MUTUAL PARTY EXTREMISM

Ivo Welch

Working Paper 34967
http://www.nber.org/papers/w34967

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2026

Mutual Party Extremism
Ivo Welch
NBER Working Paper No. 34967
March 2026
JEL No. D71, D72

ABSTRACT

With four political candidates competing first in two primaries and then in a general election, even a modestly polarized electorate can sustain (in equilibrium) much more extremist candidates. However, a party can sustain extremism only if the other side is extreme, too. A small moderation of one side's voting electorate can trigger a discontinuous collapse of candidate extremism on both sides — a "moderation export" effect. The converse is also true: minute increases in voter polarization on the more moderate side can trigger radical candidate extremism on both sides. Principled candidates can destroy party electability. Distance-related voter abstention favors extremism.

Ivo Welch
University of California, Los Angeles
and NBER
ivo.welch@anderson.ucla.edu

It has long been recognized that office-seeking candidates gain an advantage by converging to the median voter (Hotelling, 1929; Downs, 1957). It has also long been recognized that partisan primaries can pull candidates away from the centrist positions that could serve them better in the general election (Key, 1949; Owen and Grofman, 2006; Polsby, 1983). Yet, American candidates — perhaps unusually strongly so more recently — seem to have polarized even more than their electorates. This raises questions about why this has been happening and what underlying aspects could be favoring persistent simultaneous party and candidate extremism.

The paper here develops a spatial voting model in which two candidates from each party choose positions on a left-right line, compete in (within-party) primaries, and then face off in a general election. Voters are not strategic: they simply support the nearest candidate. The model has only two parameters. The first parameter α measures voters' polarization ($\alpha < 0$: single-modal; $\alpha = 0$: uniform; $\alpha > 0$: bimodal). The second parameter measures the inclination of voters to go to the polls. A voter at distance d from a candidate turns out with probability $(1-\tau)^d$, where $\tau \in (0,1)$ measures turnout decay. Less aligned distant candidates can therefore lose some voter turnout.

In such a model, with some voter polarization (bimodality) and some turnout decay, more extreme candidates can sustain general election viability even against more moderate candidates: their concentrated base turns out at higher rates, compensating for the loss of more distant moderate voters. Furthermore, in the model, party and candidate extremism on one side *requires* party and candidate extremism on the other. The mechanism has three links:

1. An extreme equilibrium requires that the extremist is viable in the general election.
2. Viability is sustained by the primary "trap" (moderating loses the primary to the extreme co-partisan).
3. The trap binds only when the current payoff is positive, which requires the other side to be extreme enough to keep the general election competitive.

If three candidates are moderate, the fourth candidate has no choice but to moderate, too. This candidate may win the primary by being more extreme, but would lose the general election. This maintains the moderate equilibrium. Conversely, if three candidates are extreme, the fourth candidate has no choice but to be extreme, too. This candidate would never survive the primary. This maintains the extremist equilibrium.

In short, the primary trap can amplify modest voter polarization into a set of extreme candidate positions, and the mutual extremism across parties can lock these positions into place. The amplification can be dramatic: in a seven-position example with modest turnout decay ($\tau = 0.25$), even $\alpha \approx 0.001$ — barely perceptible bimodality — sustains the semi-extreme equilibrium $(-2, -2, +2, +2)$.[1]

The model begins by developing the necessary conditions for multiple equilibria when voters' locations are symmetric. When the electorate is at least somewhat polarized and voters are less likely to turn out for distant candidates, a band of (Nash) equilibria coexist simultaneously, some centrist, others extreme. The band is contiguous: if centrist and extreme positions are both equilibria, then every position in between is also an equilibrium. Therefore, the same voters with the same polarization level can sustain politicians with different degrees of *mutual* extremism. Furthermore, it is also the case that the candidates can be far more extreme than the voters *in equilibrium*. This effect becomes stronger when voters are more inclined to abstain, even though centrist and polarized voters have the same tendency to abstain when politicians are distant.

The multiplicity of equilibria and the fact that politicians can be more extreme than voters seem consistent with some empirical observations. Congressional polarization, measured by DW-NOMINATE scores (Poole and Rosenthal, 1997), has risen to levels not seen since the Civil War (McCarty et al., 2006). Yet the voters who elect these legislators have not moved nearly as much. Fiorina et al. (2005) document a "disconnect between an unrepresentative political class and the citizenry it purports to represent": the American public remains "nonideological, moving

---

[1] There is only one more extreme equilibrium, but it appears only at almost complete bimodality. Once $\alpha \approx 0.981$, the fully extreme $(-3, -3, +3, +3)$ becomes viable. Here turnout decay is essential: at $\tau = 0$ (full participation), the fully extreme equilibrium is never viable, because a centrist primary challenger captures enough moderate voters to dislodge the extreme incumbent. Positive decay attenuates these moderate primary voters, protecting extreme incumbents from centrist challenges.

rightward on some issues, leftward on others, and not moving much at all on still others," while elected officials, activists, and donors have sorted into sharply opposed camps. Nevertheless, Abramowitz (2010) shows that *engaged* voters — those who actually turn out — have polarized substantially, with the correlation between ideology and party identification rising from 0.47 in 1972 to 0.77 in 2004. But he also concedes that disengaged citizens cluster at the center; the moderate middle exists, it is simply less likely to vote. The model accommodates both views. The voter-polarization parameter (α in the model) can be modest (Fiorina) or moderate among the engaged electorate (Abramowitz), and turnout based on distance can be high or low (τ in the model). Either way, a band of equilibria can emerge — some centrist, some extreme.

The model also makes it possible to explore shifts not only in candidate and party positioning, but also shifts in voter locations. A small flattening of one party's electorate can trigger a discontinuous collapse of all extreme equilibria — a "moderation export" effect. A 1% change in one side's voter distribution can shift the other side's equilibrium positions by several notches and indeed even past the center, with all candidates flocking to the flatter side of the spectrum. The collapse is sharp: it is a discontinuous shift, not a gradual adjustment. The reverse is also possible: a small polarization of voters on the more moderate side can push politicians from the two parties far apart.

The model also shows that a principled candidate (Section C) who insists on an extreme (or sometimes only somewhat extreme) position even when doing so has no chance of winning can make the general election unwinnable for her party — even when a more self-interested politician (and a more moderate) position could. The model's two parameters — voter polarization (α) and turnout decay (τ) — are in principle estimable from data on voter surveys and election turnout rates. Finally, under symmetric voter distributions (and somewhat beyond), Theorem 1 shows that equilibria must take the mirror-symmetric co-located form $(-q, -q, +q, +q)$, so all candidates win with equal probability. This restricted geometry reduces the four-dimensional candidate choice space to a single dimension q, which can be visualized via a two-dimensional "heatmap chessboard."

The remainder of this paper is organized as follows. Section I reviews the related literature and positions the present contribution. Section II sets up the model. Section III develops the full equilibrium analysis using a seven-position example. This is enough to illustrate how candidates on the left or the right can choose more centrist or more extremist positions. With fewer positions, it is easier to understand than the general model — although there is nothing particularly special about seven positions. Section IV analyzes asymmetric voter distributions (best thought of as a scenario in which the voters on one side become just slightly less polarized than those of the other side) and the resulting stronger moderation export effect. Section V generalizes the results to finer choices, such as (near-)continuous distributions. Section VI discusses extensions and limitations. Section VII concludes. All proofs and more technical aspects are collected in the Appendix, making this paper more easily accessible to non-theorists.

I Related Literature

The starting points for spatial models of electoral competition were Hotelling (1929) and Downs (1957). In their now-standard two-candidate models, office-seeking candidates converge to the median voter. Key (1949) established the empirical foundation of looking at states in which primaries took over the role of general elections. The model developed below permits this aspect — when the general election is assured, the primary becomes the binding constraint, and candidates are trapped at extreme positions by within-party competition rather than moderated by cross-party competition.

Since its early days, the literature has exploded. The review here focuses on similar work and notes how the model here is different.

In earlier work, Calvert (1985) and Wittman (1983) showed theoretically that even policy-motivated candidates converge almost fully under uncertainty. The present model does not have such convergence for two reasons: four candidates organized into two primaries (and they cannot reposition between elections), and voter turnout is distance-dependent.

The literature on whether primaries cause polarization remains mixed. Serra (2015) proved a "no polarization despite primaries" result: with Calvert-type convergence assumptions and no turnout decay, adding a nomination stage still yields convergence to the median voter. His model has no distance-based abstention, so bimodal voter distributions cannot generate the disproportionate turnout that sustains extreme positions in the present framework. Empirically, Hirano et al. (2010) and Cintolesi (2022) found little or even negative effects of direct primaries on polarization. The present model is consistent with both: the primary trap — developed formally below — only binds when voter distributions are sufficiently bimodal *and* turnout is sufficiently distance-dependent. Without either ingredient, extreme equilibria cannot survive. (However, moderate non-centrist equilibria can. For example, if voters are uniformly distributed, the equilibrium band spans positions up to roughly half the spectrum.) Under these conditions, the model is consistent with their findings: Primaries would appear to moderate candidates.

Adams and Merrill (2008) study a close institutional setup: office-seeking candidates in a primary-then-general election with abstention due to alienation. Their equilibrium involves moderate divergence, driven by valence screening. The model here has no valence in the main text. (The appendix does add some.) The primary's role is not to reveal quality but to *trap* candidates at extreme positions.

In a related model, Hummel (2013) analyzes primaries followed by a general election when candidates differ in quality. His equilibria involve strategic positioning to signal or screen quality. In the present model, candidates are homogeneous, and divergence arises purely from the spatial coordination failure. Casas and Gross (2021) show that primary candidates can signal privately observed quality through their platform choice, producing either an extremist or a centrist separating equilibrium depending on parameters. Their extremism is informational — a credible signal of valence — whereas the present model's extremism is spatial: candidates are trapped by within-party competition regardless of quality.

Adams et al. (2005) integrate alienation into a unified theoretical framework of party competition, showing that it pushes candidates away from the center — the same force that drives some of the results below. Callander and Wilson (2007) further shows that alienation-based abstention can sustain divergent equilibria in a two-candidate setup. The present model extends this to four candidates: the primary stage creates the within-party coordination failure absent from any two-candidate model.

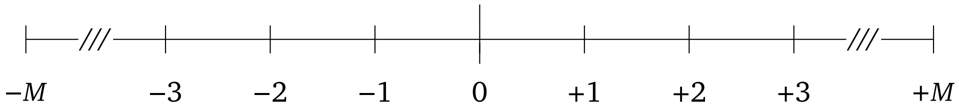
Hirsch (2023) develops another Downsian model in which policy-motivated candidates use both platforms and campaign spending to win the median voter. The unique equilibrium exhibits platform divergence, polarization, and electoral uncertainty — candidates randomize over platforms and spending. Campaign spending enables extreme platforms: candidates spend to "buy" support for positions far from the median. The present model generates polarization through a different channel — primaries and distance-dependent turnout rather than spending — but shares the core finding that office-seeking incentives alone are insufficient to ensure convergence once the Downsian baseline is enriched by institutional structure (primaries, no repositioning) and behavioral features (distance-dependent turnout)—even though candidates' only strategic choice remains their platform position.

Owen and Grofman (2006) study two-stage electoral competition with persistent divergence, but with *parties* as strategic actors. The present framework treats individual *candidates* as the strategic actors, which is what generates the collective action failure: no individual candidate has an incentive to moderate, even though the party collectively would benefit. Snyder and Ting (2011) model electoral selection with parties that screen candidates before the general election. Their primaries serve a gatekeeping function rather than generating spatial competition among candidates. The present model's candidates choose positions simultaneously, and the primary acts as a trap rather than a screen. Castanheira et al. (2010) model intra-party competition as a selection and incentive device: parties choose organizational structures (including primaries) that screen candidates before the general election. Their "hurdle effect" — where the primary adds a competitive barrier — is closest in spirit to the primary trap, but they frame it as beneficial screening rather than as a coordination failure that locks in extremism. Carroll and Nalepa (2020) show more broadly that electoral rules shape intra-party discipline and cohesion, reinforcing the view that within-party competition can constrain candidate positioning.

Empirically, Hall (2015) provided direct causal evidence for the general election penalty of primary extremism — exactly the tension this model formalizes. In a laboratory experiment, Woon (2018) showed that primaries produce divergence even when standard theory predicts convergence. In another experiment, Grosser and Palfrey (2019) found that extreme citizens enter as candidates more often than moderates, producing polarization through an entry channel that complements the present positioning-based mechanism.

The contribution of the present paper is the exploration of a specific mechanism — the *intra-party primary trap* — that sustains *mutually polarized* candidate equilibria as a coordination failure. It requires three ingredients: within-party primaries, no repositioning between stages, and distance-dependent turnout. These ingredients are enough to produce the results. Candidates can be far more polarized than voters. However, this requires configurations where candidates become equally polarized on both sides. That is, parties “export” their moderation or extremism across the aisle. If one party becomes less polarized or, conversely, the more moderate side returns to equal polarization, then the set of viable equilibria can change dramatically. Thus, when left and right voters have differential polarization, it becomes more difficult to sustain extremism. The side with more moderate voters usually always wins, and, if they are allowed to, candidates from the otherwise-losing party can be drawn to very different positions (on the more moderate but opposite side of the voter spectrum) in order to maintain electability.

II The Model



Consider a discrete line of $N = 2 \cdot M + 1$ positions, labeled $-M, -(M-1), \ldots, 0, \ldots, +M-1, +M$, where position 0 is the center. Four candidates simultaneously choose positions on this line. Two are assigned to party $\mathcal{D}$ and two to party $\mathcal{R}$. Party identity is *exogenous* — it is fixed before positions are chosen and does not change regardless of where a candidate locates.[2]

Each party holds a primary: $\mathcal{D}_1$ faces $\mathcal{D}_2$, and $\mathcal{R}_1$ faces $\mathcal{R}_2$. The two primary winners then face each other in a general election (GE) at their primary positions.

There are v_k voters at each position k. In both the primary and the general election, each voter chooses the nearer of the two relevant candidates, splitting 50/50 if equidistant. Turnout depends on distance: a voter at distance d from their preferred candidate turns out with probability

$$f(d) = (1-\tau)^d, \quad \tau \in (0,1). \tag{1}$$

At $\tau = 0$, all voters turn out regardless of distance (the standard Downsian case); as τ increases, distant voters abstain more readily. This captures the empirical regularity that voters are less

---

[2] This is equivalent to assuming a prohibitively high cost of repositioning from primary to general election, as in Adams and Merrill (2008) and Agranov (2016), who showed that when primary positions are publicly observable, repositioning is detected and punished.

likely to participate when they perceive all candidates as ideologically distant (Adams et al., 2006; Callander and Wilson, 2007).

Unlike candidates, voters are not party-bound. Each voter participates in the primary of the nearest candidate. If equidistant between the nearest D and R candidates, the voters' weights are split across primaries. (Thus, politicians can "cannibalize" one another.) Similarly, within a primary, the voter supports the closer of the two same-party candidates, splitting equally when the nearest candidates are equidistant.

Payoffs are binary: winning the general election yields 1; losing (primary or general) yields 0. In expectation, a general election tie yields 0.5; a primary election tie gives each candidate a probability 0.5 of being the nominee. All four candidates seek to maximize their individual expected payoff.

All four candidates observe the positions that the other candidates have chosen. The solution concept is pure-strategy Nash equilibrium.[3,4]

The voter distribution is fixed and described by a one-parameter family indexed by $\alpha \in [-1,1]$:

$$v_k(\alpha) \propto (1-\alpha)\cdot 100 + \alpha \cdot b_k, \qquad \text{where } b_k = 10 \cdot |k|. \tag{2}$$

The proportionality means that only relative weights matter: the normalization cancels in all vote-share ratios, so the specific coefficients 100 and 10 serve only to fix a convenient ratio between the uniform and bimodal components.[5] At $\alpha = 0$, every position has equal weight (uniform distribution). Positive α creates a *bimodal* distribution: the extremes gain weight and the center loses it, with $\alpha = 1$ putting zero mass at the center. Negative α creates a *unimodal* distribution: the center gains weight and the extremes lose it, with $\alpha = -1$ giving the maximally

---

[3] Mixed-strategy Nash equilibria seem quite unattractive. A candidate who randomized over, say, "I might run as a moderate or as an extremist" would lose credibility — voters reward commitment, and the primary electorate can directly observe the chosen position.

[4] At some parameter values (the degeneracy–replacement gap, Appendix $\mathcal{A}$.3), no pure-strategy NE exists. The main results on equilibrium bands and discontinuous shifts are stated for the parameter regions where pure NE exist. The degeneracy-replacement gap where only degenerate equilibria obtain is characterized in Appendix $\mathcal{A}$.3 (Lemmas 4-5).

[5] The qualitative results — the Nash equilibrium (NEQ) band, its contiguity, equilibrium multiplicity, the primary trap — hold under alternative bimodal shapes (e.g., $b_k = 10 \cdot |k|^2$), although the boundary formulas change.

center-peaked member of the family. In sum, the parameter α controls the degree of *voter polarization*: positive for bimodal, negative for unimodal, and zero for uniform.

[Insert Table 1 here: **Glossary of notation.**]

Table 1 collects the notation used in the main text.

III Equilibrium Analysis With Seven Voter Positions

The baseline analysis uses $N = 7$ positions (thus labeled -3 to $+3$), with voter decay of $\tau = 0.25$. Although the results generalize to arbitrarily many positions (Section V), the reader can obtain all the intuition from this example. At $\alpha = 1$, the bimodal distribution is $v = (30, 20, 10, 0, 10, 20, 30)$, totaling 120 voters. At $\alpha = 0$, every position has equal weight. The baseline uses $\tau = 0.25$ throughout this section.

A Best Responses

A.1 Uniform Voters

[Insert Figure 1 here: **Optimal-Response Diagram for Uniform Voter Distribution (**$\alpha = 0$**), Modest Voter Decay (**$\tau = 0.25$**), and** $N = 7$]

The four candidates' simultaneous choices create a four-dimensional strategy space, a 4-tuple $(\mathcal{D}_1, \mathcal{D}_2, \mathcal{R}_1, \mathcal{R}_2)$. To make the game intuitively understandable, we use a "best-response figure."

Figure 1 introduces this figure with $\alpha = 0$ (uniform voters) as a baseline. It contains three sub-panels. The top left subpanel shows the voter distribution — here, seven positions with equal numbers of voters.

The left bottom (later middle) subpanel shows the "GE only" ("general election only") benchmark: it shows where a single $\mathcal{D}$ (or coordinated Democratic party that selects the most viable candidate, rather than the candidate most capable to win a primary) would locate against a known $\mathcal{R}$ opponent at a position 0 to +3. Again, this obtains only *if* there were no $\mathcal{D}$ primary. In this $N = 7$ case, the best choice is center (0) if the $\mathcal{R}$ candidate chooses 0 or 1. If $\mathcal{R}$ chooses 0, it's a tie. If $\mathcal{R}$ chooses $+1$, the $\mathcal{D}$ candidate wins. A "g" below means a win in the general election with 50% probability, a "G" means a 100% win. If $\mathcal{R}$ chooses $+2$, $\mathcal{D}$ can choose any position from −1 to +1 and win 100% of the time ("G"). If $\mathcal{R}$ chooses $+3$, $\mathcal{D}$ can choose any position from −2 to +2 and win 100% of the time ("G").

