Divide and Diverge*

Giampaolo Bonomi†

April 6, 2026

Abstract

This paper shows that competing political parties can benefit from polarization. I study a probabilistic voting model with ex-ante identical parties and decreasing marginal utility from political power. In equilibrium, parties offer different platforms: by targeting different voter blocks, they insure against electoral risk and stabilize power. With one policy dimension, each party gains when voters on opposite sides of the median become more divided, even if only the opponent's supporters become more extreme. With multiple dimensions, parties gain from ideological coherence, the alignment of disagreement across issues. The results speak to polarizing communication, party identity, and electoral institutions.

JEL Codes: D72, D74, D78, C72, D81.

Keywords: polarization; conflict; elections; zero-sum thinking; identity.

*I am grateful to Renee Bowen, Peter Buisseret, Carlo Cusumano, Germán Gieczewski, Gleason Judd, Anton Kolotilin, John Londregan, Simone Galperti, Nicola Gennaioli, Aram Grigoryan, Matias Iaryczower, Federica Izzo, Freddie Papazyan, Kris Ramsay, Joel Sobel, Guido Tabellini, and participants at the Stony Brook International Conference on Game Theory for helpful comments and suggestions.

†Princeton University. Email: bonomi@princeton.edu

1 Introduction

Political polarization has become a defining feature of contemporary democratic politics. In many countries, and most notably in the United States, voters increasingly sort into two opposing factions that take systematically different positions across a wide range of policy issues, a pattern mirrored by political parties' platforms and rhetoric. A large empirical literature documents rising ideological distance, partisan sorting, and cultural separation in both the electorate and among political elites (Gentzkow, 2016; Gentzkow et al., 2019; McCarty, 2016; Bertrand and Kamenica, 2023; Bonomi et al., 2021). Understanding why political competition produces these outcomes—and why political actors may want to exacerbate societal divisions—remains a central question in political economy.

A wide range of mechanisms have been linked to these developments. Some contributions emphasize demand-side forces, including structural economic shocks, identity-based belief formation, and information frictions that generate disagreement among voters (e.g., Autor et al., 2020; Alesina et al., 2020; Bonomi et al., 2021; Gennaioli and Tabellini, 2025; Bowen et al., 2023; Stantcheva, 2021). Others show that political actors and media outlets shape political conflict via selective mobilization, framing, and the emphasis of divisive issues (e.g., Martin and Yurukoglu, 2017; DellaVigna and Kaplan, 2007; Kim et al., 2022; Flores et al., 2022; Cohen, 2003; Feddersen and Adams, 2022).

But why would political parties want to divide voters? This paper shows that electoral competition can itself generate incentives for polarization and, crucially, incentives to split the electorate into two opposing and loyal blocs. Deeper voter divisions reshape the electoral constraint: they let parties move apart without fear of large electoral reversals, making polarization electorally profitable when returns to power are decreasing. In short, divisions stabilize power.

I focus on a deliberately disciplined setting in which parties are ex ante identical, motivated purely by political power, and face no ideological differences or asymmetric voter bases. Aggregate political power is constant-sum, so polarization cannot be driven by changes in total power. In this benchmark, any polarization that emerges must be traced to strategic incentives within electoral competition itself. While party differences are likely important in the real world, the analysis shows that polarization need not originate in them: competitive incentives alone can make a divided electorate attractive to parties, with existing asymmetries serving only to amplify a force that is already present under full symmetry.

The analysis builds on a probabilistic voting model with aggregate popularity shocks, which makes electoral outcomes sensitive to random forces such as economic news, scandals, or candidate-specific valence. The key departure from the standard framework is the assumption that parties have strictly concave utility over political power. Importantly, this is a restriction on how political power is translated into party payoffs, not on how electoral support maps into political power itself, so the setting allows for (super)majority premia and other discontinuities in the mapping from votes to power. The concavity assumption captures a broad class of environments in which marginal gains in influence, office control, or agenda-setting capacity yield diminishing returns to the party as an organization,[1] as formalized and discussed in appendix A.2.

The mechanism underlying the results of the paper is an insurance motive. When parties offer similar policy platforms, electoral outcomes depend heavily on aggregate shocks, exposing each party to substantial payoff risk. By contrast, platform differentiation allows parties to build loyal voter blocs and reduce the volatility of realized political power. When the utility from power rents is concave, the insurance value of differentiation outweighs its competitive cost, pushing parties to offer different platforms. This insurance logic delivers a tight link between equilibrium platform distance and parties' expected payoffs, which I then use to study how changes in the distribution and structure of voter conflict shape equilibrium outcomes. Indeed, recent evidence is consistent with this insurance channel: polarization makes voters less responsive to shocks and less likely to cross party lines (Moreira, 2025; Buttice and Stone, 2012; Davis and Mason, 2016).[2]

The model has a stark implication: the same force that makes party platforms diverge also gives both parties a joint incentive to harden voter divisions. I formalize this *divide-*

[1] In particular, the curvature is intended as a reduced-form property of how organizational control translates into rents/influence within parties, rather than intrinsic candidate risk aversion.

[2] Polarization makes voters less likely to punish their party after negative economic news (Moreira, 2025); larger ideological contrasts dampen the electoral effect of candidate quality (Buttice and Stone, 2012); and partisan-ideological sorting reduces split-ticket voting (Davis and Mason, 2016).

and-diverge logic through three results: platform divergence and two comparative-statics results on polarization and ideological coherence.

First, pure-strategy equilibrium convergence does not arise. Even when parties are identical and voters vote sincerely, equilibrium platforms diverge. This divergence is driven by parties' incentives to reduce exposure to aggregate electoral risk—a force that then shapes which changes in voter conflict raise equilibrium separation and payoffs. The result is complementary to Bernhardt et al. (2020), who show that payoff nonlinearities can generate platform divergence, provided that one candidate has a sufficiently strong primitive valence disadvantage. My mechanism, which relies on elections shaping the allocation of political power and decreasing returns to power, does not build on primitive differences between parties: it explains platform divergence in the context of more contested elections and delivers insights into voter polarization.

Second, I show that parties benefit not only from platform divergence, but also from an increase in voter polarization. In a unidimensional policy space, both parties' equilibrium payoffs rise when voters on opposite sides of the median become more extreme. Notably, a party may benefit from changes in the policy views of its opponent's supporters that move those voters further away from the party—even when this directly reduces its electoral support—because such changes induce the opponent to shift its platform in their direction. This strategic response pulls the opponent away from the party's own core supporters, increasing equilibrium platform separation and reducing the party's exposure to electoral risk. What follows is one of the most distinctive implication of the paper: each party has an incentive to create voter divisions, even in the absence of different, exogenously given, voter bases.[3]

Third, leveraging the potential-game nature of the game, I extend the equilibrium analysis to a multidimensional policy space, showing that parties benefit most when political disagreement is aligned across issues, so that voters sort into ideologically cohesive factions. Holding fixed the marginal distribution of opinions on each issue, greater align-

[3]Gennaioli and Tabellini (2025) show that in the presence of primitive party-voter associations, if voters trust the policy announcements of their party more than the ones of the opposite party, then politicians benefit from making their own supporters more extreme. In this paper, I show that parties can benefit from voter polarization even in the absence of primitive asymmetries, and even if the voters becoming more extreme are those more likely to support, in equilibrium, the opponent.

ment of voter views across dimensions allows parties to diverge further in equilibrium and raises expected payoffs. This result provides a formal explanation for why political conflict increasingly takes the form of bundled ideological divides rather than isolated disagreements on individual issues, as documented in recent empirical work (Gentzkow, 2016; Bonomi et al., 2021; Bertrand and Kamenica, 2023).

These results suggest a tension between political competition and voter welfare. While polarization increases parties' expected payoffs by reducing electoral risk, it exposes voters to more policy uncertainty. I show how intervening on institutional features, such as increasing the majority premium in the mapping from votes to political power, can attenuate polarization incentives by strengthening parties' incentives to appeal to pivotal voters. The analysis highlights a trade-off between representation and policy stability.

Finally, the framework is used to shed light on several features of contemporary political conflict. It provides a rationalization for why political actors emphasize divisive, zero-sum issues over common-interest policies; why polarizing political communication may be politically convenient; and why party identity and ideological alignment across issues play a central role in modern polarization. By isolating a supply-side mechanism in a symmetric benchmark, the model provides a natural foundation for studying feedback between party strategies and voter polarization.

While this paper focuses on strategic incentives, related work (Bonomi, 2026) provides experimental evidence that political communication can generate precisely the kind of cross-issue polarization and ideological alignment that, in equilibrium, amplify the incentives characterized here.

2 Related Literature

This work contributes to the literature on political polarization and electoral competition by isolating a supply-side mechanism through which parties have incentives to increase voter conflict, even in the absence of ex ante asymmetries across parties.

A central benchmark is the Downsian model, in which office-motivated parties competing for votes converge to the policy preferred by pivotal voters (Downs, 1957). Similar convergence results obtain in probabilistic voting models with risk-neutral candidates and constant-sum payoffs (Lindbeck and Weibull, 1987; Persson and Tabellini, 2002). Departures from convergence typically rely on ex ante asymmetries between parties, such as candidates' intrinsic policy preferences (Wittman, 1977; Hansson and Stuart, 1984; Lindbeck and Weibull, 1993), asymmetries in voter followings or historical party affiliations (Glaeser et al., 2005; Gennaioli and Tabellini, 2025), or candidate selection through primary elections (Brady et al., 2007; Hirano et al., 2009).

The mechanism studied here is complementary to these approaches. In a deliberately symmetric environment, diminishing marginal returns to power make platform differentiation an insurance device against aggregate popularity shocks. The perspective relates to work on the role of payoff nonlinearities in probabilistic voting environments (Zakharov and Sorokin, 2014; Sorokin and Zakharov, 2018; Bernhardt et al., 2020); here, the constant-sum mapping from votes to power and the symmetric strategic environment discipline when and why such nonlinearities translate into divergence and let the paper isolate a clean, bilateral, divide-and-diverge force. I bring to light a tight link between platform distance and equilibrium payoffs to derive comparative statics on voter conflict: in one dimension, equilibrium payoffs rise with polarization—including polarization driven by shifts among the opponent's supporters. In multiple dimensions, equilibrium payoffs rise with cross-issue alignment (ideological coherence).

These comparative statics connect to the growing literature on political polarization. Empirical work documents increasing polarization in the electorate and among political elites (Gentzkow, 2016; McCarty, 2016; Gentzkow et al., 2019), as well as rising ideological sorting and cultural distance (Bonomi et al., 2021; Bertrand and Kamenica, 2023). Existing theoretical explanations highlight both demand-side forces—such as belief distortions, identity-based reasoning, and learning from information—and supply-side forces, including strategic extremism and selective mobilization (Glaeser et al., 2005; Gennaioli and Tabellini, 2025). My results show that parties may benefit from exacerbating voter divisions even when doing so alienates voters, because sharper divisions relax competitive pressure and permit greater equilibrium differentiation.

Finally, the discussion connects to related themes on identity politics, zero-sum thinking, and institutions. Formal approaches emphasize how group identities shape political beliefs and behavior (Greene, 1999; Shayo, 2009; Bonomi et al., 2021; Gennaioli and Tabellini, 2025). Recent work studies perceptions of zero-sum conflict and their political consequences (Carvalho et al., 2023; Chinoy et al., 2026; Ali et al., 2024; Fiorina et al., 2011; Matsusaka and Kendall, 2022; Bueno de Mesquita and Dziuda, 2023). The institutional implications connect to evidence linking electoral rules to platform polarization (Matakos et al., 2016).

The remainder of the paper is organized as follows. In section 3, I introduce the model and establish the equilibrium divergence result. In sections 4 and 5, I derive the paper's main comparative statics—how voter polarization, alienation, and cross-issue alignment affect equilibrium platforms and party payoffs. Sections 6 and 7 are devoted to applications and welfare implications. Section 8 is a conclusion.

3 Model

Two parties, A and B, compete in an election. The electorate is described by a finite distribution of voter types $(x_i, s_i)_{i \in I}$, where $i = 1, \ldots, N$ indexes a voter type, $x_i \in \mathbb{R}^K$ denotes the vector of policy bliss points of type i, and s_i is the share of voters of that type in the population. The electorate is diverse, in the sense that the support of the voter distribution contains at least two voter types, and $x_i \neq x_j$ if $i \neq j$.[4] Voter i's utility from policy $x \in \mathbb{R}^K$ is given by

$$v(x, x_i) = \sum_{k \leq K} h^k(|x^k - x_i^k|),$$

where each $h^k : \mathbb{R} \to \mathbb{R}$ is differentiable, strictly concave, and strictly decreasing in $|x^k - x_i^k|$, with $h^{k\prime}(0) = 0$. This specification ensures smooth, single-peaked preferences.

The timing of the election is as follows. First, the candidates of each party simultaneously announce their policy platforms $x_A, x_B \in \mathbb{R}^K$. After platforms are announced, but

[4]If the electorate were perfectly homogeneous, platform convergence would obtain in equilibrium. The analysis therefore focuses on non-degenerate voter distributions. However, the comparative statics derived below imply that parties strictly benefit from dividing voters and hold even with an initially homogeneous electorate, providing incentives to create heterogeneity through political communication or identity-based strategies.

before voters cast their ballots, a popularity shock $\epsilon \in \mathbb{R}$ in favor of the candidate of party B is realized. The shock has a continuous cumulative distribution function F with a continuous density f symmetric around 0 such that $f(0) > 0$, and captures random factors affecting voters' relative assessment of the two candidates, such as economic news, political scandals, or differences in personal appeal. Voters vote sincerely, taking into account the announced platforms and the realization of the shock.[5]

For any pair of platforms (x_A, x_B), define

$$\Delta_i(x_A, x_B) = v(x_A, x_i) - v(x_B, x_i).$$

Voters of type i vote for party A if and only if $\Delta_i(x_A, x_B) \geq \epsilon$, which occurs with probability $\pi_i(x_A, x_B) = F(\Delta_i(x_A, x_B))$. As standard in probabilistic voting models, I assume that the support of F is wide enough that corner probabilities $\pi_i \in \{0, 1\}$ do not arise in equilibrium.[6] To isolate the main mechanism clearly, I do not include idiosyncratic valence shocks in the baseline model, although the main results of the paper are robust to such an addition, as shown in appendix A.5.

Politicians are exclusively motivated by political power, broadly understood as control over offices, agenda-setting authority, influence over policy and economic outcomes. Let $\rho : [0, 1] \to \mathbb{R}$ denote the mapping from a party's vote share into political power, interpreted as office rents or influence. I impose the following assumptions.

Assumption 1 *The power allocation function ρ satisfies two properties: (i) ρ is strictly increasing on $[0, 1]$; (ii) there exists a constant $\bar{\rho}$ such that $\rho(s) + \rho(1 - s) = \bar{\rho}$ for all $s \in [0, 1]$.*

Assumption 1 (i) captures the basic institutional regularity that a higher vote share generally yields more seats, offices, or bargaining leverage, potentially with discontinuities at salient thresholds. Assumption 1 (ii) is natural when political power is the allocation of a fixed stock of influence—e.g., a fixed number of legislative seats, committee

[5]I restrict attention to sincere voting. With atomistic voters, strategic voting is weakly dominated, so this assumption rules out arbitrary voting behavior while keeping the focus on platform competition. In this setup, I assume a finite number of voter types, not a finite number of voters.

[6]Formally, letting $x_\vee = (\max_{i \in I} x_i^k)_{k \leq K}$ and $x_\wedge = (\min_{i \in I} x_i^k)_{k \leq K}$, I assume that both $v(x_\wedge, x_\vee) - v(x_\vee, x_\vee)$ and $v(x_\vee, x_\vee) - v(x_\wedge, x_\vee)$ lie in the support of F.

chairs, or agenda-setting opportunities—so that one party's gains mechanically reduce the other's and the analysis cleanly isolates distributive (rather than surplus-creating) incentives. A useful interpretation is that s is seat share, voter type i summarize a district or constituency block worth a share s_i of seats, and $\rho(s)$ is legislative or organizational influence of a party with seat share s. Larger caucuses typically imply more committee representation, agenda control, leadership positions, and affiliated placements, so political power need not be winner-take-all even in a two-party setting.[7]

Parties derive utility from power according to a payoff function $U : \mathbb{R} \to \mathbb{R}$.

Assumption 2 *The utility function U satisfies $U' > 0$ and $U'' < 0$.*

Assumption 2 posits that a party's utility from political power is increasing and strictly concave. A natural source of concavity is that political power can be though of as the number of valuable placements and influence opportunities that a party can allocate—such as offices, agenda control, or affiliated positions—which are assigned first to the most loyal or effective insiders, and only later to less valuable ones (Kopecký et al., 2012; Grindle, 2012; Lewis, 2008). Under this interpretation concavity is not a preference assumption: it arises mechanically as the solution to an optimal assignment problem with heterogeneous affiliates, since additional units of power are allocated to progressively lower-value individuals.[8]

A complementary force is rent sharing: political power ultimately generates monetary and quasi-monetary benefits—such as salaries, staff budgets, fundraising capacity, access to donors, and post-office opportunities—that are consumed by individuals associated with the party (Strøm, 1990; Müller and Strøm, 1999; Katz and Mair, 1995). Since individual consumption marginal utility is diminishing, aggregating these benefits across party insiders yields a concave utility from additional power. These forces are reinforced by organizational frictions, as larger power requires greater coordination, monitoring, and internal governance, which come at a cost (McCubbins et al., 1987; Huber

[7]An alternative interpretation is that voters make campaign contribution decisions rather than voting decisions, in which case parties' funds naturally increase with the size of their supporter base.

[8]For example, if political power grants ρ identical positions and affiliates have heterogeneous values, the party optimally assigns the ρ highest-value affiliates. Total rents equal the sum of the top ρ values, so the marginal value of power is the value of the ρ-th order statistic, which is weakly decreasing in ρ.

and Shipan, 2002; Martin and Vanberg, 2011).

Appendix A.2 provides formal microfoundations along these lines, and emphasizes that the curvature is a reduced-form property of how power translates into organizational payoffs rather than an assumption that candidates are intrinsically risk-averse.

Let $\nu = U \circ \rho$ denote the reduced-form mapping from vote shares to party payoffs, and normalize payoffs so that $\nu(1) - \nu(0) = 1$. Keeping ρ and U separate is substantively useful. The function ρ isolates the institutional margin: majority premia, seat bonuses, and other threshold effects belong in ρ. Concavity is imposed on U, not on ρ, and therefore not on vote shares per se. In particular, the model is compatible with environments in which crossing particular vote-share thresholds—for example, majority or supermajority thresholds—is especially valuable.

Combining assumptions 1 and 2 yields the following implication.

Remark 1 *For any $s, s' \in [0, \frac{1}{2}]$ with $s' > s$, it holds that $\nu(s') - \nu(s) > \nu(1-s) - \nu(1-s')$.*

Remark 1 implies that winning an additional group of voters is more valuable to the minority party than to the majority party. This asymmetry will play a key role in the equilibrium analysis that follows.[9] Having described the model, I now turn to the analysis of equilibrium behavior in the party competition game.

3.1 Platform Divergence as Insurance

A strategy for party $p \in \{A, B\}$ is a probability distribution $\sigma_p \in \Delta(\mathbb{R}^K)$ over policy platforms. A strategy profile $\sigma^\star = (\sigma_A, \sigma_B)$ is a Nash equilibrium of the party competition game if, for each $p \in \{A, B\}$, $\sigma_p^\star$ maximizes party p's expected utility given the strategy of the opposing party.[10]

[9] While assumption 1 makes political power constant-sum, assumption 2 implies that party payoffs from power not constant-sum. This is why both parties value reductions in the volatility of political power.

[10] Formally, $\sigma_p^\star$ solves

$$\max_{\sigma_p \in \Delta(\mathbb{R}^K)} \int_x \int_{x'} \int_\epsilon \nu(s_p(x, x', \epsilon))\, \sigma_p(x) \sigma_{-p}^\star(x') f(\epsilon)\, dx\, dx'\, d\epsilon,$$

where $s_p(x, x', \epsilon)$ denotes party p's vote share when platforms (x, x') are offered and the popularity shock takes value ϵ.

I first establish that policy convergence does not arise in equilibrium.

Proposition 1 (Party Divergence) *If a pure-strategy equilibrium exists, the two candidates offer different policy platforms.*

The intuition for proposition 1 is straightforward. If parties offer identical platforms, candidates are ex ante indistinguishable from the voters' perspective, and the electoral outcome is entirely determined by the popularity shock. In this case, each party faces substantial payoff risk: it either obtains all political power or none. By contrast, when parties offer different platforms, each can better match the preferences of some voter groups, thereby securing a positive share of electoral support and political power with higher probability. Although differentiation may worsen a party's expected electoral performance, it reduces the volatility of electoral outcomes. With a decreasing marginal utility from political power, this reduction in risk increases expected utility.

This mechanism highlights the role of decreasing marginal utility from political power. More generally, it can be shown that party divergence arises for all feasible voter distributions if and only if

$$\nu(s) + \nu(1-s) > \nu(1) + \nu(0) \quad \text{for all } s \in (0,1),$$

a condition implied by remark 1 but not satisfied, for instance, when $U'' \geq 0$.[11] In particular, as the curvature of U vanishes and ν approaches linearity, the insurance value of differentiation disappears, and the model approaches the standard risk-neutral probabilistic-voting benchmark.

In the sections that follow, I impose additional structure on voters' preferences and on the distribution of popularity shocks. These assumptions are standard in the literature

[11] More precisely, one can show that if candidates converge when each candidate maximizes expected rents $\rho(s_p)$, then if $U'' > 0$ they also converge when maximizing $U(\rho(s_p))$. To see this, let $x^\star = (x,x)$ be any pure-strategy equilibrium when parties maximize expected rents. Since platforms coincide, party p obtains $\bar{\rho}$ with probability $\frac{1}{2}$ and 0 otherwise, with expected rent $\frac{\bar{\rho}}{2}$. Now suppose party p maximizes $U(\rho(s_p))$ with $U'' > 0$, and consider the same candidate equilibrium. If p deviates, its expected rent is at most $\frac{\bar{\rho}}{2}$, but rents follow some distribution G with support different from $\{0,\bar{\rho}\}$. Let H be a mean-preserving spread of G supported on $\{0,\bar{\rho}\}$. Because $U'' > 0$, party p is better off under H than under G. But H is either equal to or first-order dominated by the rent distribution induced by $x^\star$, so the deviation cannot be profitable.

and serve two purposes. First, they guarantee the existence of pure-strategy equilibria. Second, they allow a transparent analysis of parties' incentives to polarize voters and to shape the distribution of political conflict.

