve Representation? An Experimental Test of California's 2012 Top-Two Primary". In: *Legislative Studies Quarterly* 41.2, pp. 237–268.

Aldrich, J. H. (1983). "A Downsian Spatial Model with Party Activism". In: *American Political Science Review* 77.4, pp. 974–990.

Amitai, Y. (2024). "The Activists Who Divide Us: A Cross-Country Analysis of Party Activists' Influence on Polarization and Representation". In: *Comparative Political Studies* 57.3, pp. 453–485.

Ansolabehere, S., J. Rodden, and J. M. Snyder (2008). “The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting”. In: *American Political Science Review* 102.2, pp. 215–232.

Ansolabehere, S. and J. M. Snyder Jr. (2000). “Valence Politics and Equilibrium in Spatial Election Models”. In: *Public Choice* 103.3/4, pp. 327–336.

Aragonès, E. and T. R. Palfrey (2002). “Mixed Equilibrium in a Downsian Model with a Favored Candidate”. In: *Journal of Economic Theory* 103.1, pp. 131–161.

Aranson, P. H. and P. C. Ordeshook (1972). “Spatial Strategies for Sequential Elections”. In: *Probability Models of Collective Decision Making*. Ed. by R. G. Niemi and H. F. Weisberg. Columbus, OH: Charles E. Merrill, pp. 298–331.

Bafumi, J. and M. C. Herron (2010). “Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress”. In: *American Political Science Review* 104.3, pp. 519–542.

Bernstein, R. A. (1977). “Divisive Primaries Do Hurt: U.S. Senate Races, 1956–1972”. In: *American Political Science Review* 71.2, pp. 540–545.

Bullock, W. and J. D. Clinton (2011). “More a Molehill than a Mountain: The Effects of the Blanket Primary on Elected Officials’ Behavior from California”. In: *Journal of Politics* 73.3, pp. 915–930.

Callander, S. and C. H. Wilson (2007). “Turnout, Polarization, and Duverger’s Law”. In: *Journal of Politics* 69.4, pp. 1047–1056.

Calvert, R. L. (1985). “Robustness of the Multidimensional Voting Model: Candidate Motivations, Uncertainty, and Convergence”. In: *American Journal of Political Science* 29.1, pp. 69–95.

Cintolesi, A. (2022). “Political Polarization and Primary Elections”. In: *Journal of Economic Behavior & Organization* 200, pp. 596–617.

Coleman, J. S. (1971). “Internal Processes Governing Party Positions in Elections”. In: *Public Choice* 11, pp. 35–60.

Comanor, W. S. (1976). “The Median Voter Rule and the Theory of Political Choice”. In: *Journal of Public Economics*.

Dal Bó, E., P. Dal Bó, and E. Eyster (2018). “The Demand for Bad Policy when Voters Underappreciate Equilibrium Effects”. In: *Review of Economic Studies* 85.2, pp. 964–998.

Diamond, D. W. and P. H. Dybvig (1983). “Bank Runs, Deposit Insurance, and Liquidity”. In: *Journal of Political Economy* 91.3, pp. 401–419.

DiMaggio, P., J. Evans, and B. Bryson (1996). “Have Americans’ Social Attitudes Become More Polarized?” In: *American Journal of Sociology* 102.3, pp. 690–755.

Downs, A. (1957). *An Economic Theory of Democracy*. Harper & Row.

Fernandes, J. M. et al. (2025). “The Electoral Implications of Legislative Candidate Selection Democratization”. In: *British Journal of Political Science* 55, e125, pp. 1–24.

Fiorina, M. P. and S. J. Abrams (2008). “Political Polarization in the American Public”. In: *Annual Review of Political Science* 11, pp. 563–588.

Fiorina, M. P., S. J. Abrams, and J. C. Pope (2008). “Polarization in the American Public: Misconceptions and Misreadings”. In: *Journal of Politics* 70.2, pp. 556–560.

Gans, J. S. and M. Smart (1996). “Majority Voting with Single-Crossing Preferences”. In: *Journal of Public Economics* 59, pp. 219–237.

Gerber, E. R. and R. B. Morton (1998). “Primary Election Systems and Representation”. In: *Journal of Law, Economics, and Organization* 14.2, pp. 304–324.

Grose, C. R. (2020). “Reducing Legislative Polarization: Top-Two and Open Primaries Are Associated with More Moderate Legislators”. In: *Journal of Political Institutions and Political Economy* 1.2, pp. 267–287.

Groseclose, T. (2001). “A Model of Candidate Location When One Candidate Has a Valence Advantage”. In: *American Journal of Political Science* 45.4, pp. 862–886.

Gul, F. and W. Pesendorfer (2009). “Partisan Politics and Election Failure with Ignorant Voters”. In: *Journal of Economic Theory* 144.1, pp. 146–174.

Hall, A. B. (2015). “What Happens When Extremists Win Primaries?” In: *American Political Science Review* 109.1, pp. 18–42.

Hall, A. B. (2019). *Who Wants to Run? How the Devaluing of Political Office Drives Polarization*. University of Chicago Press. Esp. ch. 4, “Polarization and the Devaluing of Office”.

Hare, C. and K. T. Poole (2014). “The Polarization of Contemporary American Politics”. In: *Polity* 46.3, pp. 411–429.

Hill, S. J. (2015). “Institution of Nomination and the Policy Ideology of Primary Electorates”. In: *Quarterly Journal of Political Science* 10.4, pp. 461–487.

Hinich, M. J., J. O. Ledyard, and P. C. Ordeshook (1973). “A Theory of Electoral Equilibrium: A Spatial Analysis Based on the Theory of Games”. In: *Journal of Politics* 35.1, pp. 154–193.

Hinich, M. J. and P. C. Ordeshook (1969). “Abstentions and Equilibrium in the Electoral Process”. In: *Public Choice* 7, pp. 81–106.

Hirano, S. and J. M. Snyder Jr. (2019). *Primary Elections in the United States*. Cambridge University Press. Esp. ch. 11, “Primaries and Polarization,” pp. 258–306.

Hirano, S. et al. (2010). “Primary Elections and Partisan Polarization in the U.S. Congress”. In: *Quarterly Journal of Political Science* 5.2, pp. 169–191.

Hummel, P. (2010). “On the Nature of Equilibria in a Downsian Model with Candidate Valence”. In: *Games and Economic Behavior* 70.2, pp. 425–445.

Iversen, T. (1994). “The Logics of Electoral Politics: Spatial, Directional, and Mobilizational Effects”. In: *Comparative Political Studies* 27.2, pp. 155–189.

Iyengar, S., G. Sood, and Y. Lelkes (2012). “Affect, Not Ideology: A Social Identity Perspective on Polarization”. In: *Public Opinion Quarterly* 76.3, pp. 405–431. DOI: 10.1093/poq/nfs038.

Iyengar, S. et al. (2019). “The Origins and Consequences of Affective Polarization in the United States”. In: *Annual Review of Political Science* 22, pp. 129–146. DOI: 10.1146/annurev-polisci-051117-073034.

Jones, M. I., A. D. Sirianni, and F. Fu (2022). “Polarization, Abstention, and the Median Voter Theorem”. In: *Humanities and Social Sciences Communications* 9.

Levendusky, M. (2009). *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*. University of Chicago Press.