The right subpanel makes it possible to consider both the primary and the general election. Unfortunately, visualization becomes more difficult when there are three other candidates involved. Fortunately, there is a simplification possibility. The "chessboard" heatmap on the right (which will be below the other two subpanels in subsequent figures) describes, *given* the other three candidates' positions, where a single $\mathcal{D}$ candidate (w.l.o.g., the first candidate in the 4-tuple) should locate. The answer depends on the $\mathcal{D}_2$ opponent's position (y-axis) and the $\mathcal{R}$ opponents positions (x-axis) — which may not necessarily be known yet because the $\mathcal{R}$ primary has not yet decided the winning candidate; however, it is deducible from the four positions in our deterministic model. The in-square text and cell colors show the best-response position; red means extreme (−3), blue means centrist (0). Where multiple positions are equally good, cells are striped in multiple colors.

For example, consider cell $(\mathcal{R}_1,\mathcal{D}_2) = (+1,0)$. The +1 could come about, e.g., if $\mathcal{R}_1$ and $\mathcal{R}_2$ both sit at +1 (guaranteeing the same location also in the GE), or if only the $\mathcal{R}$ primary winner sits at +1 and the loser chose an unelectable position. In the same cell, the same-party opponent $\mathcal{D}_2$ has chosen 0. The chessboard best response for our $\mathcal{D}_1$ is either −1 or 0, each yielding an expected payoff of 1/2.[6] There are two paths to this payoff. The first is to position at −1. $\mathcal{D}_1$ then captures all of the voters from −3 to −1; $\mathcal{D}_2$ wins only the voters at 0.[7] $\mathcal{D}_1$ thus wins the primary, and in the general election it is a toss-up: −1 and +1 are mirror-symmetric about the center, so with uniform voters the turnout-weighted totals are identical. The compact label is

---

[6] This completes the compact code description: P = wins primary, p = primary tie, G = wins GE, g = GE tie.

[7] With $\tau = 0.25$, $\mathcal{D}_1$ at −1 attracts voters at −3, −2, and −1 with turnout rates 0.75^2, 0.75^1, and 0.75^0, for a turnout-weighted total of about 2.31 units. $\mathcal{D}_2$ at 0 captures only position 0 (1.0 unit). Voters at +1 through +3 enter the $\mathcal{R}$ primary.

"Pg" (primary win, GE tie; $1 \times 1/2$). The second path is for $\mathcal{D}_1$ to choose 0, tying $\mathcal{D}_2$ in the primary. The $\mathcal{D}$ winner then beats $\mathcal{R}$ (who is at +1) in the general election. The compact label is "pG" (primary tie, GE win; $1/2 \times 1$).

Some of the full model's insights already appear in this simple response figure. It takes a more extremist $\mathcal{R}$ candidate to make it worthwhile for our $\mathcal{D}_1$ candidate to choose a more extremist position, too. If the $\mathcal{R}$ candidate is moderate, being too extreme leads to a general election loss for a non-moderate $\mathcal{D}$. However, this is not enough insight. It is not yet clear what the sustainable equilibrium outcomes are. Not all cells here are equilibria.

Finally, the figure also has cells with bold borders. These highlight the diagonal cells where the best response coincides with the partner's position, and the outcomes are "pg." They will play a special role, explained below.

A.2 Bimodal Voters

Figure 2 shows how the best-response landscape transforms as voter bimodality increases through four levels spanning some key transitions. Recall that we are still holding voter decay at $\tau = 0.25$ in our illustrations.

[Insert Figure 2 here: **Optimal Response Diagrams at Four Voter Polarization Levels, Modest Voter Decay ($\tau = 0.25$), and $N = 7$**]

The "GE only" subpanel (between the voter distribution and the lower "chessboard" subpanels) reveals the standard Hotelling-Downs result. At low α (uniform voters), more centrist positions win the GE against more extremist opponents. As α increases, extreme positions become GE-viable: the bimodal voter peaks can generate more turnout to overcome the centrist advantage. At $\alpha = 1$, the extreme position (−3) beats any other position except for the mirror-symmetric opponent at +3.[8]

---

[8] At $\alpha = 1$, $\upsilon = (30, 20, 10, 0, 10, 20, 30)$. [1] Consider $\mathcal{D}$ at −3 vs. $\mathcal{R}$ at +2: $\mathcal{D}$ earns $30{\cdot}1 + 20{\cdot}0.75 + 10{\cdot}0.75^2 = 50.625$; $\mathcal{R}$ earns $10{\cdot}0.75 + 20{\cdot}1 + 30{\cdot}0.75 = 50$. $\mathcal{D}$ wins because her peak (30 voters at −3) turns out fully, while $\mathcal{R}$'s peak (30 at +3) turns out at only 75% (one step from +2). Against +3, mirror symmetry gives an exact tie.

The chessboard in the bottom subpanel that adds the primary tells a different story. The primary adds a constraint: to benefit from winning the GE, a candidate must first survive the primary against their own co-partisan. This constraint is visible in the off-diagonal cells: at moderate α, $\mathcal{D}_1$ cannot profitably deviate to a more centrist position (-1 or 0) if $\mathcal{D}_2$ sits at -2, because $\mathcal{D}_2$ would then win the primary.[9] The GE-optimal centrist position is blocked by intra-party competition.

It does not require a lot of bimodality among voters to deliver interesting results — but it does require some. Without at least minimal polarization, extremist cannot beat centrist candidates.

B Equilibrium Constraints

The analysis of this simple model becomes even simpler due to the following theorem:

Theorem 1 (Payoff Equality) *Under any symmetric voter distribution ($v_k = v_{-k}$), every Nash equilibrium has the mirror-symmetric form* $(-q, -q, +q, +q)$ *and gives each candidate payoff exactly* $1/4$.

The proof is straightforward. In a symmetric distribution, every candidate faces identical incentives up to relabeling. If two same-party candidates occupied different positions with different profitability, one candidate would win the primary outright. However, the loser can always profitably choose a better location and, in particular, simply relocate to the same place — so candidates within a party would do so, giving profiles of the form (q, q, r, r). If such a profile is not mirror-symmetric ($q + r \neq 0$), one party wins the general election while the other gets a payoff of zero. But a candidate from the losing party can deviate to the mirror of the opponent's position, restoring GE balance and guaranteeing strictly positive payoff (the formal argument is in Appendix $\mathcal{A}$.1, Step 2). Only mirror-symmetric co-located profiles $(-q, -q, +q, +q)$ survive.

---

[2] Consider $\mathcal{D}_1$ at -2 instead of -3. Moving inward costs $30 \times 0.25 = 7.5$ votes at -3 (turnout drops from 1 to 0.75) but gains only $20 \times 0.25 = 5$ at -2 and $10 \times (0.75 - 0.5625) = 1.875$ at -1 (turnout rises from 0.75^2 to 0.75). The net is $-7.5 + 5 + 1.875 = -0.625$: a net loss, because the bimodal peaks outweigh the interior.

[9] For example, at $\alpha = 0.85$ the voter weights are approximately $(41, 32, 24, 15, 24, 32, 41)$. If $\mathcal{D}_2$ sits at -2 and $\mathcal{D}_1$ deviates to -1, the primary electorate spans -3 through 0. $\mathcal{D}_2$ captures the two heaviest left-side positions (-3 and -2) while $\mathcal{D}_1$ gets only the lighter ones (-1 and 0). $\mathcal{D}_2$ wins the primary with about 64% of the turnout-weighted vote.

At these, the general election is tied by symmetry — a property we call *GE Neutrality* — and the primary is tied by co-location, giving each candidate payoff $1/2 \times 1/2 = 1/4$. Degenerate profiles where one party always loses can be weak equilibria at knife-edge parameter values (no improving deviation exists when all payoffs are zero), but these are non-generic. The full proof of Theorem 1 is in Appendix $\mathcal{A}.1$.

GE Neutrality implies that deviations from a symmetric equilibrium have a clean two-sided structure:

- **GEC (General Election Condition):** At the **extreme edge**, if $\mathcal{D}_1$ deviates one step further from center (e.g., from -2 to -3), she may win the primary — capturing the concentrated bimodal peak — but may still lose the general election because the nominee is now farther from the median. With the primary won, the binding constraint is *GE viability*.
- **PEC (Primary Election Condition):** At the **centrist edge**, if $\mathcal{D}_1$ deviates one step toward center (e.g., from -2 to -1), then $\mathcal{D}_1$ would win the GE — the nominee is closer to center. However, $\mathcal{D}_1$ would lose the primary to $\mathcal{D}_2$ still at -2, who captures the more polarized voters. With *GE viability* assured, the binding constraint is *primary survival*.

A symmetric equilibrium exists precisely when *neither* deviation is profitable. At the equilibrium boundaries where equilibria appear and disappear, one constraint is marginal: the extreme deviation is GE-blocked (GEC) and the centrist deviation is primary-blocked (PEC). At intermediate α, well inside the existence region, both deviations may fail at the primary itself.

C Viable Equilibria

It is useful to illustrate these two conditions for $(-2,-2,+2,+2)$ at $\alpha = 0.85$ (and of course still at $\tau = 0.25$). If $\mathcal{D}_1$ deviates to -1, she would win the GE — position -1 is closer to the center — but loses the PEC: $\mathcal{D}_2$ at -2 captures the bimodal left peak and wins the $\mathcal{D}$ primary outright. Conversely, $\mathcal{D}_1$ deviating to -3 also loses the primary: $\mathcal{D}_2$ at -2 still captures the centrist positions (-2, -1, center) with enough turnout to outweigh the single bimodal peak at -3.[10] Neither deviation is profitable, so the equilibrium holds.

Being an equilibrium, at $\alpha = 0.85$, at $(-2,-2,+2,+2)$, each candidate gets payoff 0.25. If $\mathcal{D}_1$ deviates toward center — to position -1 or 0 — $\mathcal{D}_2$ at -2 captures the bimodal left peak and wins the $\mathcal{D}$ primary outright. $\mathcal{D}_1$ is eliminated. Every centrist deviation loses the primary to the extremist partner: the party-optimal centrist position is blocked by intra-party competition. This is the *primary trap*.[11] The trap binds only when the current payoff is positive — which requires the other side to be extreme enough to keep the general election competitive. With the parameters here, if $\mathcal{R}$ were more centrist, then $\mathcal{D}$'s more extreme position would lose the GE regardless, making the trap irrelevant. *Extremism on one side requires extremism on the other.*

[Insert Table 2 here: **Cascading Thresholds for Modest Voter Decay ($\tau = 0.25$) and $N = 7$.**]

Table 2 shows how the set of viable equilibria changes as voter polarization increases, creating a cascading progression from centrist to extreme across the full range $\alpha \in [-1, 1]$.

For all negative α (unimodal and center-peaked voters), regardless of how peaked the distribution is, only the two most centrist equilibria exist: $(0,0,0,0)$ and $(-1,-1,+1,+1)$. The more extreme $(-2,-2,+2,+2)$ fails: a centrist deviation from -2 to -1 ties the primary and wins the general election, giving a profitable deviation. This holds for *all* $\alpha < 0$, regardless of τ.

---

[10] The extreme deviator $\mathcal{D}_1$ at -3 captures only position -3 in the primary. $\mathcal{D}_2$ at -2 captures positions -2, -1, and half of 0, which at $\alpha = 0.85$ gives turnout-weighted total 53.8 vs. 40.5 for the deviator. At very high α ($\gtrsim 0.98$), the concentrated bimodal peak does tip the primary to the extremist, at which point the GEC becomes the binding constraint.

[11] The primary trap is a genuine collective action failure. If both $\mathcal{D}$ candidates could coordinate and move to centrist positions simultaneously, both would be better off. But neither can do so unilaterally — the first mover loses the primary. It is a prisoner's dilemma within each party.

For any $\alpha > 0$ (even marginally bimodal voters), $(-2, -2, +2, +2)$ becomes viable, creating three coexisting equilibria. The extreme equilibrium $(-3, -3, +3, +3)$ still fails: a deviation from -3 to -2 wins both the primary and the general election. At $\alpha^* \approx 0.801$, the centrist equilibrium $(0, 0, 0, 0)$ collapses: a deviation from 0 to -1 wins both the primary and the GE. By $\alpha = 0.85$, only $(-1, -1, +1, +1)$ and $(-2, -2, +2, +2)$ remain. At $\alpha^* \approx 0.874$, the moderate equilibrium $(-1, -1, +1, +1)$ also collapses. Only $(-2, -2, +2, +2)$ survives until $\alpha^* \approx 0.992$. The boundary equilibrium $(-3, -3, +3, +3)$ appears at $\alpha^* \approx 0.981$ — just before $(-2, -2, +2, +2)$ collapses — creating a narrow window $[0.981, 0.992]$ where both coexist. At $\alpha = 1$ (fully bimodal), only the most extreme equilibrium survives. The bold borders in Figures 1 and 2 mark these equilibria directly.

The sequential collapse and sequential appearance of equilibria are general, but they may interleave differently at different N and τ. The same mechanism drives every transition: a candidate who would deviate one step more extreme would now win the primary (the PEC is satisfied), and still win the general election (the GEC fails). A candidate who would deviate one step more moderate would lose the primary (the PEC fails).

At the centrist edge, the binding constraint is always the PEC. At the extreme edge, it is always the GEC. In any case, regardless of N, for positive τ, multiple equilibria are pervasive. In our specific numerical example with $\tau = 0.25$, two equilibria coexist for all $\alpha \leq 0$ (unimodal and uniform voters). Three coexist for $\alpha \in (0, 0.801)$, two for $\alpha \in (0.801, 0.874)$, one for $\alpha \in (0.874, 0.981)$, and two again briefly in $[0.981, 0.992)$. Uniqueness arrives only *after* polarization has already won: the unique-equilibrium region ($\alpha > 0.992$) contains only the extreme equilibrium $(-3, -3, +3, +3)$.

Note that there is something intrinsically different about even the smallest amount of bimodality, compared to any uniform or center-peaked voter distribution. The latter can sustain *only* moderate candidates, regardless of voter turnout. The former, when paired with stronger voter distaste for further distant candidates, can very early on procure situations in which (only) extremist candidates can thrive. This will become even clearer in the case where there are more than $N = 7$ positions.

D Equilibrium Phase Diagram

The cascading thresholds of Table 2 for α depend on τ. Figure 3 shows the full picture: for each symmetric equilibrium, it shows the region of (α, τ) parameter space where the $N = 7$ equilibrium exists.

[Insert Figure 3 here: **Phase Diagram for Equilibria at** $N = 7$]

Reading horizontally at $\tau = 0.25$ recovers Table 2. Reading vertically: lower τ (less decay) expands the existence region for centrist equilibria and shrinks it for extreme equilibria. At high τ (high abstention rates for distant candidates), extreme equilibria dominate: base turnout matters, and extreme positions are sustained by the concentrated bimodal tails.

Each boundary in Figure 3 corresponds to a GEC or PEC threshold. Writing $\lambda \equiv 1 - \tau$, the centrist equilibrium $(0,0,0,0)$ collapses when the extreme deviation $0 \to -1$ passes the GEC — the deviator wins the general election. The GEC tie condition gives:

$$\text{Most Centrist Equilibrium } (0,0,0,0) \qquad -1 \underbrace{\leq}_{\text{PEC}} \alpha \underbrace{\leq}_{\text{GEC}} \frac{10\,\lambda^3}{1+\lambda+\lambda^2+7\,\lambda^3}\,. \tag{3}$$

At the other end, the extreme equilibrium $(-3,-3,+3,+3)$ becomes viable when the centrist deviation $-3 \to -2$ fails the PEC — the partner at -3 retains the primary. The PEC tie condition gives:

$$\text{Most Extreme Equilibrium } (-3,-3,+3,+3)\text{:} \qquad \frac{10\,\lambda\,(1+\lambda)}{1+9\,\lambda+10\,\lambda^2} \underbrace{\leq}_{\text{PEC}} \alpha\,. \tag{4}$$

Here, for $N = 7$, at $\tau = 0.25$ ($\lambda = 0.75$), these two thresholds evaluate to $\alpha^*_{\text{GEC}} \approx 0.801$ and $\alpha^*_{\text{PEC}} \approx 0.981$: the centrist equilibrium collapses first, the extreme equilibrium appears last. Both formulas are positive for all $\lambda \in (0,1)$, confirming that the centrist and extreme boundaries lie in the bimodal range $\alpha > 0$. The intermediate boundaries also have this property: the lower boundary of $(-2, -2, +2, +2)$ is $\alpha = 0^+$ (strictly positive), independent of τ.[12]

IV Asymmetric Voter Distributions: Moderation and Extremism Export for Seven Positions

We already discussed how multiple equilibria are viable, and moderation by one party can block deviations to extremism by the other. However, moderation among voters can also play an interesting role.

The symmetric analysis assumed $v_k = v_{-k}$. This section allows the two sides to have different voter shapes, using separate bimodality parameters α_L (left half) and α_R (right half):

$$v_k = \begin{cases} (1 - \alpha_L) \cdot 100 + \alpha_L \cdot b_k & k = -3, \ldots, -1 \\ (1 - \bar{\alpha}) \cdot 100 + \bar{\alpha} \cdot b_0 & k = 0 \\ (1 - \alpha_R) \cdot 100 + \alpha_R \cdot b_k & k = 1, \ldots, 3 \end{cases} \tag{5}$$

where $\bar{\alpha} = (\alpha_L + \alpha_R)/2$. At $\alpha_L = \alpha_R$, this reduces to the symmetric family. As before, we consider the best response of $\mathcal{D}_1$, but now in response to voter moderation on the right.

A Less Voter Polarization on One Side Can Block Extremism on Both Sides

Throughout this section, we fix $\tau = 0.25$ and analyze $\mathcal{D}$'s equilibrium response when $\mathcal{R}$'s electorate flattens. Consider the extreme case $\alpha_R = 0$ (uniform $\mathcal{R}$ voters) and $\alpha_L = 1$ (fully bimodal $\mathcal{D}$ voters). Every centrist $\mathcal{R}$ position (+1, +2, or center) beats every $\mathcal{D}$ position in the general election. $\mathcal{D}$'s bimodal voters cannot sustain extremism because any $\mathcal{D}$ extremist would lose the GE to the centrist $\mathcal{R}$ nominee. All equilibria have winners at moderate positions.[13]

---

[12] At $\alpha = 0$ (uniform voters), a centrist deviation from -2 to -1 exactly ties the primary but wins the general election, giving expected payoff $0.5 > 0.25$—so the deviation is strictly profitable and $(-2, -2, +2, +2)$ is *not* an equilibrium at $\alpha = 0$. Any positive α tips the primary toward the more extreme partner, blocking the deviation; any negative α tips it toward the centrist deviator, enabling it.