4 Unidimensional Conflict

I begin by analyzing the case $K = 1$, where parties compete on a single policy dimension.[12] I assume that the popularity shock is uniformly distributed, $\epsilon \sim U(-\phi, \phi)$, and that voter utility from policy distance is quadratic,

$$h(|x - x_i|) = -\alpha(x - x_i)^2,$$

with $\alpha > 0$. Without loss of generality, set $\alpha = 1$. This specification admits a tractable equilibrium characterization and captures a standard unidimensional conflict, such as redistribution between left- and right-wing voters.

For the remainder of this section, voter types are ordered so that $x_1 \leq x_2 \leq \cdots \leq x_N$, and x_m denotes the median voter type.[13] I show that the game admits a pure-strategy equilibrium and that exactly two policy platforms are offered in equilibrium. One platform lies to the left of the risk-neutral benchmark

$$x^{RN} = \sum_{i=1}^{N} \frac{\rho(s_i + \sum_{j<i} s_j) - \rho(\sum_{j<i} s_j)}{\rho(1) - \rho(0)} x_i,$$

with $s_0 = 0$, while the other lies to the right.

Proposition 2 (Equilibrium) *The following properties characterize the equilibria of the game.*

(i) There exist two platforms $\underline{x}, \bar{x} \in \mathbb{R}$ with $x_1 < \underline{x} < x^{RN} < \bar{x} < x_N$ such that, in every pure-strategy equilibrium, one candidate offers $\underline{x}$ and the other offers $\bar{x}$. Each platform is a

[12]With one policy dimension, the dimension index is redundant and omitted.

[13]If there exist types x_i and x_j such that $\sum_{n\leq i} s_n = \frac{1}{2}$ and $\sum_{n\geq j} s_n = \frac{1}{2}$, then $x_m = \frac{x_i+x_j}{2}$.

convex combination of voter bliss points,

$$\underline{x} = \sum_{i=1}^{N} \underline{w}_i x_i, \qquad \bar{x} = \sum_{i=1}^{N} \bar{w}_i x_i,$$

where $\underline{w}_i, \bar{w}_i \in (0,1)$ *and weights sum to one.*

(ii) The equilibrium weights reflect the marginal payoff from attracting each voter group:

$$\underline{w}_i = \frac{\nu(\underline{s}_i) - \nu(\underline{s}_{i-1})}{\nu(1) - \nu(0)}, \qquad \bar{w}_i = \frac{\nu(\bar{s}_i) - \nu(\bar{s}_{i+1})}{\nu(1) - \nu(0)},$$

where $\underline{s}_i = \sum_{j \le i} s_j$ *and* $\bar{s}_i = \sum_{j \ge i} s_j$ *for each* $i = 1, \ldots, N$*, and* $\underline{s}_0 = \bar{s}_{N+1} = 0$.

(iii) One candidate assigns greater weight to voters to the left of the median, while the other assigns greater weight to voters to the right. In particular,

$$x_i > (<)\, x_m \implies \bar{w}_i > (<)\, \underline{w}_i.$$

(iv) Both parties obtain the same equilibrium expected payoff, which is increasing in the squared distance between the two equilibrium platforms $(\bar{x} - \underline{x})^2$.

Remark 2 *If the voter distribution is symmetric or rents are proportional to vote shares, then* $\underline{x} < x^o < \bar{x}$*, where* x^o *maximizes aggregate voter welfare.*

Proposition 2 shows that equilibrium polarization arises endogenously. Even though parties are identical, they target opposite sides of the electorate. Unlike models in which polarization reflects ideological motivations or exogenous party asymmetries, divergence here follows directly from decreasing marginal utility from political power.

When the risk-neutral benchmark coincides with the utilitarian optimum, equilibrium divergence reduces voter welfare. Voters benefit from electing their preferred candidate, but bear the risk that the opposing platform is implemented. At the same time, greater platform differentiation increases parties' expected payoffs by securing core supporters and reducing electoral volatility. This tension motivates the analysis that follows.

4.1 Voter Alienation and Polarization

I now study how changes in voters' policy views affect equilibrium outcomes. Throughout this section, without loss of generality, I assume that party A offers the right-wing platform $\bar{x}$ and party B offers the left-wing platform $\underline{x}$. For a given voter distribution $(\mathbf{x}, \mathbf{s})$, let $\hat{V}_p(\mathbf{x}, \mathbf{s})$ denote the equilibrium expected payoff of party $p \in \{A, B\}$, and let $\pi_i^p(\mathbf{x}, \mathbf{s})$ be the probability that voter group i votes for party p when equilibrium platforms are played.

Before stating the main results, it is useful to clarify what it means for a party to benefit from attracting or alienating a voter group.

Definition 1 *Fix* $(\mathbf{x}, \mathbf{s})$. *Party p benefits from attracting voter group i if*

$$\frac{\partial \pi_i^p(\mathbf{x}, \mathbf{s})}{\partial x_i} \frac{\partial \hat{V}_p(\mathbf{x}, \mathbf{s})}{\partial x_i} > 0,$$

and benefits from alienating voter group i if

$$\frac{\partial \pi_i^p(\mathbf{x}, \mathbf{s})}{\partial x_i} \frac{\partial \hat{V}_p(\mathbf{x}, \mathbf{s})}{\partial x_i} < 0.$$

That is, a party benefits from attracting (alienating) a voter group if a marginal change in that group's policy views that makes it more (less) likely to vote for the party increases the party's equilibrium expected payoff.

At first glance, it may seem surprising that a party could ever benefit from alienating voters. After all, alienation reduces the probability that the group supports the party in the new equilibrium. The next result shows that this intuition is incomplete.

Proposition 3 *Let* x_m *denote the median voter type. Party A benefits from attracting voter group i if* $x_i > \max\{\underline{x}(\mathbf{x}, \mathbf{s}), x_m\}$ *and benefits from alienating group i if* $x_i < \min\{\underline{x}(\mathbf{x}, \mathbf{s}), x_m\}$. *Symmetrically, party B benefits from attracting group i if* $x_i < \min\{\bar{x}(\mathbf{x}, \mathbf{s}), x_m\}$ *and benefits from alienating group i if* $x_i > \max\{\bar{x}(\mathbf{x}, \mathbf{s}), x_m\}$.

Remark 3 *If* $x_1 < x_m < x_N$, *party A benefits from attracting the most right-wing voter group N and alienating the most left-wing group* 1, *while party B benefits from attracting group* 1 *and alienating group N.*

To understand proposition 3, it is useful to examine how a party's equilibrium payoff changes when the bliss point of a voter group becomes marginally more extreme. Consider, for instance, the effect on party B's equilibrium expected payoff of a marginal increase in x_N, the bliss point of the most right-wing voter group. Let $V_B(x_A, x_B)$ denote party B's expected payoff when platforms (x_A, x_B) are offered. Since in equilibrium $x_A = \bar{x}(\mathbf{x}, \mathbf{s})$ and $x_B = \underline{x}(\mathbf{x}, \mathbf{s})$, the derivative of $\hat{V}_B$ with respect to x_N can be written as[14]

$$\frac{\partial \hat{V}_B(\mathbf{x}, \mathbf{s})}{\partial x_N} = \underbrace{\frac{\partial V_B(\bar{x}, \underline{x})}{\partial x_N}}_{\text{direct effect}<0} + \underbrace{\frac{\partial V_B(\bar{x}, \underline{x})}{\partial \bar{x}} \frac{\partial \bar{x}(\mathbf{x}, \mathbf{s})}{\partial x_N}}_{\text{indirect effect}>0}. \tag{1}$$

The first term captures the direct effect of a marginal increase in x_N holding platforms fixed at their equilibrium values. As group N becomes more extreme, it is less likely to vote for party B, reducing B's expected payoff. However, this effect is quantitatively limited. Group N only votes for party B in states of the world where party B is already winning a large share of voters and power, so the marginal loss associated with alienating this group is small.

The second term captures the indirect effect operating through party A's equilibrium platform adjustment. Since $\partial \bar{x}(\mathbf{x}, \mathbf{s})/\partial x_N = \bar{w}_N > 0$, a marginal increase in x_N induces party A to move its platform further to the right. This movement pushes party A away from the core supporters of party B, increasing their loyalty and raising party B's equilibrium expected payoff.[15] The indirect effect dominates the direct effect, implying that party B benefits from alienating group N. An analogous argument applies to party A and to the most left-wing voter groups.

More generally, it follows that for any voter distribution $(\mathbf{x}, \mathbf{s})$ and any party $p \in \{A, B\}$,

$$x_i > (<)\, x_m \quad \Longrightarrow \quad \frac{\partial \hat{V}_p(\mathbf{x}, \mathbf{s})}{\partial x_i} > (<)\, 0.$$

[14] Although the expected payoff function is not everywhere differentiable, lemma A.1 in the appendix shows that equilibrium platforms satisfy the first order conditions, so that the envelope theorem applies.

[15] Party B's platform also adjusts in response to x_N, but this change bears no effect on its payoff and drops out of equation 1 by the envelope theorem.

That is, both parties benefit when voters to the left of the median become more left-wing and when voters to the right become more right-wing.

This insight extends beyond marginal changes. Parties are better off when the distribution of voter preferences becomes more polarized in the sense of a spread, as defined below.

Definition 2 *For a given voter distribution* $(\mathbf{x}, \mathbf{s})$, *let* $(\mathbf{x}_L, \mathbf{s}_L)$ *and* $(\mathbf{x}_R, \mathbf{s}_R)$ *denote the distributions of voters below and above the median, respectively. A distribution* $(\mathbf{x}', \mathbf{s}')$ *is a spread of* $(\mathbf{x}, \mathbf{s})$ *if* $(\mathbf{x}_L, \mathbf{s}_L)$ *first-order stochastically dominates* $(\mathbf{x}'_L, \mathbf{s}'_L)$ *and* $(\mathbf{x}'_R, \mathbf{s}'_R)$ *first-order stochastically dominates* $(\mathbf{x}_R, \mathbf{s}_R)$, *with at least one relation strict.*

A spread increases disagreement between voters on opposite sides of the median. For any distribution $(\mathbf{x}, \mathbf{s})$, let $\mathbf{x}_m \in \mathbb{R}^I$ be the vector with all entries equal to the median x_m.

Proposition 4 *Let* $(\mathbf{x}', \mathbf{s}')$ *and* $(\mathbf{x}, \mathbf{s})$ *be two voter distributions. If* $(\mathbf{x}' - \mathbf{x}'_m, \mathbf{s}')$ *is a spread of* $(\mathbf{x} - \mathbf{x}_m, \mathbf{s})$, *then* $\hat{V}_p(\mathbf{x}', \mathbf{s}') > \hat{V}_p(\mathbf{x}, \mathbf{s})$ *for* $p = A, B$.

The intuition behind proposition 4 closely parallels the derivative-based logic developed above, but applies to larger, non-marginal changes in the voter distribution. Holding the median constant, a spread of the distribution increases the distance between voters on opposite sides of the median Greater separation between voter groups allows parties to diverge more in equilibrium: a spread of the voter distribution changes the weighted averages defining $\underline{x}$ and $\bar{x}$ in a way that pulls the two equilibrium platforms farther apart. As voter and platform separation increases, each party builds a more distinct electoral base and reduces its exposure to aggregate popularity shocks. As the utility from power is concave, this power stabilization benefits both parties.

Taken together, the results of this section show that decreasing marginal utility from power generates incentives not only for party divergence, but also for voter polarization. Parties benefit from alienating the extreme supporters of their opponent and, more broadly, from a more polarized electorate.

5 Multidimensional Conflict

I now consider the case with $K \geq 2$ policy instruments, so that party platforms x_A and x_B are vectors in $\mathbb{R}^K$. While the incentives identified in the unidimensional case continue to operate, multidimensional competition raises additional technical difficulties. In particular, parties' expected payoffs are not everywhere differentiable at platform profiles (x_A, x_B) that induce ties in voters' relative preferences, and best responses need not be continuous. As a result, standard fixed-point arguments do not directly apply.

To make the analysis tractable, I impose additional working assumptions. First, I require the reduced-form payoff function v to be strictly concave.[16] Second, for the Nash equilibrium existence results, I focus on symmetric voter distributions[17], which allows me to exploit potential-game techniques to establish equilibrium existence. The set of symmetric distributions is denoted by $\mathcal{V}$, and symmetry is not used in our results unless stated explicitly. Finally, I assume that voters assign equal weight to all policy dimensions, so that voter i's utility takes the form

$$v(x, x_i) = -\alpha \|x - x_i\|^2,$$

with $\alpha > 0$ (normalized to $\alpha = 1$). This assumption simplifies the exposition and can be relaxed to allow for dimension-specific salience.[18]

Section 5.1 characterizes equilibrium existence and properties. Section 5.2 derives comparative statics and studies how multidimensional voter conflict affects equilibrium payoffs.

[16]As shown in the appendix, strict concavity guarantees that best responses do not occur at points of non-differentiability. The assumption can be relaxed, for instance by allowing for payoff jumps at majority thresholds, provided a median voter type exists.

[17]Formally, a voter distribution $(\mathbf{x}, \mathbf{s})$ with type set I is symmetric if for each type $i \in I$ there exist a type $\bar{i} \in I$ such that: $x_{\bar{i}} = 2x_m - x_i$ and $s_{\bar{i}} = s_i$, for x_m the vector of mean ideal policies of distribution $(\mathbf{x}, \mathbf{s})$.

[18]Assigning weight α^k to dimension k is equivalent to rescaling voter ideal points and party platforms accordingly.

5.1 Equilibrium Analysis

I begin by introducing notation that generalizes the unidimensional case. Let $(\mathbf{x}, \mathbf{s})$ be a voter distribution and with set of voter types $I = \{1, 2, \ldots, N\}$, and let $\mathcal{R}(I)$ be the set of permutations of I. For $R \in \mathcal{R}(I)$ and $j = 1, \ldots, N$, let R_j denote the type in position j of permutation R.[19] For each permutation R, define platform $\bar{x}(R), \underline{x}(R) \in \mathbb{R}^K$ by

$$\bar{x}(R) = \sum_{i=1}^{N} \bar{w}_{R_i} x_{R_i}, \qquad \underline{x}(R) = \sum_{i=1}^{N} \underline{w}_{R_i} x_{R_i},$$

where $\bar{w}_{R_i} = \nu(\bar{s}_{R_i}) - \nu(\bar{s}_{R_{i+1}})$, $\underline{w}_{R_i} = \nu(1 - \bar{s}_{R_{i+1}}) - \nu(1 - \bar{s}_{R_i})$, and $\bar{s}_{R_i} = \sum_{j=i}^{N} s_{R_j}$, with $\bar{s}_{R_{N+1}} = 0$. As in the unidimensional case, these weights reflect the marginal payoff from attracting successive voter groups along a given ranking.

For any pair of platforms (x, x'), let $\nabla(x, x') \subset \mathcal{R}(I)$ denote the set of strict rankings of voter types induced by sincere voting, ordered from voters more likely to support party B to voters more likely to support party A:

$$\nabla(x, x') = \Big\{R \in \mathcal{R}(I) : i > j \Rightarrow \Delta_{R_i}(x, x') > \Delta_{R_j}(x, x')\Big\}.$$

The set $\nabla(x, x')$ contains at most one element, by construction, and is empty if the platform profile induces ties in relative preferences.

I now define the notion of local equilibrium, which is instrumental to the characterization of pure-strategy Nash equilibria.

Definition 3 (Local Equilibrium) *A platform pair $(x_A, x_B) \in \mathbb{R}^{2K}$ is a local equilibrium if there exists $u > 0$ such that, for each $p \in \{A, B\}$, x_p is a best response to x_{-p} when party p's feasible deviations are restricted to the open ball $B_u(x_p)$.*

Let $\mathcal{L}$ denote the set of local equilibria of the game. The next result characterizes $\mathcal{L}$ and shows that local equilibria can be ranked according to the degree of platform differentiation.

[19] For example, if $I = \{1, 2, 3, 4\}$ and $R = (2, 3, 1, 4)$, then $R_3 = 1$ and $R_1 = 2$.

Lemma 1 *The set of local equilibria satisfies*

$$\mathcal{L} = \Big\{(x_A, x_B) \in \mathbb{R}^{2K} : \exists R \in \nabla(x_A, x_B), (x_A, x_B) = (\bar{x}(R), \underline{x}(R))\Big\}.$$

For any $x \in \mathcal{L}$, $V_A(x_A, x_B) = V_B(x_A, x_B)$. Moreover, for any $x, x' \in \mathcal{L}$,

$$V_p(x_A, x_B) > V_p(x'_A, x'_B) \quad \textit{if and only if} \quad \|x_A - x_B\|^2 > \|x'_A - x'_B\|^2,$$

for party $p = A, B$.

Lemma 1 implies that local equilibria correspond to self-consistent segmentations of the electorate. Given a ranking of voter types, each party optimally chooses a platform that places greater weight on its most loyal supporters. As in the unidimensional case, party payoffs increase with platform distance, reflecting the insurance motive: greater differentiation secures support from distinct voter blocs and reduces electoral volatility.

All Nash equilibria are local equilibria, but local equilibria need not be Nash equilibria when large platform changes are feasible. I therefore turn to the characterization of pure-strategy Nash equilibria.

Let $\mathcal{E}$ denote the set of pure-strategy Nash equilibria. Since parties obtain the same expected payoff in equilibrium, I focus on the set of party-preferred equilibria,

$$\mathcal{E}^\star := \arg \max_{(x_A, x_B) \in \mathcal{E}} V_p(x_A, x_B),$$

which does not depend on the identity of party p. Not only does the set of party-preferred equilibria have a neat characterization, but the common equilibrium payoff allows for a Pareto-ranking of Nash equilibria, making party-preferred equilibria focal.

Proposition 5 (Nash Equilibrium Existence) *Let $(\mathbf{x}, \mathbf{s}) \in \mathcal{V}$. The set of party preferred Nash equilibria $\mathcal{E}^\star$ is non-empty and coincides with the set of local equilibria that maximize platform differentiation:*

$$\mathcal{E}^\star = \arg \max_{(x_A, x_B) \in \mathcal{L}} \|x_A - x_B\|^2 \neq \emptyset.$$

Besides equilibrium existence, proposition 5 shows that equilibrium selection favors maximal differentiation. Among all local equilibria, those that allow parties to diverge the most are also party-preferred Nash equilibria. Other equilibria—if existing—are pareto dominated. This result generalizes the unidimensional finding that parties share a common interest in platform separation as a means of securing core voters and insuring against aggregate shocks.

5.2 Divide and Diverge

I now study how changes in the multidimensional distribution of voter preferences affect equilibrium outcomes. A salient feature of contemporary political conflict is that voters' policy views have become increasingly aligned across issues: individuals who hold extreme positions on one policy dimension are more likely to hold congruent positions on others (e.g., Gentzkow, 2016; Bonomi et al., 2021; Pew Research Center, 2023).[20] I refer to this phenomenon as an increase in *ideological coherence*. In this paper, ideological coherence corresponds to an increase in the correlation of voter preferences across policy dimensions, rather than to greater marginal dispersion on any single issue. This subsection shows why such changes benefit the political supply side.

To fix ideas, consider a two-dimensional policy space. Figure 1 illustrates two voter distributions with identical marginal distributions on each policy dimension. In the first, voters' policy views are weakly correlated across dimensions, so ideological factions are diffuse. In the second, voter preferences are more strongly aligned across issues, so that the electorate sorts into two opposing and internally consistent ideological blocs.

In a unidimensional policy space, increased disagreement takes the form of a spread of voter preferences away from the median. In a multidimensional space, disagreement need not increase marginally on any single issue. Instead, it manifests through greater alignment of positions across issues, so that voters sort into cohesive ideological factions.

[20] Recent work studies how voter learning can generate issue alignment and polarization (Vaeth, 2024). At the same time, the extent of such alignment in the mass public, as opposed to political elites or highly engaged citizens, remains debated. In this paper, *ideological coherence* refers to the cross-issue alignment of politically relevant disagreement in the voter distribution, not to a claim that most voters hold fully constrained ideological belief systems; see Kinder and Kalmoe (2017).

The notion of a d-spread introduced in this section captures this idea: it is the multidimensional analogue of a unidimensional spread, measured along the endogenous direction of party differentiation. From the perspective of electoral competition, both changes increase effective ideological disagreement between opposing blocs and relax competitive pressure in the same way.

As in section 4.1, I express equilibrium objects as functions of the voter distribution. Let $\mathcal{L}(\mathbf{x},\mathbf{s})$ and $\mathcal{E}^\star(\mathbf{x},\mathbf{s})$ denote the sets of local equilibria and party-preferred equilibria associated with voter distribution $(\mathbf{x},\mathbf{s})$. I begin by characterizing the types of opinion changes that increase parties' equilibrium payoffs locally.

Fix a local equilibrium $x \in \mathcal{L}$ with associated ranking $R \in \nabla(x)$. Let m_R be any index such that $\sum_{i<m_R} s_{R_i} \leq \frac{1}{2}$ and $\sum_{i>m_R} s_{R_i} \leq \frac{1}{2}$. Such an index always exists. For $i > m_R$, voter group R_i is a relative supporter of party A, while for $i < m_R$ voter group R_i is a relative supporter of party B.[21] For each party $p \in \{A,B\}$, define its ideological direction by $x_p - x_{-p}$, and define the ideological direction on policy dimension k by $x_p^k - x_{-p}^k$. Finally, let $\hat{V}_p^R(\mathbf{x},\mathbf{s})$ denote party p's expected payoff when the voter distribution is $(\mathbf{x},\mathbf{s})$ and the local equilibrium inducing ranking R is played.

Proposition 6 *Fix $(\mathbf{x},\mathbf{s})$ and a corresponding local equilibrium ordering R. It holds:*

(i) For each policy dimension $k = 1,\dots,K$, both parties benefit if the ideal policy k of a relative supporter of party $p \in \{A,B\}$ changes marginally in p's ideological direction. Formally, for each $i \in I$,

$$\frac{\partial \hat{V}_p^R}{\partial x_{R_i}^k}(\bar{x}^k(R) - \underline{x}^k(R)) \geq (\leq)\, 0 \quad \text{if } i > (<)\, m_R,$$

with strict inequality whenever $\bar{x}^k(R) \neq \underline{x}^k(R)$.

(ii) Parties benefit the most from marginal opinion changes when these changes are aligned with the party's overall ideological direction.