Mason, L. (2018). *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press.

McGhee, E. et al. (2014). “A Primary Cause of Partisanship? Nomination Systems and Legislator Ideology”. In: *American Journal of Political Science* 58.2, pp. 337–351.

Owen, G. and B. Grofman (2006). “Two-Stage Electoral Competition in Two-Party Contests: Persistent Divergence of Party Positions”. In: *Social Choice and Welfare* 26, pp. 547–569.

Priest, G. L. and B. Klein (1984). “The Selection of Disputes for Litigation”. In: *Journal of Legal Studies* 13.1, pp. 1–55.

Rogowski, J. C. and S. Langella (2015). “Primary Systems and Candidate Ideology: Evidence from Federal and State Legislative Elections”. In: *American Politics Research* 43.4, pp. 846–871.

Serra, G. (2011). “Why Primaries? The Party’s Tradeoff between Policy and Valence”. In: *Journal of Theoretical Politics* 23.1, pp. 21–51.

Serra, G. (2015). “No Polarization in Spite of Primaries”. In: *The Political Economy of Governance*. Ed. by N. Schofield and G. Caballero. Springer, pp. 211–229.

Sides, J. et al. (2020). “On the Representativeness of Primary Electorates”. In: *British Journal of Political Science* 50.2, pp. 677–685.

Smithies, A. (1941). “Optimum Location in Spatial Competition”. In: *Journal of Political Economy* 49.3, pp. 423–439.

Thomsen, D. M. (2017). *Opting Out of Congress: Partisan Polarization and the Decline of Moderate Candidates*. Cambridge University Press. Esp. ch. 4, “Ideological Moderates Won’t Run for Congress,” pp. 81–97.

Xefteris, D. (2013). “Equilibrium in a Discrete Downsian Model Given a Non-Minimal Valence Advantage and Linear Loss Functions”. In: *Mathematical Social Sciences* 65, pp. 150–153.

Table 1: Glossary of Notation

Symbol	Description
PE	Primary election (fought within a party)
GE	General election (across the two nominees).
closed	Primary rule: voters split at voter location party line $x = 0$.
open	Primary rule: voters split at candidate midpoints
	Party Names
$\mathcal{D}, \mathcal{R}$	The two parties: $\mathcal{D}$ on the left, $\mathcal{R}$ on the right.
$\mathcal{D}_1, \mathcal{D}_2, \mathcal{R}_1, \mathcal{R}_2$	The candidates' fixed-label positions (wlog $\mathcal{D}_1 \le \mathcal{D}_2 \le \mathcal{R}_1 \le \mathcal{R}_2$).
$(\mathcal{D}_1, \mathcal{D}_2, \mathcal{R}_1, \mathcal{R}_2)$	A position profile, e.g. the symmetric co-located $(-q, -q, +q, +q)$.
	Specific Functional Parameterization
x	A position on the spectrum, $x \in [-1, +1]$, with the center at 0.
q	A candidate's distance from the center — the half-separation of the co-located profile $(-q, -q, +q, +q)$.
$g(x; \alpha)$	Voter density $[\,(1-\alpha) + 2\alpha\lvert x\rvert\,]/2$ on $[-1, +1]$ (1); $\alpha = g(1) - g(0)$.
α	Voter polarization, $\alpha \in [-1, 1]$: single-peaked ($\alpha < 0$), uniform ($\alpha = 0$), bimodal ($\alpha > 0$).
α_c	Centrist threshold $\tau/(2e^{\tau} - 2 - \tau) \approx 0.41$ at τ_0; the lower edge lifts off the center only for $\alpha > \alpha_c$.
$w(d) = e^{-\tau d}$	Turnout weight of a voter at distance d from her nearest candidate (2).
τ	Turnout-decay rate. Baseline $\tau_0 = 1$; $\tau = 0$ is full turnout.
	Useful Functions
$m_{\text{out}}(q; \tau, \alpha)$	Outer mass: turnout-weighted electorate between $-q$ and the left edge (3).
$m_{\text{in}}(q; \tau, \alpha)$	Inner mass: turnout-weighted electorate between $-q$ and the center (4).
q_{hi}	PE (upper) edge — the turnout-weighted half-median, where $m_{\text{out}} = m_{\text{in}}$.
q_{lo}	GE (lower) edge — where the outward deviation stops winning; 0 below α_c.
$[\,q_{\text{lo}}, q_{\text{hi}}\,]$	The equilibrium band: Nash eqbm iff q lies in it (Theorem 1).

Explanations: The notation of the main text, grouped as the parties and the game, the two voter primitives (density and turnout), and the objects that define the equilibrium band. All positions lie on $[-1, +1]$ with the center at 0.

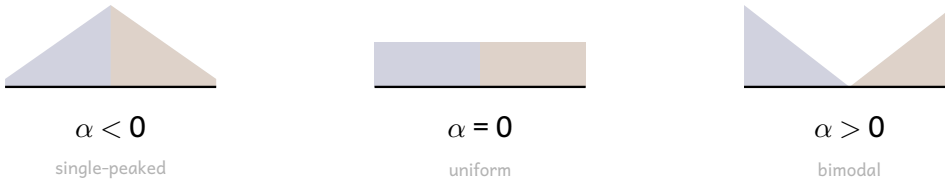
Table 2: Primary Systems as Voter-to-Primary Assignment Rules

States P, C	System	Voter → primary assignment	In model?	Effect on the band
18, 13	Closed (*baseline*)	registration, not proximity	yes	cross-over blocked; no collapse
16, 18	Open	nearest candidate (proximity)	yes	cross-over enabled; collapse possible
9, 11	Semi-closed / semi-open	partisans fixed; indep. choose	partial	between closed and open
0, 2	Top-two "jungle"	one ballot, top two advance	no	same-party finalists possible
0, 1	Top-four + RCV	top four, ranked-choice general	no	larger finalist set; moderating
—	Runoff / two-round	majority runoff	tweak	modest departure
—	Caucus / convention	activist-weighted selection	reweight	activist-skewed electorate
—	No primary	party fixes one nominee	yes	degenerate Downsian

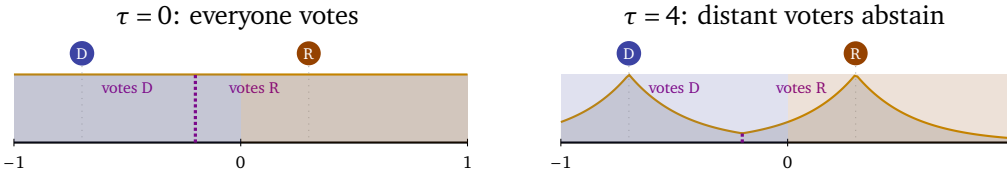
Explanations: The first column gives the approximate number of U.S. states using the system for presidential (P) and congressional (C) primaries (2026; from Table $\mathcal{B}.1$ and the NCSL/Ballotpedia counts). The figures are approximate and do not sum to fifty, because some states run different systems for the two parties and because runoff and caucus selection (marked "—") are layers on top of an open or closed primary rather than separate categories. "In model": *yes* = derived here; *partial* = a mixture of two derived cases; *tweak/reweight* = a documented one-parameter change (a runoff stage, an electorate reweighting); *no* = a different stage structure not derived here.