[13] At $\alpha_L = 1$, $\alpha_R = 0$: the left-side voter weights are $(30, 20, 10)$ while each right-side position carries weight 100. Any $\mathcal{D}$ candidate at a negative position loses the GE decisively — e.g., $\mathcal{D}$ at -2 vs. $\mathcal{R}$ at $+2$ gives turnout-weighted

More generally, as α_R decreases (holding $\alpha_L = 1$), the most extreme $\mathcal{D}$ position in any equilibrium drops monotonically. A centrist electorate on one side *exports* its moderation to the other side, forcing both parties toward the center.

B Sudden Onset of Flattening Collapse

The moderation export is not gradual — it is a *discontinuous shift*. Figure 4 shows what happens when one holds the $\mathcal{D}$'s electorate fixed at $\alpha_L = 0.90$ (bimodal) and flattens $\mathcal{R}$'s electorate from $\alpha_R = 0.90$ to 0.89.

[Insert Figure 4 here: **Asymmetric Voter Discrete Phase Transition ($N = 7$)**]

At $\alpha_R = 0.90$ (left panel), the distribution is symmetric and the semi-extreme equilibrium $(-2, -2, +2, +2)$ exists. By GE Neutrality, the general election is tied at 50%. $\mathcal{D}$ candidates are pinned at position −2 by the primary trap.

At $\alpha_R = 0.89$ (right panel), $\mathcal{R}$'s electorate has flattened by 1%. The symmetric equilibrium now collapses. GE Neutrality broke: flattening $\mathcal{R}$'s distribution adds centrist mass to the right side (the gap between uniform and bimodal weights at each position), which generates disproportionate turnout in the GE — these added centrist voters are close to $\mathcal{R}$'s nominee and turn out at high rates, tipping the balance. At the previously viable profile $(-2, -2, +2, +2)$, $\mathcal{R}$ would now win the GE, and both $\mathcal{D}$ candidates would receive payoff 0.

There is only one way out that allows $\mathcal{D}$ candidates to still win. The replacement equilibria are visible above the dashed center line: $(+1, +1, +1, +1)$ (degenerate, all four at +1) and $(+1, +1, +2, +2)$ ($\mathcal{D}$ at +1, $\mathcal{R}$ at +2; $\mathcal{D}$ wins the GE at 100%). $\mathcal{D}$ jumped from position −2 to +1 — a three-position shift that crosses center, onto $\mathcal{R}$'s side of the spectrum — not because $\mathcal{D}$'s voters changed, but because the slightly flatter $\mathcal{R}$ electorate destroyed the primary trap.

---

totals of roughly 64 vs. 264. $\mathcal{D}$ can only compete by positioning on the right side of the spectrum, where the large voter pool provides enough GE support. The surviving equilibria therefore have all candidates at non-negative positions — an extreme case of the moderation export.

To verify $(+1,+1,+2,+2)$: at $\alpha_R = 0.89$, $\mathcal{D}$ at +1 earns 62.0 turnout-weighted GE votes vs. $\mathcal{R}$'s 57.1 at +2, winning the GE at 100%. $\mathcal{D}_1$ deviating to 0 would win the primary but lose the GE (the asymmetric voter distribution tips the balance to $\mathcal{R}$); deviating to +2 loses the primary to $\mathcal{D}_2$, who captures the entire left side plus center. On the $\mathcal{R}$ side: $\mathcal{R}_1$ deviating from +2 to +3 wins the primary (capturing the bimodal peak) but loses the GE to $\mathcal{D}$ at +1 (too far from center); $\mathcal{R}_1$ deviating to +1 ties $\mathcal{D}$ in the GE but loses the $\mathcal{R}$ primary to $\mathcal{R}_2$ at +2 (the $\mathcal{R}$-side peak voters favor +2). Neither deviation is profitable.[14]

If $\mathcal{D}$ candidates were barred from positive positions — either by partisan loyalty or closed-primary rules — the cross-party defection cannot occur, and the degenerate equilibrium persists. We return to this point in the Remark in Section C below.

The change from 0.90 to 0.89 is a *discontinuous shift*: the 1% change in one party's voter shape shifts all equilibria across the center line. The $\mathcal{D}$ electorate is identical across both panels. The shift is driven by the other side's flattening.

C Summary of Causal Chain

The results from both the symmetric and asymmetric analyses establish a three-link causal chain.

Link 1: GE viability is necessary. An extreme equilibrium (q,q,r,r) requires $\mathrm{GE}(q,r) \geq 0.5$: the extremist must be competitive in the general election. For symmetric distributions, this is guaranteed by GE Neutrality. For asymmetric distributions, it is the binding constraint.

Link 2: The primary trap is the mechanism. At any symmetric equilibrium, each candidate gets payoff 0.25. Deviating toward the center loses the primary, dropping the payoff to 0. The trap only binds when the current payoff is positive (Link 1).

---

[14] For $(+1,+1,+1,+1)$: all four candidates co-locate, giving each payoff 1/4. With $\alpha_R < \alpha_L$, $\mathcal{R}$'s flatter electorate generates more centrist turnout near +1, so $\mathcal{R}$ would win any GE against a $\mathcal{D}$ at a different position. No candidate can profitably deviate because any other position either wins the primary but loses the GE (if the deviator moves further right—the right-tail voters enter the deviator's primary, but the resulting nominee is too far from center to beat the opponent at +1 given $\alpha_R < \alpha_L$) or loses the GE outright (if the deviator moves left).

Link 3: Cross-party enabling. The trap on one side requires extremism on the other. If the opposing party were centrist, the extremist would lose the GE and get payoff 0 regardless — making the primary trap irrelevant.

The three-link chain is the general mechanism behind all equilibria in the model. For symmetric distributions, GE Neutrality ensures Link 1 automatically, and the binding constraints are the primary trap (Link 2) and the two-boundary decomposition. For asymmetric distributions, the chain also explains the moderation export:

Remark (Moderation Export). A centrist electorate on one side destroys GE viability for the other side's extremists. This breaks the primary trap on the other side, forcing moderation for *both* parties simultaneously. Polarization is not a coordination game between parties. It is a collective action failure *within* each party, sustained by the primary system and enabled by the opposing party's extremism. Conversely, a modest increase in polarization by the less polarized side can collapse moderate equilibria and leave only extremist equilibria.

Remark (Crossing The Spectrum). The equilibrium-breaking deviation at $\alpha_R = 0.89$ is not a within-party move. Under the model's proximity-based primary participation rule (analogous to an open or blanket primary), a $\mathcal{D}$ candidate who relocates to a positive position is not switching parties—party identity remains exogenous—but still *attracts $\mathcal{R}$-side voters into the $\mathcal{D}$ primary*. The institutional analogue is a system where any voter may participate in any primary (as in California's pre-2010 blanket primary or Louisiana's jungle primary). Under closed primaries—where the primary electorate is fixed by party registration—this cross-party voter attraction cannot occur, and the collapse mechanism is blocked (see the footnote below and Appendix $\mathcal{A}$.3). The scenario can be viewed as a *cross-party ideology defection*: a candidate from the losing side relocates to the winning side's position, and wins both primary and GE. If voters had an inherent loyalty to their own party's candidates — even a small bonus ϵ penalizing other-party candidates — the cross-party defection would be harder. A small ϵ shifts the collapse threshold upward; a large ϵ blocks the collapse entirely, locking in a degenerate equilibrium where the flatter party permanently wins the GE.[15]

V Generalizations to Large N

A A Smoother Spectrum ($N = 1{,}001$) of Voter Positions

The seven-position example captures almost all of the important intuition. However, a natural question is whether the results are artifacts of a coarse grid. Thus, this subsection extends the best-response analysis of Section III.A from $N = 7$ to $N \geq 1{,}001$ positions.

One can approximate a near-continuous voter distribution by increasing N — say to 1,001 — and rescaling the turnout decay so that total decay across the line is held constant. The per-step retention at $N = 7$ is $1 - \tau_7 = 0.75$; spreading the same end-to-end decay over 1,000 steps gives $(1-\tau_{1{,}001}) = 0.75^{6/1000}$, i.e., $\tau_{1{,}001} \approx 0.00173$.

[Insert Figure 5 here: **Optimal Response Diagrams at $N = 1{,}001$**]

Figure 5 shows the best-response landscape at $N = 1{,}001$ for four levels of voter polarization. The same qualitative patterns as for $N = 7$ emerge: centrist positions (bluish) dominate at low α; extreme positions (reddish) become viable at high α; at $\alpha = 1$, only the extreme equilibrium survives. There is also a new feature — a large white area with X, soon to be discussed (Section V.C).

[Insert Figure 6 here: **Phase Diagram For Three Equilibria ($N = 1{,}001$)**]

Because it is more difficult to visualize equilibrium existence diagrams with overlapping equilibria and thousands of choices, Figure 6 shows the parameters that can sustain one of three equilibria:

---

[15] Under closed primaries (party-based voter assignment), cross-party deviation is foreclosed by construction. The flattening collapse cannot occur, and the degenerate zero-payoff equilibrium persists. See Appendix $\mathcal{A}$.3 for a formal treatment.

1. $(0,0,0,0)$ is the most centrist candidate equilibrium. If there is no voter polarization, it exists regardless of turnout decay. Similarly, if there is no turnout decay, it exists regardless of polarization. However, even modest voter polarization can quickly destabilize this centrist equilibrium when there is turnout decay.

2. $(-0.6,-0.6,+0.6,+0.6)$ is a mid-extremist candidate equilibrium. Without turnout decay, it spans (can appear for) a wide range of strong bimodality. With more turnout decay, it spans ever-smaller slices but at lower bimodality.

3. $(-0.8,-0.8,0.8,0.8)$ is an extremist candidate equilibrium. It exists only for a very thin slice of strong turnout decay and bimodality.

More extreme equilibria require higher bimodality or turnout decay. The differently overlapping regions illustrate that the progression of equilibria for $N = 7$ — where all three moderate to mid-extreme equilibria were available early on and fell off one by one, (almost) until the most extreme equilibrium appeared — is not generalizable. Equilibria's appearances and disappearances interleave irregularly.

[Insert Figure 7 here: **NEQ Band Diagram (*N* = 1,001)**]

Figure 7 traces the equilibrium band's two boundaries as α varies continuously from uniform to bimodal, holding $\tau = 0.25$ fixed. (Think of this diagram as a horizontal cross-section of the equilibrium existence diagram in Figure 6, re-expressed with equilibrium position on the vertical axis.) The dashed general election boundary drops as α increases; the solid primary boundary rises. The four vertical markers (a)–(d) correspond to the four panels of Figure 5: at $\alpha = 0$ (uniform), the band spans from center to $s \approx 0.50$ — the most extreme positions are excluded because a centrist primary challenger captures nearly the entire half-line of uniformly distributed voters, winning the primary;[16] by $\alpha = 1$ (full bimodal), only 139 positions survive near the

---

[16] At $\alpha = 0$ the band's extreme boundary is at $s \approx 0.50$, not at $s = 1$. The most extreme position requires $\alpha \geq 0.37$ before bimodality concentrates enough voters at the tails to protect it from a centrist primary challenge. In the true continuous limit ($N \to \infty$), the PE boundary at the extreme edge diverges: a point-mass incumbent cannot hold off a challenger with the full half-line, so the band never quite reaches $s = 1$.

extreme edge. The band squeezes to a narrow sliver at high α, confirming that polarization drives candidates to extreme positions.

The flattening collapse is sharp at $N = 1{,}001$. Figure 8 holds $\alpha_L = 0.90$ and shows three snapshots: $\alpha_R = 0.90$, 0.88, and 0.86. When $\mathcal{R}$ flattens by 2%, $\mathcal{D}$'s equilibrium positions jump from $s \approx -0.35$ (extreme left) past center to $s \approx +0.10$ (right of center) — a discontinuous shift triggered by a small change in the other side's distribution.

[Insert Figure 8 here: **Flattening Collapse** ($N = 1{,}001$)]

B The Equilibrium Band Theorem

This subsection generalizes the equilibrium existence results and equilibrium existence diagram of Sections III.B–III.D from $N = 7$ to arbitrary N. The patterns observed at $N = 7$ and 1,001 hold for all N.

Theorem 2 (Nash Equilibrium (NEQ) Band Contiguity) *For $N = 2{\cdot}M + 1$ positions with symmetric bimodal voter distribution $v_k(\alpha) = (1-\alpha)\cdot 100 + \alpha \cdot 10 \cdot |k|$ and turnout decay $\tau \in (0,1)$:*

*(i) At any (α,τ), the set of $j \in \{0,\dots,M\}$ such that $(-j,-j,j,j)$ is a Nash equilibrium forms a **contiguous interval** $\{j_{\min},\dots,j_{\max}\}$.*

*(ii) Each NEQ $(-j,-j,j,j)$ exists for $\alpha \in [\alpha^-(j,\tau),\ \alpha^+(j,\tau)]$, where $\alpha^+(j)$ is the **GE boundary** (non-decreasing in j) and $\alpha^-(j)$ is the **primary boundary** (non-decreasing in j).*

(iii) Contiguity follows: if $j_1 < j_2$ are both NEQ at some α, then every $j_1 \le j \le j_2$ is also NEQ.

The theorem says that at any level of voter polarization and turnout decay, the equilibria form a *contiguous band* of positions. If both a relatively moderate equilibrium $(-j_1,-j_1,+j_1,+j_1)$ and a more extreme one $(-j_2,-j_2,+j_2,+j_2)$ survive, then so does every equilibrium in between. Polarization is not all-or-nothing: at most parameter values, a range of symmetric equilibria coexist, from relatively moderate to relatively extreme. The political question is which equilibrium from the band obtains.

Each equilibrium position j faces two constraints. The *GE boundary* $\alpha^+(j)$ is the maximum polarization at which position j can survive the general election: beyond it, a more extreme challenger wins the GE. The *primary boundary* $\alpha^-(j)$ is the minimum polarization needed to sustain position j: below it, a centrist challenger wins the within-party primary. Both boundaries are non-decreasing in j — more extreme positions need more bimodal voters to sustain — and this monotonicity is what delivers contiguity.[17]

The boundaries have closed-form expressions (derived in Appendix $\mathcal{A}$.2, equations (6)–(7)). A notable feature is that the GE boundary's numerator depends only on turnout decay and grid geometry, not on the specific bimodal voter weights. This *universality* means the qualitative structure of the equilibrium existence diagram — which equilibria exist at which parameters — is robust to the precise shape of voter bimodality.[18]

Table 3 summarizes how the equilibrium structure scales with N at $\tau = 0.25$. As N grows, the centrist equilibrium becomes fragile ($\alpha^+(0) \to 0$), the number of coexisting equilibria increases, and for $M \geq 4$ the most extreme viable equilibrium (j^*) falls short of the grid boundary M. Proposition 1 (Appendix $\mathcal{A}$.4) proves these limiting behaviors: $\alpha^+(0) \to 0$, $j^*/M \to 1$, and $\alpha^-(j^*) \to 0$ as $N \to \infty$.

[Insert Table 3 here: **Scaling of NEQ Structure with Coarseness.**]

Figure 6 shows the continuous-limit equilibrium existence diagram at $N = 1{,}001$. The left panel plots the PE lower-boundary curves for five equilibrium positions: as α decreases, the primary deviation to a more centrist position becomes profitable and the equilibrium collapses. The right panel shades the existence region for the centrist ($s = 0$) and semi-extreme ($s = 0.6$) equilibria, bounded above by the GEC and below by the PEC.

---

[17] Figures 6–7 plot the inverse mapping $j^*(\alpha)$ (the most extreme viable equilibrium as a function of α), which is non-increasing. This is consistent with $\alpha^+(j)$ being non-decreasing in j: the two are inverse functions. When the text says the GE boundary "drops" in the figures, it refers to $j^*(\alpha)$ declining as α increases, not to $\alpha^+(j)$ declining in j.

[18] This universality holds for the GE boundary numerator at fixed N. The large-N limiting claims ("even the faintest bimodality destroys centrist equilibria") depend on rescaling τ with N to preserve end-to-end decay; under fixed τ, the conclusions differ (Section IV.B). The boundary formulas themselves are specific to the linear voter family (2) and geometric turnout (1). Alternative functional forms preserve the qualitative band structure but shift the quantitative boundaries.

The scaling table reveals a striking contrast between small and large electorates. At $N = 7$, the cascade from centrist to extreme plays out over the full range $\alpha \in [0, 1]$: the centrist equilibrium survives until $\alpha \approx 0.80$, and the extreme equilibrium does not appear until $\alpha \approx 0.98$. At $N = 1{,}001$, both thresholds have collapsed toward zero: the centrist equilibrium is destroyed by any detectable bimodality ($\alpha^+(0) < \varepsilon$), and the near-extreme equilibrium ($j^*/M = 0.97$) already appears at $\alpha \approx 0.34$. In the limit, $\alpha^+(0) \propto \lambda^M$ vanishes exponentially, while $\alpha^-(j^*) = O(1/M)$ vanishes algebraically (Proposition 1). Thus, the entire progression that required substantial bimodality at small N compresses into a vanishing neighborhood of $\alpha = 0$ at large N. For any electorate of realistic size, even the faintest bimodality simultaneously destroys centrist equilibria and enables near-extreme ones.

With perfectly uniform voters ($\alpha = 0$), the equilibrium band spans approximately $[0, 0.50]$ at large N. The mechanism is straightforward: a centrist primary challenger at position $j-1$ captures roughly j of the uniformly distributed primary positions while the incumbent at $-j$ captures roughly $M - j + 1$. When j exceeds the midpoint of the half-line, the challenger has a territorial majority and wins the primary. Turnout decay ($\tau > 0$) modifies the exact boundary — distant voters on both sides contribute less — but the approximate $s \approx 0.50$ cutoff is robust. In both cases, absent bimodality, no extreme equilibrium is possible: there are no concentrated tails to protect an extremist from a centrist challenger.

C The Curse of the Principled Candidate

At $N = 1{,}001$, a new phenomenon appeared: white cells marked with X, where *no* $\mathcal{D}$ position yields positive payoff. The candidate is caught between winning the primary (must be extreme enough) and winning the GE (must be moderate enough) — no position bridges both.[19]

These unwinnable configurations have a precise geometric characterization: they lie strictly southwest of the equilibrium diagonal in the heatmap grid (just defined as a band in Theorem 2)—that is, $\mathcal{D}_2$ must be more extreme than the mirror image of the $\mathcal{R}$ opponent (Proposition 2).

---

[19] At $\alpha = 0.80$, there are 65 such unwinnable cells; at $\alpha = 1$, none. This squeeze does not arise at $N \leq 9$ under symmetric distributions: the coarse grid means a single step can bridge the primary–GE gap.

Northeast of the diagonal, where $\mathcal{D}$ is more centrist than $\mathcal{R}$'s mirror, the co-location deviation always yields positive payoff (at least 1/4), so the Squeeze cannot arise.