Proposition 6 generalizes the unidimensional alienation logic. At a given local equilibrium, both parties benefit when their relative supporters move further in the direction

[21] In exact-half cases m_R need not be unique; the results below apply to the groups strictly to either side of the chosen split.

of the party's platform, increasing disagreement with the opposing faction across policy dimensions. These changes strengthen voter loyalty and allow parties to differentiate more without losing support.

I now extend this logic beyond marginal changes and beyond local equilibria. To do so, I formalize the notion of ideological coherence using the concept of a d-spread.

Definition 4 *Let $d \in \mathbb{R}^K$. A voter distribution $(\mathbf{x}', \mathbf{s}')$ is a d-spread of $(\mathbf{x}, \mathbf{s})$ if the projected distribution $(\mathbf{x}'^{\top} d, \mathbf{s}')$ is a spread of $(\mathbf{x}^{\top} d, \mathbf{s})$.*

A d-spread captures an increase in disagreement along direction d, which corresponds to the equilibrium direction of party differentiation. Importantly, a d-spread may leave the marginal distribution of preferences on each policy dimension unchanged, while increasing the alignment of preferences across dimensions.

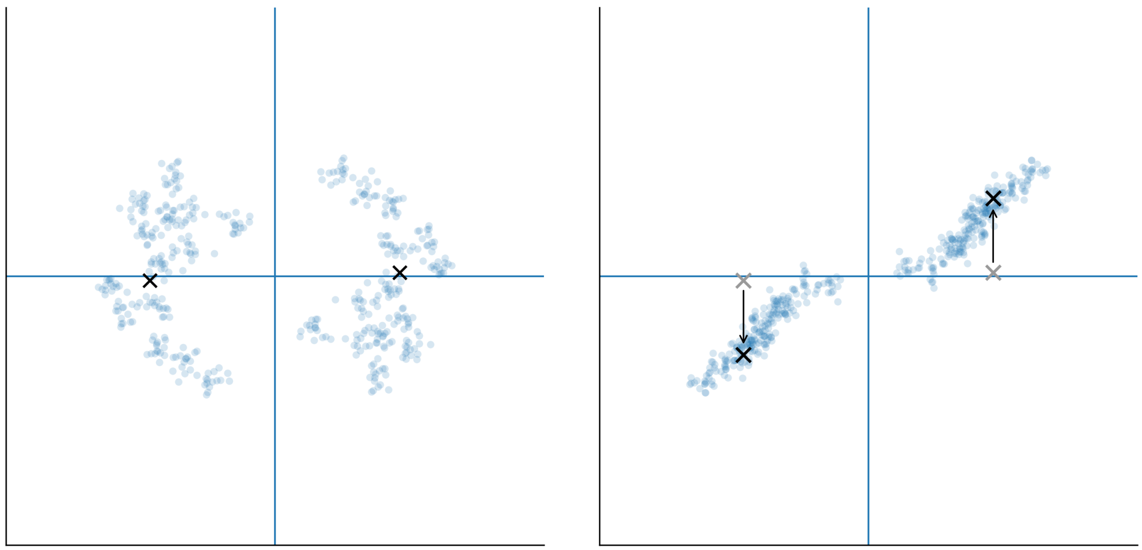
Let $(\mathbf{x} - \mathbf{x}_m, \mathbf{s})$ denote the voter distribution obtained by centering each policy dimension around its median. Let $\hat{V}_p(\mathbf{x}, \mathbf{s})$ denote party p's equilibrium payoff in a party-preferred equilibrium.

Proposition 7 *Consider two voter distributions $(\mathbf{x}, \mathbf{s}), (\mathbf{x}', \mathbf{s}') \in \mathcal{V}$ and let $\tilde{x} \in \mathcal{E}^{\star}(\mathbf{x}, \mathbf{s})$. If $(\mathbf{x}', \mathbf{s}')$ is a d-spread of $(\mathbf{x}, \mathbf{s})$ for $d = \tilde{x}_A - \tilde{x}_B$, then*

$$\hat{V}_p(\mathbf{x}', \mathbf{s}') > \hat{V}_p(\mathbf{x}, \mathbf{s}) \quad \textit{for } p = A, B.$$

Proposition 7 shows that what benefits parties in a multidimensional setting is not simply greater dispersion of voter preferences along individual policy dimensions. Rather, parties benefit from increases in ideological coherence that sharpen disagreement between two opposing factions along the equilibrium direction of differentiation. As illustrated in Figure 1, even when marginal distributions remain unchanged, greater alignment of voter views across issues allows parties to diverge more in equilibrium and increases expected payoffs.

Remark 4 *Increases in the ideological coherence of opposing voter factions increase parties' preferred equilibrium payoffs.*



(a) Low ideological coherence (b) High ideological coherence

Figure 1: Increase in ideological coherence holding marginal distributions fixed. Greater alignment of voter preferences across dimensions generates more cohesive ideological factions and allows parties to diverge further in equilibrium. Panel (b) is obtained from panel (a) by applying a d-spread, where d is the equilibrium ideological direction in panel (a). Party-preferred equilibrium platforms are represented by dark x markers, and are identified building on the method of appendix A.4

Taken together, the results of this subsection show that parties have incentives to divide the electorate along multiple dimensions simultaneously. By fostering ideological coherence within factions, parties relax competitive pressures and secure higher equilibrium payoffs, providing a supply-side explanation for the growing multidimensional nature of political polarization.

6 Extensions and Applications

In this section, I illustrate how the insurance motive identified in the baseline model operates through different channels — issue framing, information provision, and party identity — without introducing new strategic forces. In each application, parties benefit insofar as political communication or identity activation increases effective polarization: the stability and separation of voter blocs across realizations of aggregate electoral shocks. The goal is not to provide independent models of communication or identity, but to show

how the same incentive to insure against aggregate electoral risk manifests across domains. All results below rely on the same mechanism as in the core analysis: actions that reduce voter substitutability relax competitive pressure and increase equilibrium payoffs under concave utility from political power.

6.1 Common-Interest versus Conflict-of-Interest Issues

A recent stream of literature stresses that politicians often focus on divisive, zero-sum issues over common-interest policies, even when the latter could benefit a broad share of voters (e.g., Fiorina et al., 2011; Matsusaka and Kendall, 2022; Bueno de Mesquita and Dziuda, 2023). This tendency is potentially amplified by voters' propensity toward zero-sum thinking (e.g., Carvalho et al., 2023; Chinoy et al., 2026; Ali et al., 2024). I show how the incentives identified in this paper provide a simple rationale for these patterns.

Consider a stylized version of the baseline model with two voters, one religious (R) and one secular (S), and one policy issue x (e.g., government spending aimed at tackling climate change). In the common-interest scenario, preferences are aligned, so that $x_R = x_S$. In the conflict-of-interest scenario, preferences differ, so that $x_R < x_S$. One interpretation is that the religious voter views such spending as conflicting with divine control over nature (Pew Research Center, 2022).

The baseline logic applies immediately. If $x_R = x_S$, there is no heterogeneity in bliss points, so parties converge on the common ideal policy and obtain expected payoff $\frac{\nu(1)+\nu(0)}{2}$. If instead $x_R < x_S$, the electorate is diverse and the equilibrium platforms satisfy $\underline{x} < \bar{x}$, generating expected payoff

$$\frac{\nu(1)+\nu(0)}{2} + \frac{(\bar{x}-\underline{x})^2}{2\phi} > \frac{\nu(1)+\nu(0)}{2}.$$

Thus, politicians are better off when voters perceive the policy as defining a conflict of interest.

Remark 5 *Parties benefit if voters perceive a policy issue as defining a conflict of interest in society, even when in reality the same policy could benefit everyone.*

The same reasoning implies that parties benefit when conflict-of-interest issues are more salient than common-interest issues. To illustrate, consider two policy instruments, (x,q), and voter utility

$$v(x,q,x_i,q_i) = -\alpha(x_i - x)^2 - (1-\alpha)(q_i - q)^2, \tag{2}$$

where $\alpha \in (0,1)$ captures the relative payoff importance of issue x. Assume $x_R \neq x_S$ and $q_R = q_S = q^\star$. In equilibrium, parties converge on the common-interest policy, $q_A = q_B = q^\star$, but diverge on x. Equilibrium payoffs equal

$$\frac{\nu(1)+\nu(0)}{2} + \alpha\frac{(\bar{x}-\underline{x})^2}{2\phi},$$

which is increasing in α.

Remark 6 *Parties have an incentive to render conflict-of-interest policy issues more salient than common-interest ones.*

6.2 Information Provision

I now examine information provision incentives. The key point is that information affects party payoffs only insofar as it changes equilibrium platform differentiation.

Assume utility 2. Suppose that there is uncertainty about both $q^\star$ and about which group benefits from the conflict-of-interest issue x. Let $q^\star \in \{0,1\}$ be the common-interest state, with prior $\Pr(q^\star = 1) = \pi_q \in (0,1)$. Let the conflict state be $\omega_x \in \{0,1\}$, where $\omega_x = 1$ corresponds to $(x_R, x_S) = (1,0)$ and $\omega_x = 0$ corresponds to $(x_R, x_S) = (0,1)$, with prior $\Pr(\omega_x = 1) = \pi_x \in (0,1)$. Assume $q^\star$ and ω_x are independent. Before parties choose platforms, either party can provide public information on $q^\star$ and/or ω_x to voters, at a small cost.[22] For the purpose of this application, when parties run with platforms (q_A, x_A) and (q_B, x_B), suppose that the policies (q,x) resulting from the political process belong to $\{q_A, q_B\} \times \{x_A, x_B\}$.

[22]This is a reduced-form way to capture costly information acquisition or disclosure.

Common-interest issue. If $q^\star$ is revealed, voters learn the socially optimal policy and equilibrium convergence obtains at $q_A = q_B = q^\star$. If $q^\star$ is not revealed, voters hold the prior and parties still converge at $q_A = q_B = \pi_q$. In both cases, parties do not differentiate on q, so party payoffs are unaffected by learning $q^\star$. Voter welfare, however, is higher when $q^\star$ is revealed: under quadratic loss, expected welfare increases by $(1-\alpha)\pi_q(1-\pi_q)$ per voter relative to $q = \pi_q$. Therefore, parties have no incentive to provide information about $q^\star$ even though it is welfare-improving.

Remark 7 *Parties have no incentive to keep the electorate informed about common-interest issues, even when such information is welfare-improving.*

Conflict-of-interest issue. Now consider information about ω_x. Let $\hat{\pi}_x$ denote the posterior belief that $\omega_x = 1$ at the time platforms are chosen. If ω_x is revealed, then $\hat{\pi}_x \in \{0, 1\}$; otherwise $\hat{\pi}_x = \pi_x$.

To make the link with the baseline unidimensional analysis explicit, note that under belief $\hat{\pi}_x$, the two voters' bliss points for x are $\hat{\pi}_x$ and $1 - \hat{\pi}_x$. As in section 4, equilibrium platforms are weighted averages of the two bliss points. In particular, letting party A target voter R without loss of generality, one can write the equilibrium platforms as

$$\bar{x} = \left(\nu(\tfrac{1}{2}) - \nu(0)\right)\hat{\pi}_x + \left(\nu(1) - \nu(\tfrac{1}{2})\right)(1 - \hat{\pi}_x), \qquad \underline{x} = \left(\nu(\tfrac{1}{2}) - \nu(0)\right)(1 - \hat{\pi}_x) + \left(\nu(1) - \nu(\tfrac{1}{2})\right)\hat{\pi}_x,$$

so that $\bar{x} - \underline{x} = \gamma(2\hat{\pi}_x - 1)$ and $\gamma := 2\nu(\tfrac{1}{2}) - \left(\nu(1) + \nu(0)\right) > 0$.

Revealing ω_x sets $|2\hat{\pi}_x - 1| = 1$, maximizes platform separation, and therefore maximizes equilibrium party payoffs. In contrast, when ω_x is not revealed, platform separation is smaller unless π_x is already close to 0 or 1. Voter welfare moves in the opposite direction: revealing ω_x pushes equilibrium platforms further from the utilitarian optimum $x^o = \frac{1}{2}$ and increases the risk of extreme policy implementation.

Remark 8 *Parties have an incentive to keep the electorate informed about conflict-of-interest issues, even when such information is welfare detrimental.*

Taken together, the two cases imply systematic distortions in political communication: politicians underprovide information on common-interest issues and overprovide information on divisive issues.

6.3 Party Identity

My last application is to party identity. Party identity shapes voters' perceptions and behavior by associating policy views with group membership (Greene, 1999). Formal models of identity show that identification with a group can distort beliefs and preferences toward positions that are more representative of the in-group relative to the outgroup (Shayo, 2009; Bonomi et al., 2021; Gennaioli and Tabellini, 2025). I show that, in this environment, party identity is an effective tool for increasing ideological coherence and therefore equilibrium differentiation.

Following Bonomi et al. (2021), suppose that when voter i identifies with group $G \in \{A, B\}$, her identity-adjusted bliss point takes the reduced-form form

$$\tilde{x}_i = x_i + \theta(x_G - x_{\bar{G}}), \tag{3}$$

where x_G is the policy typical of group G, $x_{\bar{G}}$ is the policy typical of the outgroup, and $\theta > 0$ captures the strength of identity effects.

To connect this to the baseline model, consider a two-stage interaction. Parties start from a status quo voter distribution $(\mathbf{x}, \mathbf{s}) \in \mathcal{V}$ and an associated party-preferred equilibrium $\hat{x} \in \mathcal{E}^{\star}(\mathbf{x}, \mathbf{s})$. In stage 1, parties can engage in communication that strengthens partisan identity. In stage 2, parties compete in the election game described in the paper. If identity is activated, voter i's bliss points shift from x_i to $\tilde{x}_i$, where for each voter i, x_G ($x_{\bar{G}}$) is the status quo equilibrium platform $\hat{x}_p$ closest to (furthest from) x_i. This shifts voters in the direction of their preferred party across all policy dimensions, effectively implementing a d-spread along the party conflict direction $\hat{x}_A - \hat{x}_B$, increasing ideological coherence and allowing greater platform differentiation. Hence $\hat{V}_p(\tilde{\mathbf{x}}, \mathbf{s}) > \hat{V}_p(\mathbf{x}, \mathbf{s})$ for $p = A, B$.

Remark 9 *Party identity increases disagreement between opposing-party supporters in a way that benefits parties.*

As in the multidimensional analysis, the key effect of party identity is not only to increase dispersion along individual issues, but to align disagreements across issues, reducing substitutability between parties and relaxing competition.

6.3.1 A Dynamic Application: Self-reinforcing Polarization

I conclude this section by illustrating how party identity can generate a feedback loop between party and voter polarization over time. For simplicity, I focus on a setting with two voters, L and R, and a single policy instrument x, with rational bliss points $x_R < x_L$.

Elections are repeated at dates $t = 0, 1, 2, \ldots$. Let x_A^t and x_B^t denote the platforms offered by parties A and B at date t. In each period, before platforms are chosen, parties may invest in an identity-based communication campaign at cost $c > 0$, which increases the strength of identity effects. Voters' identity-adjusted bliss points evolve according to

$$\tilde{x}_i^t = x_i + \theta_t \Big(x_G^{t-1} - x_{\bar{G}}^{t-1} \Big),$$

where $G \in \{A, B\}$ denotes the party that minimizes $|x_G^{t-1} - x_i|$, $\bar{G}$ denotes the other party, and $\theta_t \geq 0$ captures the strength of identity effects at date t. If both parties invest in identity-building communication, then $\theta_t = \theta_H > 0$; if only one party invests, $\theta_t = \theta_L$ with $0 < \theta_L < \theta_H$; and if neither invests, $\theta_t = 0$.

After identity investments are made, parties choose platforms myopically, taking voters' current bliss points $\tilde{x}_i^t$ as given and playing the equilibrium of the unidimensional election game described in section 4. For simplicity, I assume $\tilde{x}_i^0 = x_i$ for $i \in \{L, R\}$.

Let $\gamma := 2\nu(\frac{1}{2}) - \Big(\nu(1) + \nu(0)\Big) > 0$, as defined in the information provision application. The following condition ensures that identity investment is self-reinforcing:

$$(x_L - x_R)^2 > \frac{\phi c}{2\gamma^3} \max \left\{ \frac{1}{(\theta_H - \theta_L)\Big[1 + \gamma(\theta_H + \theta_L)\Big]}, \frac{1}{\theta_L(1 + \gamma\theta_L)} \right\}. \tag{4}$$

When condition 4 holds, investing in identity-based communication is a best response for both parties in every period $t \geq 1$. As a result, both party platforms and voter bliss points become increasingly polarized over time. In particular, platform divergence and voter polarization evolve, respectively, according to

$$|x_A^t - x_B^t| = |x_A^0 - x_B^0| \sum_{s=0}^{t} (2\theta_H \gamma)^s, \qquad |\tilde{x}_L^t - \tilde{x}_R^t| = |x_L - x_R| \sum_{s=0}^{t} (2\theta_H \gamma)^s.$$

Thus, party identity generates a feedback loop: greater platform differentiation amplifies voter polarization, which in turn sustains further party divergence.

Remark 10 *The incentive to divide voters through party identity can generate a feedback loop between party and voter polarization.*

7 Welfare and the Majority Premium

Several results of sections 6 and 4 point to welfare losses associated with politicians' incentives to polarize voters. In this section, I evaluate those losses under the power-sharing implementation rule formalized in appendix A.3. When parties target voters at opposite ends of the political spectrum, equilibrium platforms move away from policies that maximize aggregate voter welfare, and greater platform differentiation increases the dispersion of implemented policy outcomes.

To illustrate, consider the unidimensional setting of section 4 and assume that the voter distribution has a median voter type x_m such that $s_m > 0$. In this case, power-maximizing parties converge to the policy that maximizes utilitarian welfare, $x^{RN} = x^o$, under a fully proportional power allocation rule $\rho(s) = s\bar{\rho}$. By contrast, when parties have decreasing marginal utility from political power, equilibrium platforms satisfy $\underline{x} < x^o < \bar{x}$, generating a welfare loss relative to the utilitarian benchmark.

One way to mitigate this loss could be to reduce the dispersion of voter preferences, bringing voters' bliss points closer to the median. However, I have shown that parties' incentives point in the opposite direction: politicians benefit from increasing voter conflict and will exploit communication strategies and identity politics to achieve this goal.

An alternative lever is the institutional mapping from votes to political power. In the model, this mapping is summarized by the rent allocation function $\rho(s)$. In particular, let

$$\rho_m := \lim_{s \to \frac{1}{2}^+} \rho(s) - \lim_{s \to \frac{1}{2}^-} \rho(s)$$

denote the majority premium, that is, the additional power associated with winning a narrow majority. The equilibrium weights $\bar{w}_m$ and $\underline{w}_m$ that parties assign to the median voter's policy views are increasing in ρ_m, reflecting the greater payoff from securing the median voter. As ρ_m increases, equilibrium platforms move closer to the median voter's bliss point. In the limit, as $\rho_m \to \bar{\rho}^-$, both platforms converge to x_m. This proves welfare-improving if x_m and x^o are close (e.g., symmetric voter distributions).

Remark 11 *Increasing the majority premium reduces equilibrium platform divergence. It can improve utilitarian welfare if the median voter's policy views are close to the welfare optimum.*

Adjusting the majority premium ρ_m provides an institutional lever that can mitigate the welfare costs of equilibrium divergence, but it involves a trade-off between policy variance reduction (due to decreased polarization) and bias reduction. Appendix A.3 provides a detailed formalization of these arguments for when the implemented policy is a power-weighted compromise supported on $[\underline{x}, \bar{x}]$. Convergence of $(\underline{x}, \bar{x})$ toward the median type x_m mechanically reduces the variance of the implemented policy by shrinking the interval of feasible outcomes. At the same time, convergence toward x_m can shift the *average* implemented policy away from the utilitarian optimum x^o when $x^o \neq x_m$, increasing bias even as variance falls. The appendix formalizes this bias–variance decomposition (lemma A.5) and shows that maximal majority premia eliminate outcome variance but generically deliver a residual welfare loss of $(x_m - x^o)^2$ when $x_m \neq x^o$ (corollary 1). In general, the optimal $\rho_m^\star \in [0, \bar{\rho}]$ can be at an interior point.[23]

These observations suggest that institutional design plays an important role in shaping polarization incentives. Greater disproportionality between votes and power can, somewhat paradoxically, reduce policy divergence by strengthening parties' incentives to appeal to pivotal voters, consistent with empirical evidence in Matakos et al. (2016).

[23] A numerical example proving that $\rho_m^\star \notin \{0, \bar{\rho}\}$ can be made available by the author upon request.

Of course, broader considerations—such as checks and balances or the risk of majority tyranny—also matter. The key takeaway is that electoral institutions influence not only representation but also the intensity of political polarization. If power concentration itself affects voter welfare, the normative implications will be more ambiguous.

8 Conclusion

This paper proposes a supply-side explanation for political polarization based on electoral competition among office-motivated parties with decreasing marginal utility from political power. In the model, parties benefit from differentiating their platforms to insure against aggregate electoral shocks. This incentive leads not only to party divergence but also to strategic efforts to increase voter polarization.

The analysis shows that parties benefit from alienating the extreme supporters of their opponent and, more generally, from an electorate divided into two opposing factions. In a multidimensional setting, what matters for parties is not simply extremism on individual issues, but the alignment of disagreements across issues—greater ideological coherence within factions—which reduces voter switching and relaxes competitive pressure.

These incentives help rationalize several features of contemporary political conflict: the emphasis on divisive, zero-sum issues over common-interest policies, selective information provision, and the strategic use of party identity. While these strategies may reduce voter welfare, they increase equilibrium payoffs for competing parties.

Highlighting the supply-side incentives to divide voters is only a first step toward a more complete understanding of polarization. An important direction for future work is to study how politicians influence voter beliefs and how effective different communication strategies are in shaping multidimensional disagreement. Another is to examine how institutional design—such as electoral rules or campaign finance regulations—can attenuate or amplify the forces identified here.

More broadly, understanding the interaction between political communication, voter belief formation, and electoral incentives is essential for designing policies that limit polarization without undermining democratic competition.

References

Alesina, A., Miano, A., and Stantcheva, S. (2020). The polarization of reality. *AEA Papers and Proceedings*, 110:324–28.

Ali, S. N., Mihm, M., and Siga, L. (2024). The political economy of zero-sum thinking.

Autor, D., Dorn, D., Hanson, G., and Majlesi, K. (2020). Importing political polarization? the electoral consequences of rising trade exposure. *American Economic Review*, 110(10):3139–83.