Figure 1: Voter Polarization, Turnout Decay, and the Two Elections

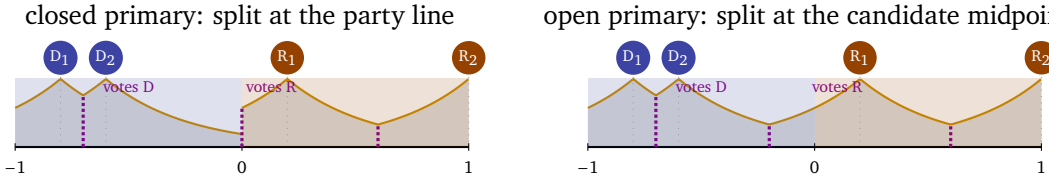
Panel A: Voter Distribution: the one-parameter family $g(x;\,\alpha) = [\,(1-\alpha) + 2{\cdot}\alpha|x|\,]/2$ indexed by α



Panel B: General Election: turnout to the nearer nominee ($e^{-\tau|x-c|}$ to nominee c), voters split at the midpoint



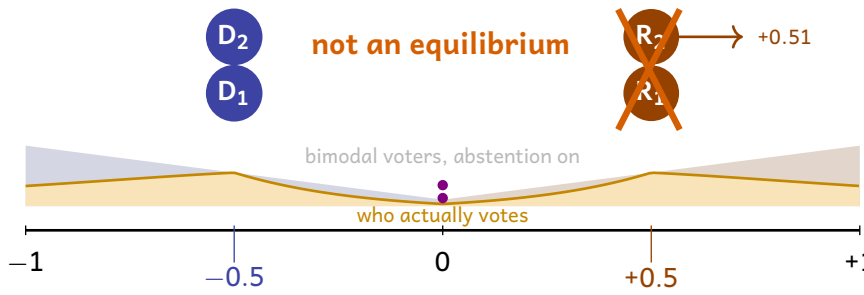
Panel C: Primaries: turnout to the nearest co-partisan (closed) or nearest candidate (open)



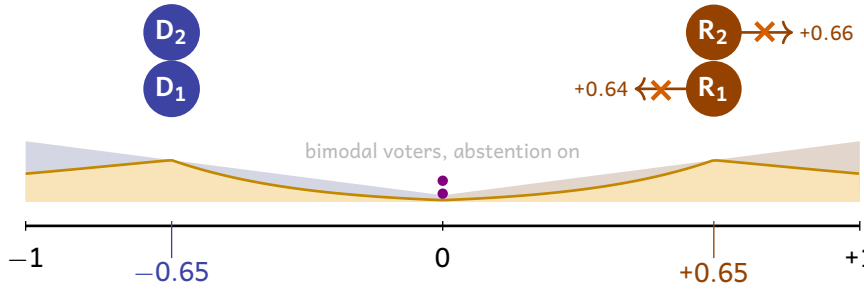
Explanations: Shapes — and every voter line in the paper — are lightly tinted toward the party each side leans to ($\mathcal{D}$ blue, $\mathcal{R}$ red). Panels B and C build the turnout-weighted electorate (orange) over the gray density. **Panel B**, the *general election* between two nominees (−0.7 and +0.3). **Panel C**, the *closed primary* over the four-candidate slate (−0.8, −0.6, +0.2, +1.0). In an *open primary* (right), any voter may back the nearest candidate of *either* party, so the dividers move to the candidate midpoints (−0.7, −0.2, +0.6), the jump at $x = 0$ vanishes, and a voter just left of center now turns out for $\mathcal{R}_1$ (the cross-over channel of Section III.B.1).

Figure 2: The Band's Two Edges on the Candidate Line ($\alpha = 2/3$, $\tau_0 = 1$)

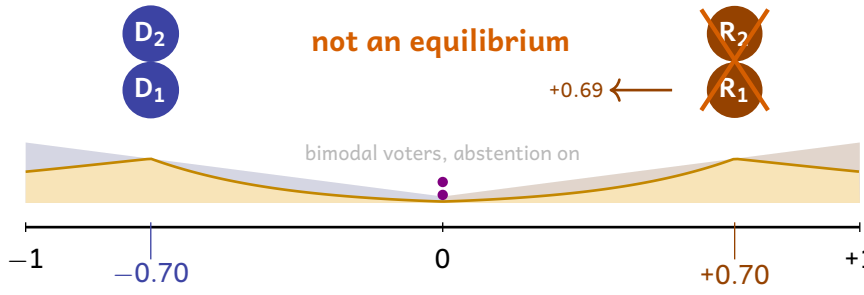
Panel A: $q = 0.50$: just below the lower edge — the outward step wins both



Panel B: $q = 0.65$: inside the band — both deviations fail



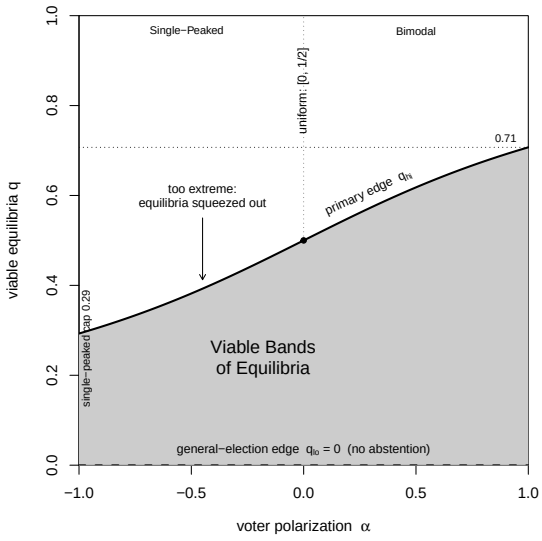
Panel C: $q = 0.70$: above the upper edge — the centrist undercut wins both



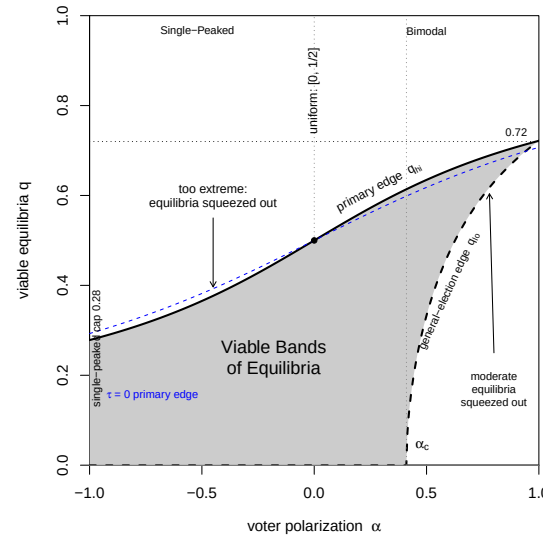
Explanations: This illustrates potential symmetric choices under a bimodal-voter scenario ($\alpha = 2/3$, gray) with turnout decay at the baseline ($\tau_0 = 1$). The orange area is the turnout-weighted electorate. In **Panel A**, an outward deviation wins both the PE and the GE. In **Panel C**, an inward deviation wins both the PE and the GE. Only **Panel B** shows an equilibrium. Not shown, the viable range (for $\alpha = 2/3$, $\tau_0 = 1$) is from 0.536 to 0.669.