The preceding analysis concerned symmetric voter distributions. We can preview how the asymmetric (flattening) mechanism extends to large N. The asymmetric case — where $\mathcal{R}$'s electorate flattens and this forces $\mathcal{D}$ to reposition — operates at $N = 1{,}001$ just as at $N = 7$. The flattening collapse at $N = 1{,}001$ is illustrated in Figure 8 (Section V.A), the direct analog of the $N = 7$ collapse in Figure 4 (Section IV). The causal chain is the same as in Section IV.C: GE Neutrality breaks (Link 1), the primary trap loses its bite (Link 2), and the cross-party enabling collapses (Link 3). Theorem 3 in Section V.D proves the result for arbitrary N.

An interesting aspect here is that the principled candidate curse does not facilitate extremism on the other side. The logic remains reversed: the principled candidate destroys the viability *because* the other side can remain moderate. Arguably, this has been true in some states, such as California where neither Jerry Brown nor Gavin Newsom were extremist candidates. If they had been, the Republican candidates may have become electable.

D Asymmetric Voters and Moderation/Extremism Exporting Equilibria

This subsection generalizes the flattening collapse of Section IV ("The Moderation Export") from $N = 7$ to arbitrary N. The proof (Appendix $\mathcal{A}$.3) proceeds through five lemmas.

Theorem 3 (Flattening Collapse) *Consider the symmetric bimodal distribution with* $\alpha_L = \alpha_R = \alpha$ *sufficiently high that* $(-j, -j, +j, +j)$ *with* $j > 0$ *is a Nash equilibrium. Then for any perturbation that reduces* α_R *to* $\alpha - \delta$ *with* $\delta > 0$*:*

(i) ***GE Neutrality breaks.*** $\mathrm{GE}(-j, +j) < 1/2$ *under the perturbed distribution:* $\mathcal{D}$ *loses the general election at the symmetric profile.*

(ii) ***Primary trap persists.*** *No* $\mathcal{D}$*-side deviation simultaneously wins the* $\mathcal{D}$ *primary and the GE. More extreme deviations win the primary but lose the GE; centrist deviations lose the primary.*

(iii) ***Replacement NEQ.*** *For $\delta \geq \delta^*(j, \alpha, \tau, M)$, there exists a replacement NEQ with both $\mathcal{D}$ candidates on the $\mathcal{R}$ side of the spectrum: $(+s, +s, +j, +j)$ for some $s > 0$.*

The intuition is that any asymmetry ($\alpha_R < \alpha_L$) breaks GE Neutrality: the flatter $\mathcal{R}$ side has a "deep bench" of moderate voters who generate broader turnout, tipping the GE in $\mathcal{R}$'s favor. $\mathcal{D}$ candidates at the symmetric profile now receive payoff 0 (they win the primary but lose every GE). With zero payoff, a $\mathcal{D}$ candidate can profitably defect to the $\mathcal{R}$ side of the spectrum. The center equilibrium ($j = 0$) is immune because both nominees sit at the same position, so GE Neutrality holds trivially regardless of voter asymmetry.

The replacement equilibria have a striking property: $\mathcal{D}$ politicians, whose own voters are still bimodal and concentrated at negative positions, stake out *positive* positions on $\mathcal{R}$'s side of the spectrum. They do this not because $\mathcal{D}$'s voters changed, but because the slightly flatter $\mathcal{R}$ electorate destroyed the primary trap that kept $\mathcal{D}$ extreme. Both parties' candidates crowd onto the flatter side, where moderate positions are electorally viable.

The proof, which proceeds through five lemmas, is in Appendix $\mathcal{A}$.3.

VI Discussion

A The Roles of Both Parameters

The model has only two parameters: voter polarization (α) and turnout decay (τ). Both are necessary for the main results.

Without bimodality ($\alpha = 0$), voter distributions are uniform. Extreme positions cannot win the primary: the centrist partner draws from a broader and equally dense voter base. Without concentrated bases at the extremes, the within-party competition that drives the primary trap has no teeth. Multiple equilibria still exist under uniform distributions (the equilibrium band at $\alpha = 0$ spans from center to $s \approx 0.50$ at $\tau = 0.25$), but they are all moderate — no extreme equilibrium is possible.

Without turnout decay ($\tau = 0$), all voters turn out regardless of distance. The primary trap still operates — partisan primary assignment restricts the primary electorate, so a centrist deviator captures fewer primary voters than the extreme incumbent — and inner equilibria survive: at $N = 7$, three equilibria coexist for all $\alpha \in (0,1)$: $(0,0,0,0)$, $(-1,-1,+1,+1)$, and $(-2,-2,+2,+2)$. The two most centrist persist even into the unimodal range ($\alpha < 0$); standard Hotelling–Downs convergence does *not* obtain. But the *most extreme* equilibrium $(-3,-3,+3,+3)$ is absent: at full participation, a centrist challenger captures enough moderate primary voters to dislodge the extreme incumbent. It is the *interaction* of turnout decay and bimodal voters that creates the conditions for extreme equilibria.

Both equilibrium thresholds decrease with decay (for $\tau \in (0,1]$):

$$\frac{\partial \alpha^*_{\text{GEC}}}{\partial \tau} < 0\,, \qquad \frac{\partial \alpha^*_{\text{PEC}}}{\partial \tau} < 0\,.$$

As τ increases (less participation), α^*_{GEC} falls — the centrist equilibrium collapses at a *lower* α — and α^*_{PEC} falls — the extreme equilibrium becomes viable at a lower α. Higher decay attenuates moderate voters, allowing the concentrated extreme base to block centrist challengers. The equilibrium existence diagram makes this visible: the equilibrium boundaries are one-dimensional curves in (α, τ) space, and it is intermediate values of both parameters that distinguish the solutions.

B Partisan Loyalty

The flattening collapse relies on cross-party defection: a $\mathcal{D}$ candidate repositioning to the $\mathcal{R}$ side of the spectrum. Two institutional features can block this mechanism, and both lead to the same conclusion: the steeper side permanently loses the general election.

First, crossing over may not be permitted at all. Under closed primaries — where party registration, not spatial position, determines primary assignment — a $\mathcal{D}$ candidate cannot enter $\mathcal{R}$'s side of the electoral contest. Without this escape route, the primary trap locks both $\mathcal{D}$ candidates at extreme positions. GE Neutrality is broken (the flatter $\mathcal{R}$ electorate generates broader turnout), so $\mathcal{R}$ wins every general election. The steeper side gets payoff zero with no recourse.

Second, even if crossing over is spatially permitted, voters may retain a minimal preference for their own party's candidates — a loyalty bonus ϵ that penalizes other-party candidates by adding ϵ to their effective distance. A $\mathcal{D}$ candidate at a positive position then faces a headwind in attracting $\mathcal{R}$-side voters. A small ϵ shifts the collapse threshold upward (requiring a larger asymmetry δ to trigger the cross-party defection). A sufficiently large ϵ (exceeding one grid step) blocks cross-party attraction entirely, tipping the effective distance in favor of the $\mathcal{R}$ nominee. Without $\mathcal{R}$-side votes, the cross-party candidate cannot win the GE, and the defection is futile. Again, the steeper side permanently loses.[20]

This creates a paradox: partisan identification — which one might expect to protect a party's electoral prospects — instead *sustains* the worst outcome for the losing party. Strong loyalty blocks the self-correcting collapse that would otherwise reposition both parties at moderate positions — even if those moderate positions still face the opposing party.

---

[20] Symmetric equilibria are *unaffected* by partisan loyalty. Under symmetric voters and a symmetric profile $(-j, -j, +j, +j)$, the loyalty bonuses cancel by symmetry: $\mathcal{D}$-affiliated voters penalize $\mathcal{R}$ candidates by ϵ, and vice versa. GE Neutrality is preserved, and the within-party primary trap survives (both co-partisans share the same label).

C Model Estimability

The parameters of this model are, in principle, estimable from data. The voter distribution shape (α) can be recovered from survey data on ideal points or from the geographic distribution of voters across constituencies. The turnout decay (τ) can be estimated from the relationship between candidate–voter distance and turnout rates, controlling for other factors. Adams et al. (2006) estimate alienation and indifference effects jointly in a spatial voting model. Their alienation parameter is closely related to our τ.

VII Conclusion

This paper was based on realistic premises for today's political landscape: competitive primaries, competitive general elections, voter polarization, and the ability of voters to abstain when candidates are too distant. These features were sufficient to generate interesting results:

1. **Voter Polarization.** As far as candidate positioning is concerned, it is unimportant whether voters are steeply or shallowly single-modal. It is bimodality (voter polarization) that can enable qualitatively and quantitatively very different and extremist candidate choices *in equilibrium*.
2. **The primary trap.** In a four-candidate spatial model with primaries, distance-dependent turnout, and bimodal voter distributions, individual candidates can be trapped at extreme positions by within-party competition *in equilibrium*. Moderating loses the primary to the more extreme co-partisan. The party-optimal centrist position is blocked by intra-party competition — a within-party collective action failure.
3. **A band of equilibria.** A contiguous band of symmetric equilibria can coexist at any given level of voter polarization — ranging from centrist to extreme. Candidates cannot break equilibria by deviating individually. It would require multiple-candidate deviations to

switch equilibria. Uniqueness arrives only after polarization has pushed all equilibria to the boundary; that is, uniqueness confirms that polarization has already won.

4. **Cross-party enabling.** Extremism on one side requires extremism on the other. The primary trap binds only when the general election is competitive, which requires the opposing party to be extreme enough that the general election is not a foregone conclusion.

5. **The voter/party export.** A small moderation of one party's electorate can trigger a discontinuous collapse of all extreme equilibria, with the more extreme party's candidates relocating past the center towards the flatter side of the spectrum. The collapse is a discontinuous shift, not a gradual adjustment. Conversely, a small polarization of voters on the more moderate side can shift the equilibrium from candidates that are moderate and similar in a cluster, to candidates that are extreme and dissimilar and on opposite sides of the spectrum.

6. **The "Unwinnable Configuration" Curse.** When a co-partisan is exogenously committed to an extreme position, the remaining candidate may face an unwinnable configuration: no position may be able to simultaneously survive the primary and the general election. This is not modeled. (Candidate could have preference for principles; the model here assumes pure office-seeking.) However, it captures the strategic consequence of having a co-partisan whose position is effectively fixed.

Many open questions remain.

First, the model has no dynamics or equilibrium selection mechanism that determines which equilibrium from the band plays out in real life. The model shows that multiple equilibria coexist, but is silent on which one obtains. Historical path dependence, focal points, elite coordination, or candidate preferences (and their dynamics) could all play a role.

Second, the model could integrate valence (which is already sketched in Appendix $\mathcal{B}$). The two features can work together to offer richer implications.

Third, voters on both sides had the same tolerance for candidates more distant from them. A richer model could explore the effects of different degrees of tolerance for candidates further from voters' preferred positions (e.g., τ parameters different for left, right, or centrist voters).

Fourth, candidates could obtain some utility from winning the primary even if they lose the general election. This could greatly increase the incentives of candidates to locate to a more extremist position, where more strident party voters convey primary but not general election victories.

Fifth, the two parameters in the model — voter distribution and preference-based turnout — could themselves be endogenous and link back to other empirically measurable quantities.

Sixth, there could be more than two candidates competing in primaries.

Seventh, there are also other electoral systems — such as open primaries where the top two candidates square off in the general election, even if they are from the same party — that could be analyzed similarly.

The parameters of the model and many of the just-mentioned extensions are intrinsically measurable and indeed often measured. The model had no asymmetric information, which often renders game-theoretic analysis difficult to explore empirically. It's testable. The model can explore some widely recognized and sometimes puzzling phenomena in a rigorous but still quite intuitive manner. It's a good step in a promising direction.

References

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

Adams, James, Samuel Merrill III, and Bernard Grofman. 2005. *A Unified Theory of Party Competition.* Cambridge: Cambridge University Press.

Adams, James, Jay Dow, and Samuel Merrill III. 2006. "The Political Consequences of Alienation-Based and Indifference-Based Voter Abstention." *Political Behavior* 28(1): 65–86.

Adams, James, and Samuel Merrill III. 2008. "Candidate and Party Strategies in Two-Stage Elections Beginning with a Primary." *American Journal of Political Science* 52(2): 344–359.

Agranov, Marina. 2016. "Flip-Flopping, Primary Visibility, and the Selection of Candidates." *American Economic Journal: Microeconomics* 8(2): 61–85.

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

Benoit, Kenneth. 2004. "Models of Electoral System Change." *Electoral Studies* 23(3): 363–389.

Boix, Carles. 1999. "Setting the Rules of the Game: The Choice of Electoral Systems in Advanced Democracies." *American Political Science Review* 93(3): 609–624.

Callander, Steven, and Catherine H. Wilson. 2007. "Turnout, Polarization, and Duverger's Law." *Journal of Politics* 69(4): 1047–1056.

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

Carroll, Royce, and Monika Nalepa. 2020. "The Personal Vote and Party Cohesion: Modeling the Effects of Electoral Rules on Intraparty Politics." *Journal of Theoretical Politics* 32(1): 36–69.

Casas, Agustín, and Till Gross. 2021. "Signaling Valence in Primary Elections." *Games and Economic Behavior* 131: 59–71.

Castanheira, Micael, Benoît S. Y. Crutzen, and Nicolas Sahuguet. 2010. "Party Organization and Electoral Competition." *Journal of Law, Economics, and Organization* 26(2): 212–242.

Cintolesi, Andrea. 2022. "Political Polarization and Primary Elections." *Journal of Economic Behavior & Organization* 200: 596–617.

Cox, Gary W., and Jonathan N. Katz. 2002. *Elbridge Gerry's Salamander: The Electoral Consequences of the Reapportionment Revolution.* Cambridge: Cambridge University Press.

Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper.

Fiorina, Morris P., Samuel J. Abrams, and Jeremy C. Pope. 2005. *Culture War? The Myth of a Polarized America.* New York: Pearson Longman.

Grosser, Jens, and Thomas R. Palfrey. 2019. “Candidate Entry and Political Polarization: An Experimental Study.” *American Political Science Review* 113(1): 209–225.

Hall, Andrew B. 2015. “What Happens When Extremists Win Primaries?” *American Political Science Review* 109(1): 18–42.

Hirano, Shigeo, James M. Snyder Jr., Stephen Ansolabehere, and John Mark Hansen. 2010. “Primary Elections and Partisan Polarization in the U.S. Congress.” *Quarterly Journal of Political Science* 5(2): 169–191.

Hirsch, Alexander V. 2023. “Polarization and Campaign Spending in Elections.” *Journal of Politics* 85(1): 240–249.

Hotelling, Harold. 1929. “Stability in Competition.” *Economic Journal* 39(153): 41–57.

Hummel, Patrick. 2013. “Candidate Strategies in Primaries and General Elections with Candidates of Heterogeneous Quality.” *Games and Economic Behavior* 78: 85–102.

Key, V. O., Jr. 1949. *Southern Politics in State and Nation.* New York: Knopf.

Kim, Jonghoon, Hyeong-Chai Jeong, and Seung Ki Baek. 2022. “Symmetric Nash Equilibrium of Political Polarization in a Two-Party System.” *Physica A* 607(C): 128207.

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

Owen, Guillermo, and Bernard Grofman. 2006. “Two-Stage Electoral Competition in Two-Party Contests: Persistent Divergence of Party Positions.” *Social Choice and Welfare* 26(3): 547–569.

Polsby, Nelson W. 1983. *Consequences of Party Reform.* New York: Oxford University Press.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting.* New York: Oxford University Press.

Serra, Gilles. 2015. “No Polarization in Spite of Primaries.” In *The Political Economy of Governance*, ed. Norman Schofield and Gonzalo Caballero, 211–229. Springer.

Snyder, James M., Jr., and Michael M. Ting. 2011. “Electoral Selection with Parties and Primaries.” *American Journal of Political Science* 55(4): 782–796.

Stokes, Donald E. 1963. “Spatial Models of Party Competition.” *American Political Science Review* 57(2): 368–377.

Wittman, Donald A. 1983. “Candidate Motivations: A Synthesis of Alternative Theories.” *American Political Science Review* 77(1): 142–157.

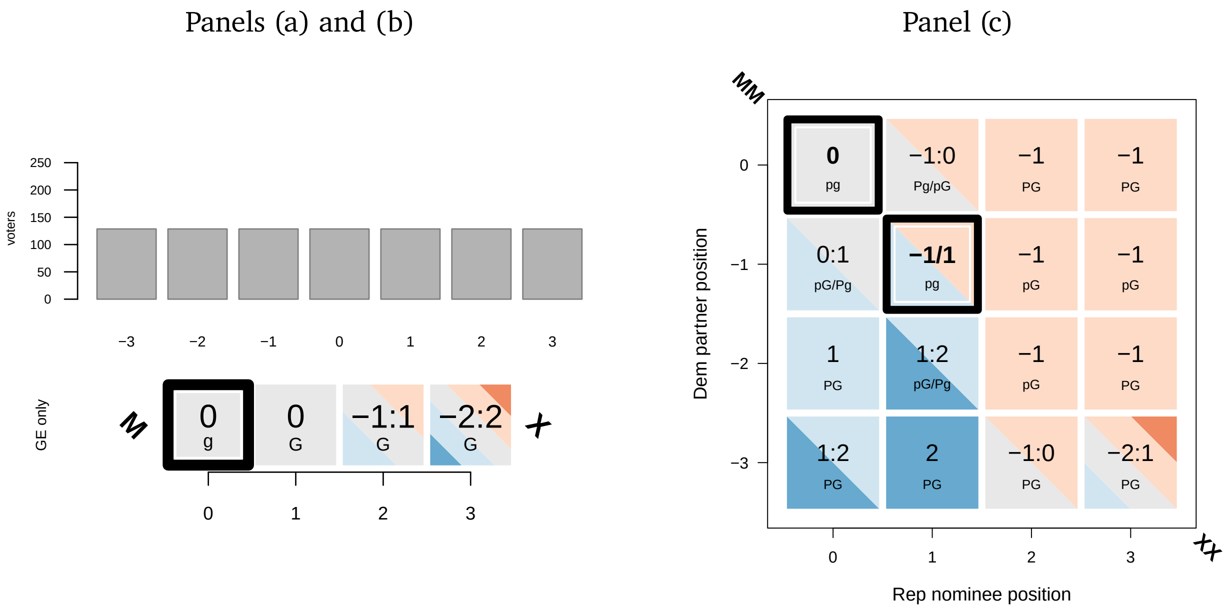
Woon, Jonathan. 2018. “Primaries and Candidate Polarization: Behavioral Theory and Experimental Evidence.” *American Political Science Review* 112(4): 826–843.

EXHIBITS

Table 1: Glossary of notation.