Bernhardt, D., Buisseret, P., and Hidir, S. (2020). The race to the base. *American Economic Review*, 110(3):922–42.

Bertrand, M. and Kamenica, E. (2023). Coming apart? cultural distances in the united states over time. *American Economic Journal: Applied Economics*, 15(4):100–141.

Bonomi, G. (2026). Disagreement spillovers. *Mimeo*.

Bonomi, G., Gennaioli, N., and Tabellini, G. (2021). Identity, Beliefs, and Political Conflict. *The Quarterly Journal of Economics*, 136(4):2371–2411.

Bowen, T. R., Dmitriev, D., and Galperti, S. (2023). Learning from Shared News: When Abundant Information Leads to Belief Polarization*. *The Quarterly Journal of Economics*, 138(2):955–1000.

Brady, D. W., Han, H., and Pope, J. C. (2007). Primary elections and candidate ideology: Out of step with the primary electorate? *Legislative Studies Quarterly*, 32(1):79–105.

Bueno de Mesquita, E. and Dziuda, W. (2023). Partisan traps. Working Paper 31827, National Bureau of Economic Research.

Buttice, M. K. and Stone, W. J. (2012). Candidates matter: Policy and quality differences in congressional elections. *The Journal of Politics*, 74(3):870–887.

Carvalho, J.-P., Bergeron, A., Henrich, J., Nunn, N., and Weigel, J. L. (2023). Zero-sum thinking, the evolution of effort-suppressing beliefs, and economic development. Working Paper 31663, National Bureau of Economic Research.

Chinoy, S., Nunn, N., Sequeira, S., and Stantcheva, S. (2026). Zero-sum thinking and the roots of U.S. political differences. *American Economic Review*. Forthcoming.

Cohen, G. L. (2003). Party over policy: The dominating impact of group influence on political beliefs. *Journal of personality and social psychology*, 85(5):808.

Davis, N. T. and Mason, L. (2016). Sorting and the split-ticket: Evidence from presidential and subpresidential elections. *Political Behavior*, 38(2):337–354.

DellaVigna, S. and Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting*. *The Quarterly Journal of Economics*, 122(3):1187–1234.

Downs, A. (1957). *An economic theory of democracy*. Harper and Row, New York.

Feddersen, A. and Adams, J. F. (2022). Public opinion backlash in response to party messages: A case for party press releases. *Electoral Studies*, 80:102532.

Fiorina, M., Abrams, S., and Pope, J. (2011). *Culture War?: The Myth of a Polarized America*. Great questions in politics series. Longman.

Flores, A., Cole, J. C., Dickert, S., Eom, K., Jiga-Boy, G. M., Kogut, T., Loria, R., Mayorga, M., Pedersen, E. J., Pereira, B., Rubaltelli, E., Sherman, D. K., Slovic, P., Västfjäll, D., and Boven, L. V. (2022). Politicians polarize and experts depolarize public support for covid-19 management policies across countries. *Proceedings of the National Academy of Sciences*, 119(3):e2117543119.

Gennaioli, N. and Tabellini, G. (2025). Presidential address: Identity politics. *Econometrica*, 93(6):1937–1967.

Gentzkow, M. (2016). Polarization in 2016.

Gentzkow, M., Shapiro, J. M., and Taddy, M. (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, 87(4):1307–1340.

Glaeser, E. L., Ponzetto, G. A. M., and Shapiro, J. M. (2005). Strategic extremism: Why republicans and democrats divide on religious values. *The Quarterly Journal of Economics*, 120(4):1283–1330.

Greene, S. (1999). Understanding party identification: A social identity approach. *Political Psychology*, 20(2):393–403.

Grindle, M. S. (2012). *Jobs for the Boys: Patronage and the State in Comparative Perspective*. Harvard University Press, Cambridge, MA.

Hansson, I. and Stuart, C. (1984). Voting competitions with interested politicians: Platforms do not converge to the preferences of the median voter. *Public Choice*, 44(3):431–441.

Hirano, S., Snyder, J. M., and Ting, M. M. (2009). Distributive politics with primaries. *The Journal of Politics*, 71(4):1467–1480.

Huber, J. D. and Shipan, C. R. (2002). *Deliberate Discretion?: The Institutional Foundations of Bureaucratic Autonomy*. Cambridge University Press, Cambridge.

Katz, R. S. and Mair, P. (1995). Changing models of party organization and party democracy: The emergence of the cartel party. *Party Politics*, 1(1):5–28.

Kim, E., Lelkes, Y., and McCrain, J. (2022). Measuring dynamic media bias. *Proceedings of the National Academy of Sciences*, 119(32):e2202197119.

Kinder, D. R. and Kalmoe, N. P. (2017). *Neither Liberal nor Conservative: Ideological Innocence in the American Public*. University of Chicago Press.

Kopecký, P., Mair, P., and Spirova, M., editors (2012). *Party Patronage and Party Government in European Democracies*. Oxford University Press, Oxford.

Lewis, D. E. (2008). *The Politics of Presidential Appointments: Political Control and Bureaucratic Performance*. Princeton University Press, Princeton, NJ.

Lindbeck, A. and Weibull, J. (1993). A model of political equilibrium in a representative democracy. *Journal of Public Economics*, 51(2):195–209.

Lindbeck, A. and Weibull, J. W. (1987). Balanced-budget redistribution as the outcome of political competition. *Public Choice*, 52(3):273–297.

Martin, G. J. and Yurukoglu, A. (2017). Bias in cable news: Persuasion and polarization. *American Economic Review*, 107(9):2565–99.

Martin, L. W. and Vanberg, G. (2011). *Parliaments and Coalitions: The Role of Legislative Institutions in Multiparty Governance*. Oxford University Press, Oxford.

Matakos, K., Troumpounis, O., and Xefteris, D. (2016). Electoral rule disproportionality and platform polarization. *American Journal of Political Science*, 60(4):1026–1043.

Matsusaka, J. G. and Kendall, C. (2022). The common good and voter polarization. Working paper.

McCarty, N. (2016). View of polarization, congressional dysfunction, and constitutional change. *Indiana Law Review*, 50:223–245.

McCubbins, M. D., Noll, R. G., and Weingast, B. R. (1987). Administrative procedures as instruments of political control. *The Journal of Law, Economics, and Organization*, 3(2):243–277.

Moreira, T. M. Q. (2025). Is it still the economy? economic voting in polarized politics. *British Journal of Political Science*, 55:e45.

Müller, W. C. and Strøm, K., editors (1999). *Policy, Office, or Votes?: How Political Parties in Western Europe Make Hard Decisions*. Cambridge University Press, Cambridge.

Persson, T. and Tabellini, G. (2002). *Political Economics: Explaining Economic Policy*, volume 1 of *MIT Press Books*. The MIT Press.

Pew Research Center (2022). How religion intersects with americans' views on the environment. Technical report.

Pew Research Center (2023). Inflation, health costs, partisan cooperation among the nation's top problems. Technical report.

Shayo, M. (2009). A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political Science Review*, 103(2):147–174.

Sorokin, C. and Zakharov, A. (2018). Vote-motivated candidates. *Journal of Economic Theory*, 176:232–254.

Stantcheva, S. (2021). Understanding tax policy: How do people reason?*. *The Quarterly Journal of Economics*, 136(4):2309–2369.

Strøm, K. (1990). A behavioral theory of competitive political parties. *American Journal of Political Science*, 34(2):565–598.

Vaeth, M. (2024). Rational voter learning, issue alignment, and polarization. Working paper.

Wittman, D. (1977). Candidates with policy preferences: A dynamic model. *Journal of Economic Theory*, 14(1):180–189.

Zakharov, A. V. and Sorokin, C. S. (2014). Policy convergence in a two-candidate probabilistic voting model. *Social Choice and Welfare*, 43(2):429–446.

A Appendix

A.1 Proofs

Proof of Proposition 1 We prove the statement by contradiction. Assume that there exists a pure strategy equilibrium $x^\star \in \mathbb{R}^{2K}$ such that $x_A^\star = x_B^\star$. Since the electorate is diverse (not homogeneous), there exists a policy issue k such that $x_j^k \neq x_i^k$ for some $i, j \in I$. Fix k. Without loss of generality, assume that voter types are indexed in such a way that $x_1^k \leq x_2^k \leq ... \leq x_N^k$. Note that at least one inequality must hold strictly. Consider now deviation x_A for party A, with $x_A^k > x_A^{k\star}$ and $x_A^{k'} = x_A^{k'\star}$ for each $k' \neq k$. It is easy to see that, at platform profile $(x_A, x_B^\star)$, the expected utility of party A is

$$V_A(x_A, x_B^\star) = \nu(0) + \sum_{i=1}^{N} \pi_i(x_A, x_B^\star)\bar{w}_i$$

where, for each i, $\bar{w}_i = \nu(\bar{s}_i) - \nu(\bar{s}_{i+1})$ and $\bar{s}_i = \sum_{j \geq i} s_j$, with $\bar{s}_{N+1} = 0$. The partial derivative of the expected utility with respect to x_A^k at $(x_A, x_B^\star)$ exists and is equal to

$$\frac{\partial V_A(x_A, x_B^\star)}{\partial x_A^k} = \sum_{i=1}^{N} \frac{\partial \pi_i(x_A, x_B^\star)}{\partial x_A^k}\bar{w}_i.$$

Letting x_A^k approach $x_A^{k\star}$ from above, the right derivative with respect to x_A^k at $x^\star$ is

$$\lim_{x_A^k \to x_A^{k\star+}} \frac{\partial V_A(x_A, x_B^\star)}{\partial x_A^k} = \sum_{i=1}^{N} \frac{\partial \pi_i(x^\star)}{\partial x_A^k}\bar{w}_i.$$

Consider now the left derivative. Formally, let x_A be such that $x_A^k < x_A^{k\star}$ and $x_A^{k'} = x_A^{k'\star}$ for each $k' \neq k$. By following steps similar to the ones above, it is easy to see that

$$\lim_{x_A^k \to x_A^{k\star-}} \frac{\partial V_A(x_A, x_B^\star)}{\partial x_A^k} = \sum_{i=1}^{N} \frac{\partial \pi_i(x^\star)}{\partial x_A^k}\underline{w}_i.$$

where, for each i, $\underline{w}_i = \nu(1 - \bar{s}_{i+1}) - \nu(1 - \bar{s}_i)$.

We now show that $\lim_{x_A^k \to x_A^{k\star -}} \frac{\partial V_A(x_A, x_B^\star)}{\partial x_A^k} - \lim_{x_A^k \to x_A^{k\star +}} \frac{\partial V_A(x_A, x_B^\star)}{\partial x_A^k} < 0$, which proves that $x_A^\star$ is not a best response to $x_B^\star$. Note that the difference between limits is equal to

$$\sum_{i=1}^{N} \frac{\partial \pi_i(x^\star)}{\partial x_A^k}(\underline{w}_i - \bar{w}_i) = \sum_{i=1}^{N} \frac{\partial \pi_i(x^\star)}{\partial x_A^k}\left[(\nu(1-\bar{s}_{i+1}) + \nu(\bar{s}_{i+1})) - (\nu(1-\bar{s}_i) + \nu(\bar{s}_i))\right]$$
$$= \left(\frac{\partial \pi_N(x^\star)}{\partial x_A^k} - \frac{\partial \pi_1(x^\star)}{\partial x_A^k}\right)(\nu(1) + \nu(0)) + \sum_{i=2}^{N}\left(\frac{\partial \pi_{i-1}(x^\star)}{\partial x_A^k} - \frac{\partial \pi_i(x^\star)}{\partial x_A^k}\right)\Big(\nu(1-\bar{s}_i) + \nu(\bar{s}^i)\Big).$$

Next, note that, given the concavity of $v^k(x_A^k, x_i^k)$ in $|x_A^k - x_i|$ and $\frac{\partial \pi_i(x^\star)}{\partial x_A^k} = f(0)\frac{\partial v^k(x_A^k, x_i^k)}{\partial x_A^k}$, it has

$$\frac{\partial \pi_i(x^\star)}{\partial x_A^k} \geq \frac{\partial \pi_{i-1}(x^\star)}{\partial x_A^k} \text{ for } i = 2, ..., N$$

which holds strictly for at least some $i \in I$. Finally, using remark 1, it holds that $\nu(1-\bar{s}_i) + \nu(\bar{s}_i) > \nu(1) + \nu(0)$ for $i \geq 2$. It follows that

$$\sum_{i=1}^{N} \frac{\partial \pi_i(x^\star)}{\partial x_A^k}(\underline{w}_i - \bar{w}_i) < \underbrace{\left[\frac{\partial \pi_N(x^\star)}{\partial x_A^k} - \frac{\partial \pi_1(x^\star)}{\partial x_A^k} + \sum_{i=2}^{N}\left(\frac{\partial \pi_{i-1}(x^\star)}{\partial x_A^k} - \frac{\partial \pi_i(x^\star)}{\partial x_A^k}\right)\right]}_{=0}(\nu(1) + \nu(0))$$

which proves that $x_A^\star$ is not a best response to $x_B^\star$. Hence $x^\star$ is not a Nash equilibrium. ■

Remark 12 *Because the argument above only uses the equality $x_A^\star = x_B^\star$, it proves the stronger statement that for any $x \in \mathbb{R}^K$, if one party offers x, then x is not a best response for the opponent.*

Proof of Proposition 2 We start by proving equilibrium existence and part (i). Fix a distribution $(\mathbf{x}, \mathbf{s})$ and without loss of generality let $i, ..., N$ be such that $x_1 \leq x_2 \leq ... \leq x_N$. Fix $p \in \{A, B\}$ and fix $x_{-p} \in \mathbb{R}$. As a first step we show that, if $x_p \in \mathbb{R}$ is a best response to x_{-p}, then either $x_p > x_{-p}$ and $x_p = \sum_{i=1}^{N} \bar{w}_i x_i$, for $\bar{w}_i = \frac{\nu(\bar{s}_i) - \nu(\bar{s}_{i+1})}{\nu(1) - \nu(0)}$, or $x_p < x_{-p}$ and $x_p = \sum_{i=1}^{N} \underline{w}_i x_i$, for $\underline{w}_i = \frac{\nu(1-\bar{s}_{i+1}) - \nu(1-\bar{s}_i)}{\nu(1) - \nu(0)}$.

Assume that x_p is a best response to x_{-p}. By remark 12, $x_p \neq x_{-p}$. There are two cases. The first is $x_p > x_{-p}$. Subject to the constraint $x_p > x_{-p}$, V_p is strictly concave in p's platform x_p. Hence, x_p must satisfy the first-order condition $\frac{\partial V_p}{\partial x_p} = 0$. When $p = A$, the first order condition is

$$\frac{\partial V_A(x_A,x_B)}{\partial x_A}=\sum_{i=1}^{N}\frac{\partial \pi_i(x_A,x_B)}{\partial x_A}\bar{w}_i=0. \tag{5}$$

Note that $\pi_i(x_A,x_B)=\frac{1}{2}+\frac{\Delta_i}{2\phi}$ due to the uniform distribution assumption, so it has

$$\frac{\partial \pi_i(x_A,x_B)}{\partial x_A}=\frac{x_i-x_A}{\phi}\text{ for } i=1,...,N.$$

Plugging the above expression in 5, we obtain,

$$\frac{\partial V_A(x_A,x_B)}{\partial x_A}=0 \iff \sum_{i=1}^{N}\frac{x_i-x_A}{\phi}\bar{w}_i=0 \iff x_A=\sum_{i=1}^{N}\bar{w}_i x_i$$

Next, consider the second case, when the best response to x_{-p} is strictly less than x_{-p}. Because of strict quasi-concavity of V_p in party p's platform x under the constrain $x<x_{-p}$, x_p must satisfy the first order condition $\frac{\partial V_p}{\partial x_p}=0$. When $p=A$, this requires

$$\frac{\partial V_A(x_A,x_B)}{\partial x_A}=\sum_{i=1}^{N}\frac{\partial \pi_i(x_A,x_B)}{\partial x_A}\underline{w}_i=0. \tag{6}$$

Substituting for $\frac{\partial \pi_i(x_A,x_B)}{\partial x_A}=\frac{x_i-x_A}{\phi}$ in 6, we obtain,

$$\frac{\partial V_A(x_A,x_B)}{\partial x_A}=0 \iff \sum_{i=1}^{N}\frac{x_i-x_A}{\phi}\underline{w}_i=0 \iff x_A=\sum_{i=1}^{N}\underline{w}_i x_i.$$

Note that we could follow the exact same steps for $p=B$, and, as parties are identical with respect of the primitives of the game, we would indeed find the same best responses, $x_p=\sum_{i=1}^{N}\bar{w}_i x_i$ if $x_p>x_{-p}$, and $x_p=\sum_{i=1}^{N}\underline{w}_i x_i$ if $x_p<x_{-p}$. Hence, the game admits at most two best responses, $\underline{x}=\sum_{i=1}^{N}\underline{w}_i x_i$ and $\bar{x}=\sum_{i=1}^{N}\bar{w}_i x_i$.

Next, note that for each $p\in\{A,B\}$ and each $x_{-p}\in\mathbb{R}$, p has at least one best response to any x_{-p}. Proving this requires two simple steps. First, note that any platform choice outside $[x_1,x_N]$ is strictly dominated by either x_1 or x_N. We can therefore reduce the pure strategy set to $[x_1, x_N]$. Second, note that $V_A(\cdot, x_B)$ and $V_A(x_A, \cdot)$ are continuous on $[x_1, x_N]$. Hence, by the extreme value theorem applies.

Given that the best response correspondence is never empty valued, to show that a pure strategy Nash equilibrium exists it suffices to show that $\underline{x} < \bar{x}$. Define the risk-neutral weights

$$r_i := \frac{\rho(\bar{s}_i) - \rho(\bar{s}_{i+1})}{\rho(1) - \rho(0)} = \frac{\rho(\underline{s}_i) - \rho(\underline{s}_{i-1})}{\rho(1) - \rho(0)}, \qquad i = 1, \ldots, N,$$

where the equality follows from $1 - \bar{s}_i = \underline{s}_{i-1}$, $1 - \bar{s}_{i+1} = \underline{s}_i$, and assumption 1 (ii). Then

$$x^{RN} = \sum_{i=1}^{N} r_i x_i.$$

For each $i = 1, \ldots, N$, let $\lambda_i := \frac{\rho(\bar{s}_i) - \rho(0)}{\rho(1) - \rho(0)} \in [0, 1]$. Since $\rho(\bar{s}_i) = (1 - \lambda_i)\rho(0) + \lambda_i \rho(1)$, strict concavity of U implies $\nu(\bar{s}_i) = U(\rho(\bar{s}_i)) \geq (1 - \lambda_i)\nu(0) + \lambda_i \nu(1)$, with strict inequality whenever $0 < \lambda_i < 1$, i.e. whenever $i > 1$. Using telescoping,

$$\sum_{j \geq i}^{N} \bar{w}_j = \frac{\nu(\bar{s}_i) - \nu(0)}{\nu(1) - \nu(0)} \geq \lambda_i = \sum_{j \geq i}^{N} r_j,$$

strictly for every $i > 1$.

Next, assumption 1 (ii) gives $\rho(1 - \bar{s}_i) = \rho(1) + \rho(0) - \rho(\bar{s}_i)$, so that $1 - \lambda_i = \frac{\rho(1 - \bar{s}_i) - \rho(0)}{\rho(1) - \rho(0)}$. Applying the same concavity argument to $U(\rho(1 - \bar{s}_i))$ yields $\nu(1 - \bar{s}_i) \geq \lambda_i \nu(0) + (1 - \lambda_i)\nu(1)$, hence

$$\sum_{j \geq i}^{N} r_j = \lambda_i \geq \frac{\nu(1) - \nu(1 - \bar{s}_i)}{\nu(1) - \nu(0)} = \sum_{j \geq i}^{N} \underline{w}_j,$$

again strictly for every $i > 1$.

Hence, $(\bar{w}_i, x_i)_{i \in I}$ strictly first-order stochastically dominates $(r_i, x_i)_{i \in I}$, which strictly first-order stochastically dominates $(\underline{w}_i, x_i)_{i \in I}$. Since all three distributions have common support $\{x_1, \ldots, x_N\}$, it follows that

$$\underline{x} < x^{RN} < \bar{x}.$$

Since the only best responses are $\underline{x}$ and $\bar{x}$, and $\underline{x} < \bar{x}$, the pure-strategy equilibrium profiles are exactly $(\bar{x}, \underline{x})$ and $(\underline{x}, \bar{x})$. Moreover, $\underline{w}_i, \bar{w}_i > 0$ for all i because ν is strictly increasing and $s_i > 0$, so both $\underline{x}$ and $\bar{x}$ are strict interior convex combinations of $x_1, \ldots, x_N$. Hence

$$x_1 < \underline{x} < x^{RN} < \bar{x} < x_N.$$

This concludes the proof of equilibrium existence and of part (i).

To prove part (ii), note that for each $i \in I$, it has $\bar{w}_i = \nu(\bar{s}_i) - \nu(\bar{s}_{i+1})$ and $\underline{w}_i = \nu(1 - \bar{s}_{i+1}) - \nu(1 - \bar{s}_i)$. The former weight represents the payoff gain of the party offering $\bar{x}$ when she wins type i. Voter group i will vote for the party offering $\bar{x}$ only if all $j > i$ – that is, a share $\bar{s}_{i+1}$ of the distribution, also vote for that party. If voter group i also votes for the party offering $\bar{x}$, the vote share of that party increases from $\bar{s}_{i+1}$ to $\bar{s}_i = \bar{s}_{i+1} + s_i$, with an increase in payoff equal to $\nu(\bar{s}_i) - \nu(\bar{s}_{i+1}) = \bar{w}_i$. Using the same argument, it is easy to show that $\underline{w}_i = \nu(1 - \bar{s}_{i+1}) - \nu(1 - \bar{s}_i)$ is the payoff gain of the party offering $\underline{x}$ from winning voter group i.