Figure 3: The Continuum Equilibrium Band as a Function of Voter Polarization

Panel A: Full participation ($\tau = 0$)



Panel B: Baseline decay ($\tau_0 = 1$)



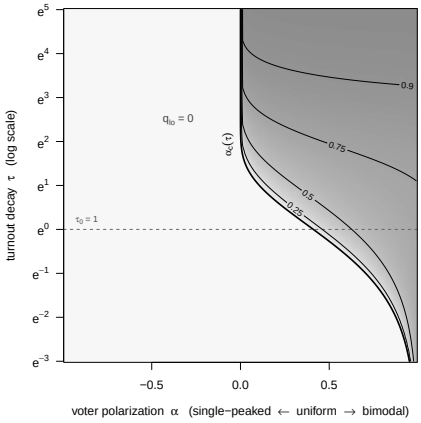
Explanations: In **Panel A**, under full participation, the most moderate candidate equilibria are viable for the entire family of density distributions. More extreme candidate equilibria become progressively more viable for higher voter polarization. In **Panel B**, under turnout decay (more discriminating voters, less likely to turn out for candidates if they are farther from them), moderate equilibria become non-viable if voter polarization is high enough.

Interpretation: Voter turnout has only a modest effect on the primary, but a strong effect on moderate equilibria for very high levels of bimodality.

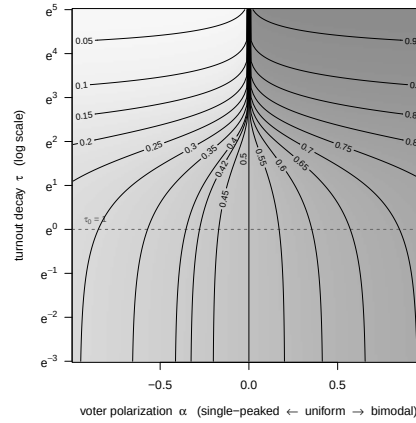
Source: work/plot-cont-band-full-tau0.R, work/plot-cont-band-full.R [noted date: 2026-07-15]

Figure 4: The Band Edges and Their Width over the (α, τ) Plane

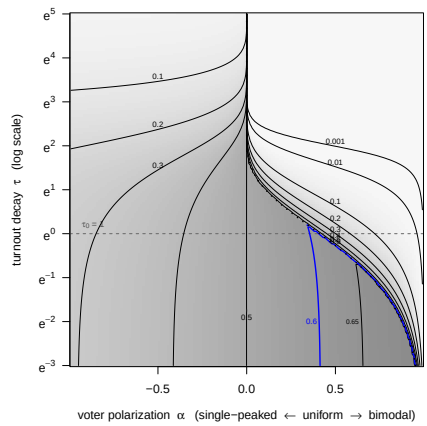
Panel A: Lower edge q_{lo} (general-election floor)



Panel B: Upper edge q_{hi} (primary-election ceiling)



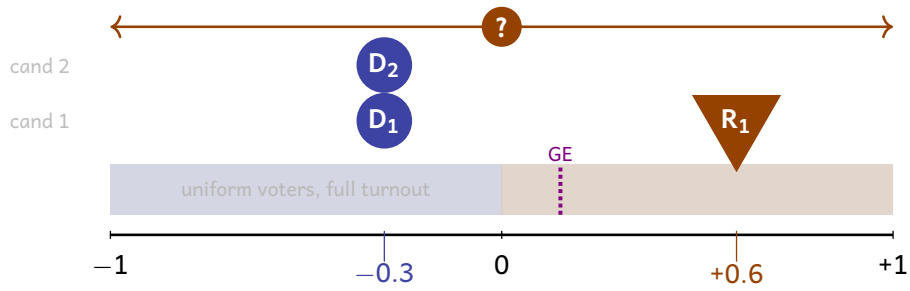
Panel C: Band width $q_{\text{hi}} - q_{\text{lo}}$ (separation of the two edges)



Explanations: This shows the contours of the two band edges over the whole (α, τ) plane, with voter polarization α on the horizontal axis (single-peaked $\alpha < 0$, uniform $\alpha = 0$, bimodal $\alpha > 0$) and turnout decay τ on a logarithmic vertical axis running from e^{-3} to e^{5}. The dashed horizontal line marks the baseline $\tau_0 = 1$. **Panel A** maps the lower general-election floor q_{lo}: it is identically 0 (the flat, unlabelled region) to the left of the centrist threshold $\alpha_c(\tau) = \tau/(2e^{\tau} - 2 - \tau)$ (heavy curve) and lifts off to the right of it. **Panel B** maps the primary ceiling q_{hi}, the turnout-weighted half-electorate median: the $q_{\text{hi}} = 1/2$ contour is exactly $\alpha = 0$ (the uniform cap, invariant in τ), and the edge rises toward the extremes with α. As τ grows, the ceiling is pushed farther out for bimodal voters ($\alpha > 0$) and back toward the center only for single-peaked ones ($\alpha < 0$). **Panel C** maps the band width $q_{\text{hi}} - q_{\text{lo}}$. The maximum band is about 0.70 in width, at $\alpha \approx 0.95$ with very little decay — not at $\alpha = 1$, where the band pinches to a point for every $\tau > 0$.

Source: work/plot-cont-edges-contour.R [noted date: 2026-07-15]

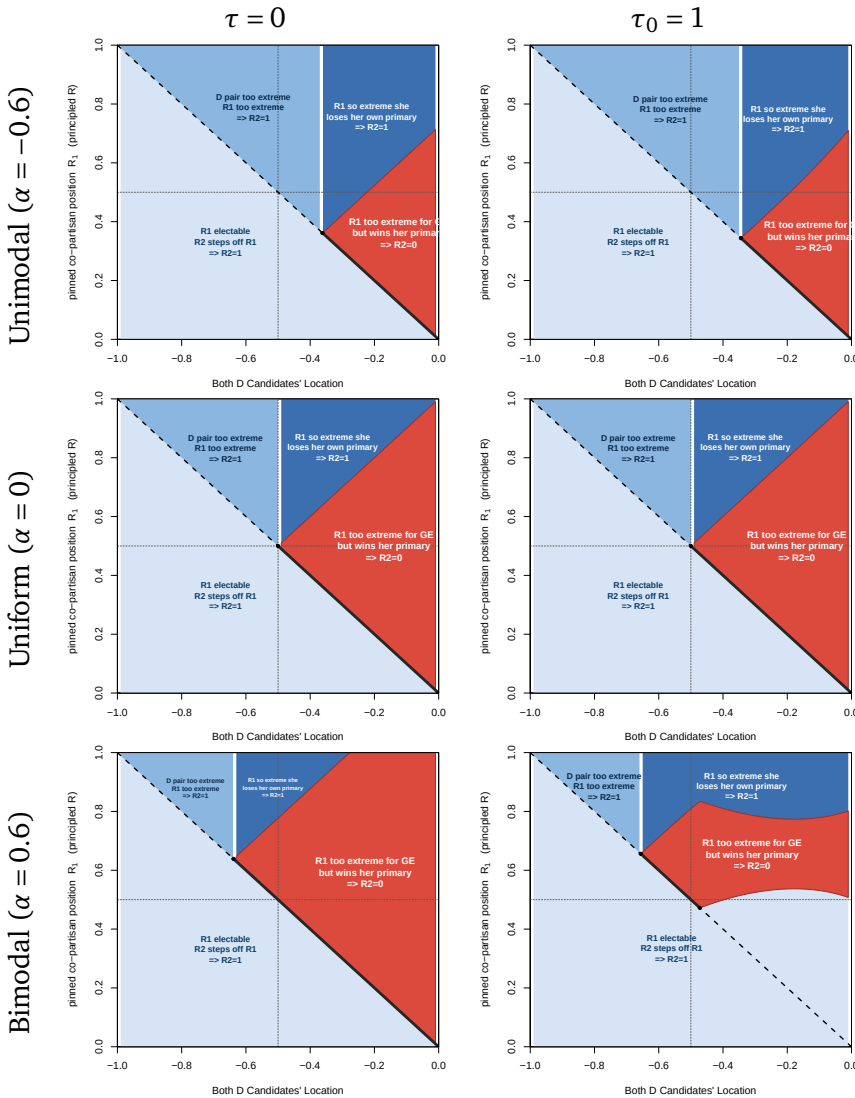
Figure 5: The Principled-Candidate Curse (Uniform $\alpha = 0$, No Decay $\tau = 0$)