Symbol	Description
N	Number of positions (odd); $N = 2{\cdot}M + 1$
k	A position on the line ($-M$ to $+M$); 0 is center
v_k	Number of voters at position k
b_k	Bimodal base distribution: $b_k = 10 \cdot \lvert k \rvert$
α	Voter polarization parameter ($\alpha \in [-1,1]$) unimodal at $\alpha<0$, uniform at $\alpha = 0$, bimodal at $\alpha>0$
τ	Turnout decay: voter at distance d turns out with probability $(1-\tau)^d$ $\tau = 0$ is full turnout; $\tau = 1$ vote only for perfect candidate
α^*_{GEC}	Threshold α from the General Election Condition (GEC)
α^*_{PEC}	Threshold α from the Primary Election Condition (PEC)
$\alpha^+(j)$, $\alpha^-(j)$	GEC upper bound and PEC lower bound for NEQ at position j (general N)
$\mathcal{D}_1, \mathcal{D}_2$	Location choices of two $\mathcal{D}$ party candidates (w.l.o.g. $\mathcal{D}_1 \le \mathcal{D}_2$)
$\mathcal{R}_1, \mathcal{R}_2$	Location choices of two $\mathcal{R}$ party candidates (w.l.o.g. $\mathcal{R}_1 \le \mathcal{R}_2$)
$\text{GE}(q,r)$	General election $\mathcal{D}$'s vote share when $\mathcal{D}$ nominee is at q, $\mathcal{R}$ nominee at r
$(\mathcal{D}_1, \mathcal{D}_2, \mathcal{R}_1, \mathcal{R}_2)$	Position of four candidates, e.g. $(q, -3, 3, 3)$ (q can be an endogenous equilibrium choice against 3 given candidates)

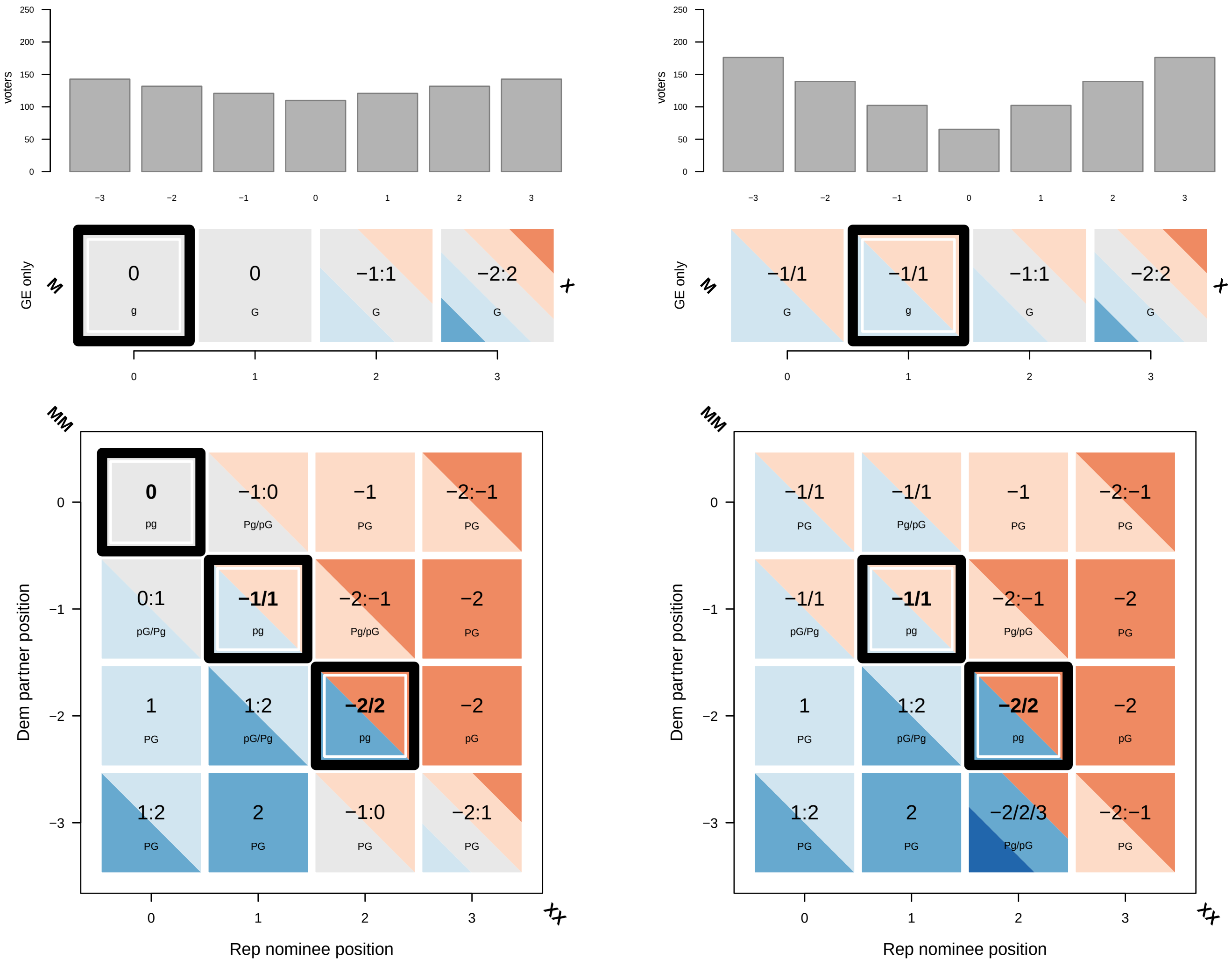
Figure 1: Optimal-Response Diagram for Uniform Voter Distribution ($\alpha = 0$), Modest Voter Decay ($\tau = 0.25$), and $N = 7$



Explanations: Optimal Response Diagram for $\alpha = 0$ (uniform voters), $\tau = 0.25$, $N = 7$. **Panel (a):** The voter distribution here is uniform across all 7 positions. **Panel (b)** (GE only): This shows the best *hypothetically coordinated* $\mathcal{D}$ location against an $\mathcal{R}$ candidate at a specific position in the general election (GE). If the $\mathcal{R}$ candidate chooses the center, the best (coordinated) $\mathcal{D}$ response is also to choose the center, and each candidate wins the general election with 50% probability ("g": the compact code below each position label decomposes the payoff: G = wins the general election, g = GE tie.) If the $\mathcal{R}$ candidate chooses +1, the best $\mathcal{D}$ response is still to choose the center, which then always wins ("G"). If the $\mathcal{R}$ candidate chooses +2, any choice from −1 to +1 wins, and so on. **Panel (c):** This shows the best individual candidate location against both a $\mathcal{R}$ candidate in the GE and against the intra-party competitor (y axis). The cell color visualizes the range (red = extreme −3, blue = centrist 0). For compact codes here: P = wins primary outright, p = primary tie (50%), The bold borders mark NEQ. MM indicates the "moderate-moderate" candidate configuration, XX indicates the "extreme-extreme" candidate configuration.

Interpretation: Panel (b) shows that the centrist position dominates (Hotelling-Downs). Panel (c): The probability of winning is 1/4 only in the two most centrist diagonal cells. (By symmetry, this means that only these two cells are equilibria.)

Figure 2: Optimal Response Diagrams at Four Voter Polarization Levels, Modest Voter Decay ($\tau = 0.25$), and $N = 7$

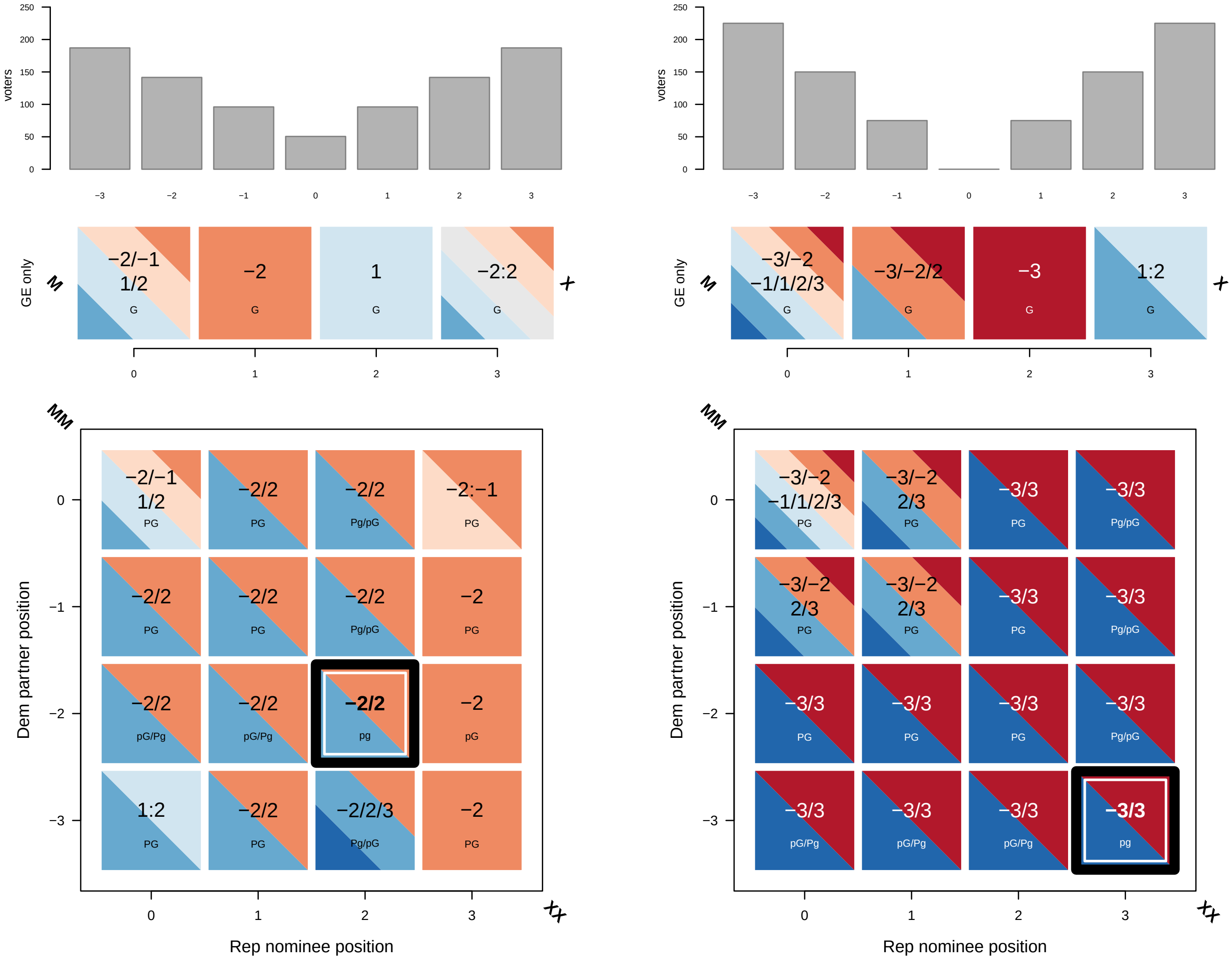


(a) $\alpha = 0.50$

(b) $\alpha = 0.85$

(continues)

(continued)



(c) $\alpha = 0.90$

(d) $\alpha = 1.00$

Explanations: Best-response chessboards, layout and color coding as in Figure 1. (Bold borders in the "chessboards" mark symmetric equilibria.)

Interpretation: For modest voter polarization, three equilibria are simultaneously viable. For higher voter polarization, viable equilibria become more extreme.

Note that at $\alpha = 0.90$, if $\mathcal{R}$ chooses +1, $\mathcal{D}$ wins by going a little bit more extreme (−2). However, if $\mathcal{R}$ chooses +2, $\mathcal{D}$ flips to the other side (+1) and thus captures more of the center and (less of) the left side. But in this case, R would want to change at the very least to "1" to prevent an outright loss. There is no pure strategy Nash equilibrium in this *GE-only* (two-player, coordinated-party) benchmark. (The full four-candidate primary+GE game does have equilibria at this α; the cycling here is specific to the hypothetical GE-only subgame.) Candidates would constantly try to outflip one another.

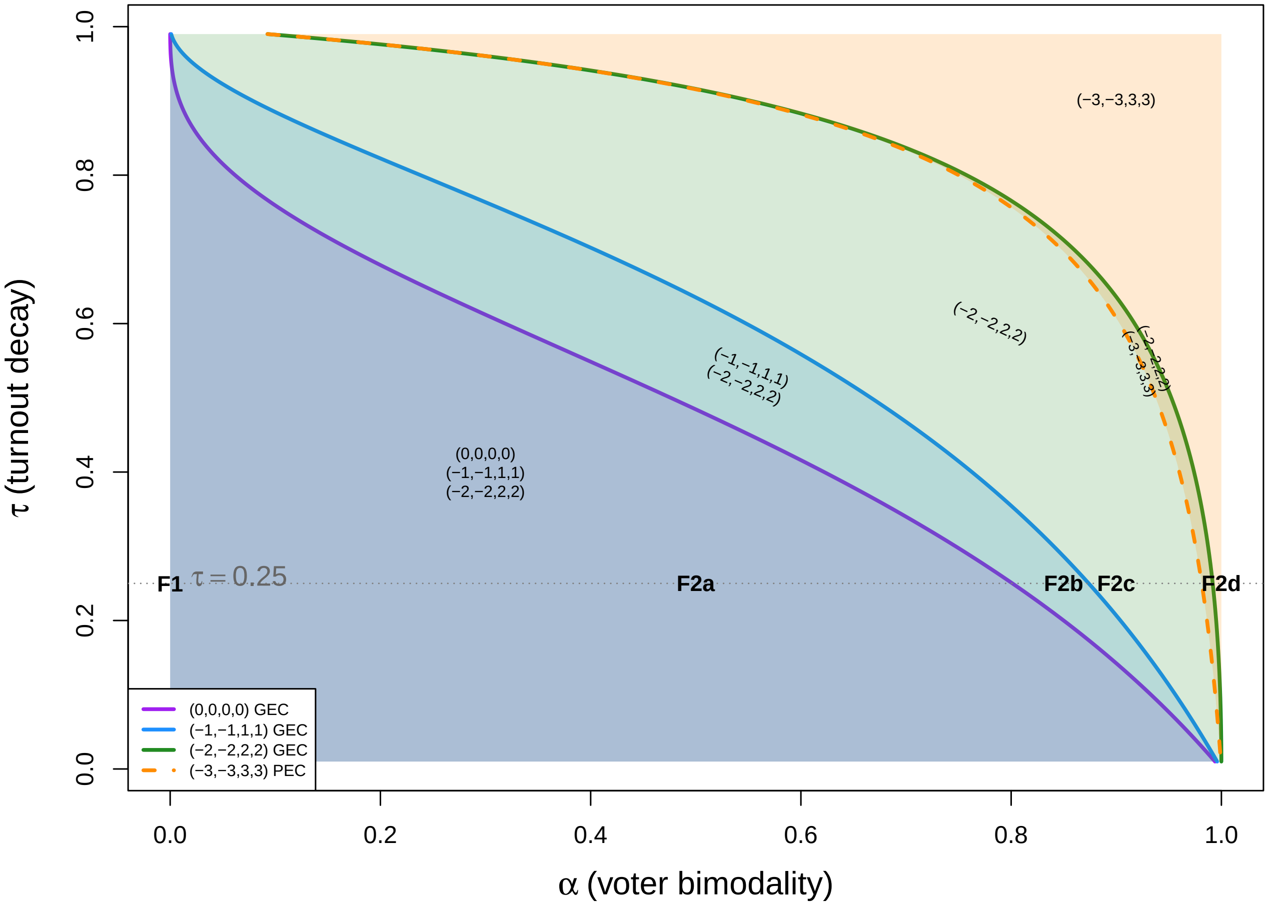
Table 2: Cascading Thresholds for Modest Voter Decay ($\tau = 0.25$) and $N = 7$.

α^*	Event	equilibria that exist
< 0	Single-Modal Voters	$(0,0,0,0)$; $(-1,-1,+1,+1)$
0	$(-2,-2,+2,+2)$ appears	$(0,0,0,0)$; $(-1,-1,+1,+1)$; $(-2,-2,+2,+2)$
0.801	$(0,0,0,0)$ collapses	$(-1,-1,+1,+1)$; $(-2,-2,+2,+2)$
0.874	$(-1,-1,+1,+1)$ collapses	$(-2,-2,+2,+2)$
0.981	$(-3,-3,+3,+3)$ appears	$(-2,-2,+2,+2)$; $(-3,-3,+3,+3)$
0.992	$(-2,-2,+2,+2)$ collapses	$(-3,-3,+3,+3)$ only

Explanations: Equilibria appearances and collapses as voter polarization α increases.

Interpretation: With higher voter polarization, more centrist equilibria collapse and more extreme equilibria appear.

Figure 3: Phase Diagram for Equilibria at $N = 7$



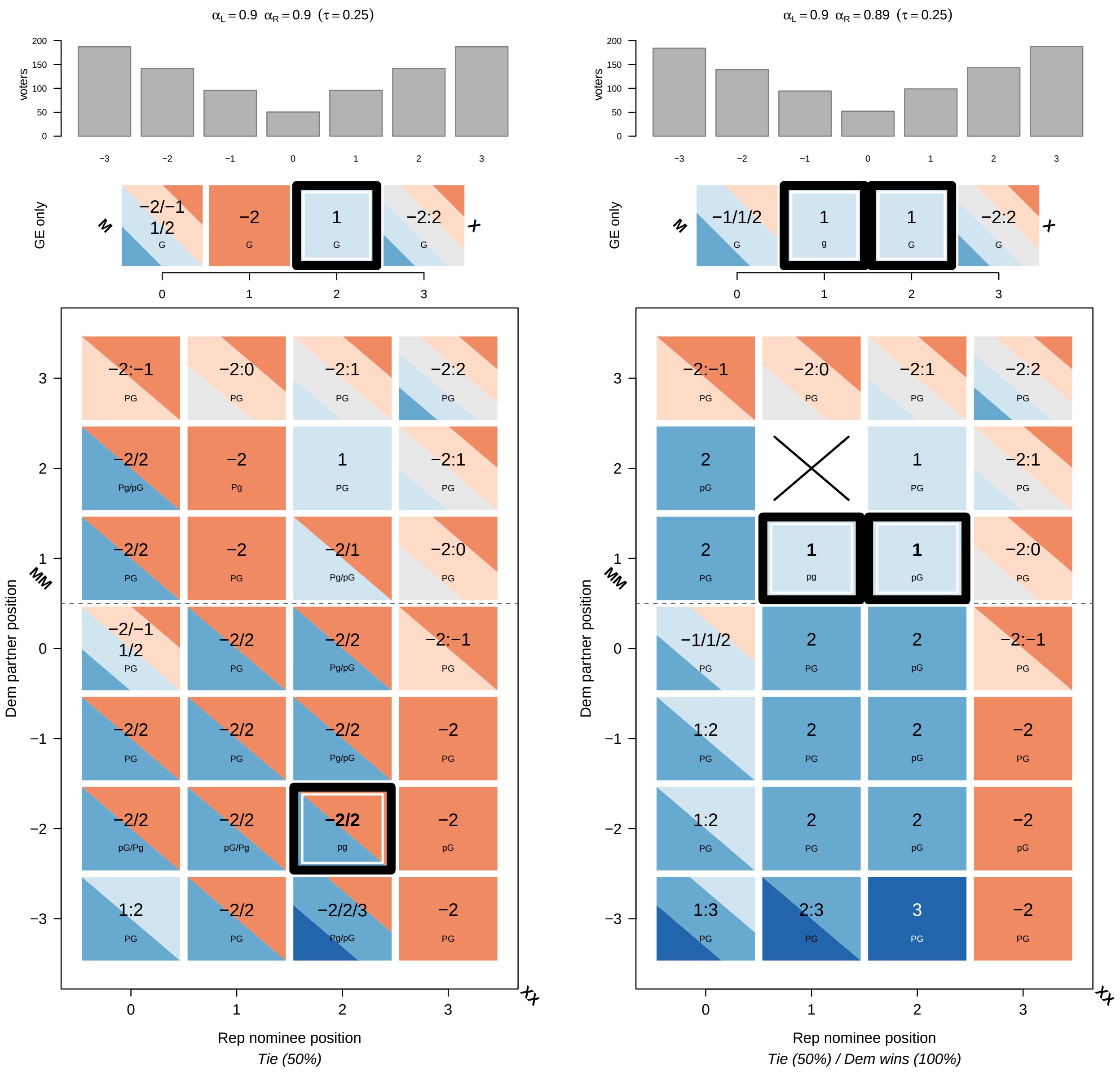
Explanations: Existence regions for symmetric equilibria in (α, τ) space. The narrow band near $\alpha \approx 0.98$–0.99 shows where $(-2,-2,2,2)$ and $(-3,-3,3,3)$ briefly coexist before the $(-2,-2,2,2)$ semi-extreme equilibrium vanishes. The dotted line marks $\tau = 0.25$ (baseline), which was graphed in preceding figures.