Part (iii) follows from the location of the interval $[\bar{s}_{i+1}, \bar{s}_i]$ relative to $\frac{1}{2}$. Recall

$$\bar{w}_i = \nu(\bar{s}_i) - \nu(\bar{s}_{i+1}), \qquad \underline{w}_i = \nu(1 - \bar{s}_{i+1}) - \nu(1 - \bar{s}_i).$$

By remark 1, for any $a > b$ with $a \leq \frac{1}{2}$,

$$\nu(a) - \nu(b) > \nu(1 - b) - \nu(1 - a),$$

and for any $a > b$ with $b \geq \frac{1}{2}$,

$$\nu(a) - \nu(b) < \nu(1 - b) - \nu(1 - a).$$

Applying this with $a = \bar{s}_i$ and $b = \bar{s}_{i+1}$ yields

$$\bar{s}_i \leq \frac{1}{2} \implies \bar{w}_i > \underline{w}_i, \qquad \bar{s}_{i+1} \geq \frac{1}{2} \implies \bar{w}_i < \underline{w}_i.$$

If $x_i > x_m$, then at most half of voters lie weakly to the right of x_i, so $\bar{s}_i < \frac{1}{2}$. If $x_i < x_m$, then at least half of voters lie strictly to the right of x_i, so $\bar{s}_{i+1} \geq \frac{1}{2}$. Therefore $\bar{w}_i > \underline{w}_i$ for $x_i > x_m$

and $\bar{w}_i < \underline{w}_i$ for $x_i < x_m$.

Finally, assume, without loss of generality that $x^A = \bar{x}$ in equilibrium. It has

$$V_A(\bar{x},\underline{x}) = \nu(0) + \sum_{i=1}^{N} \pi_i(\bar{x},\underline{x})\bar{w}_i = \frac{\nu(1)+\nu(0)}{2} + \sum_{i=1}^{N} \frac{\underline{x}^2 - \bar{x}^2 - 2x_i(\underline{x}-\bar{x})}{2\phi}\bar{w}_i$$
$$= \frac{\nu(1)+\nu(0)}{2} + \frac{(\bar{x}-\underline{x})^2}{2\phi}$$

and

$$V_B(\bar{x},\underline{x}) = \nu(1) - \sum_{i=1}^{N} \pi_i(\bar{x},\underline{x})\underline{w}_i = \frac{\nu(1)+\nu(0)}{2} + \sum_{i=1}^{N} \frac{\bar{x}^2 - \underline{x}^2 - 2x_i(\bar{x}-\underline{x})}{2\phi}\underline{w}_i$$
$$= \frac{\nu(1)+\nu(0)}{2} + \frac{(\bar{x}-\underline{x})^2}{2\phi},$$

which concludes the proof of the proposition. ■

Proof of Proposition 3 Fix $(\mathbf{x},\mathbf{s}) = (x_i,s_i)_{i\in I}$, and, without loss, let indexes be ordered such that $x_1 \leq x_2 \leq ... \leq x_N$. Recall that $\hat{V}_p(\mathbf{x},\mathbf{s}) = V_p(\bar{x},\underline{x})$ by definition. Using this definition and plugging the expressions for $\bar{x}$ and $\underline{x}$ it is easy to see that $\frac{\partial \hat{V}_p(\mathbf{x},\mathbf{s})}{\partial x_i}$ is equal to $(\bar{x}-\underline{x})(\bar{w}_i - \underline{w}_i)$ multiplied by a proportionality constant. Hence, the derivative is strictly positive if $x_i > x_m$ and strictly negative if $x_i < x_m$, by proposition 2.

Next, note that

$$\frac{\partial \pi_i(\mathbf{x},\mathbf{s})}{\partial x_i} = \frac{(\underline{x}\underline{w}_i - \bar{x}\bar{w}_i) + (\bar{x}-\underline{x}) + x_i(\bar{w}_i - \underline{w}_i)}{\phi}.$$

If $x_i > \max\{\underline{x}, x_m\}$ then

$$(\underline{x}\underline{w}_i - \bar{x}\bar{w}_i) + (\bar{x}-\underline{x}) + x_i(\bar{w}_i - \underline{w}_i) > (\underline{x}\underline{w}_i - \bar{x}\bar{w}_i) + (\bar{x}-\underline{x}) + \underline{x}(\bar{w}_i - \underline{w}_i)$$
$$= (1-\bar{w}_i)(\bar{x}-\underline{x}) > 0,$$

and, given that $x_i > x_m \implies \frac{\partial \hat{V}_p(\mathbf{x},\mathbf{s})}{\partial x_i} > 0$, it has $\frac{\partial \hat{V}_p(\mathbf{x},\mathbf{s})}{\partial x_i}\frac{\partial \pi_i(\mathbf{x},\mathbf{s})}{\partial x_i} > 0$.

Similarly, if $x_i < \min\{\underline{x}, x_m\}$,

$$\begin{aligned}(\underline{x}\underline{w}_i - \bar{x}\bar{w}_i) + (\bar{x} - \underline{x}) + x_i(\bar{w}_i - \underline{w}_i) &= (1 - \bar{w}_i)\bar{x} - (1 - \underline{w}_i)\underline{x} + x_i(\bar{w}_i - \underline{w}_i) \\ &> (1 - \bar{w}_i)\bar{x} - (1 - \underline{w}_i)\underline{x} - (\underline{w}_i - \bar{w}_i)\underline{x} = (1 - \bar{w}_i)(\bar{x} - \underline{x}) > 0\end{aligned}$$

and, given that $x_i < x_m \implies \frac{\partial \hat{V}_p(\mathbf{x},\mathbf{s})}{\partial x_i} < 0$, it has $\frac{\partial \hat{V}_p(\mathbf{x},\mathbf{s})}{\partial x_i}\frac{\partial \pi_i(\mathbf{x},\mathbf{s})}{\partial x_i} < 0$. By the same argument, it is easy to show that if $x_i < \min\{\bar{x}, x_m\}$ then $\frac{\partial \hat{V}_p(\mathbf{x},\mathbf{s})}{\partial x_i}\frac{\partial (1-\pi_i(\mathbf{x},\mathbf{s}))}{\partial x_i} > 0$ and if $x_i > \max\{\bar{x}, x_m\}$ then $\frac{\partial \hat{V}_p(\mathbf{x},\mathbf{s})}{\partial x_i}\frac{\partial (1-\pi_i(\mathbf{x},\mathbf{s}))}{\partial x_i} < 0$, concluding the proof. Since $\underline{w}_i, \bar{w}_i \in (0,1)$ for each $i = 1, \dots, N$, we have $x_1 < \underline{x}$ and $x_N > \bar{x}$. Hence, $x_1 < \min\{\underline{x}, x_m\} < \max\{\bar{x}, x_m\} < x_N$ whenever $x_m \notin \{x_1, x_N\}$, which proves the remark ■

Proof of Proposition 4 Fix two voter distributions $(\mathbf{x}, \mathbf{s})$ and $(\mathbf{x}', \mathbf{s}')$ such that $(\mathbf{x}' - \mathbf{x}'_m, \mathbf{s}')$ is a spread of $(\mathbf{x} - \mathbf{x}_m, \mathbf{s})$ in the sense of definition 2. Translating all bliss points by a constant translates both equilibrium platforms by the same constant, so equilibrium platform distance and equilibrium payoffs are translation invariant. Hence, without loss of generality, recenter and assume $x_m = x'_m = 0$.

If either distribution assigns positive mass to the median bliss point 0, split that mass into two virtual types x_{m^-} and x_{m^+}, with $x_{m^-} = x_{m^+} = 0$, and choose their masses so that exactly half of the population lies weakly to each side of the split. Concretely, for $(\mathbf{x}, \mathbf{s})$ set

$$s_{m^-} = \tfrac{1}{2} - \sum_{x_i < 0} s_i, \qquad s_{m^+} = \tfrac{1}{2} - \sum_{x_i > 0} s_i,$$

and keep all other masses unchanged; define s'_{m^+} and s'_{m^-} using $(\mathbf{x}', \mathbf{s}')$ analogously. This operation leaves equilibrium platforms unchanged, since both virtual types have bliss point 0 and therefore do not affect any weighted average of bliss points.

Let $\bar{x} > \underline{x}$ denote the equilibrium platforms under $(\mathbf{x}, \mathbf{s})$, and $\bar{x}' > \underline{x}'$ those under $(\mathbf{x}', \mathbf{s}')$. We compare $\bar{x} - \underline{x}$ to $\bar{x}' - \underline{x}'$.

Define the common grid $\mathcal{X}$ as follows. Let $z_1 \le \cdots \le z_M$ be the weakly ordered list of all distinct bliss point values appearing in either distribution, except that if the median is split we include the value 0 twice, as $z_{m^-} = z_{m^+} = 0$, where z_{m^-} immediately precedes

$z_{m^+} = 0$ in the ordering. All other bliss point values appear only once in the grid. Thus, the grid is weakly ordered, with possible equality only at value 0:

$$z_1 \le z_2 \le \cdots \le z_M,$$

and $z_k = z_{k+1}$ can occur only when $(z_k, z_{k+1}) = (z_{m^-}, z_{m^+}) = (0,0)$. For each grid point $z_k \neq 0$, define $s(z_k) = \sum_{i:x_i=z_k} s_i$ and $s'(z_k) = \sum_{i:x'_i=z_k} s'_i$. If 0 belongs to the grid, then the median was split, so it appears twice. In this case, set $s(z_{m^-}) = s_{m^-}$, $s(z_{m^+}) = s_{m^+}$, $s'(z_{m^-}) = s'_{m^-}$ and $s'(z_{m^+}) = s'_{m^+}$.

For each $k = 1, \dots, M$, define tail masses

$$\bar{s}(z_k) := \sum_{\ell \ge k} s(z_\ell), \qquad \bar{s}'(z_k) := \sum_{\ell \ge k} s'(z_\ell).$$

By proposition 2 (ii), equilibrium platforms can be written as weighted averages of bliss points. On the grid $\{z_k\}$, define the equilibrium weights for $(\mathbf{x}, \mathbf{s})$ by $\bar{w}(z_k) = \nu(\bar{s}(z_k)) - \nu(\bar{s}(z_{k+1}))$ and $\underline{w}(z_k) = \nu(1 - \bar{s}(z_{k+1})) - \nu(1 - \bar{s}(z_k))$, with $\bar{s}(z_{M+1}) = 0$. Then

$$\bar{x} = \sum_{k=1}^{M} \bar{w}(z_k) z_k, \qquad \underline{x} = \sum_{k=1}^{M} \underline{w}(z_k) z_k, \qquad \bar{x} - \underline{x} = \sum_{k=1}^{M} \Big(\bar{w}(z_k) - \underline{w}(z_k)\Big) z_k.$$

Let $A_k := \sum_{\ell \ge k} \Big(\bar{w}(z_\ell) - \underline{w}(z_\ell)\Big)$ and $\Gamma(s) := \nu(s) + \nu(1-s) - \nu(1) - \nu(0)$. Note that $A_k = \Gamma(\bar{s}(z_k))$. Hence,

$$\bar{x} - \underline{x} = \sum_{k=1}^{M} \Gamma(\bar{s}(z_k))\,(z_k - z_{k-1}), \qquad z_0 := z_1.$$

The same construction for $(\mathbf{x}', \mathbf{s}')$ gives $\bar{x}' - \underline{x}' = \sum_{k=1}^{M} \Gamma(\bar{s}'(z_k))\,(z_k - z_{k-1})$.

We now use remark 1. For any $0 \le s < s' \le \frac{1}{2}$, it implies $\nu(s') - \nu(s) > \nu(1-s) - \nu(1-s')$. Equivalently,

$$\begin{aligned} \Gamma(s') > \Gamma(s) &\quad \text{for all } 0 \le s < s' \le \tfrac{1}{2}, \\ \Gamma(s') > \Gamma(s) &\quad \text{for all } \tfrac{1}{2} \le s' < s \le 1. \end{aligned}$$

Because $(\mathbf{x}', \mathbf{s}')$ is a spread of $(\mathbf{x}, \mathbf{s})$ around the common median 0, the right conditional distribution shifts to the right (first-order stochastic dominance) and the left conditional distribution shifts to the left. Evaluated on the common grid, this implies that for all k,

$$z_k > 0 \implies \bar{s}'(z_k) \geq \bar{s}(z_k), \qquad z_k < 0 \implies \bar{s}'(z_k) \leq \bar{s}(z_k),$$

with at least one strict inequality for some k with $z_k \neq 0$. Moreover, by construction of the median split,

$$\bar{s}(z_{m^+}) = \bar{s}'(z_{m^+}) = \tfrac{1}{2}, \qquad \bar{s}(z_{m^-}) \geq \tfrac{1}{2}, \qquad \bar{s}'(z_{m^-}) \geq \tfrac{1}{2}.$$

Fix k. If $z_k > 0$, then by construction $\bar{s}(z_k) \leq \frac{1}{2}$ and $\bar{s}'(z_k) \leq \frac{1}{2}$. Since $(\mathbf{x}', \mathbf{s}')$ is a spread of $(\mathbf{x}, \mathbf{s})$, we have $\bar{s}'(z_k) \geq \bar{s}(z_k)$. This implies $\Gamma(\bar{s}'(z_k)) \geq \Gamma(\bar{s}(z_k))$, with strict inequality whenever $\bar{s}'(z_k) > \bar{s}(z_k)$. If $z_k < 0$, then $\bar{s}(z_k) > \frac{1}{2}$ and $\bar{s}'(z_k) > \frac{1}{2}$, and the spread property implies $\bar{s}'(z_k) \leq \bar{s}(z_k)$. This again implies $\Gamma(\bar{s}'(z_k)) \geq \Gamma(\bar{s}(z_k))$, with strict inequality whenever $\bar{s}'(z_k) < \bar{s}(z_k)$. If $z_k = z_{m^-}$, the spread condition implies $s_{m^-} \geq s'_{m^-}$, because of the first order dominance relation between left conditional distributions.[24] Hence $\bar{s}(z_k) \geq \bar{s}'(z_k) \geq \frac{1}{2}$, which implies $\Gamma(\bar{s}'(z_k)) \geq \Gamma(\bar{s}(z_k))$. If $z_k = z_{m^+}$, then $\bar{s}(z_k) = \bar{s}'(z_k) = \frac{1}{2}$ by construction, so the coefficients coincide.

Since the spread is strict, the left or right conditional distribution changes strictly at some nonzero grid point; otherwise the two conditional distributions would coincide. Equivalently, there exists $k^\star \in \{1, \ldots, M\}$ with $z_{k^\star} \neq 0$ such that $\bar{s}'(z_{k^\star}) \neq \bar{s}(z_{k^\star})$. By the case analysis above, this inequality goes in the direction that raises Γ, so

$$\Gamma(\bar{s}'(z_{k^\star})) > \Gamma(\bar{s}(z_{k^\star})).$$

Because $z_{k^\star} \neq 0$, we have $z_{k^\star} - z_{k^\star - 1} > 0$. Since for every k we already established $\Gamma(\bar{s}'(z_k)) \geq \Gamma(\bar{s}(z_k))$ and every increment $z_k - z_{k-1}$ is nonnegative, it follows that

$$\bar{x}' - \underline{x}' = \sum_{k=1}^{M} \Gamma(\bar{s}'(z_k))\,(z_k - z_{k-1}) > \sum_{k=1}^{M} \Gamma(\bar{s}(z_k))\,(z_k - z_{k-1}) = \bar{x} - \underline{x}.$$

[24] FOSD of the left conditional distributions implies that their cumulative distribution functions satisfy $\lim_{t \to 0^-} F_L(t) \leq \lim_{t \to 0^-} F'_L(t)$. Since $\lim_{t \to 0^-} F_L(t) = 1 - 2s_{m^-}$ and $\lim_{t \to 0^-} F'_L(t) = 1 - 2s'_{m^-}$, it has $s_{m^-} \geq s'_{m^-}$.

Finally, proposition 2 (iv) implies that equilibrium expected payoffs satisfy

$$\hat{V}_p(\mathbf{x},\mathbf{s}) = \frac{\nu(1)+\nu(0)}{2} + \frac{(\bar{x}-\underline{x})^2}{2\phi}, \qquad p \in \{A,B\},$$

which is strictly increasing in $\bar{x}-\underline{x}$. Therefore $\hat{V}_p(\mathbf{x}',\mathbf{s}') > \hat{V}_p(\mathbf{x},\mathbf{s})$ for $p = A,B$. ■

Proof of Lemma 1 Throughout this proof and the following, we maintain the quadratic Euclidean loss and uniform shock assumptions stated in section 4 and the assumption of strict concavity of ν made in section 5. Before starting the proof of the lemma, it is useful to state and prove the following intermediate result.

Definition 5 *A policy platform $x_A \in \mathbb{R}^K$ is a* local best response *to $x_B \in \mathbb{R}^K$ if there exists $u > 0$ such that $V_A(x_A,x_B) \geq V_A(x'_A,x_B)$ for all $x'_A \in \mathbb{R}^K$ with $\|x'_A - x_A\| < u$. Similarly, a policy platform $x_B \in \mathbb{R}^K$ is a* local best response *to $x_A \in \mathbb{R}^K$ if there exists $u > 0$ such that $V_B(x_A,x_B) \geq V_B(x_A,x'_B)$ for all $x'_B \in \mathbb{R}^K$ with $\|x'_B - x_B\| < u$.*

Lemma A.1 *A policy platform $x_A \in \mathbb{R}^K$ for party A is a local best response to platform $x_B \in \mathbb{R}^K$ for party B if and only if $\frac{\partial V_A(x_A,x_B)}{\partial x_A^k} = 0$ for each $k = 1,\ldots,K$. A policy platform $x_B \in \mathbb{R}^K$ for party B is a local best response to platform $x_A \in \mathbb{R}^K$ for party A if and only if $\frac{\partial V_B(x_A,x_B)}{\partial x_B^k} = 0$ for each $k = 1,\ldots,K$. Moreover, for any $(x_A,x_B) \in \mathbb{R}^{2K}$, if x_A is a local best response to x_B or x_B is a local best response to x_A, then $\Delta_i(x_A,x_B) \neq \Delta_j(x_A,x_B)$ for each $i,j \in I, i \neq j$.*

Proof of Lemma A.1 First, we prove that if x_A is a local best response to x_B, then there cannot be $i,j \in I$, $i \neq j$, such that $\Delta_i(x) = \Delta_j(x)$. The proof is by contradiction.

Assume that $x_A \in \mathbb{R}^K$ is a local best response to $x_B \in \mathbb{R}^K$ such that $\exists i,j \in I, i \neq j$ with $\Delta_i(x) = \Delta_j(x)$. Since $x_i \neq x_j$, one can find $k \leq K$ such that $x_i^k \neq x_j^k$. Fix such k. Now consider the maximal subset $G \subset I$ such that (i) for each $i,j \in G$, $\Delta_i(x) = \Delta_j(x)$ and (ii) $\exists i,j \in G$ such that $x_i^k \neq x_j^k$. Note that G is non-empty by construction. Next, let $R \in \mathcal{R}(G)$ be any order over types in G such that for every two positions $i,j \in \{1,2,\ldots,|G|\}$, $i > j \implies x_{R_i}^k \geq x_{R_j}^k$, where R_j denotes the type in position j of the ranking. Next, let $\Delta = \max_{i\in G}\Delta_i(x)$, and define $\bar{s} = \sum_{j\in I:\Delta_j(x)>\Delta} s_j$ to be the total share of voters that are more likely to vote A than voters in G. Define, for each $i = 1,...,|G|$, $\bar{s}_{R_i} = \bar{s} + \sum_{j=i}^{|G|} s_{R_j}$. We will show that

$$\lim_{\hat{x}_A^k \to x_A^{k+}} \sum_{i=1}^{|G|} \frac{\partial \pi_{R_i}(x)}{\partial x_A^k} \bar{w}_{R_i}^{+} - \lim_{\hat{x}_A^k \to x_A^{k-}} \sum_{i=1}^{|G|} \frac{\partial \pi_{R_i}(x)}{\partial x_A^k} \bar{w}_{R_i}^{-} > 0,$$

where $\bar{w}_{R_i}^{+} = \nu(\bar{s}_{R_i}) - \nu(\bar{s}_{R_{i+1}})$ and $\bar{w}_{R_i}^{-} = \nu(\bar{s}+\bar{s}_{R_1}-\bar{s}_{R_{i+1}}) - \nu(\bar{s}+\bar{s}_{R_1}-\bar{s}_{R_i})$, and $\bar{s}_{R_{|G|+1}} = \bar{s}$. Note that the above difference of limits is equal to

$$\frac{1}{\phi}\left(\sum_{i=1}^{|G|} \bar{w}_{R_i}^{+} x_{R_i}^k - \sum_{i=1}^{|G|} \bar{w}_{R_i}^{-} x_{R_i}^k\right) > 0 \tag{7}$$

and such inequality holds true because $(x_{Ri}^k, \frac{\bar{w}_{R_i}^{+}}{\nu(\bar{s}_{R_1})-\nu(\bar{s})})_{i=1}^{|G|}$ strictly first order stochastically dominates $(x_{R_i}^k, \frac{\bar{w}_{R_i}^{-}}{\nu(\bar{s}_{R_1})-\nu(\bar{s})})_{i=1}^{|G|}$ by the strict concavity of ν.[25] To see this last result, fix $i \in \{1,\ldots,|G|\}$. Using definitions,

$$\sum_{j=i}^{|G|} \bar{w}_{R_j}^{+} = \nu(\bar{s}_{R_i}) - \nu(\bar{s}), \qquad \sum_{j=i}^{|G|} \bar{w}_{R_j}^{-} = \nu(\bar{s}_{R_1}) - \nu(\bar{s} + \bar{s}_{R_1} - \bar{s}_{R_i}).$$

Since $\bar{s}_{R_i} \in [\bar{s}, \bar{s}_{R_1}]$ and $\bar{s}_{R_i} + (\bar{s}+\bar{s}_{R_1}-\bar{s}_{R_i}) = \bar{s}+\bar{s}_{R_1}$, strict concavity of ν implies $\nu(\bar{s}_{R_i})-\nu(\bar{s}) \geq \nu(\bar{s}_{R_1})-\nu(\bar{s}+\bar{s}_{R_1}-\bar{s}_{R_i})$, strictly for some i because G contains at least two distinct x^k-values. This proves the dominance claim and inequality 7.