Explanations: Uniform voters with full turnout ($\alpha = 0$, $\tau = 0$). The $\mathcal{D}$ pair sits at -0.3. One principled $\mathcal{R}$ candidate sits at $+0.6$ and is indifferent to winning. The remaining $\mathcal{R}$ candidate ("?") has no viable position. To win her primary she must stand in $(+0.4, +0.6)$. To win the general election she must stand in $(-0.3, +0.3)$. The two windows are disjoint. See also Theorem W3.2.

Interpretation: One candidate fixed at an extreme position can make the general election unwinnable for her whole party. This does not rely on bimodality or turnout decay.

Figure 6: The Principled-Candidate Curse $\mathcal{R}_1$: Winnable and Unwinnable Regions



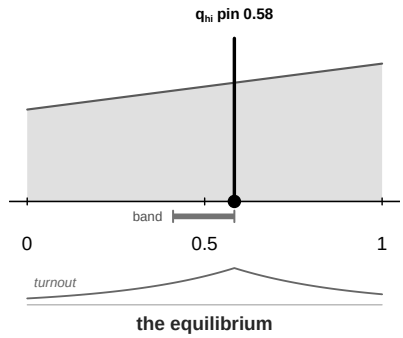
Explanations: Both $\mathcal{D}$ candidates are co-located at $d \leq 0$ (on the horizontal axis). The principled co-partisan $\mathcal{R}_1$ is exogenously pinned at $r_1 \geq 0$ (on the vertical axis). The plots show the outcome of the free candidate $\mathcal{R}_2$ under best-response over $[-1, 1]$. Every *blue* region is one where $\mathcal{R}_2$ *can* win ($\mathcal{R}_2 = 1$), and the *red* region is the principled curse, where $\mathcal{R}_2$ has *no* winning position, so $\mathcal{R}_2$ *and* $\mathcal{R}_1$ both lose ($\mathcal{R}_2 = 0$). Three blue shades mark *why* $\mathcal{R}_2$ wins (Appendix W3): (1) *light blue*, where $\mathcal{R}_1$ is moderate enough to be general-election-viable and $\mathcal{R}_2$ steps off her — taking the larger flank of the primary while still clearing the general; (2) *deep blue* above the curse, where $\mathcal{R}_1$ is so extreme that she loses her own primary and $\mathcal{R}_2$ holds the base; and (3) *medium blue* at the upper left, where the $\mathcal{D}$ pair is itself so extreme that a moderate $\mathcal{R}_2$ beats it outright. (The curse's lower edge is the general-election tie (the most extreme $\mathcal{R}_2$ that still beats $\mathcal{D}$ at d) and its upper edge is the primary tie.) The dashed line is $r_1 = -d$, the Theorem W3.2 guarantee boundary: at or below it $\mathcal{R}_2$ can always secure at least 1/4, so the curse can occur only above it. The thick segment on it marks the symmetric equilibrium band.

Interpretation: When a party's principled candidate is fixed at a very extreme position, the other candidate can be left with no winning position — but only when that fixed position is more extreme than the opponents' mirror image. With bimodal voters (lower right) the general-election edge is pushed *above* the mirror diagonal, opening the curse strictly inside the guarantee region.

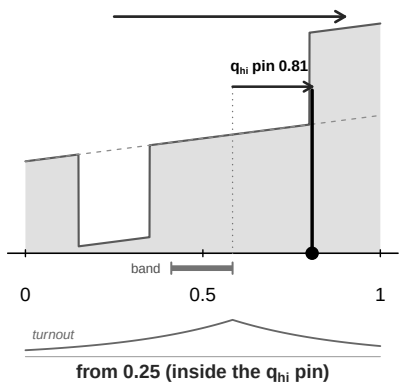
Source: work/plot-cont-xregion-grid.R [noted date: 2026-07-18]

Figure 7: Increasing Polarization Moving Probability Mass

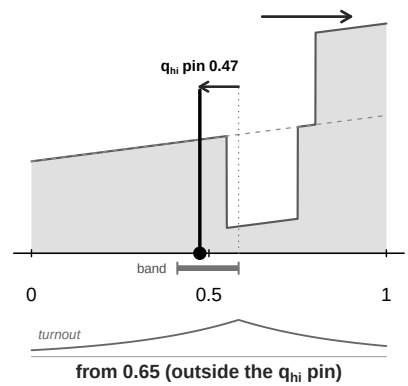
Panel A: The Original Voters and q_{hi}



Panel B: From *inside* the q_{hi} pin



Panel C: From *outside* the q_{hi} pin



Explanations: The electorate is g_α with $\alpha = 0.2$ and the turnout decay is $\tau = 3$, steeper than the $\tau_0 = 1$ baseline so that the inward move is visible. The baseline q_{hi} is 0.584. **Panel B** and **Panel C** show the effects of polarizing an equal mass of $\mathcal{R}$ voters on q_{hi}: the same blob (mass 0.16 on the drawn per-half-mass-one scale, i.e. 8% of the electorate; half-width 0.10) is sent to the same destination 0.90 (in Panel B from 0.25 and in Panel C from 0.65, moving q_{hi} to 0.808 and to 0.474 respectively). The total voter mass on the $\mathcal{R}$ side remains the same. In Panel B, the moving mass is weighted by turnout the same at its origin as at its destination. Thus, the q_{hi} increases. In Panel C, although the mass moves to the right, the effective turnout-weighted mass does not. Thus, the q_{hi} decreases.