Interpretation: Without voter polarization ($\alpha \leq 0$), the same two equilibria are always viable. Higher α favors more extreme equilibria and disfavors more moderate equilibria. Without voter decay ($\tau = 0$), the same three equilibria are always viable. More voter decay τ favors more extreme equilibria and disfavor more moderate equilibria.

Figure 4: Asymmetric Voter Discrete Phase Transition ($N = 7$)

(a) Symmetric Bimodal

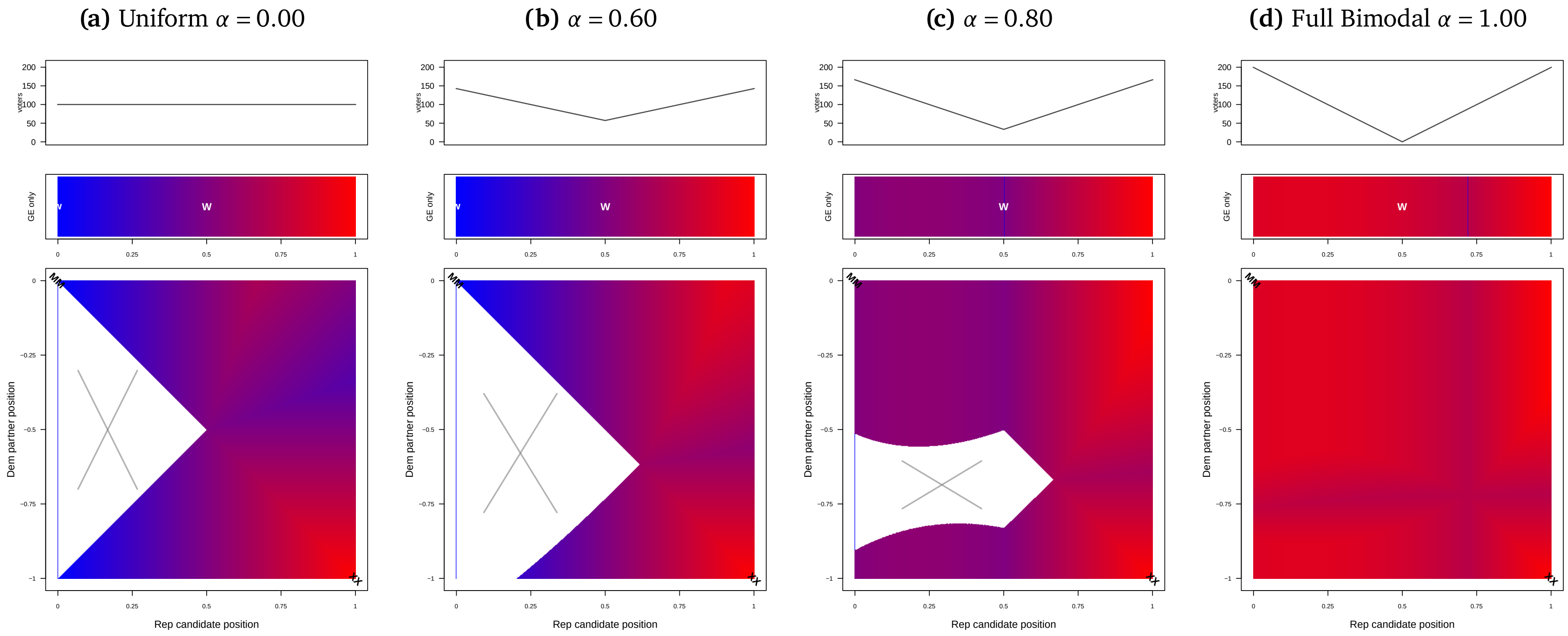
(b) Rep Flattens 1%



Explanations: Left panel: $\alpha_L = \alpha_R = 0.90$ (symmetric). The viable equilibrium is semi-extreme $(-2, -2, +2, +2)$. Right panel: $\alpha_R = 0.89$ ($\mathcal{R}$ flattens by 1%). The symmetric equilibrium collapses. Instead, two non-symmetric replacement equilibria appear on the $\mathcal{R}$ side of center: $(+1, +1, +1, +1)$ and $(+1, +1, +2, +2)$ (visible above the dashed center line, bold borders). Italicized text at the bottom shows the GE winner.

Interpretation: $\mathcal{D}$ jumps from -2 to $+1$ — a three-position shift past center — not because $\mathcal{D}$'s voters changed, but because the slightly flatter $\mathcal{R}$ electorate has destroyed the conditions sustaining extreme $\mathcal{D}$ positions.

Figure 5: Optimal Response Diagrams at $N = 1{,}001$

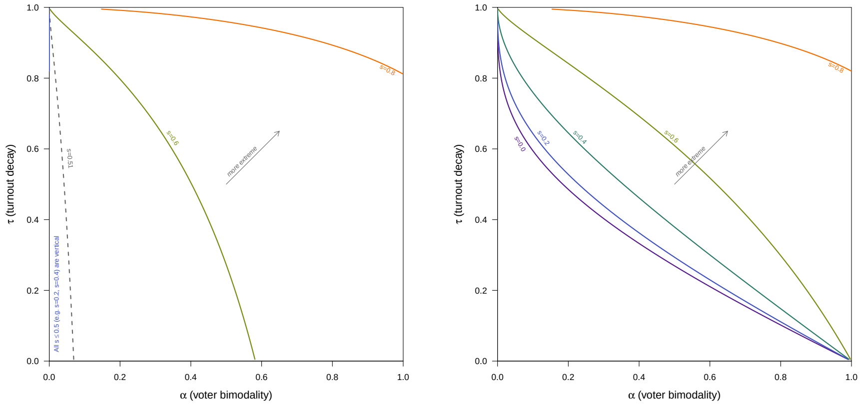


(a) Uniform $\alpha = 0.00$ **(b)** $\alpha = 0.60$ **(c)** $\alpha = 0.80$ **(d)** Full Bimodal $\alpha = 1.00$

Explanations: Best-response chessboards at $N = 1{,}001$ for four levels of voter bimodality (α) and (near-)continuous voter distributions. Positions are normalized to $[-1, +1]$, where ± 1 are the most extreme positions ($\pm M = \pm 500$) and 0 is center. Layout as in Figure 1. In this figure, when there are multiple equilibria, the moderate one is selected. The color ramp runs from red (extreme) to blue (centrist); when multiple positions tie as best responses, the most centrist is shown. Decay is scaled from $\tau = 0.25$ at $N = 7$ to be comparable. White cells with X marks are unwinnable at any position (the Principled-Candidate Squeeze).

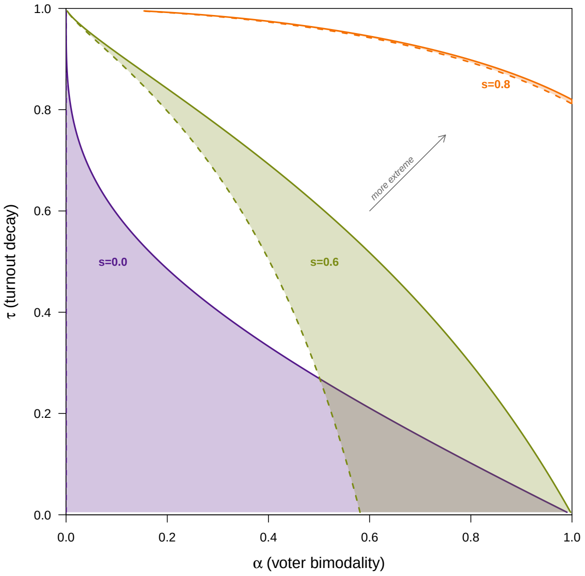
Interpretation: Under uniform voter distributions with modest voting decay ($\alpha = 0, \tau = 0.25$), the optimal response to many choices by the opposite party and the same-party opponent can be moderate (bluish). As voter polarization increases, the optimal response to many choices becomes ever more extreme.

Figure 6: Phase Diagram For Three Equilibria ($N = 1{,}001$)



(a) PE boundary curves

(b) GE boundary curves

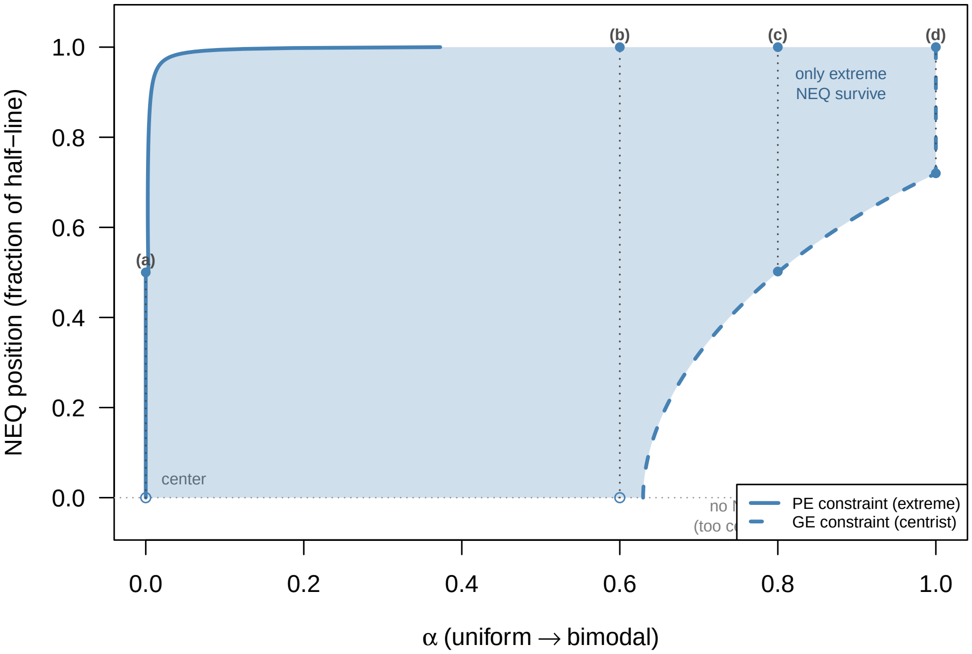


(c) equilibrium existence regions

Explanations: Phase diagram at $N = 1{,}001$ (near-continuous limit) in $(\alpha, \tau_{\text{ref}})$ space. Positions are normalized as fractions $s \in [0, 1]$ of the half-line ($s = 0$ is the center, $s = 1$ is the most extreme position $\pm M$). **(a)** PE lower-boundary curves for $s = 0.2, 0.4, 0.6, 0.8, 1.0$: each equilibrium $(-s, -s, +s, +s)$ first becomes viable above its curve (the centrist primary deviation ceases to win). **(b)** GE upper-boundary curves for $s = 0, 0.2, 0.4, 0.6, 0.8$: the equilibrium ceases to be viable above its curve (a more extreme deviation wins the GE); also the PE boundary for $s = 1.0$ (dashed). **(c)** Shaded existence regions for $s = 0$ and $s = 0.6$: each equilibrium exists between its PEC lower boundary and its GEC upper boundary. The vertical axis is τ_{ref}, the $N = 7$ reference decay; the actual τ is scaled so that λ^{N-1} is constant across all N, preserving candidates' effective reach.

Interpretation: When voter polarization increases, centrist equilibria disappear (the GE boundary drops) and extreme equilibria become viable (the PE boundary rises). Higher turnout decay τ shifts the entire landscape toward extremism: centrist equilibria collapse at *lower* α, and extreme equilibria appear at *lower* α. At any given (α, τ), the existence regions overlap substantially, producing a wide band of coexisting equilibria — the continuous-limit analog of the multiplicity observed at $N = 7$.

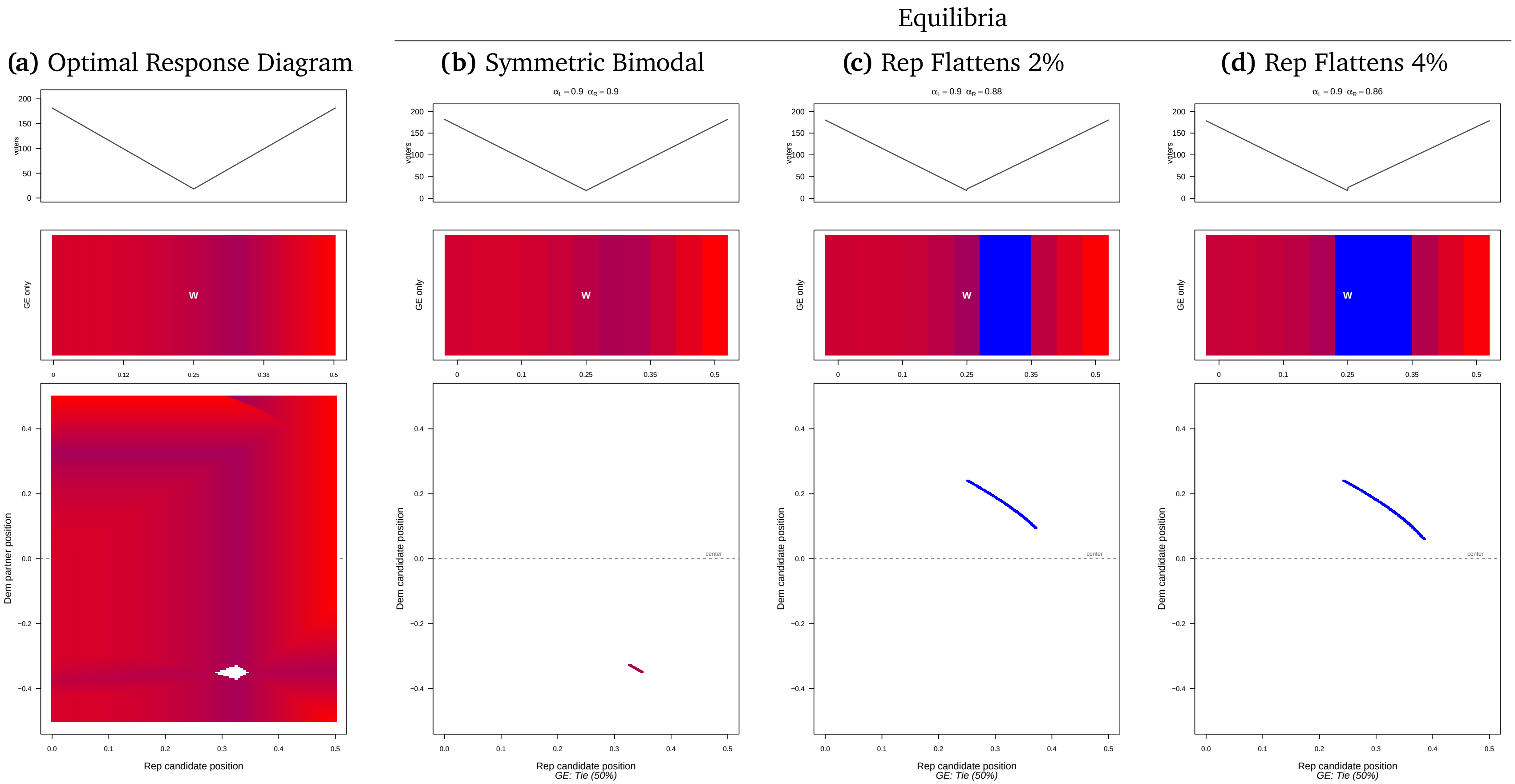
Figure 7: NEQ Band Diagram ($N = 1{,}001$)



Explanations: Equilibrium band diagram at $N = 1{,}001$, $\tau = 0.25$ (with λ scaled to preserve effective reach). This is a cross-section of the equilibrium existence diagram (Figure 6) at fixed τ, re-expressed with equilibrium position on the vertical axis. The y-axis measures each symmetric equilibrium position as a fraction of the half-line: 0 = center, 1 = most extreme ($\pm M$). Dashed = GE boundary (above which an extreme deviation wins the general election); solid = primary boundary (below which a centrist deviation wins the primary). The shaded band is the set of viable equilibria at each α. Vertical dotted segments mark four representative α values: at $\alpha = 0$ the band spans $[0, 0.50]$; at $\alpha = 0.60$ the full range $[0, 1]$; at $\alpha = 0.80$ only $[0.50, 1]$; and at $\alpha = 1$ only $[0.72, 1]$.

Interpretation: Follow the four dotted verticals from left to right. At $\alpha = 0$ (uniform voters), equilibria span the bottom half of the spectrum — all moderate. At $\alpha = 0.60$ (moderate bimodality), the band covers the entire range from center to extreme: centrist and extreme candidates coexist as equilibria. At $\alpha = 0.80$, the center equilibrium is gone — an extreme deviation now wins the GE against a centrist nominee — and only the upper half survives. At $\alpha = 1$, the band has narrowed to $[0.72, 1]$: the most moderate surviving equilibrium is already nearly three-quarters of the way to the extreme. The pattern is that extreme equilibria appear early and never leave, while centrist equilibria are progressively eliminated.

Figure 8: Flattening Collapse ($N = 1{,}001$)


Equilibria

(a) Optimal Response Diagram

(b) Symmetric Bimodal

(c) Rep Flattens 2%

(d) Rep Flattens 4%

Explanations: (a) The optimal response diagram. (b) The equilibria for the same configuration of (a). (c-d) More flattened (less polarized right) voters.

Interpretation: (b) Note how the equilibria are just north-east of the white region. (c) The flip from extremist −0.35 to completely different and more moderate equilibria (of about +0.10) given a slight flattening of voters on the positive issue domain. (d) For even more flattening, ever more centrist equilibria become feasible, too.

Table 3: Scaling of NEQ Structure with Coarseness.