Next, we show that for any $i \in I$ such that $\nexists j, \Delta_i(x) = \Delta_j(x), x_i^k \neq x_j^k$, then the relative position of type i in the ordering of types in I based on $\Delta_i(x)$ does not change for small changes in x_A^k. Let $R \in \mathcal{R}$ be a (strict) type ordering such that $i > j \implies \Delta_{R_i}(x) \geq \Delta_{R_j}(x)$. Let $i \in \{1,...,|I|\}$ be such that $\nexists j \in \{1,...,|I|\}, j \neq i$ such that $\Delta_{R_i}(x) = \Delta_{R_j}(x)$ and $x_{R_i}^k \neq x_{R_j}^k$. Let $\hat{x} = (\hat{x}_A, x_B)$ be such that $\hat{x}_A^{k'} = x_A^{k'}$ for each $k' \neq k$. It has

$$\lim_{\hat{x}_A^k \to x_A^{k+}} \frac{\partial \pi_{R_i}(\hat{x})}{\partial x_A^k} \bar{w}_{R_i}^{+} - \lim_{\hat{x}_A^k \to x_A^{k-}} \frac{\partial \pi_{R_i}(\hat{x})}{\partial x_A^k} \bar{w}_{R_i}^{-} = 0,$$

where $\bar{w}_{R_i}^{+} = \bar{w}_{R_i}^{-} = \nu(\bar{s}_{R_i}) - \nu(\bar{s}_{R_{i+1}}) = \bar{w}_{R_i}$. To see why w_{R_i} is the same for the right and left limits, note that it only depends on $\bar{s}_{R_i} = \bar{s}_{R_{i+1}} + s_{R_i}$ and $\bar{s}_{R_{i+1}} = \sum_{j=i+1}^{|I|} s_{R_j}$. By construction

[25] Note that $\sum_{i=1}^{|G|} \bar{w}_{R_i}^{+} = \sum_{i=1}^{|G|} \bar{w}_{R_i}^{-} = \nu(\bar{s}_{R_1}) - \nu(\bar{s})$.

$\nexists j \in \{1,...,|I|\}, j \neq i$ such that $\Delta_{R_i}(x) = \Delta_{R_j}(x)$ and $x^k_{R_i} \neq x^k_{R_j}$, so $\bar{s}_{R_{i+1}}$ is constant in a neighborhood of x^k_A: the set of types in positions strictly above R_i is unaffected by small enough changes in x^k_A.

It follows that

$$\lim_{\hat{x}^k_A \to x^{k+}_A} \frac{\partial V_A(\hat{x})}{\partial x^k_A} - \lim_{\hat{x}^k_A \to x^{k-}_A} \frac{\partial V_A(\hat{x})}{\partial x^k_A}$$
$$= \lim_{\hat{x}^k_A \to x^{k+}_A} \sum_{i=1}^{|I|} \frac{\partial \pi_{R_i}(\hat{x})}{\partial x^k_A} \bar{w}^+_{R_i} - \lim_{\hat{x}^k_A \to x^{k-}_A} \sum_{i=1}^{|I|} \frac{\partial \pi_{R_i}(\hat{x})}{\partial x^k_A} \bar{w}^-_{R_i} > 0.$$

At a local optimum, the left derivative must be weakly nonnegative and the right derivative weakly nonpositive (subgradient condition); because of this, the strict inequality above proves that party A has a profitable marginal deviation from x_A. Hence x_A is not a local best response to x_B. This proves that if $x_A \in \mathbb{R}^K$ is a local best response to $x_B \in \mathbb{R}^K$, then for each $i, j \in I, i \neq j$, it has $\Delta_i(x) \neq \Delta_j(x)$.

Now, assume that $x_A \in \mathbb{R}^K$ is a local best response to $x_B \in \mathbb{R}^K$. Let $R \in \mathcal{R}(I)$ represent the strict ordering of types in I such that $i > j \implies \Delta_{R_i}(x) > \Delta_{R_j}(x)$. First, such an ordering is well-defined because of our previous result. Second, since the ordering is strict, it is preserved in a neighborhood of x: because of the continuity of $\Delta_i(x)$, $\Delta_{R_1}(x) < \Delta_{R_2}(x) < ... < \Delta_{R_{|I|}}(x)$ implies $\Delta_{R_1}(x') < \Delta_{R_2}(x') < ... < \Delta_{R_{|I|}}(x')$ for x' in a neighborhood of x. For each $i \in \{1,...,|I|\}$, let $\bar{s}_{R_i} = \sum_{j=i}^{|I|} s_{R_j}$ and $\bar{w}_{R_i} = \nu(\bar{s}_{R_i}) - \nu(\bar{s}_{R_{i+1}})$. The above implies that, for each i, k, k', p', $\frac{\partial \pi_{R_i}(x)}{\partial x^k_A} \bar{w}_{R_i}$ and $\frac{\partial^2 \pi_{R_i}(x)}{\partial x^k_A \partial x^{k'}_{p'}} \bar{w}_{R_i}$ exist and have the same sign as $\frac{\partial \pi_{R_i}(x)}{\partial x^k_A}$ and $\frac{\partial^2 \pi_{R_i}(x)}{\partial x^k_A \partial x^{k'}_{p'}}$, respectively. In particular, given our assumptions on payoffs, $\frac{\partial^2 \pi_{R_i}(x)}{\partial x^k_A \partial x^{k'}_{p'}} < 0$ if $A = p'$ and $k = k'$, and $\frac{\partial^2 \pi_{R_i}(x)}{\partial x^k_A \partial x^{k'}_{p'}} = 0$ otherwise. Note that $\frac{\partial^2 \pi_{R_i}(x)}{\partial x^k_A \partial x^k_A} < 0$ implies that, under our assumptions, $V_A(x)$ is strictly concave in x_A on any profile inducing a strict ranking.

Since the ranking induced by (x_A, x_B) is strict, it is locally preserved, and $V_A(x)$ is locally strictly concave in x_A. Hence, a local maximum exists if and only if the first-order conditions hold. It follows that, for each $k = 1,..,K$, it has $\frac{\partial V_A(x)}{\partial x^k_A} = 0$, $\frac{\partial^2 V_A(x)}{\partial x^k_A \partial x^k_A} < 0$, and $\frac{\partial^2 V_A(x)}{\partial x^k_A \partial x^{k'}_{p'}} = 0$ for $k' \neq k$ or $p' \neq A$. Hence, if x^A is a local best response, it satisfies the FOC.

To see why the reverse direction holds, observe that the argument used to show inequality 7 did not rely on x_A being a local best response, but only on the existence of a tie block G containing at least two types with distinct x^k values. In particular, whenever there exist $i \neq j$ such that $\Delta_i(x_A, x_B) = \Delta_j(x_A, x_B)$, the directional derivatives of $V_A(\cdot, x_B)$ with respect to x_A^k from the right and from the left differ, as shown in 7. Hence $V_A(\cdot, x_B)$ is not differentiable at any profile inducing Δ-ties. Therefore, if $\frac{\partial V_A(x_A,x_B)}{\partial x_A^k}$ exists for all k, the induced ranking of types must be strict. Under our maintained assumptions, strict rankings are locally preserved and $V_A(\cdot, x_B)$ is locally strictly concave in x_A. It follows that $\frac{\partial V_A(x_A,x_B)}{\partial x_A^k} = 0$ for all k is sufficient for x_A to be a local best response to x_B.

Following the same steps as above, one can show that (i) if $x_B \in \mathbb{R}^K$ is a local best response to $x_A \in \mathbb{R}^K$, then for each $i, j \in I, i \neq j$, it has $\Delta_i(x) \neq \Delta_j(x)$; and (ii) for each $k = 1,..,K$, it has $\frac{\partial V_B(x)}{\partial x_B^k} = 0$, $\frac{\partial^2 V_B(x)}{\partial x_B^k \partial x_B^k} < 0$, and $\frac{\partial^2 V_B(x)}{\partial x_B^k \partial x_{p'}^{k'}} = 0$ for $k' \neq k$ or $p' \neq B$. This concludes the proof of lemma A.1. ■

Note that, $(x_A, x_B) \in \mathcal{L}$ if and only if x_p is a local best response to x_{-p} for $p = A, B$. Lemma A.1 implies that every local equilibrium $x = (x_A, x_B)$ must induce a strict ordering over types, based on their $(\Delta_i(x))_{i \in I}$, and that the payoff function V_p is differentiable at any profile inducing such a strict ordering. Hence, for each $p = A, B$, x_p must satisfy the first order condition $\frac{\partial V_p(x)}{\partial x_p^k} = 0$, for $k = 1, ..., K$. But given that the payoff function is strictly concave in a neighborhood of every differentiability point, the first order condition is also sufficient for x_p to be a local best response to x_{-p}.

We now prove lemma 1. We start by showing that $\mathcal{L} = \mathcal{L}_{\mathcal{R}}$, for

$$\mathcal{L}_{\mathcal{R}} := \Big\{(x_A, x_B) \in \mathbb{R}^{2K} : \exists R \in \nabla(x_A, x_B), (x_A, x_B) = (\bar{x}(R), \underline{x}(R))\Big\}.$$

Let $\hat{x} \in \mathcal{L}$. First, we prove that $\hat{x}_A = \bar{x}(R)$ for $\{R\} = \nabla(\hat{x}_A, \hat{x}_B)$. Since $\hat{x} \in \mathcal{L}$, $\hat{x}_A$ is a local best response to $\hat{x}_B$. Let $R \in \mathcal{R}(I)$ be such that for each $i, j \in \{1, 2, .., |I|\}$, $i > j$ implies $\Delta_{R_i}(\hat{x}) > \Delta_{R_j}(\hat{x})$. By lemma $A.1$, R exists. Moreover, R is unique by construction. By A.1, the first order condition $\frac{\partial V_A(\hat{x})}{\partial x_A^k} = 0$ must hold for each $k = 1, ..., K$. Hence, for each

$k = 1, ..., K,$

$$
\begin{aligned}
&\frac{\partial V_A(\hat{x})}{\partial x_A^k} = 0 \\
\iff &\sum_{i=1}^{|I|} \frac{\partial \pi_{R_i}(\hat{x})}{\partial \hat{x}_A^k} \bar{w}_{R_i} = 0 \\
\iff &\hat{x}_A^k = \frac{\sum_{i=1}^{|I|} \bar{w}_{R_i} x_{R_i}^k}{\sum_{i=1}^{|I|} \bar{w}_{R_i}} = \frac{\sum_{i=1}^{|I|} \bar{w}_{R_i} x_{R_i}^k}{\nu(1) - \nu(0)} = \bar{x}^k(R),
\end{aligned}
$$

where $\bar{w}_{R_i} = \nu(\bar{s}_{R_i}) - \nu(\bar{s}_{R_{i+1}})$, $\bar{s}_{R_i} = \sum_{j=i}^{|I|} s_{R_j}$, with $\bar{s}_{R_{|I|+1}} = 0$, and the earlier normalization $\nu(1) - \nu(0) = 1$ was used. Hence, $\hat{x}_A = \bar{x}(R)$.

Next, we prove that $\hat{x}_B = \underline{x}(R)$ for $\{R\} = \nabla(\hat{x}_A, \hat{x}_B)$. Since $\hat{x} \in \mathcal{L}$, $\hat{x}_B$ is a local best response to $\hat{x}_A$. Let $R \in \mathcal{R}(I)$ be such that for each $i, j \in \{1, 2, .., |I|\}$, then $i > j$ implies $\Delta_{R_i}(\hat{x}) > \Delta_{R_j}(\hat{x})$. By lemma $A.1$, R exists. Moreover, R is unique by construction. By A.1, the first order condition $\frac{\partial V_B(\hat{x})}{\partial x_B^k} = 0$ must hold for each $k = 1, ..., K$. Hence, for each $k = 1, ..., K$,

$$
\begin{aligned}
&\frac{\partial V_B(\hat{x})}{\partial x_B^k} = 0 \\
\iff &\sum_{i=1}^{|I|} \frac{\partial \pi_{R_i}^B(\hat{x})}{\partial x_B^k} \underline{w}_{R_i} = 0 \\
\iff &\hat{x}_B^k = \frac{\sum_{i=1}^{|I|} \underline{w}_{R_i} x_{R_i}^k}{\sum_{i=1}^{|I|} \underline{w}_{R_i}} = \frac{\sum_{i=1}^{|I|} \underline{w}_{R_i} x_{R_i}^k}{\nu(1) - \nu(0)} = \underline{x}^k(R)
\end{aligned}
$$

where $\underline{w}_{R_i} = \nu(1 - \bar{s}_{R_{i+1}}) - \nu(1 - \bar{s}_{R_i})$, and the earlier normalization $\nu(1) - \nu(0) = 1$ was used. As a result, $\hat{x}_B = \underline{x}(R)$. We proved that $\{R\} = \nabla(\hat{x}_A, \hat{x}_B)$, $\hat{x}_A = \bar{x}(R)$, and $\hat{x}_B = \underline{x}(R)$. Hence $R \in \nabla(\bar{x}(R), \underline{x}(R))$ and $\hat{x} \in \mathcal{L}_{\mathcal{R}}$. This proves $\mathcal{L} \subseteq \mathcal{L}_{\mathcal{R}}$.

Before showing that $\mathcal{L}_{\mathcal{R}} \subseteq \mathcal{L}$, note that our proof that $\hat{x}_A = \bar{x}(R)$ only relied on $\hat{x}_A$ being a local best response to $\hat{x}_B$. Similarly, our proof that $\hat{x}_B = \underline{x}(R)$ only relied on $\hat{x}_B$ being a local best response to $\hat{x}_A$. Hence, the following remark holds.

Remark 13 *Let $x, x' \in \mathbb{R}^K$. If x' is a local best response to x, then there exists $R \in \nabla(x', x)$ such that $x' \in \{\bar{x}(R), \underline{x}(R)\}$.*

We now show $\mathcal{L}_{\mathcal{R}} \subseteq \mathcal{L}$. Let $\hat{x} \in \mathcal{L}_{\mathcal{R}}$. By construction, there exists $R \in \mathcal{R}(I)$ such that $\hat{x} = (\bar{x}(R), \underline{x}(R))$ and $R \in \nabla(\hat{x})$. Since $R \in \nabla(\hat{x})$ exists, $\Delta_{R_1}(\hat{x}) < \Delta_{R_2}(\hat{x}) < \ldots < \Delta_{R_{|I|}}(\hat{x})$, then V_A and V_B are differentiable at $\hat{x}$ by the same argument of the previous lemma. Note that, for each $k = 1, \ldots, K$

$$\frac{\partial V_A(\hat{x})}{\partial x_A^k} \propto \sum_{i=1}^{|I|} \bar{w}_{R_i} x_{R_i}^k - \hat{x}_A^k = \sum_{i=1}^{|I|} \bar{w}_{R_i} x_{R_i}^k - \bar{x}(R)^k = 0$$
$$\frac{\partial V_B(\hat{x})}{\partial x_B^k} \propto \sum_{i=1}^{|I|} \underline{w}_{R_i} x_{R_i}^k - \hat{x}_B^k = \sum_{i=1}^{|I|} \underline{w}_{R_i} x_{R_i}^k - \underline{x}(R)^k = 0$$

so that, by lemma A.1, platforms $\hat{x}_A$ and $\hat{x}_B$ are local best responses to each other. Hence $\hat{x} \in \mathcal{L}$. This proves that $\mathcal{L} = \mathcal{L}_{\mathcal{R}}$.

Finally, let $x \in \mathcal{L}$, and let $R \in \nabla(x)$ such that $x = (\bar{x}(R), \underline{x}(R))$ simple algebra yields

$$V_p(x) = \frac{\nu(1) + \nu(0)}{2} + \frac{\|x_A - x_B\|^2}{2\phi} = \frac{\nu(1) + \nu(0)}{2} + \frac{\|\bar{x}(R) - \underline{x}(R)\|^2}{2\phi}$$

for $p = A, B$, which concludes the proof of lemma 1. ■

Two Instrumental Lemmas I now prove two lemmas instrumental to the proofs of the next propositions.

Lemma A.2 *Let* $(\mathbf{x}, \mathbf{s}) \in \mathcal{V}$ *be symmetric around the origin,* $x, x' \in \mathbb{R}^K$*, and let* $R \in \nabla(x, x')$. *Then* R *is symmetric and*

$$\bar{x}(R) = -\underline{x}(R).$$

Proof. Symmetry around the origin implies that for each $i \in I$, there exists a conjugate type $\bar{i} \in I$ such that $x_{\bar{i}} = -x_i$ and $s_{\bar{i}} = s_i$.

Note that,

$$\Delta_i(x, x') = (\|x'\|^2 - \|x\|^2) + 2x_i^\top (x - x').$$

Thus R ranks types by $x_i^\top d$, where $d := x - x'$. Since $x_{\bar{i}}^\top d = -x_i^\top d$, strictness yields $R_j =$

$i \iff R_{|I|+1-j} = \bar{i}$. Let $\bar{s}_{R_j} := \sum_{\ell \geq j}^{|I|} s_{R_\ell}$ and define

$$\bar{w}_{R_j} := \nu(\bar{s}_{R_j}) - \nu(\bar{s}_{R_{j+1}}), \quad \underline{w}_{R_j} := \nu(1 - \bar{s}_{R_{j+1}}) - \nu(1 - \bar{s}_{R_j}).$$

Symmetry gives $\bar{s}_{R_j} = 1 - \bar{s}_{R_{|I|+2-j}}$, hence $\underline{w}_{R_j} = \bar{w}_{R_{|I|+1-j}}$. Therefore

$$\underline{x}(R) = \sum_j \underline{w}_{R_j} x_{R_j} = -\sum_j \bar{w}_{R_{|I|+1-j}} x_{R_{|I|+1-j}} = -\bar{x}(R),$$

using $x_{\bar{i}} = -x_i$. ■

Lemma A.3 *Let* $(\mathbf{x}, \mathbf{s}) \in \mathcal{V}$, *and let* $x_0 \in \mathcal{L}$, $x_0 = (x_{A0}, x_{B0})$. *For* $n \in \mathbb{N}$ *let* x_n *be the profile obtained in the* n^{th} *step of the best response dynamics initiated at the local equilibrium* x_0. *Then* $||x_{An} - x_{Bn}||^2 \geq ||x_{A0} - x_{B0}||^2$. *The inequality holds strictly for* $n \geq 2$ *if* x_0 *is not a Nash equilibrium.*

Proof. For any $x \in \mathbb{R}^K$, let $\mathcal{B}(x)$ be the set of (global) best responses to x. The payoff function is continuous and the usual dominance argument of probabilistic voting models implies that best responses must lie in $\times_{k \leq K}[x_\wedge^k, x_\vee^k]$, defined in section 5. Hence, for each $x \in \mathbb{R}^K$, $\mathcal{B}(x)$ is non-empty by the extreme value theorem.

Next, note that for any $x, x' \in \mathbb{R}^K$ whenever party $p = A, B$ plays x, party $-p$ plays x', and x is a local (or global) best response to x', party p's payoff is $\frac{\nu(1)+\nu(0)}{2} + \frac{||x-x'||^2}{2\phi}$. This means that if (x, y) is a local equilibrium, $y \notin \mathcal{B}(x)$, and $z \in \mathcal{B}(x)$, then $||x - z||^2 > ||x - y||^2$.

Define a best response sequence to be $\{x_n\}_{n=0}^\infty \in (\mathbb{R}^{2K})^{\mathbb{N}_0}$, $\{x_n\}_{n=0}^\infty = \{(x_{An}, x_{Bn})\}_{n=0}^\infty$ with the following properties:

(i) $(x_{A0}, x_{B0}) \in \mathbb{R}^{2K}$

(ii) for $n > 0$, n odd: set $x_{An} \in \mathcal{B}(x_{Bn-1})$, with $x_{An} = x_{An-1}$ if $x_{An-1} \in \mathcal{B}(x_{Bn-1})$; $x_{Bn} = x_{Bn-1}$.

(iii) for $n > 0$, n even: set $x_{Bn} \in \mathcal{B}(x_{An-1})$, with $x_{Bn} = x_{Bn-1}$ if $x_{Bn-1} \in \mathcal{B}(x_{An-1})$; $x_{An} = x_{An-1}$.

Note that we are requiring player A to best respond first, but nothing would change was B to move first.

Let $(\mathbf{s},\mathbf{x}) \in \mathcal{V}$. Without loss of generality, assume that the median voter type vector is $\mathbf{x}_m = 0$. Let $(x_{A0}, x_{B0}) \in \mathcal{L}$, and let $\{x_n\}_{n=0}^{\infty} \in (\mathbb{R}^{2k})^{\mathbb{N}_0}$ be a best response sequence initiated at $x_0 = (x_{A0}, x_{B0})$. If $x_0 \in \mathcal{E}$, $\|x_{An} - x_{Bn}\| = \|x_{A0} - x_{B0}\|$ holds by construction for each $n \geq 0$. It remains to prove that, if $x_0 \notin \mathcal{E}$, then $\|x_{An} - x_{Bn}\| > \|x_{A0} - x_{B0}\|$ for each $n > 1$. First, we prove the following lemma. Let $x^\top$ denote the transposition of vector x.

Lemma A.4 *Let* $(\mathbf{s},\mathbf{x}) \in \mathcal{V}$, $(x_{A0}, x_{B0}) \in \mathcal{L}$, *and* $\{x_n\}_{n=0}^{\infty} \in (\mathbb{R}^{2K})^{\mathbb{N}_0}$ *be a best response sequence initiated at* $x_0 = (x_{A0}, x_{B0})$. *Then*

$$\|x_{An}\|^2 - \|x_{Bn}\|^2 \geq \frac{1}{2}\|x_{A0} - x_{B0}\|^2 + 2x_{An}^\top x_{Bn} \qquad \forall n = 1, 3, 5, \ldots \tag{8}$$

and

$$\|x_{Bn}\|^2 - \|x_{An}\|^2 \geq \frac{1}{2}\|x_{A0} - x_{B0}\|^2 + 2x_{Bn}^\top x_{An} \qquad \forall n = 2, 4, 6, \ldots \tag{9}$$

Moreover, if $x_0 \notin \mathcal{E}$, *then both inequalities hold strictly for* $n > 1$

The proof of lemma A.4 is by induction. First, we show that if 8 holds for some $n \geq 1$ odd, then 9 holds for $n+1$ and 8 holds for $n+2$. Then, we show that 8 holds for $n = 1$.