Source: work/plot-cont-illus.R [noted date: 2026-07-17]

Table 3: Moderating One Side's Voters: Help or Harm — and Even the Direction — Depends on the Electorate

Panel A: Bimodal baseline ($\alpha_{\mathcal{D}} = \alpha_{\mathcal{R}} = 0.6$): the same flattening forces $\mathcal{D}$ *inward*

	voters (after shift)							
τ	$\alpha_{\mathcal{D}}$	$\alpha_{\mathcal{R}}$	q^*	$q_{\mathcal{D},2}$	$V_{\mathcal{D}}$	$V_{\mathcal{R}}$	$V_{\mathcal{D}} - V_{\mathcal{R}}$	Moderation on $\mathcal{D}$
0.25	0.55	0.60	0.6427	0.6325	0.9461	0.9434	0.0027	helps ($\mathcal{D}$ wins)
0.50	0.55	0.60	0.6470	0.6367	0.8945	0.8930	0.0015	helps ($\mathcal{D}$ wins)
1.00	0.55	0.60	0.6559	0.6456	0.8030	0.8035	−0.0004	hurts ($\mathcal{R}$ wins)
2.00	0.55	0.60	0.6748	0.6645	0.6580	0.6611	−0.0032	hurts ($\mathcal{R}$ wins)
4.00	0.55	0.60	0.7142	0.7048	0.4701	0.4755	−0.0055	hurts ($\mathcal{R}$ wins)

Panel B: Centrist baseline ($\alpha_{\mathcal{D}} = \alpha_{\mathcal{R}} = -0.6$): the same flattening forces $\mathcal{D}$ *outward*

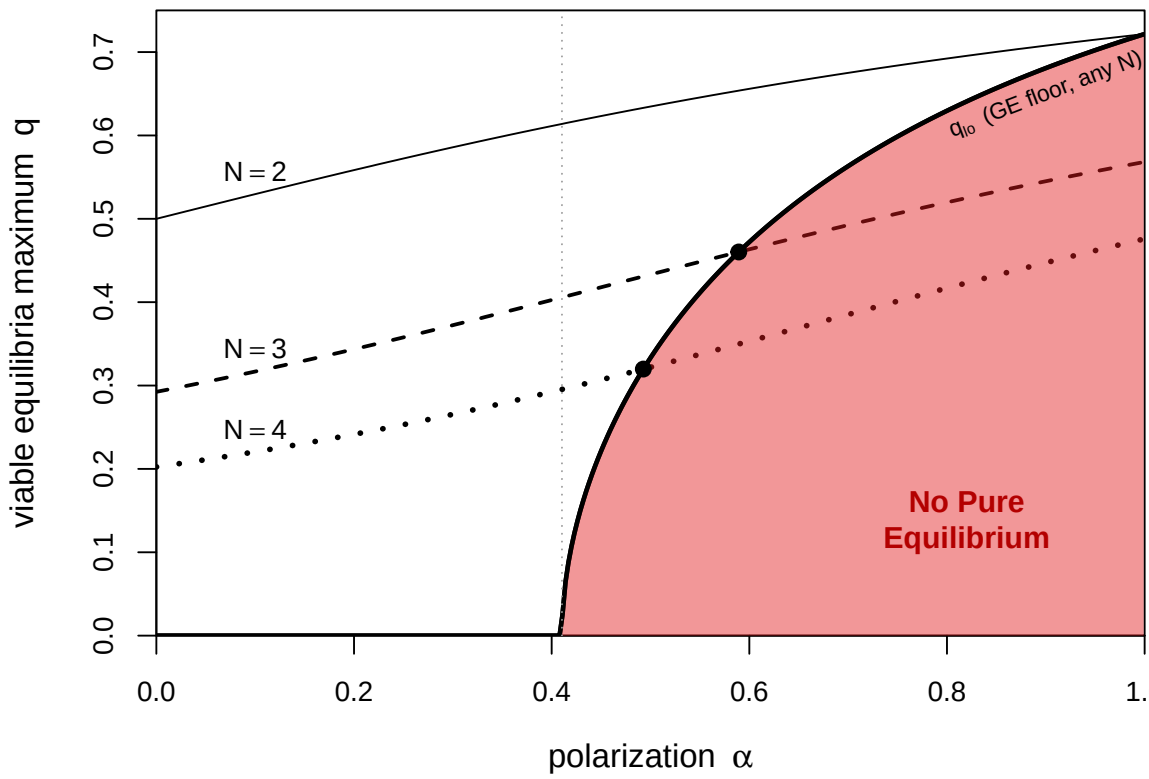
	voters (after shift)							
τ	$\alpha_{\mathcal{D}}$	$\alpha_{\mathcal{R}}$	q^*	$q_{\mathcal{D},2}$	$V_{\mathcal{D}}$	$V_{\mathcal{R}}$	$V_{\mathcal{D}} - V_{\mathcal{R}}$	Effect on $\mathcal{D}$
0.25	−0.55	−0.60	0.3573	0.3675	0.9372	0.9523	−0.0151	pinned outward; $\mathcal{R}$ win
0.50	−0.55	−0.60	0.3530	0.3633	0.8864	0.9011	−0.0147	pinned outward; $\mathcal{R}$ win
1.00	−0.55	−0.60	0.3441	0.3544	0.7963	0.8102	−0.0139	pinned outward; $\mathcal{R}$ win
2.00	−0.55	−0.60	0.3252	0.3355	0.6533	0.6658	−0.0125	pinned outward; $\mathcal{R}$ win
4.00	−0.55	−0.60	0.2858	0.2952	0.4677	0.4779	−0.0102	pinned outward; $\mathcal{R}$ win

Explanations: Both panels start from a symmetric electorate at an exact-tie co-located equilibrium pinned at the ceiling $q^* = q_{\text{hi}}$, then moderate only $\mathcal{D}$'s voters — the same operation of flattening $\mathcal{D}$'s density toward uniform, with that half renormalized to preserve its mass and $\mathcal{R}$ left untouched. $\mathcal{R}$'s candidates stay at q^*, pinned by their own untouched primary. Because the pre-shift baseline is an exact tie in every row, the sign of $V_{\mathcal{D}} - V_{\mathcal{R}}$ (the turnout-weighted general-election votes) is the entire effect of the moderation. (The vote columns use a per-half-mass-one scale, twice that of (1); no further rescaling is applied, because under (1) each half already carries mass 1 at every α.) **Panel A** starts bimodal ($\alpha_{\mathcal{D}} = \alpha_{\mathcal{R}} = 0.6$): flattening lowers $\alpha_{\mathcal{D}}$ (to 0.55), which lowers $\mathcal{D}$'s ceiling, so q^* is no longer viable and $\mathcal{D}$ is forced *inward* to the new ceiling $q_{\mathcal{D},2} = q_{\text{hi}}(\,\alpha_{\mathcal{D}} = 0.55\,)$. **Panel B** starts centrist ($\alpha_{\mathcal{D}} = \alpha_{\mathcal{R}} = -0.6$): the same flattening *raises* $\alpha_{\mathcal{D}}$ (to −0.55), because for a single-peaked density flatter means less concentrated at the center. Since q_{hi} rises with α, $\mathcal{D}$'s old position survives and the primary's outward pull re-pins it *outward* to $q_{\mathcal{D},2} > q^*$. The pivot is the uniform electorate ($\alpha = 0$, $q_{\text{hi}} = 1/2$): nudging toward uniform lowers the ceiling from above and raises it from below. The turnout-decay rate τ is the environment, not the treatment: the identical shift is applied in every row.