Positions	$\alpha^{+}(0)$: alpha where most centrist NEQ (0) dies	Most Extreme NEQ	$= j^{*}/M$	$\alpha^{-}(j^{*})$: alpha where most extreme NEQ appears
7	0.801	3/3	= 1.000	0.981
9	0.683	3/4	= 0.750	0.634
11	0.560	4/5	= 0.800	0.786
15	0.345	6/7	= 0.857	0.975
21	0.149	8/10	= 0.800	0.823
51	0.002	21/25	= 0.840	0.956
101	<0.001	43/50	= 0.860	0.386
501	ε	238/250	= 0.952	0.197
1,001	ε	486/500	= 0.972	0.340
2,001	ε	983/1000	= 0.983	0.040
5,001	ε	2480/2500	= 0.992	0.022
10,001	ε	4978/5000	= 0.996	0.045
100,001	ε	49970/50000	= 0.999	0.004
∞	0		$\to 1$	0

Explanations: $\tau = 0.25$ throughout (unscaled $\lambda = 0.75$). Each symmetric equilibrium has the form $(-j, -j, +j, +j)$ for some position $j \in \{0, \dots, M\}$ where $M = (N-1)/2$. Column descriptions: $\alpha^{+}(0)$: highest α at which the centrist equilibrium ($j = 0$) exists. j^{*}/M: most extreme position that is an equilibrium at some α, expressed as a fraction of the half-line (1.000 means the boundary position $j = M$). $\alpha^{-}(j^{*})$: lowest α at which this most extreme equilibrium first appears. The final row shows the proven limiting behavior (Proposition 1): all three quantities converge, with $\alpha^{-}(j^{*}) = O(1/M)$. The α^{-} column is non-monotonic because j^{*} changes in discrete steps, but the envelope trends to zero. Meaningful large-N results at moderate α require scaling λ with N to preserve the candidates' effective reach (Section A).

Interpretation: As the electorate grows, the centrist equilibrium becomes exponentially fragile: $\alpha^{+}(0) \propto \lambda^{M}$ falls below machine precision by $N = 501$. Even the faintest bimodality destroys centrism. At the same time, the most extreme viable equilibrium approaches the grid boundary ($j^{*}/M \to 1$) but never quite reaches it: the most extreme position always requires sufficient bimodal concentration to survive a centrist primary challenge. The implication is that at realistic electorate sizes, the equilibrium structure is qualitatively different from $N = 7$: there is no meaningful "centrist era" — polarization wins almost immediately.

$\mathcal{A}$ Appendix

$\mathcal{A}$.1 Proof of Equilibrium Payoff Equality (Theorem 1)

This proof supports Theorem 1 (Section III.B), which reduces the four-dimensional candidate strategy space to the one-dimensional family $(-q, -q, +q, +q)$, enabling the equilibrium analysis in Sections III.C–III.D and the two-boundary decomposition.

Theorem 1 (Payoff Equality). Under any symmetric voter distribution ($v_k = v_{-k}$), every Nash equilibrium has the mirror-symmetric form $(-q, -q, +q, +q)$ and gives each candidate payoff exactly $1/4$.

Proof of Theorem 1. The proof establishes that every NEQ under symmetric voters ($v_k = v_{-k}$) must have the mirror-symmetric co-located form $(-q, -q, +q, +q)$ for some $q \in \{0, \ldots, M\}$.

Step 1: Co-location. Suppose $\mathcal{D}_1$ is at position a and $\mathcal{D}_2$ at position $b \in \{-M, \ldots, +M\}$, with $a < b$. With distance-dependent turnout and symmetric bimodal voters, the $\mathcal{D}$ primary votes are generically unequal: the candidate closer to the bimodal peak wins outright (a tied primary requires co-location). The loser receives payoff 0 and can profitably deviate to any configuration yielding a positive payoff — which generically exists. Therefore, at any non-degenerate equilibrium, both candidates within each party are co-located: the profile is (q, q, r, r).

Step 2: Mirror symmetry. At (q, q, r, r) with $q + r \neq 0$, GE Neutrality does not apply, and one party loses the general election. The losing party's candidates receive payoff 0. Consider $\mathcal{D}_1$ at q deviating to $-r$. Co-partisan $\mathcal{D}_2$ remains at q. *Primary survival:* If $-r < q$ (i.e., the deviator moves to the opposite side of the co-partisan), then $\mathcal{D}_1$ at $-r$ is closer to the left-side base and wins the $\mathcal{D}$ primary outright. If $-r = q$, the two co-partisans tie the primary ($1/2$ each). If $-r > q$ (both on the same side), then $\mathcal{D}_1$ at $-r$ is at least as extreme as $\mathcal{D}_2$ at q and captures the nearer tail voters, again winning or tying the primary. *GE payoff:* In all cases the deviator reaches the GE with positive probability. At the GE matchup $(-r, r)$, GE Neutrality gives GE $= 1/2$, so expected payoff is at least $1/2 \times 1/2 = 1/4 > 0$. This strictly improves on payoff 0, breaking the equilibrium. Therefore $q + r = 0$, i.e., $r = -q$.

Step 3: Payoff calculation. At $(-q, -q, +q, +q)$, GE Neutrality gives $\mathrm{GE}(-q, +q) = 1/2$. Each candidate ties the primary (co-located) and ties the GE. Payoff: $1/2 \times 1/2 = 1/4$.

$\square$

In a recent physics paper in *Physica A*, Kim et al. (2022) have four candidates choose positions from voters who are uniformly distributed. Their analysis considers only uniform voter distributions with perfect turnout. They are interested in dynamics arising from candidates randomly sampling locations, given that the others are held fixed. This shows full convergence at 1/4 win probability equilibrium. Every point on the uniform can be an equilibrium — as long as it has the form $(-q, -q, q, q)$. This continuum of equilibria is an artifact of the continuous position space: an infinitesimal deviation toward the center barely changes primary votes but immediately makes the deviator farther from the party median, so the primary is lost. In the discrete-step model, even with $\tau = 0$ and uniform voters, a one-step centrist deviation can tie or win the primary while also winning the general election, so only positions close to the center may survive as equilibria. Indeed, for a continuous position function, with even an epsilon voter turnout decay, the Kim et al equilibria collapse from completely covering the entire spectrum (i.e., indeterminacy) to a single point at the center.

$\mathcal{A}$.2 Proof of the Equilibrium Band Theorem (Theorem 2)

This proof supports Theorem 2 (Section V.B), which establishes the contiguous band structure underlying the equilibrium existence diagrams in Figure 3 ($N = 7$) and Figure 6 ($N = 1{,}001$).

Symbol	Description
$V_{\mathcal{D}}, V_{\mathcal{R}}$	Turnout-weighted vote totals for the $\mathcal{D}$ and $\mathcal{R}$ nominees $V_{\mathcal{D}} = \sum_{k<0} v_k\,\lambda^{\lvert k+j\rvert} + \frac{1}{2}\,v_0\,\lambda^j;\quad V_{\mathcal{R}} = \sum_{k>0} v_k\,\lambda^{\lvert k-j\rvert} + \frac{1}{2}\,v_0\,\lambda^j$
$c_k = 100 - 10 \cdot \lvert k\rvert$	Gap between uniform and bimodal densities at position k
$D_{\mathrm{GE}}(j, \lambda)$	Denominator of the GE boundary $\alpha^+(j)$
$D_{\mathrm{PE}}(j, \lambda)$	Denominator of the PE boundary $\alpha^-(j)$

Theorem 2 (NEQ Band Contiguity). For $N = 2{\cdot}M + 1$ positions with symmetric bimodal voter distribution $v_k(\alpha) = (1 - \alpha) \cdot 100 + \alpha \cdot 10 \cdot \lvert k\rvert$ and turnout decay $\tau \in (0,1)$:

(i) At any (α, τ), the set of $j \in \{0, \ldots, M\}$ such that $(-j, -j, j, j)$ is a Nash equilibrium forms a **contiguous interval** $\{j_{\min}, \ldots, j_{\max}\}$.

(ii) Each NEQ $(-j, -j, j, j)$ exists for $\alpha \in [\alpha^-(j, \tau), \alpha^+(j, \tau)]$, where $\alpha^+(j)$ is the **GE boundary** (non-decreasing in j) and $\alpha^-(j)$ is the **primary boundary** (non-decreasing in j).

(iii) Contiguity follows: if $j_1 < j_2$ are both NEQ at some α, then every $j_1 \le j \le j_2$ is also NEQ.

Proof of Theorem 2. Throughout this proof, write $\lambda \equiv 1 - \tau$ for brevity. Recall that each symmetric NEQ has the form $(-j, -j, +j, +j)$ for position $j \in \{0, \ldots, M\}$, where $M = (N - 1)/2$. We prove that the GEC upper bound $\alpha^+(j)$ and the PEC lower bound $\alpha^-(j)$ are both non-decreasing in j. Contiguity then follows by the sandwich argument in the main text.

GE Boundary $\alpha^+(j)$. At $(-j, -j, +j, +j)$, the extreme deviation from $-j$ to $-(j+1)$ produces a GE between the $\mathcal{D}$ nominee at $-(j+1)$ and the $\mathcal{R}$ nominee at $+j$. Write $V_{\mathcal{D}}$ and $V_{\mathcal{R}}$ for the turnout-weighted vote totals of the $\mathcal{D}$ and $\mathcal{R}$ nominees, respectively. The midpoint between $-(j+1)$ and $+j$ is at $-1/2$, so positions ≤ -1 vote $\mathcal{D}$ and positions ≥ 0 vote $\mathcal{R}$.

The key identity: the uniform contribution to $V_{\mathcal{R}} - V_{\mathcal{D}}$ equals exactly $100\lambda^{M-j}$, independent of bimodal weights. This follows from the telescoping sum $\sum_{d=0}^{M} \lambda^{|d-j|} - \sum_{d=1}^{M} \lambda^{|d-(j+1)|} = \lambda^{M-j}$.

Setting $V_{\mathcal{D}} = V_{\mathcal{R}}$ and collecting the α-dependent terms:

$$\alpha^+_{\mathrm{GE}}(j) = \frac{100\,\lambda^{M-j}}{D_{\mathrm{GE}}(j,\lambda)}, \tag{6}$$

where $D_{\mathrm{GE}}(j,\lambda) = \sum_{k=0}^{M} c_k \lambda^{|k-j|} - \sum_{i=1}^{M} c_i \lambda^{|i-(j+1)|}$ and $c_k = 100 - 10 \cdot |k|$.

As j increases: the numerator $100\lambda^{M-j}$ increases (smaller exponent), while D_{GE} decreases (the deficit terms shift toward positions where c_k is smaller). Both effects push α^+ upward.

Primary Boundary $\alpha^-(j)$. For $j \geq 1$, the centrist deviation from $-j$ to $-(j-1)$ enters the $\mathcal{D}$ primary against the partner at $-j$. The deviator at $-(j-1)$ captures voters at the centrist positions $-(j-1), \ldots, 0$; the partner at $-j$ captures the extreme positions $-M, \ldots, -j$.

For $j \leq \lfloor M/2 \rfloor$, the partner has more positions at uniform, so $\alpha^-(j) = 0$ (the centrist deviation always loses the primary, even at $\alpha = 0$). For $j > \lfloor M/2 \rfloor$, bimodality is needed to protect the partner, and the primary tie gives

$$\alpha^-_{\mathrm{PE}}(j) = \frac{100\left|\lambda^j - \lambda^{M-j+1}\right|}{(1-\lambda)\, D_{\mathrm{PE}}(j,\lambda)}. \tag{7}$$

As j increases, the numerator grows (the partner has fewer nearby positions), and D_{PE} shrinks (the partner's extreme voters are more heavily weighted by bimodality). Both effects push α^- upward.

Contiguity. Suppose $(-j_1, -j_1, j_1, j_1)$ and $(-j_2, -j_2, j_2, j_2)$ are both equilibria at some α, with $j_1 < j_2$. For any j with $j_1 \leq j \leq j_2$: $\alpha^+(j) \geq \alpha^+(j_1) \geq \alpha$ (non-decreasing upper bound), and $\alpha^-(j) \leq \alpha^-(j_2) \leq \alpha$ (non-decreasing lower bound). Therefore $\alpha \in [\alpha^-(j), \alpha^+(j)]$, and $(-j, -j, j, j)$ is also an equilibrium. □

$\mathcal{A}$.3 Proof of the Asymmetric Voter Flattening Collapse (Theorem 3)

This proof supports Theorem 3 (Section V.D), which generalizes the flattening collapse illustrated for $N = 7$ in Section IV and for $N = 1{,}001$ in Section V.A.

Theorem 3 (Flattening Collapse). Consider the symmetric bimodal distribution with $\alpha_L = \alpha_R = \alpha$ sufficiently high that $(-j, -j, +j, +j)$ with $j > 0$ is a Nash equilibrium. Then for any perturbation that reduces α_R to $\alpha - \delta$ with $\delta > 0$:

(i) **GE Neutrality breaks.** $\mathrm{GE}(-j, +j) < 1/2$ under the perturbed distribution: $\mathcal{D}$ loses the general election at the symmetric profile.

(ii) **Primary trap persists.** No $\mathcal{D}$-side deviation simultaneously wins the $\mathcal{D}$ primary and the GE. More extreme deviations win the primary but lose the GE; centrist deviations lose the primary.

(iii) **Replacement NEQ.** For $\delta \geq \delta^*(j, \alpha, \tau, M)$, there exists a replacement NEQ with both $\mathcal{D}$ candidates on the $\mathcal{R}$ side of the spectrum: $(+s, +s, +j, +j)$ for some $s > 0$.

The proof of Theorem 3 proceeds through five lemmas. Throughout, $j \in \{0, \ldots, M\}$ denotes the NEQ position (so the profile is $(-j, -j, +j, +j)$), and $\lambda \equiv 1 - \tau$.

Symbol	Description
$\tilde{v}$	Perturbed voter distribution: $\alpha_R = \alpha - \delta$ with $\delta > 0$
Δv_k	Change in voter count at position k from the perturbation

Lemma 1 (GE Neutrality is a knife-edge) *For any symmetric distribution ($v_k = v_{-k}$) and any $j \in \{0, \ldots, M\}$:* $\mathrm{GE}(-j, +j) = 1/2$.

Proof. Write the turnout-weighted vote totals for the $\mathcal{D}$ nominee at $-j$ and the $\mathcal{R}$ nominee at $+j$: $V_{\mathcal{D}} = \sum_{k<0} v_k \lambda^{|k+j|} + \frac{1}{2} v_0 \lambda^j$ and $V_{\mathcal{R}} = \sum_{k>0} v_k \lambda^{|k-j|} + \frac{1}{2} v_0 \lambda^j$. The substitution $k \to -k$ in $V_{\mathcal{D}}$, using $v_{-k} = v_k$ and $|-k+j| = |k-j|$, gives $V_{\mathcal{D}} = V_{\mathcal{R}}$. □

Lemma 2 (Perturbation breaks GE) *Let $\tilde{v}$ be the asymmetric distribution obtained by setting $\alpha_R = \alpha - \delta$ for $\delta > 0$ (flattening $\mathcal{R}$'s voter distribution toward uniform). Then* $\mathrm{GE}(-j, +j) < 1/2$ *under $\tilde{v}$ for all $j \in \{1, \ldots, M\}$.*

Proof. The perturbation changes the voter count at each $\mathcal{R}$-side position $k > 0$ by $\Delta v_k = \delta \cdot c_k$, where $c_k = 100 - 10 \cdot |k|$ is the gap between uniform and bimodal densities. Let $k^* = \lfloor 100/10 \rfloor = 10$ be the cutoff: for $|k| < k^*$, $c_k > 0$ (flattening adds voters); for $|k| \geq k^*$, $c_k \leq 0$ (flattening removes voters). The net change in $\mathcal{R}$'s GE total is $\Delta V_{\mathcal{R}} = \delta \sum_{k=1}^{M} c_k \lambda^{|k-j|}$. To sign this sum, decompose it as $S^+ + S^-$ where $S^+ = \sum_{k<k^*} c_k \lambda^{|k-j|} > 0$ and $S^- = \sum_{k \geq k^*} c_k \lambda^{|k-j|} \leq 0$. The positive terms have large c_k values (up to $c_1 = 90$) and contribute even when far from the nominee, while the negative terms have small $|c_k|$ values (at most $10(M - k^*)$) and carry exponentially decaying turnout weights for positions far from the nominee. Thus $S^+ + S^- > 0$, giving $\Delta V_{\mathcal{R}} > 0$, while $V_{\mathcal{D}}$ is unchanged. Therefore $\text{GE}(-j, +j) < 1/2$.[21] □

Lemma 3 (Primary trap on own side) *At $(-j, -j, +j, +j)$ under the perturbed distribution ($\alpha_R < \alpha_L$), no unilateral $\mathcal{D}$ deviation yields positive payoff.*

Proof. Let $p \in \{-M, \ldots, +M\}$ denote the deviation position for $\mathcal{D}_1$ (with $\mathcal{D}_2$ fixed at $-j$). (a) More extreme deviations ($p \leq -j$): win/tie the primary but $\text{GE}(p, +j) < \text{GE}(-j, +j) < 1/2$ (Lemma 2). Payoff 0. (b) Centrist deviations ($-j < p \leq 0$): the partner at $-j$ captures the extreme-left bimodal peak and wins the primary. Payoff 0. (c) Cross-party deviations ($p > 0$): the partner at $-j$ still holds the entire left-side base and wins the primary. Payoff 0. □

Lemma 4 (Payoff degeneracy) *Under the perturbed distribution, $(-j, -j, +j, +j)$ is a weak equilibrium with payoff vector $(0, 0, 1/2, 1/2)$ for all $j > 0$.*

Proof. $\mathcal{D}$ loses the GE (Lemma 2), so both $\mathcal{D}$ candidates get payoff 0. No $\mathcal{D}$ deviation yields positive payoff (Lemma 3). $\mathcal{R}$ ties the primary and wins the GE, giving payoff 1/2; the perturbation only helps $\mathcal{R}$ in the GE, leaving $\mathcal{R}$'s primary incentives unchanged. □

Lemma 5 (Replacement NEQ) *There exists a threshold $\delta^* > 0$ such that for $\delta \geq \delta^*$, a replacement equilibrium of the form $(+s, +s, +j, +j)$ exists, where $s \in \{1, \ldots, j-1\}$ is a centrist $\mathcal{D}$ position on the $\mathcal{R}$ side.*

Proof sketch. At $(+s, +s, +j, +j)$, both $\mathcal{D}$ candidates are co-located at $+s$, so they tie the primary. The $\mathcal{D}$ nominee at $+s$ is more centrist than the $\mathcal{R}$ nominee at $+j$; for large enough δ, the added centrist mass on the $\mathcal{R}$ side (from the flattening perturbation) tips the GE to $\mathcal{D}$. Leftward

---

[21] The "near the nominee" intuition is imprecise when j is large, because at $k \approx j$ one has $c_k < 0$ (voters are removed, not added). The formal sign claim requires verifying $|S^-| < S^+$ for all $j \in \{1, \ldots, M\}$. Since c_k is linear in $|k|$ and the turnout weights $\lambda^{|k-j|}$ decay exponentially in distance from the nominee, the positive contributions (centrist positions with large c_k) dominate the negative contributions (extreme positions with small $|c_k|$) for any $\lambda < 1$.

deviations by a $\mathcal{D}$ candidate may win the primary (capturing the bimodal base) but lose the GE; rightward deviations lose the primary to the partner. The threshold δ^* is the smallest δ such that all primary-winning deviations also lose the GE. □

Remark (degeneracy-replacement gap). Any $\delta > 0$ makes the symmetric equilibria degenerate (Lemma 4), but the replacement equilibrium requires $\delta \geq \delta^*$ (Lemma 5). In this gap, the only equilibria are the degenerate $(0, 0, 1/2, 1/2)$ profiles. The gap values are small: δ^* is 1–5% of α for $N \leq 15$.