Proof of Lemma A.4. Let $(\mathbf{s},\mathbf{x}) \in \mathcal{V}$. Let $x_0 \in \mathcal{L}$ and assume that 8 holds at $n \geq 1$, n odd. By construction $x_{Bn+1} \in \mathcal{B}(x_{An})$, hence it must hold that $V_B(x_{An}, x_{Bn+1}) \geq V_B(x_{An}, x_{Bn})$. Using the definitions of $V_B(x_{An}, x_{Bn})$ and $V_B(x_{An}, x_{Bn+1})$, and given that $x_{An} = x_{An+1}$ this requires

$$\|x_{An+1} - x_{Bn+1}\|^2 \geq \|x_{An}\|^2 - \|x_{Bn}\|^2 - 2\underaccent{\tilde}{x}(R)^\top (x_{An} - x_{Bn})$$

where $R \in \mathcal{R}(I)$ is the unique ordering of voters induced when A is playing the best response $x_{An} \in \mathcal{B}(x_{Bn})$, that is $\nabla(x_{An}, x_{Bn}) = \{R\}$. By lemma A.2, $\underaccent{\tilde}{x}(R) = -\tilde{x}(R) = -x_{An}$. Hence

$$\begin{aligned}
&\|x_{An+1} - x_{Bn+1}\|^2 \geq \|x_{An}\|^2 - \|x_{Bn}\|^2 + 2x_{An}^\top (x_{An} - x_{Bn}) \\
\iff\ &\|x_{Bn+1}\|^2 - \|x_{An+1}\|^2 \geq \|x_{An}\|^2 - \|x_{Bn}\|^2 - 2x_{An}^\top x_{Bn} + 2x_{An+1}^\top x_{Bn+1} \\
&\geq \frac{1}{2}\|x_{A0} - x_{B0}\|^2 + 2x_{An+1}^\top x_{Bn+1}
\end{aligned}$$

where the last inequality follows from the inductive assumption that 8 holds at n. Next, note that $n+1$ is even, hence $n+2$ is odd, hence $V_A(x_{An+2}, x_{Bn+2}) \geq V_A(x_{An+1}, x_{Bn+1})$. This implies

$$\begin{aligned}
&\|x_{An+2} - x_{Bn+2}\|^2 \geq \|x_{Bn+1}\|^2 - \|x_{An+1}\|^2 - 2\bar{x}(R)^\top (x_{Bn+1} - x_{An+1}) \\
\iff &\|x_{An+2} - x_{Bn+2}\|^2 \geq \|x_{Bn+1}\|^2 - \|x_{An+1}\|^2 + 2x_{Bn+1}^\top (x_{Bn+1} - x_{An+1}) \\
\iff &\|x_{An+2}\|^2 - \|x_{Bn+2}\|^2 \geq \|x_{Bn+1}\|^2 - \|x_{An+1}\|^2 - 2x_{Bn+1}^\top x_{An+1} + 2x_{Bn+2}^\top x_{An+2} \\
&\geq \frac{1}{2}\|x_{A0} - x_{B0}\|^2 + 2x_{Bn+2}^\top x_{An+2}
\end{aligned}$$

where the last inequality follows from the inductive assumption that 9 holds at $n+1$. We now show that $\|x_{A1}\|^2 - \|x_{B1}\|^2 \geq \frac{1}{2}\|x_{A0} - x_{B0}\|^2 + 2x_{A1}^\top x_{B1}$. Note that since $x_{A1} \in \mathcal{B}(x_{B0})$, we have $\|x_{A1} - x_{B1}\|^2 \geq \|x_{A0} - x_{B0}\|^2$. Since $x_0 \in \mathcal{L}$, $x_{A0} = \bar{x}(R)$ and $x_{B0} = \underline{x}(R)$ for the same ordering $R \in \mathcal{R}(I)$. Hence, by the symmetry argument used earlier in the proof, $x_{A0} = \bar{x}(R) = -\underline{x}(R) = -x_{B0}$. This implies

$$\begin{aligned}
&\|x_{A1} - x_{B1}\|^2 \geq \|x_{A0} - x_{B0}\|^2 \\
\iff &\|x_{A1} - x_{B1}\|^2 \geq \frac{1}{2}\|x_{A0} - x_{B0}\|^2 + \frac{1}{2}\|x_{A0} - x_{B0}\|^2 \\
\iff &\|x_{A1} - x_{B1}\|^2 \geq 2\|x_{B0}\|^2 + \frac{1}{2}\|x_{A0} - x_{B0}\|^2 \\
\iff &\|x_{A1}\|^2 - \|x_{B1}\|^2 \geq \frac{1}{2}\|x_{A0} - x_{B0}\|^2 + 2x_{A1}^\top x_{B1}
\end{aligned}$$

This concludes the proof by induction of 8 and 9. Finally, note that if $x_0 \notin \mathcal{E}$, then either $\|x_{A1} - x_{B1}\|^2 > \|x_{A0} - x_{B0}\|^2$, which implies $\|x_{A1}\|^2 - \|x_{B1}\|^2 > \frac{1}{2}\|x_{A0} - x_{B0}\|^2 + 2x_{A1}^\top x_{B1}$, or $\|x_{A2} - x_{B2}\|^2 > \|x_{A1} - x_{B1}\|^2$, which implies $\|x_{B2}\|^2 - \|x_{A2}\|^2 > \frac{1}{2}\|x_{A0} - x_{B0}\|^2 + 2x_{A2}^\top x_{B2}$. Following the same steps of the above proof by induction, it is clear that the strict inequality relation carries through for higher values of n, that is for every $n > 1$. This concludes the proof of lemma A.4. ■

We now complete the proof of lemma A.3. Let $x_0 \in \mathcal{L}$. Fix $n > 0$, odd. Using the construction of the best response dynamics and lemma A.2, it has

$$\begin{aligned}\|x_{An} - x_{Bn}\|^2 &\geq \|x_{Bn-1}\|^2 - \|x_{An-1}\|^2 + 2x_{Bn-1}^\top (x_{Bn-1} - x_{An-1})\\ &\geq \|x_{Bn-1} - x_{An-1}\|^2 + 2(\|x_{Bn-1}\|^2 - \|x_{An-1}\|^2).\end{aligned}$$

By lemma A.4, it has

$$\begin{aligned}\|x_{An} - x_{Bn}\|^2 &\geq \|x_{An-1} - x_{Bn-1}\|^2 + 2(\|x_{Bn-1}\|^2 - \|x_{An-1}\|^2)\\ &\geq \|x_{An-1} - x_{Bn-1}\|^2 + \|x_{A0} - x_{B0}\|^2 + 4x_{Bn-1}^\top x_{An-1}\\ \iff \|x_{An} - x_{Bn}\|^2 &\geq \|x_{An-1} + x_{Bn-1}\|^2 + \|x_{A0} - x_{B0}\|^2 \geq \|x_{A0} - x_{B0}\|^2,\end{aligned}$$

where lemma A.4 was used in the second inequality. Note that if $x_0 \notin \mathcal{E}$, our previous results imply that $\|x_{An} - x_{Bn}\|^2 > \|x_{A0} - x_{B0}\|^2$ provided that $n \geq 3$, because the inequality of lemma A.4 always holds strictly from $n = 2$ onwards.

Next, fix $n > 0$, even. Using the construction of the best response dynamics and lemma A.2, it has

$$\begin{aligned}\|x_{An} - x_{Bn}\|^2 &\geq \|x_{An-1}\|^2 - \|x_{Bn-1}\|^2 + 2x_{An-1}^\top (x_{An-1} - x_{Bn-1})\\ &\geq \|x_{An-1} - x_{Bn-1}\|^2 + 2(\|x_{An-1}\|^2 - \|x_{Bn-1}\|^2).\end{aligned}$$

By lemma A.4, it has

$$\begin{aligned}\|x_{An} - x_{Bn}\|^2 &\geq \|x_{An-1} - x_{Bn-1}\|^2 + 2(\|x_{An-1}\|^2 - \|x_{Bn-1}\|^2)\\ &\geq \|x_{An-1} - x_{Bn-1}\|^2 + \|x_{A0} - x_{B0}\|^2 + 4x_{An-1}^\top x_{Bn-1}\\ \iff \|x_{An} - x_{Bn}\|^2 &\geq \|x_{An-1} + x_{Bn-1}\|^2 + \|x_{A0} - x_{B0}\|^2 \geq \|x_{A0} - x_{B0}\|^2,\end{aligned}$$

where lemma A.4 was used in the second inequality. Note that if $x_0 \neq \mathcal{E}$, our previous results imply that $\|x_{An} - x_{Bn}\|^2 > \|x_{A0} - x_{B0}\|^2$, since the inequality of lemma A.4 holds strictly from $n = 2$ onwards. This concludes the proof of lemma A.3. ■

Proof of Proposition 5 Let $(\mathbf{x},\mathbf{s}) \in \mathcal{V}$. By translation invariance, we may normalize the center of symmetry to $x_m = 0$, so that lemma A.2 applies directly. Existence of the Nash equilibrium follows from the following two observations: (i) the game admits a potential function defined over platform pairs inducing strict rankings; and (ii) the set of best responses to pure strategies is non-empty and finite.

To see why (ii) is true, define the following two sets

$$\mathcal{B} := \{x \in \mathbb{R}^K : \exists y \in \mathbb{R}^K, x \in \mathcal{B}(y)\}$$
$$\mathcal{B}_{\mathcal{R}} := \left\{x \in \mathbb{R}^K : \exists R \in \mathcal{R}(I), x \in \{\bar{x}(R), \underline{x}(R)\}\right\}$$

Note that $\mathcal{B}$ is the set of best responses to pure strategies. As we proved at the beginning of the proof of lemma A.3, for each $x \in \mathbb{R}^K$, $\mathcal{B}(x)$ is nonempty by the extreme value theorem. Hence $\mathcal{B}$ is nonempty. Let $x' \in \mathcal{B}$. By definition, $\exists x \in \mathbb{R}^K$ such that $x' \in \mathcal{B}(x)$. Because x' is a global best response to x, it is a local best response to x. By lemma A.1, $\nabla(x', x)$ is nonempty, by remark 13, $\exists R \in \nabla(x', x) \subseteq \mathcal{R}(I)$ such that $x' \in \{\bar{x}(R), \underline{x}(R)\}$. Hence, $x' \in \mathcal{B}_{\mathcal{R}}$. This proves $\mathcal{B} \subseteq \mathcal{B}_{\mathcal{R}}$. Next, note that because $|I| < \infty$ then $|\mathcal{R}(I)| < \infty$, which implies $|\mathcal{B}_{\mathcal{R}}| < \infty$. Since $\mathcal{B} \subseteq \mathcal{B}_{\mathcal{R}}$, $|\mathcal{B}| < \infty$ also holds, which proves finiteness.

To see that (i) is true, define the set of platform profiles that induces strict rankings, $\mathcal{S} := \left\{(x, x') \in \mathbb{R}^{2K} : \nabla(x, x') \neq \emptyset\right\}$. Note that for each $(x_A, x_B) \in \mathcal{S}$, we have

$$V_A(x_A, x_B) = \frac{\nu(1) + \nu(0)}{2} + \frac{1}{2\phi}\left[\|x_B\|^2 - \|x_A\|^2 - 2\,\bar{x}(R)^\top (x_B - x_A)\right],$$
$$V_B(x_A, x_B) = \frac{\nu(1) + \nu(0)}{2} - \frac{1}{2\phi}\left[\|x_B\|^2 - \|x_A\|^2 - 2\,\underline{x}(R)^\top (x_B - x_A)\right].$$

where $\{R\} = \nabla(x_A, x_B)$. Equivalently, using A.2,

$$V_A(x_A, x_B) = P(x_A, x_B) + \frac{\|x_B\|^2}{\phi}, \qquad V_B(x_A, x_B) = P(x_A, x_B) + \frac{\|x_A\|^2}{\phi},$$

for $P : (x_A, x_B) \mapsto V_A(x_A, x_B) - \frac{\|x_B\|^2}{\phi}$, and $(x_A, x_B) \in \mathcal{S}$. Note that for any $x, x', x'' \in \mathbb{R}^K$ such that $(x,x') \in \mathcal{S}$ and $(x,x'') \in \mathcal{S}$, we have

$$V_A(x'',x) - V_A(x',x) = P(x'',x) - P(x',x), \qquad V_B(x,x'') - V_B(x,x') = P(x,x'') - P(x,x').$$

In other words, P works as an exact potential, provided that we consider profiles in $\mathcal{S}$.

We now prove the existence of a Nash equilibrium. Since any global best response is a local best response, note that any profile $(x,x') \in \mathbb{R}^{2K}$ with $x' \in \mathcal{B}(x)$ or $x \in \mathcal{B}(x')$ is such that at least one party is playing a local best response to the other party's platform, so that $(x,x') \in \mathcal{S}$ by lemma A.1. This means that any platform profile (x_{An}, x_{Bn}) reached at iteration $n \geq 1$ of the best response dynamics of lemma A.3 satisfies $(x_{An}, x_{Bn}) \in \mathcal{S}$. Hence P is non-decreasing along the best-response dynamics path after step $n = 1$. Second, by definition, any platform profile (x_{An}, x_{Bn}) reached at any iteration $n > 1$ of such best response dynamics satisfies $(x_{An}, x_{Bn}) \in \mathcal{B} \times \mathcal{B}$. Because $\mathcal{B}$ is finite, the best response dynamics exhibits a finite number of distinct platforms.

Pick any $x \in \mathbb{R}^{2K}$ and initialize the best response dynamics at x. Consider any $n > 1$. By the definition of the dynamics, $P(x_{An+2}, x_{Bn+2}) \geq P(x_{An}, x_{Bn})$. If $(x_{An+2}, x_{Bn+2}) = (x_{An}, x_{Bn})$ then no player deviated when given the chances, so the dynamics converged and $(x_{An}, x_{Bn}) \in \mathcal{E}$. If $(x_{An+2}, x_{Bn+2}) \neq (x_{An}, x_{Bn})$ then $P(x_{An+2}, x_{Bn+2}) > P(x_{An}, x_{Bn})$. In fact, by construction of the dynamics, a player changes platform only when the previous platform is not a best response, so any change yields a strict payoff improvement and hence a strict increase P. But because P is non-decreasing on the best response path and the best response path exhibits a finite number of distinct platforms, P can only increase strictly between n and $n+2$ for a finite number of iterations n. This implies that the best response sequence converges and a Nash equilibrium exists.

Next, we prove that $\mathcal{E}^\star$ is non-empty and finite. It is sufficient to note the non-emptiness and finiteness of $\mathcal{E}$, which follows from the previous step of the proof and from the finiteness of $\mathcal{B}$. Finally, we prove that

$$\mathcal{E}^\star = \arg \max_{(x_A, x_B) \in \mathcal{L}} \|x_A - x_B\|^2.$$

Let $\hat{x} \in \arg\max_{(x_A,x_B)\in\mathcal{L}} \|x_A - x_B\|^2$ and assume per contra that $\hat{x} \notin \mathcal{E}^\star$. There are two cases, $\hat{x} \in \mathcal{E}$ and $\hat{x} \notin \mathcal{E}$. If $\hat{x} \in \mathcal{E}$, it means that $\exists x \in \mathcal{E}$ such that $V_A(x) > V_A(\hat{x})$. Since $\mathcal{E} \subseteq \mathcal{L}$, lemma 1 implies $\|x_A - x_B\|^2 > \|\hat{x}_A - \hat{x}_B\|^2$, but then $\exists x \in \mathcal{L}$ such that $\|x_A - x_B\|^2 > \|\hat{x}_A - \hat{x}_B\|^2$, which is a contradiction. Hence it must be $\hat{x} \notin \mathcal{E}$. But we know that the best response sequence initiated at $\hat{x}$ must converge to some pure strategy equilibrium $x \in \mathcal{E}$. Since $\hat{x} \in \mathcal{L}$ by construction, lemma A.3 implies that, according to the best response dynamics initiated at $\hat{x}$, $\|x_{An} - x_{Bn}\|^2 > \|\hat{x}_A - \hat{x}_B\|^2$ for each $n > 1$. Hence it must be that $\|x_A - x_B\|^2 > \|\hat{x}_A - \hat{x}_B\|^2$. Since $\mathcal{E} \subseteq \mathcal{L}$, $x \in \mathcal{L}$. But this means that $\exists x \in \mathcal{L}$ such that $\|x_A - x_B\|^2 > \|\hat{x}_A - \hat{x}_B\|^2$, which is a contradiction. Hence $\hat{x} \in \mathcal{E}^\star$, and $\arg\max_{(x_A,x_B)\in\mathcal{L}} \|x_A - x_B\|^2 \subseteq \mathcal{E}^\star$.

Let $\hat{x} \in \mathcal{E}^\star$ and assume per contra that $\hat{x} \notin \arg\max_{(x_A,x_B)\in\mathcal{L}} \|x_A - x_B\|^2$. Then there exists $x = (x_A, x_B) \in \arg\max_{(x_A,x_B)\in\mathcal{L}} \|x_A - x_B\|^2$ such that $\|x_A - x_B\|^2 > \|\hat{x}_A - \hat{x}_B\|^2$. Since $\hat{x} \in \mathcal{E}^\star$, $x \notin \mathcal{E}$. But the best response sequence initiated at x converges to some pure strategy equilibrium $x' \in \mathcal{E}$. Since $x \in \mathcal{L}$ by construction, lemma A.3 implies that, if the best response sequence is initiated at x, $\|x_{An} - x_{Bn}\|^2 > \|x_A - x_B\|^2$ for each $n > 1$. Hence it must be that $\|x'_A - x'_B\|^2 > \|x_A - x_B\|^2$. Since $\mathcal{E} \subseteq \mathcal{L}$, $\exists x' \in \mathcal{L}$ such that $\|x'_A - x'_B\|^2 > \|x_A - x_B\|^2$, which is a contradiction. Hence $\hat{x} \in \arg\max_{(x_A,x_B)\in\mathcal{L}} \|x_A - x_B\|^2$, and $\mathcal{E}^\star \subseteq \arg\max_{(x_A,x_B)\in\mathcal{L}} \|x_A - x_B\|^2$. This proves $\mathcal{E}^\star = \arg\max_{(x_A,x_B)\in\mathcal{L}} \|x_A - x_B\|^2$. ■

Proof of Proposition 6 To prove both point (i) and point (ii) of the proposition it is sufficient to take the derivative of $\hat{V}_p^R$ with respect of $x_{R_i}^k$. Fix $(\mathbf{x}, \mathbf{s})$ and $k \leq K$. Let $x \in \mathcal{L}(\mathbf{x}, \mathbf{s})$ and let $R \in \nabla(x)$. Note that, for $i \in I$

$$\frac{\partial \hat{V}_p^R(\mathbf{x}, \mathbf{s})}{\partial x_{R_i}^k} = \frac{\bar{w}_{R_i} - \underline{w}_{R_i}}{\phi}\left(\bar{x}^k(R) - \underline{x}^k(R)\right).$$

The derivative is well defined because the ranking induced by x on $(\Delta_i(x))_{i\in I}$ is strict as proven in lemma A.1. Next, note that if $i > m_R$ then $\bar{w}_{R_i} > \underline{w}_{R_i}$, while if $i < m_R$ then $\bar{w}_{R_i} < \underline{w}_{R_i}$. Hence

$$\frac{\partial \hat{V}_p^R(\mathbf{x}, \mathbf{s})}{\partial x_{R_i}^k}\left(\bar{x}^k(R) - \underline{x}^k(R)\right) = \frac{\bar{w}_{R_i} - \underline{w}_{R_i}}{\phi}\left(\bar{x}^k(R) - \underline{x}^k(R)\right)^2,$$

which has the weak sign stated in part (i), and is strict whenever $\bar{x}^k(R) \neq \underline{x}^k(R)$. As for part (ii), note that, from the above, it has that,

$$\left[\frac{\partial \hat{V}_p^R(\mathbf{x},\mathbf{s})}{\partial x_{R_i}^1}, \frac{\partial \hat{V}_p^R(\mathbf{x},\mathbf{s})}{\partial x_{R_i}^2}, \ldots, \frac{\partial \hat{V}_p^R(\mathbf{x},\mathbf{s})}{\partial x_{R_i}^K}\right] =$$

$$\frac{\bar{w}_{R_i} - \underline{w}_{R_i}}{\phi}\left[\bar{x}^1(R) - \underline{x}^1(R), \bar{x}^2(R) - \underline{x}^2(R), \ldots, \bar{x}^K(R) - \underline{x}^K(R)\right]$$

so that the direction of fastest payoff improvement is $\bar{x}(R) - \underline{x}(R)$ if $i > m_R$ and $\underline{x}(R) - \bar{x}(R)$ if $i < m_R$.

Proof of Proposition 7 Fix $(\mathbf{x},\mathbf{s}), (\tilde{\mathbf{x}},\tilde{\mathbf{s}}) \in \mathcal{V}$ and let $\hat{x} \in \mathcal{E}^\star(\mathbf{x},\mathbf{s})$. Define $d \equiv \hat{x}_A - \hat{x}_B$ and assume that $(\tilde{\mathbf{x}},\tilde{\mathbf{s}})$ is a spread of $(\mathbf{x},\mathbf{s})$ in direction d. We prove that

$$\hat{V}_p(\tilde{\mathbf{x}},\tilde{\mathbf{s}}) > \hat{V}_p(\mathbf{x},\mathbf{s}).$$

$\hat{V}_p(\mathbf{x},\mathbf{s})$ is invariant to translations of the symmetric marginals $(\mathbf{x}^k,\mathbf{s})_{k \le K}$. Hence, without loss of generality, we assume that both $(\mathbf{x},\mathbf{s})$ and $(\tilde{\mathbf{x}},\tilde{\mathbf{s}})$ have marginal distributions centered at 0. By lemma A.2, any $\hat{x} \in \mathcal{E}^\star(\mathbf{x},\mathbf{s}) \subseteq \mathcal{L}(\mathbf{x},\mathbf{s})$ satisfies $\hat{x}_B = -\hat{x}_A$, so that $d = 2\hat{x}_A$. In particular, for every voter type i, the projections $x_i^\top d$ and $x_i^\top \hat{x}_A$ differ only by a positive scalar factor, and therefore they induce exactly the same weak ordering of types.

For each distribution $(\mathbf{x},\mathbf{s})$, let I index its types and let $\mathcal{R}(I)$ denote the set of strict rankings (permutations) of I. For a platform profile x, let $\nabla(x) \subseteq \mathcal{R}(I)$ be the set of rankings consistent with x. For any $R \in \nabla(x)$, let $(\bar{w}_{R_i})_{i \in I}$ denote the associated weights constructed "as usual" in the appendix, i.e. from R and $(\mathbf{x},\mathbf{s})$. Similarly, for $(\tilde{\mathbf{x}},\tilde{\mathbf{s}})$, let $\tilde{I}$ index its types, fix $\tilde{R} \in \nabla(\hat{x}) \subseteq \mathcal{R}(\tilde{I})$, and define weights $(\bar{w}_{\tilde{R}_i})_{i \in \tilde{I}}$ from $\tilde{R}$ and $(\tilde{\mathbf{x}},\tilde{\mathbf{s}})$.

Let $\tilde{V}_p(\hat{x})$ denote party p's expected payoff when parties play $\hat{x}$ under voter distribution $(\tilde{\mathbf{x}},\tilde{\mathbf{s}})$, and let $V_p(\hat{x})$ denote the corresponding payoff under $(\mathbf{x},\mathbf{s})$.