Short interpretation: *Moderating one side's voters can push its candidates either toward the center or away from it, depending on whether the electorate started bimodal or centrist — and whether the move helps or hurts is a separate question again.*

Source: work/compute-modbreak.R [noted date: 2026-07-19]

Figure 8: Symmetric Equilibria With More Than Two Candidates ($\tau_0 = 1$)



Explanations: Both edges are drawn against voter polarization α at the baseline turnout decay $\tau_0 = 1$. (A positive τ is required for lack of pure equilibria.) The centrist threshold is $\alpha_c = \tau/(\,2e^{\tau} - 2 - \tau\,) \approx 0.41$. The $N = 2$ is the equivalent of Figure 3 Panel B. A symmetric equilibrium with two candidates per party always exists, even for a fully bimodal distribution ($\alpha = 1.0$). This is not necessarily the case for three or more candidates. For example, there is no pure equilibrium in which candidates can survive both the PE and the GE if there are three candidates and bimodality is greater than 0.59. Under the other assumptions (A1)–(A4), there are no non-symmetric pure equilibria, either.

Interpretation: Under high polarization and positive voter turnout decay, pure equilibria may not exist.

Source: work/plot-cont-nonexist.R [noted date: 2026-06-30]

Appendices

- *Appendix $\mathcal{A}$* — the proofs of the body's two theorems (Theorems 1–2) and the two machinery lemmas they use (GE neutrality, nomination). The band proof is one question — what it would take for a candidate to beat 1/4 — and it never reads the shape of the density or of the turnout rule.
- *Appendix $\mathcal{B}$* — table of primary systems to support Table 2.

The Web Appendix is a separate standalone document. Its sections run in the order in which the material applies above, so a reader following the body front to back meets them in sequence.

- *Appendix W1* — the exponential instance of the parameterized model: closed forms for the two band edges, the lower-edge gradient (Lemma W1.1), the pinch at full bimodality, and the numerical check of the global primary clause of (A4) outside the exponential family.
- *Appendix W2* — the endogenous trap (Proposition W2.1): where a party's own electorate places the q_{hi} pin, when the trap is fatal and when it is harmless, with the too-thin-anchor example.
- *Appendix W3* — the principled-candidate curse and the unwinnable region: the cursed pins form an interval (Theorem W3.1), and a co-partisan who moderates hurts only herself (Theorem W3.2).
- *Appendix W4* — open primaries: the double-cross geometry (Proposition W4.1) and the cross-over collapse lemmas.
- *Appendix W5* — no sustainable cross-over (Proposition W5.1): a shut-out party's crossed pair is never both an equilibrium and a general-election win.
- *Appendix W6* — the participation proof: under the raw turnout kernel the more loyal base carries the general election.
- *Appendix W7* — nonexistence with more than two candidates per party (Proposition W7.1), where the primary ceiling falls below the general-election floor.
- *Appendix W8* — valence: the institutional primary-dividend reframe (ν_{crit}, its closed form, and its two comparative statics). A shared dividend cancels (Lemma W8.1), a one-sided one is a tie-break.
- *Appendix W9* — more than two parties: why the two-nominee general election is load-bearing, and what breaks under plurality.
- *Appendix W10* — one conjecture for each of the neglected forces listed in Section III.D.3, in the same order, each flagged as conjecture rather than result.

$\mathcal{A}$ Proofs

The body defers two proofs, and they are given here together, because they share one argument. The assumptions (A1)–(A4) are those of Section II.C, and nothing below reads the shape of the density g or of the participation weight P: the exponential instance of Section II is worked out in Appendix W1, not used here.

The regularity deferred from (A4) has two clauses, both restrictions on the *shape* of the contest rather than on the functional form of g or P. *Single-crossing along the mirror family:* along $(-q, -q, +q, +q)$ the primary margin and the first-order outward general-election margin are each single-crossing in q — the clause that makes the band a single interval. *The global primary comparison:* in a closed primary, of two co-partisans standing at or inside their half's turnout-weighted median, the more extreme is nominated. The confinement to candidates at or inside the median is deliberate. The clause is not the wider statement "the candidate nearer the median wins," which is false — the primary tie locus is the mirror ρ of Appendix W3, which is not distance-symmetric about the median.[26]

Lemma 1 (GE Neutrality) *Under (A1)–(A3), the general election between mirror nominees at $-q$ and $+q$ is a tie, for every $q \in [0,1]$.*

Proof. A mirror-symmetric slate is invariant, as a set, under $x \mapsto -x$, so by (A1)–(A3) the effective electorate $g(x)P$ is even. The substitution $x \mapsto -x$ then carries the $\mathcal{D}$ nominee's vote integral into the $\mathcal{R}$ nominee's. □

Lemma 2 (Nomination) *In any equilibrium in which all four candidates earn positive payoff, both primaries tie and each candidate is nominated with probability $1/2$.*

Proof. A candidate who strictly loses her primary is not nominated and earns 0. A positive payoff rules this out, so neither primary has a strict winner. □

[26] At $\alpha = 2/3$ and $\tau_0 = 1$, $\rho(0.5) = 0.8335$, where reflection about q_{hi} would give 0.8378.

$\mathcal{A}$.1 Theorem 1: What Would It Take to Beat 1/4

A candidate's payoff is the product of two probabilities, that she is nominated and that she then wins the general election. At the symmetric co-located profile $(-q, -q, +q, +q)$ each factor is $1/2$: co-location makes her primary a tie, and Lemma 1 makes the general election one, so her payoff is $1/4$. Both ties are *structural* rather than coincidental — each rests on an exact symmetry, of the two co-partisans to each other and of the two nominees about the center, not on a numerical accident — so neither can be broken by a small joint re-positioning. The entire question is then what it would take for one candidate to do better than $1/4$. Since each factor is at most one and both already stand at $1/2$, she must raise one factor without giving back more on the other, and because she can only move her own position, she has two things to try, inward and outward. The two boundaries of Theorem 1 are precisely the conditions under which each attempt fails.

Inward — buying the general election at the price of the primary. A step toward the center raises the second factor. Her position is now nearer the center than her mirror opponent's, so the tie of Lemma 1 breaks her party's way and she would win the general election outright. But it costs her the first factor entirely. By the global primary clause of (A4), of two co-partisans standing at or inside their half's turnout-weighted median the more extreme is nominated: her co-partisan has stayed at q and she now stands inside it, so she is not nominated at all, and the product is zero. That argument needs both candidates to lie at or inside the turnout-weighted median — which is what $q \le q_{\text{hi}}$ says. If instead $q > q_{\text{hi}}$, an inward undercut captures more than half of the party's primary electorate: the deviator is nominated with probability one and then wins the general election, for a payoff of one, and the profile is no equilibrium. The primary boundary q_{hi} is therefore the largest q at which the inward attempt still fails — clause (i).

Outward — buying the primary at the price of the general election. A step away from the center reverses the trade. Her co-partisan is now the more moderate of the two, so by the same clause of (A4) the deviator is nominated outright and the first factor rises to one. The second factor now carries the whole payoff, and it is no longer $1/2$, because she has moved off the mirror where Lemma 1 guaranteed the tie. Whether she still wins turns on one sign: that of the party's general-election margin as its nominee moves outward from q. Where that first-order margin is positive the deviation wins outright, for a payoff of one, and the profile is no equilibrium. Where

it is nonpositive she loses the general election and earns zero.[27] The general-election boundary q_{lo} is the smallest q at which that margin has turned nonpositive — clause (ii).