Remark on partisan loyalty. With a loyalty parameter $\epsilon > 0$, symmetric equilibria are unaffected (bonuses cancel by symmetry), but cross-party deviation becomes harder. Small ϵ shifts the collapse threshold upward; large ϵ blocks it entirely, sustaining the degenerate equilibrium where the flatter party permanently wins. Under closed primaries (party-based assignment), cross-party deviation is foreclosed by construction.

$\mathcal{A}$.4 Convergence of Equilibrium Structure at Large N

This appendix provides the formal limiting results summarized in the last row of Table 3 (Section V.B).

Symbol	Description
D^*	Effective reach: $\lceil 18/\lvert\log_{10}\lambda\rvert\rceil$; distance beyond which turnout is negligible
$k = M - j^*$	Distance of the most extreme viable NEQ from the grid boundary
$P(\alpha) = U + \alpha B$	Primary vote balance (partner minus deviator), linear in α

Proposition 1 *Fix $\tau \in (0,1)$ with $\lambda = 1 - \tau$. As $N = 2M{+}1 \to \infty$:*

(i) $\alpha^+(0) \to 0$*: the centrist equilibrium is destroyed by any positive bimodality.*

(ii) $j^*/M \to 1$*: the most extreme viable equilibrium approaches the grid boundary.*

(iii) $\alpha^-(j^*) \to 0$*: this most extreme equilibrium appears at vanishing bimodality.*

Proof. Throughout, write $\lambda = 1 - \tau$ and define the **effective reach** $D^* = \lceil 18/\lvert\log_{10}\lambda\rvert\rceil$, the distance beyond which turnout $\lambda^d < 10^{-18}$ is negligible. For fixed τ, D^* is a constant independent of N.

Part (i). From (6), $\alpha^+(0) = 100\lambda^M / D_{\mathrm{GE}}(0,\lambda)$. The denominator $D_{\mathrm{GE}}(0,\lambda)$ converges to a positive constant as $M \to \infty$ (it is a convergent power series in λ), while $\lambda^M \to 0$ exponentially.

Parts (ii) and (iii). Let $k = M - j^*$ denote the distance from the extreme edge. At the profile $(-j, -j, +j, +j)$ with $j = M - k$, the binding constraint is the centrist deviation: a $\mathcal{D}$ candidate at $-j$ moves to $-(j-1)$. Because $k < D^*$, the partner at $-j$ draws turnout from $k+1$ positions toward the extreme edge, while the deviator at $-(j-1)$ draws from D^*+1 positions toward center.

The primary vote balance (partner minus deviator) is linear in α: $P(\alpha) = U + \alpha B$, where the uniform component $U < 0$ (the deviator has a territorial advantage) satisfies $|U| = 100\lambda^{k+1}/(1-\lambda) + O(\lambda^M)$, and the bimodal component B captures the net effect of the bimodal weights $10 \cdot |k|$.

Decomposing B into a positive constant and an M-dependent term:

$$B = \frac{10(1+\lambda)}{(1-\lambda)^2} - \frac{10\,M\lambda^{k+1}}{1-\lambda} + O\left(\frac{1}{M}\right).$$

For the profile to be NEQ at some $\alpha > 0$, we need $B > 0$. The leading terms give:

$$M\lambda^{k+1} < \frac{1+\lambda}{1-\lambda}.$$

Taking logarithms, this forces $k > \big(\ln M - C\big)/|\ln \lambda|$ for a constant C, so $k = \Theta(\log M)$ and $j^*/M = 1 - O(\log M/M) \to 1$, proving (ii).

For (iii), at $j = j^*$ the quantity $M\lambda^{k+1}$ is bounded (by the viability condition), so $|U| = 100\lambda^{k+1}/(1-\lambda) = O(1/M)$ while B stays bounded away from zero. The threshold $\alpha^-(j^*) = |U|/B = O(1/M) \to 0$. □

Remark. The convergence $\alpha^-(j^*) \to 0$ is non-monotonic (Table 3): the distance $k = M - j^*$ takes integer values, so $M\lambda^{k+1}$ oscillates within the interval $\big(\lambda(1+\lambda)/(1-\lambda), (1+\lambda)/(1-\lambda)\big)$ as M grows, producing corresponding oscillations in α^-. The envelope, however, decays at rate $O(1/M)$.

$\mathcal{A}$.5 Geometry of the Principled-Candidate Squeeze

This appendix provides the geometric characterization of unwinnable cells (X marks) in the heatmaps of Figure 5 (Section V.A).

Proposition 2 (Squeeze Geometry) *Under symmetric voters, every unwinnable configuration lies strictly southwest of the equilibrium diagonal. That is: if $\mathcal{D}_2$ is at position p (with $p < 0$) and $\mathcal{R}_1 = \mathcal{R}_2$ at position q (with $q > 0$) and $|p| \le q$ (the $\mathcal{D}$ partner is at least as centrist as the mirror of $\mathcal{R}$), then $\mathcal{D}_1$ achieves strictly positive payoff.*

Proof. Consider $\mathcal{D}_1$ co-locating with $\mathcal{D}_2$ at position p.

(1) **Primary.** Identical positions yield identical vote totals, so $\mathcal{D}_1$ wins the primary with probability 1/2.

(2) **GE.** $\mathcal{D}$ at p vs. $\mathcal{R}$ at q. The midpoint is $(p+q)/2 \geq 0$ (since $|p| \leq q$), which is weakly right of center, so $\mathcal{D}$ captures at least half the electorate. When $|p| < q$, $\mathcal{D}$ strictly wins; when $|p| = q$, GE Neutrality gives each party share 1/2.

(3) **Payoff.** $\mathcal{D}_1$'s payoff $= 1/2 \times \Pr[\mathcal{D} \text{ wins GE}] \geq 1/2 \times 1/2 = 1/4 > 0$.

□

Remark. The proposition implies that X marks in the heatmap (Figures 5–7) are confined to the region where $\mathcal{D}_2$'s position p satisfies $|p| > q$—the southwest triangle below the equilibrium diagonal. As α increases toward 1 and the equilibrium band narrows, this southwest region shrinks, explaining the observed reduction in unwinnable cells (from 65 at $\alpha = 0.80$ to 0 at $\alpha = 1.00$).

$\mathcal{B}$ Appendix: Valence

The base model assumes all four candidates are identical except in their chosen positions. In practice, candidates differ on dimensions that all voters value equally: perceived competence, integrity, fundraising ability, name recognition, or incumbency advantage. Following Stokes (1963), these non-positional attributes are collectively called *valence*. Unlike policy position, which voters evaluate relative to their own ideal point, valence is universally preferred: every voter would rather support a more competent candidate, all else equal.

$\mathcal{B}$.1 Specification

Each candidate i has an exogenous valence $\nu_i \geq 0$. Different candidates may have different valences — a popular incumbent might have $\nu_i = 0.10$ while an unknown challenger has $\nu_j = 0$. Valence enters as a multiplicative advantage on turnout-weighted votes: if candidate i at position p_i would receive base votes V_i^{base} under the standard distance-turnout rule (1), their effective votes become

$$V_i = (1 + \nu_i) \cdot V_i^{\text{base}}. \tag{8}$$

Vote *choice* (which candidate a voter supports) is unchanged by valence; valence scales the *count*, not the *direction*, of votes. This corresponds to a mobilization interpretation: the high-valence candidate's supporters are more motivated to turn out. The multiplicative form applies identically in both the primary and the general election. At $\nu_i = 0$ for all i, the base model is recovered exactly.

$\mathcal{B}$.2 A Numerical Example

Consider the extreme equilibrium $(-3,-3,+3,+3)$ at $\alpha = 1$, $\tau = 0.25$ ($\lambda = 0.75$) with $N = 7$.

Starting point: the zero-valence equilibrium. Both $\mathcal{D}$ candidates sit at −3, both $\mathcal{R}$ candidates at +3. Since co-located candidates split the primary 50–50 and symmetric positions give GE = 1/2, each candidate's payoff is $1/2 \times 1/2 = 1/4$.

Adding valence at the equilibrium position. Suppose $\mathcal{D}_1$ acquires a 10% mobilization advantage ($\nu_1 = 0.10$) while the other three candidates retain $\nu = 0$. At the equilibrium position $(-3,-3,+3,+3)$, both $\mathcal{D}$ candidates draw from the same voters, so their base votes are identical. But $\mathcal{D}_1$'s effective votes are 10% higher, so $\mathcal{D}_1$ wins the $\mathcal{D}$ primary with certainty. In the general election, $\mathcal{D}_1$ at −3 faces $\mathcal{R}$ at +3; by symmetry the base votes are equal, but valence tips the balance. Payoffs: $\mathcal{D}_1 = 1$, $\mathcal{D}_2 = 0$, $\mathcal{R}_1 = \mathcal{R}_2 = 0$. Valence already disrupts the equal-payoff property of the equilibrium.

Deviation to −2: the primary trap. Now consider whether $\mathcal{D}_1$ deviates from −3 to −2 (one step toward center), giving the profile $(-2,-3,+3,+3)$. In the $\mathcal{D}$ primary, the partner at −3 holds the bimodal peak while the deviator at −2 draws from weaker interior voters:

$$V_{\text{partner at }-3} = \upsilon_{-3} = 30,$$
$$V_{\text{deviator at }-2} = \upsilon_{-2} + \upsilon_{-1}\lambda + \upsilon_0\lambda^2 = 20 + 10(0.75) + 0 = 27.5.$$

Without valence, the partner wins (30 > 27.5) and the deviator's payoff drops from 1/4 to 0. This is the primary trap: the extreme incumbent's positional advantage with the base is too large.

In the general election, the deviator at −2 actually *loses* against $\mathcal{R}$ at +3 at $\alpha = 1$: the GE share is ≈ 49.7%. (At lower α, it would be above 50%.) So, without valence, both constraints block the deviation.

Valence breaks the trap. With $\nu_1 = 0.10$, $\mathcal{D}_1$'s effective primary votes become $1.10 \times 27.5 = 30.25 > 30$: the primary flips. In the GE, $\mathcal{D}_1$'s effective votes become $1.10 \times 50 = 55$ against $\mathcal{R}$'s 50.625: the GE also flips. $\mathcal{D}_1$ wins both stages. Payoffs: $\mathcal{D}_1 = 1$, $\mathcal{D}_2 = 0$, $\mathcal{R}_1 = \mathcal{R}_2 = 0$—the same as staying at −3.

Since $\mathcal{D}_1$ gets payoff 1 at both the equilibrium position and after deviating, the deviation is indifferent, not strictly profitable.[22] The **critical valence** $\nu^* = 30/27.5 - 1 \approx 9.1\%$ marks the threshold where the primary trap just breaks; the GE requires only $\nu^*_{\text{GE}} = 50.625/50 - 1 = 1.25\%$, so the primary is the binding constraint. Whether 9.1% is a large or small quality differential is an empirical question: it might correspond to a former governor challenging a first-term legislator, or it might correspond to a modest difference in campaign organization.

$\mathcal{B}$.3 Critical Valence for Each Equilibrium

The question generalizes: starting from any symmetric equilibrium where all candidates have zero valence, how much valence must a single deviating candidate acquire to profitably change position? The **critical valence** ν^* measures the robustness of each equilibrium against quality differentials.

Since voter masses are linear in α — $v_k(\alpha) = 100 - c_k\alpha$ with $c = (70, 80, 90, 100, 90, 80, 70)$ — all vote totals are linear in α, and ν^* is a rational function of α and λ.

For centrist equilibria, the binding deviation goes more extreme (the deviator wins the primary but loses the GE); valence helps the deviator win the GE. For the extreme equilibrium, the binding deviation goes toward the center (the deviator wins the GE but loses the primary); valence helps the deviator win the primary. Specifically:

Equilibrium	Binding deviation	Constraint	ν^* at $\alpha = 1$
$(0,0,0,0)$	$0 \to -1$	GE$(-1,0) = 1/2$	(not a NEQ)
$(-1,-1,+1,+1)$	$-1 \to -2$	GE$(-2,+1) = 1/2$	(not a NEQ)
$(-2,-2,+2,+2)$	$-2 \to -3$	GE$(-3,+2) = 1/2$	(not a NEQ)
$(-3,-3,+3,+3)$	$-3 \to -2$	Primary tie	9.1%

The closed-form expressions are:

$$\nu^*_{44}(\alpha) = \frac{100\lambda^3 - 10\alpha(1+\lambda+\lambda^2+7\lambda^3)}{100(1+\lambda+\lambda^2) - \alpha(90+80\lambda+70\lambda^2)}, \tag{9}$$

$$\nu^*_{33}(\alpha) = \frac{100\lambda^2 - 10\alpha(1+2\lambda+7\lambda^2)}{100(1+2\lambda) - \alpha(80+160\lambda)}, \tag{10}$$

---

[22] The equilibrium still exists in the Nash sense at $\nu = \nu^*$. For $\nu > \nu^*$, the deviating candidate strictly prefers the new position (winning both primary and GE with certainty vs. sharing the primary at the original position), so the equilibrium is genuinely broken. The critical valence ν^* thus marks the boundary between strict and non-strict equilibrium, not the boundary of NE existence per se. We use it as a robustness measure.

each valid when the numerator and denominator are both positive (i.e., within the equilibrium region). For the extreme equilibrium, the primary constraint dominates:

$$\nu_{11}^*(\alpha) = \frac{v_{-3} - (v_{-2} + v_{-1}\lambda + v_0\lambda^2)}{v_{-2} + v_{-1}\lambda + v_0\lambda^2} \,. \tag{11}$$

Setting $\nu^* = 0$ in each expression recovers the base-model equilibrium boundary $\alpha^*(\lambda)$ from Section D.

$\mathcal{B}$.4 Valence Phase Diagram

Figure $\mathcal{B}$.1 shows the full picture. For each symmetric equilibrium, the shaded region below its $\nu^*(\alpha)$ curve is where the equilibrium survives. The $\nu = 0$ line (horizontal axis) recovers the base-model boundaries from Figure 3.

Three features stand out. First, the semi-extreme equilibrium $(-2, -2, +2, +2)$ is extraordinarily robust to valence: its curve runs along the top of the diagram ($\nu^* > 25\%$) across most of the α range. The binding deviation ($-2 \to -3$) must overcome both the primary and the GE, and the GE at position -3 vs $+2$ is severely disadvantaged.

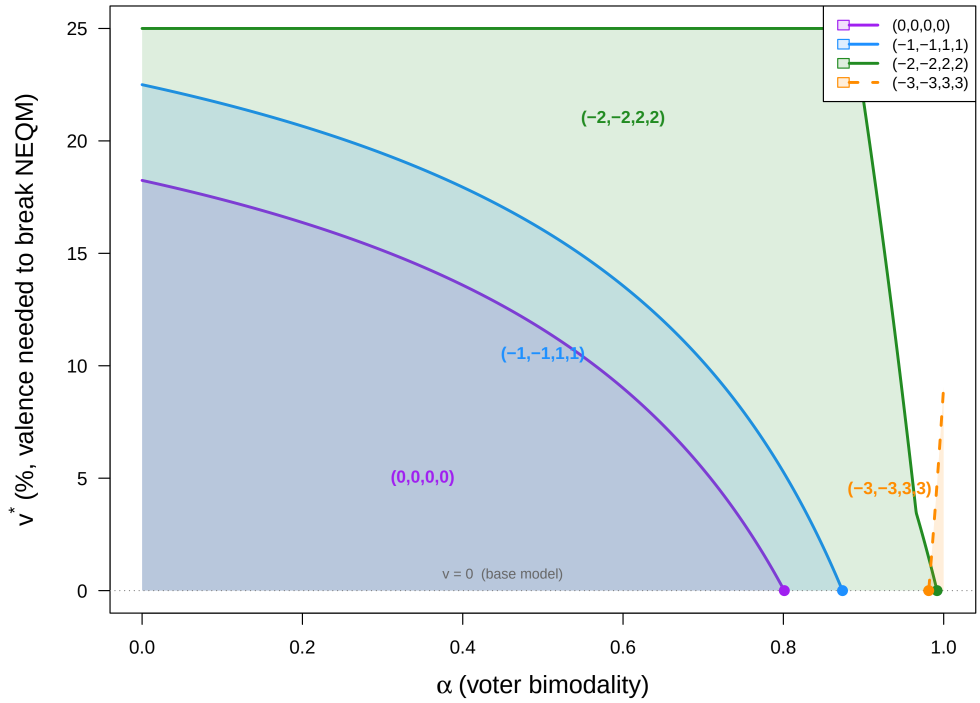
Second, the extreme equilibrium $(-3, -3, +3, +3)$ has a primary trap of moderate strength: $\nu^* \approx 9.1\%$ at $\alpha = 1$. The primary constraint dominates ($\nu^*_{\text{pri}} = 9.1\%$ vs $\nu^*_{\text{GE}} = 1.25\%$). The trap is specifically about the primary, not the general election: valence breaks it by letting the moderate win on quality despite a positional disadvantage with the base.

Third, centrist equilibria are fragile at their boundaries — $\nu^* \to 0$ as $\alpha \to \alpha^*$ — so any quality differential, however small, breaks the centrist equilibrium at the margin. But at low α (near-uniform voters), these equilibria are robust: $\nu^* \approx 18\%$ and 22% at $\alpha = 0$.

Valence thus narrows the equilibrium band from both ends. For centrist equilibria, valence helps an extreme challenger win the general election. For the extreme equilibrium, valence helps a moderate challenger win the primary. The two mechanisms are different, but the net effect is the same: fewer equilibria survive when candidates differ in quality.

fig:valence in voting1.tex, March 9, 2026

Appendix Figure $\mathcal{B}$.1: Valence Phase Diagram ($N = 7$, $\tau = 0.25$)



Explanations: Existence regions for symmetric equilibria in (α, ν) space. Each equilibrium survives in the shaded region *below* its $\nu^*(\alpha)$ curve. Colored dots on the horizontal axis mark the base-model boundaries ($\nu = 0$) from Figure 3. The semi-extreme equilibrium $(-2, -2, +2, +2)$ (green) is off-scale for most of the α range ($\nu^* > 25\%$). The extreme equilibrium $(-3, -3, +3, +3)$ (orange, dashed) requires $\nu^* \approx 9.1\%$ at $\alpha = 1$ — a moderate primary-trap strength.

Interpretation: Valence narrows the equilibrium band from both ends. Centrist equilibria are broken by extreme challengers who use valence to win the GE; extreme equilibria are broken by moderate challengers who use valence to win the primary. The semi-extreme equilibrium is the most robust to candidate quality differentials.