Why the two marginal attempts exhaust the deviations. Single-peakedness in (A4) says each candidate's turnout-weighted vote is single-peaked in her own position, the others fixed. A candidate therefore has no distant second peak to jump to: within each direction her prospects improve or worsen without reversal, so the best deviation in a direction is signed by the marginal one, and the two comparisons above cover every move available to her. This is the sole role of single-peakedness, and it is why the assumption is needed once turnout may read the whole slate. For contiguity — clause (iii) — the single-crossing clause of (A4) has each of the two margins change sign at most once along the mirror family. The positions surviving the inward test are therefore an interval $[\,0, q_{\text{hi}}\,]$ and those surviving the outward test an interval $[\,q_{\text{lo}}, 1\,]$, so the equilibrium set is their intersection $[\,q_{\text{lo}}, q_{\text{hi}}\,]$, which never splits.

Moderation is blocked, not unwanted — clause (iv). Read the outward comparison in reverse. At any $q \geq q_{\text{lo}}$ no outward move raises the party's general-election margin, so along that direction the mirror already maximizes it: the margin weakly falls as the nominee moves out, and therefore weakly rises as it moves in — strictly, above q_{lo}. The party would thus win the general election by moving *both* of its candidates toward the center, and it can do so without disturbing either tie, since the two stay co-located. No candidate can make that move alone, because alone it is the inward deviation, which forfeits her primary. The argument reads only the mirror tie of Lemma 1.

$\mathcal{A}$.2 Theorem 2: No Other Equilibria, and the Payoffs

No candidate is shut out. In any equilibrium every candidate earns a positive payoff. One who would earn zero has two escapes: co-locating with her co-partisan ties her primary, and, if her party would then still lose the general election, moving to the mirror of the opposing nominee ties that (Lemma 1). Under a symmetric electorate one of the two secures her a positive share. Hence, by Lemma 2, both primaries tie.

Shape. Under the closed primary a party's electorate is the fixed half-line on its own side, so by (A4) its two candidates tie it in only two ways: by co-locating, which is a structural tie robust to small joint moves, or by standing at two distinct positions where the vote totals happen to

[27] At q strictly above q_{lo} the margin is strictly negative and the outward step loses. At q_{lo} the first-order margin vanishes, and the single-crossing clause of (A4) — which has the margin passing from positive to negative there — makes it negative just beyond, so the step loses at the boundary too. The band is thus a closed interval.

be equal. The latter is an isolated coincidence — it forces a second vote-total equality to hold simultaneously and independently — and is excluded as a knife-edge. So each party co-locates, giving (d,d,r,r). The general election must tie as well, since a strict loser would earn zero, already ruled out. By Lemma 1 the mirror $r=-d$ ties structurally, while a non-mirror co-located tie is again an isolated equality, excluded on the same ground. Hence $r=-d=:q\geq 0$: every equilibrium is a mirror-symmetric co-located profile of Theorem 1, and none is asymmetric.

Payoffs. Both primaries tie and, by Lemma 1, so does the general election. Each candidate is therefore nominated with probability $1/2$ and then wins with probability $1/2$, for a payoff of $1/4$.

A nomination prize. Suppose finally that a candidate who wins her primary collects a prize $\epsilon\geq 0$ even when she loses the general election, office still worth one, so that her payoff is $\Pr[\text{nominated}]\cdot\big(\Pr[\text{wins the general}]+\epsilon\,\Pr[\text{loses it}]\big)$. At a symmetric co-located profile she is nominated with probability $1/2$ and the general election ties, so her payoff is $\frac{1}{2}\big(\frac{1}{2}+\frac{1}{2}\epsilon\big)=\frac{1}{4}+\frac{1}{4}\epsilon$: the prize enters symmetrically and the office probability is still $1/4$, so Theorem 2 stands. The two boundaries do not move either, since the marginal comparisons that cut them out are unchanged — an inward step still forfeits the primary, for a payoff of 0. Only the outward step is newly tempting: it wins the primary outright but loses the tied general election, for a payoff of ϵ, which beats co-location's $\frac{1}{4}+\frac{1}{4}\epsilon$ when $\epsilon>1/3$. Hence for every $\epsilon\leq 1/3$ the band of Theorem 1 and the payoffs of Theorem 2 are as stated, while for $\epsilon>1/3$ the co-located profiles cease to be equilibria, each candidate preferring the certain nomination to the shared tie.

Whether the global primary clause of (A4) follows from the remaining assumptions is open. It is verified well outside the exponential instance in Appendix W1.

$\mathcal{B}$ Primary Systems

Appendix Table $\mathcal{B}$.1: U.S. States by Primary Type (Congressional and Presidential)

State	Congress.	Presid.	State	Congress.	Presid.
AL	Open	Open	MT	Open	Open
AK[a]	Open	Open	NE	SemiCl	SemiCl
AZ	SemiCl	Closed	NV	Closed	Closed
AR	Open	Open	NH	SemiCl	SemiCl
CA	Top2	SemiCl	NJ	Closed	Closed
CO[b]	Closed	Closed	NM[b]	Closed	Closed
CT	Closed	Closed	NY	Closed	Closed
DE	Closed	Closed	NC	SemiCl	SemiCl
DC	Closed	—	ND	Open	Op(R)/Cl(D)
FL	Closed	Closed	OH	Open	Open
GA	Open	Open	OK	Cl(R)/Op(D)	Cl(R)/Op(D)
HI	Open	Op(R)/Cl(D)	OR	Closed	Closed
ID	Op(D)/Cl(R)	Op(D)/Cl(R)	PA	Closed	Closed
IL	SemiCl	SemiCl	RI	SemiCl	SemiCl
IN	Open	Open	SC	Open	Open
IA	Closed	Closed	SD	Op(D)/Cl(R)	Op(D)/Cl(R)
KS	Closed	Closed	TN	Open	Open
KY	Closed	Closed	TX	Open	Open
LA[c]	Top2	Closed	UT	Cl(R)/SC(D)	Cl(R)/SC(D)
ME[b]	Closed	Closed	VT	Open	Open
MD	Closed	Closed	VA	Open	Open
MA	SemiCl	SemiCl	WA	Top2	SemiCl
MI	Open	Open	WV	SemiCl	SemiCl
MN	Open	Open	WI	Open	Open
MS	Open	Open	WY	Closed	Closed
MO	Open	Open			

Explanations: Classification from FairVote (as of 2015), the one source giving both a congressional/state and a presidential column per state. "SemiCl" = semi-closed; "Top2" = top-two; "Op", "Cl", "SC" with (R)/(D) denote a split where the two parties run different systems (OK, e.g., is closed for Republicans and open for Democrats). Three states have since changed: [a]Alaska adopted a top-four primary with a ranked-choice general (Measure 2, 2020); [b]Colorado (Prop. 108, 2016), Maine (2021), and New Mexico (S.B. 16, 2025, effective June 2026) opened their primaries to unaffiliated voters. [c]Louisiana, long a majority "jungle" general, uses a semi-closed primary for congressional and several state offices from 2026. Aggregate 2026 counts (NCSL, Ballotpedia, MultiState) are broadly consistent — roughly 14–18 open, 13 closed, 10–12 semi-closed, 5 top-two or top-four — but the sources disagree on individual borderline states and on where the semi-open/semi-closed line falls, and runoff (most Southern states) and caucus selection are orthogonal layers not shown here. CA, WA, AK, and LA are not covered by the model.