Since $\hat{x} \in \mathcal{L}(\mathbf{x},\mathbf{s})$, lemma 1 implies

$$V_p(\hat{x}) = \frac{\nu(1) + \nu(0)}{2} + \frac{1}{2\phi}\|\hat{x}_A - \hat{x}_B\|^2.$$

Moreover, for any $\tilde{R} \in \nabla(\hat{x})$,

$$\tilde{V}_p(\hat{x}) = \frac{\nu(1) + \nu(0)}{2} + \frac{2}{\phi} \sum_{i \in \tilde{I}} \bar{w}_{\tilde{R}_i} \tilde{x}_{\tilde{R}_i}^{\top} \hat{x}_A,$$

and, for any $R \in \nabla(\hat{x})$,

$$V_p(\hat{x}) = \frac{\nu(1) + \nu(0)}{2} + \frac{2}{\phi} \sum_{i \in I} \bar{w}_{R_i} x_{R_i}^{\top} \hat{x}_A,$$

where we used $\hat{x}_A = -\hat{x}_B$, by lemma A.2. Define the projected unidimensional types

$$\tilde{z}_{\tilde{R}_i} \equiv \tilde{x}_{\tilde{R}_i}^{\top} \hat{x}_A, \qquad z_{R_i} \equiv x_{R_i}^{\top} \hat{x}_A.$$

Although projections along d may induce ties, any two rankings that differ only by permutations within a tie block generate the same weighted sums, since all types in a tie block share the same projected bliss point. Tie-breaking is therefore immaterial.

We now prove the strict inequality $\sum_{i \in \tilde{I}} \bar{w}_{\tilde{R}_i} \tilde{x}_{\tilde{R}_i}^{\top} \hat{x}_A > \sum_{i \in I} \bar{w}_{R_i} x_{R_i}^{\top} \hat{x}_A$, which immediately implies $\tilde{V}_p(\hat{x}) > V_p(\hat{x})$.

Define

$$\bar{z} := \sum_{i \in I} \bar{w}_{R_i} z_{R_i}, \qquad \underline{z} := \sum_{i \in I} \underline{w}_{R_i} z_{R_i},$$

and similarly

$$\bar{\tilde{z}} := \sum_{i \in \tilde{I}} \bar{w}_{\tilde{R}_i} \tilde{z}_{\tilde{R}_i}, \qquad \underline{\tilde{z}} := \sum_{i \in \tilde{I}} \underline{w}_{\tilde{R}_i} \tilde{z}_{\tilde{R}_i}.$$

By the same symmetry argument as in lemma A.2,

$$\underline{z} = -\bar{z}, \qquad \underline{\tilde{z}} = -\bar{\tilde{z}}.$$

Since $d = 2\hat{x}_A$, $(\tilde{\mathbf{x}}, \tilde{\mathbf{s}})$ being a d-spread of $(\mathbf{x}, \mathbf{s})$ implies that the projected one-dimensional distribution induced by $(\tilde{\mathbf{x}}, \tilde{\mathbf{s}})$ is a spread of the one induced by $(\mathbf{x}, \mathbf{s})$ in the sense of definition 2. Applying the same common-grid/tail-mass argument used in the proof of proposition 4 to these projected one-dimensional distributions yields

$$\bar{\tilde{z}} - \underline{\tilde{z}} > \bar{z} - \underline{z}.$$

Using symmetry, this reduces to $\tilde{\bar{z}} > \bar{z}$, that is,

$$\sum_{i \in \tilde{I}} \bar{w}_{\tilde{R}_i} \tilde{x}_{\tilde{R}_i}^{\top} \hat{x}_A > \sum_{i \in I} \bar{w}_{R_i} x_{R_i}^{\top} \hat{x}_A.$$

Therefore $\tilde{V}_p(\hat{x}) > V_p(\hat{x})$. Combining this with the local-equilibrium identity above gives

$$\tilde{V}_p(\hat{x}) > V_p(\hat{x}) = \frac{\nu(1) + \nu(0)}{2} + \frac{2}{\phi} \|\hat{x}_A\|^2.$$

We now complete the argument by considering the best response dynamics under $(\tilde{\mathbf{x}}, \tilde{\mathbf{s}})$ initiated at $\hat{x}$. Let x_{A1} be a best response to $\hat{x}_B$ and let $x_{B1} = \hat{x}_B$. As in the appendix proof, the inductive step of lemma A.4 applies verbatim when the distribution is $(\tilde{\mathbf{x}}, \tilde{\mathbf{s}})$; therefore it suffices to verify inequality 8 at $n = 1$.

Since x_{A1} is a best response to $\hat{x}_B$ under $(\tilde{\mathbf{x}}, \tilde{\mathbf{s}})$, lemma A.3 implies that party A's payoff at $(x_{A1}, \hat{x}_B)$ can be written as

$$\tilde{V}_A(x_{A1}, \hat{x}_B) = \frac{\nu(1) + \nu(0)}{2} + \frac{\|x_{A1} - \hat{x}_B\|^2}{2\phi}.$$

Moreover, by best-response optimality,

$$\tilde{V}_A(x_{A1}, \hat{x}_B) \geq \tilde{V}_A(\hat{x}_A, \hat{x}_B) = \tilde{V}_p(\hat{x}).$$

Combining the last two displays yields

$$\|x_{A1} - x_{B1}\|^2 \geq 2\phi \left[\tilde{V}_p(\hat{x}) - \frac{\nu(1) + \nu(0)}{2} \right].$$

Since $\tilde{V}_p(\hat{x}) > V_p(\hat{x})$ and $V_p(\hat{x}) = \frac{\nu(1)+\nu(0)}{2} + \frac{\|\hat{x}_A - \hat{x}_B\|^2}{2\phi}$, it follows that

$$\|x_{A1} - x_{B1}\|^2 > 2\phi \left[V_p(\hat{x}) - \frac{\nu(1) + \nu(0)}{2} \right] = \|\hat{x}_A - \hat{x}_B\|^2,$$

that is,

$$\|x_{A1} - x_{B1}\|^2 > \|\hat{x}_A - \hat{x}_B\|^2.$$

Following exactly the same algebraic steps as in lemma A.4, this strict distance inequality implies that inequality 8 holds at $n = 1$, i.e.

$$\|x_{A1}\|^2 - \|x_{B1}\|^2 > \frac{1}{2}\|\hat{x}_A - \hat{x}_B\|^2 + 2x_{A1}^\top x_{B1}.$$

Hence the induction argument of lemma A.3 applies under $(\tilde{\mathbf{x}}, \tilde{\mathbf{s}})$ when the best-response dynamics is initiated at $\hat{x}$, and it follows that

$$\|x_{An} - x_{Bn}\|^2 > \|\hat{x}_A - \hat{x}_B\|^2 \quad \text{for all } n \geq 2.$$

By proposition 5, the best-response dynamics converges to some $x^\star \in \mathcal{E}(\tilde{\mathbf{x}}, \tilde{\mathbf{s}})$. Since the distance along the sequence is strictly larger than $\|\hat{x}_A - \hat{x}_B\|$ from step 1 onward, we have $\|x_A^\star - x_B^\star\| > \|\hat{x}_A - \hat{x}_B\|$. Using the equilibrium payoff identity at $x^\star$ and at $\hat{x}$, we obtain

$$\tilde{V}_p(x^\star) = \frac{\nu(1) + \nu(0)}{2} + \frac{\|x_A^\star - x_B^\star\|^2}{2\phi} > \frac{\nu(1) + \nu(0)}{2} + \frac{\|\hat{x}_A - \hat{x}_B\|^2}{2\phi} = V_p(\hat{x}) = \hat{V}_p(\mathbf{x}, \mathbf{s}).$$

Finally, by definition of $\mathcal{E}^\star(\tilde{\mathbf{x}}, \tilde{\mathbf{s}})$, we have $\hat{V}_p(\tilde{\mathbf{x}}, \tilde{\mathbf{s}}) \geq \tilde{V}_p(x^\star)$. Therefore,

$$\hat{V}_p(\tilde{\mathbf{x}}, \tilde{\mathbf{s}}) > \hat{V}_p(\mathbf{x}, \mathbf{s}),$$

which proves the proposition. ■

A.2 Concavity of Party Payoffs

This appendix provides microfoundations for assumption 2, which posits that parties have strictly concave utility over political power rents. The assumption does not rely on intrinsic risk aversion at the party level, but reflects standard features of how political power is translated into effective benefits within party organizations. All mechanisms discussed below are symmetric across parties and are consistent with environments in which aggregate political rents are constant-sum.

Ranked Appointments and Affiliated Placements A pervasive feature of political organizations is that greater power translates into control over a larger number of valuable placements, broadly defined (Kopecký et al., 2012; Grindle, 2012; Lewis, 2008). These include not only formal public offices—such as cabinet positions, committee chairs, regulatory appointments, or staff slots—but also access to influential roles in the private sector, such as board seats, advisory positions, or other forms of informal influence over firms operating in regulated or politically sensitive environments. Party affiliates differ in loyalty, competence, and reliability, and placements are typically allocated by ranking insiders along these dimensions.

Formally, suppose that political power $\rho \geq 0$ determines the measure of placements the party can allocate. Let affiliates be indexed by a continuous type $k \in [0, \bar{k}]$, $\bar{k} \geq \bar{\rho}$, ordered so that lower values of k correspond to higher-value affiliates. Let $q(k)$ denote the value generated by assigning affiliate k to a placement, where q is integrable and strictly decreasing. The party chooses which affiliates to assign in order to maximize total value:

$$U(\rho) = \max_{A \subseteq [0,\bar{k}], |A| = \rho} \int_A q(k)\, dk.$$

The optimal assignment fills placements starting from the highest-value affiliates, so the solution is

$$U(\rho) = \int_0^{\rho} q(k)\, dk.$$

By the envelope theorem,

$$U'(\rho) = q(\rho) \quad \text{and} \quad U''(\rho) = q'(\rho) < 0.$$

Thus, concavity of party payoffs arises as the envelope of an optimal assignment problem with heterogeneous affiliates: marginal expansions of political power must be allocated to progressively lower-value insiders. Concavity therefore does not reflect intrinsic risk aversion or diminishing returns at the level of individual placements, but follows mechanically from heterogeneity and optimal sorting within party organizations.

Rent Sharing and Diminishing Marginal Utility of Insiders A second, independent source of concavity arises from the fact that political rents are ultimately consumed by individuals associated with the party (Strøm, 1990; Müller and Strøm, 1999; Katz and Mair, 1995). Suppose that total political rents ρ are shared among $n \geq 1$ party insiders to maximize

$$U(\rho) = \max_{\{x_i\}_{i=1}^{n}} \sum_{i=1}^{n} u(x_i) \quad \text{s.t.} \quad \sum_{i=1}^{n} x_i = \rho,$$

where $u(\cdot)$ is strictly increasing and strictly concave. By symmetry, the optimal allocation satisfies $x_i = \rho/n$ for all i, so that

$$U(\rho) = n\,u\left(\frac{\rho}{n}\right).$$

Because u is concave, $U(\rho)$ is strictly concave in ρ. This channel reflects diminishing marginal utility of monetary and quasi-monetary benefits—such as salaries, discretionary budgets, fundraising capacity, or post-office opportunities—accruing to party members. Importantly, concavity arises even if the party itself is risk-neutral and even if political rents are deterministic.

Convex Organizational and Governance Costs Concavity may also arise from organizational and governance frictions internal to parties. As political power expands, parties must coordinate a broader agenda, manage larger coalitions, monitor more officeholders, and enforce discipline among an increasing number of agents (McCubbins et al., 1987; Huber and Shipan, 2002; Martin and Vanberg, 2011). These activities entail real costs that plausibly increase more than proportionally with the scale of power.

To capture this channel, suppose that gross political rents equal ρ, but that maintaining control entails a cost $c(\rho)$, with $c'(\rho) > 0, c''(\rho) > 0$. Net party payoffs are then

$$U(\rho) = \rho - c(\rho),$$

which is strictly concave in ρ. This formulation reflects the increasing complexity of governing larger organizations and coalitions, and does not rely on uncertainty or heterogeneity in preferences.

The mechanisms described above capture distinct but complementary aspects of political organizations. Ranked appointments and affiliated placements emphasize heterogeneity among insiders and selection by quality; rent sharing reflects diminishing marginal utility at the individual level; and convex governance costs capture increasing organizational complexity. Each mechanism alone is sufficient to generate strict concavity of party payoffs. Taken together, they show that assumption 2 is a natural reduced-form representation of how political power translates into effective benefits within party organizations, rather than a behavioral or ad hoc assumption.

A.3 Welfare and Majority Premia

This appendix complements section 7 by evaluating voter welfare under a *power-sharing implementation rule*. Throughout, maintain the unidimensional assumptions of section 4 (uniform shock $\epsilon \sim U(-\phi, \phi)$ and quadratic loss).

Implemented Policy Fix any platform profile (x_A, x_B) and any shock realization ϵ. Let

$$s_A(x_A, x_B, \epsilon) := \sum_{i=1}^{N} s_i \, \mathbb{1}\{\Delta_i(x_A, x_B) \geq \epsilon\}$$

denote party A's realized vote share. The policy that results from the political process is the power-weighted compromise

$$x^{\star}(x_A, x_B, \epsilon) := \frac{\rho(s_A(x_A, x_B, \epsilon))}{\bar{\rho}} \, x_A + \frac{\rho(1 - s_A(x_A, x_B, \epsilon))}{\bar{\rho}} \, x_B.$$

Since $\rho(s) + \rho(1-s) = \bar{\rho}$, the weights sum to one. Let $\rho(\cdot) \in [0, \bar{\rho}]$, so that $x^{\star}(x_A, x_B, \epsilon) \in [\min\{x_A, x_B\}, \max\{x_A, x_B\}]$. Follow the paper's convention that in equilibrium $x_A = \bar{x}$ and $x_B = \underline{x}$. Define the induced implemented-policy random variable

$$X^{\star}(\epsilon) := x^{\star}(\bar{x}, \underline{x}, \epsilon).$$

For convenience, also define the realized *power share*

$$\Lambda(\epsilon) := \frac{\rho(s_A(\bar{x}, \underline{x}, \epsilon))}{\bar{\rho}}, \qquad D := \bar{x} - \underline{x} > 0,$$

so that

$$X^\star(\epsilon) = \underline{x} + \Lambda(\epsilon) D. \tag{10}$$

Utilitarian Benchmark Let $x^o \in \arg\max_x \sum_i s_i v(x, x_i)$ be the utilitarian optimum. Under quadratic loss, $x^o = \sum_i s_i x_i$.

Lemma A.5 (Decomposition) *Let $\mathcal{W}^\star := \mathbb{E}_\epsilon\Big[\sum_i s_i v(X^\star(\epsilon), x_i)\Big]$ be ex-ante utilitarian welfare. Under quadratic loss,*

$$\mathcal{W}^\star = \mathcal{W}(x^o) - \Big(\mathbb{E}[X^\star] - x^o\Big)^2 - \mathrm{Var}(X^\star), \tag{11}$$

where $\mathcal{W}(x^o) := \sum_i s_i v(x^o, x_i)$ is the first-best welfare level. Moreover, using 10,

$$\mathbb{E}[X^\star] = \underline{x} + D\,\mathbb{E}[\Lambda], \qquad \mathrm{Var}(X^\star) = D^2\,\mathrm{Var}(\Lambda). \tag{12}$$

Proof. Write $\sum_i s_i v(x, x_i) = C - \sum_i s_i (x - x_i)^2$ for a constant $C \in \mathbb{R}$. Under quadratic loss, $\sum_i s_i (x - x_i)^2 = (x - x^o)^2 + \sum_i s_i (x_i - x^o)^2$, hence $\sum_i s_i v(x, x_i) = \mathcal{W}(x^o) - (x - x^o)^2$. Taking expectations at $x = X^\star$ yields $\mathcal{W}^\star = \mathcal{W}(x^o) - \mathbb{E}\Big[(X^\star - x^o)^2\Big]$, and the standard identity $\mathbb{E}[(X - x^o)^2] = (\mathbb{E}X - x^o)^2 + \mathrm{Var}(X)$ gives 11. Finally, 12 follows from $X^\star = \underline{x} + \Lambda D$. ■

Varying the Majority Premium in a Proportional System To interpret comparative statics in ρ_m, it is useful to consider a one-parameter family of *proportional-with-premium* power mappings (holding $\bar{\rho}$ fixed):

$$\rho_{\rho_m}(s) := \begin{cases} (\bar{\rho} - \rho_m)s, & s < \frac{1}{2}, \\ \frac{\bar{\rho}}{2}, & s = \frac{1}{2}, \\ (\bar{\rho} - \rho_m)s + \rho_m, & s > \frac{1}{2}, \end{cases} \tag{13}$$

for $\rho_m \in [0,\bar{\rho})$.This family satisfies assumption 1 and has jump ρ_m at 1/2. Let $(\underline{x}(\rho_m), \bar{x}(\rho_m))$ be the equilibrium platforms associated with ρ^{ρ_m} (see proposition 2), and let $X^\star_{\rho_m}$.

Lemma A.6 (Maximal-premium limit of equilibrium platforms) *Assume $s_m > 0$. Under 13,*

$$\underline{x}(\rho_m) \to x_m, \qquad \bar{x}(\rho_m) \to x_m, \qquad D(\rho_m) := \bar{x}(\rho_m) - \underline{x}(\rho_m) \to 0, \quad \text{as } \rho_m \to \bar{\rho}^-.$$

Proof. As $\rho_m \to \bar{\rho}^-$, ρ^{ρ_m} converges pointwise to a step function that assigns (essentially) all power to any strict majority. Since $\nu^{\rho_m} = U \circ \rho^{\rho_m}$ inherits this limit, the increment formulas for the equilibrium weights in proposition 2 (ii) imply weight concentration on the (type) median group: all weights except the median's converge to 0, while the median weight converges to 1. Therefore both equilibrium platforms converge to x_m and the distance converges to 0. ■

Corollary 1 (Full Majority Limit) *Under the assumptions of Lemma A.6,*

$$X^\star_{\rho_m} \xrightarrow[\rho_m \to \bar{\rho}^-]{} x_m \quad \text{a.s.}, \qquad \operatorname{Var}(X^\star_{\rho_m}) \to 0, \qquad \mathcal{W}^\star(\rho_m) \to \mathcal{W}(x^o) - (x_m - x^o)^2.$$

Hence, if $x_m = x^o$ (e.g. under the conditions in remark 2), then $\lim_{\rho_m \to \bar{\rho}^-} \mathcal{W}^\star(\rho_m) = \mathcal{W}(x^o)$ (first-best welfare).

Proof. $X^\star_{\rho_m}(\epsilon) \in [\underline{x}(\rho_m), \bar{x}(\rho_m)]$ for all ϵ, so $|X^\star_{\rho_m}(\epsilon) - \underline{x}(\rho_m)| \le D(\rho_m)$. Lemma A.6 implies $\underline{x}(\rho_m) \to x_m$ and $D(\rho_m) \to 0$, hence $X^\star_{\rho_m} \to x_m$ a.s. and $\operatorname{Var}(X^\star_{\rho_m}) \to 0$. The welfare limit follows from lemma A.5 and $\mathbb{E}[X^\star_{\rho_m}] \to x_m$. ■

Discussion Lemma A.5 shows that voter welfare depends on (i) the *bias* of the mean implemented policy relative to x^o and (ii) the *variance* of the implemented policy. Increasing ρ_m affects both terms through (i) endogenous platform responses $(\underline{x}(\rho_m), \bar{x}(\rho_m))$ and (ii) the induced distribution of the power share $\Lambda(\epsilon)$ via ρ^{ρ_m}. If $x_m \neq x^o$, then maximizing welfare via the design of ρ_m can entail a trade-off between variance reduction and bias reduction.

A.4 Algorithmic Equilibrium Search

The task of finding the local equilibria is simplified by lemma 1. The following procedure constructs $\mathcal{L}$ in a finite number of steps. First, let $\mathcal{L}_0 = \emptyset$, construct $\mathcal{R}(I)$ as the set of permutations of I, and let $R^1, R^2, \ldots, R^{|\mathcal{R}(I)|}$ denote its elements. Second, for each $r = 1, \ldots, |\mathcal{R}(I)|$, perform the following steps: (i) compute $\bar{x}(R^r)$ and $\underline{x}(R^r)$; (ii) compute $\nabla(\bar{x}(R^r), \underline{x}(R^r))$; (iii) if $R^r \in \nabla(\bar{x}(R^r), \underline{x}(R^r))$, set

$$\mathcal{L}_r = \mathcal{L}_{r-1} \cup \{(\bar{x}(R^r), \underline{x}(R^r))\},$$

and otherwise set $\mathcal{L}_r = \mathcal{L}_{r-1}$. By lemma 1, $\mathcal{L}_{|\mathcal{R}(I)|} = \mathcal{L}$. Furthermore, if $(\mathbf{x}, \mathbf{s})$ is symmetric, proposition 5 implies that the party-preferred Nash equilibria can be found by looking for the elements of $\mathcal{L}$ that maximize V_p.

A.5 Robustness to Voter-Level Idiosyncratic Shocks

In the environment of section 4, suppose that, in addition to the aggregate shock $\epsilon \sim U(-\phi, \phi)$, each voter draws an iid idiosyncratic shock $\eta \sim U[-\kappa, \kappa]$, independent of ϵ. A voter of type i votes for party A if and only if

$$\Delta_i(x_A, x_B) \geq \epsilon + \eta.$$

Let

$$G_\kappa(t) = \begin{cases} 0 & \text{if } t \leq -\kappa, \\ \frac{t+\kappa}{2\kappa} & \text{if } -\kappa < t < \kappa, \\ 1 & \text{if } t \geq \kappa. \end{cases}$$

Conditional on ϵ, the vote share of party A is

$$S_A^\kappa(\epsilon; x_A, x_B) = \sum_{i=1}^{N} s_i G_\kappa(\Delta_i(x_A, x_B) - \epsilon),$$

so that

$$V_A^\kappa(x_A, x_B) = \frac{1}{2\phi}\int_{-\phi}^{\phi} \nu(S_A^\kappa(\epsilon; x_A, x_B))\, d\epsilon.$$

Define V_B^κ analogously.

Proposition A.1 (Small voter-level noise) *Fix a voter distribution* $(\mathbf{x}, \mathbf{s})$ *and let* $(\underline{x}, \bar{x})$ *be the equilibrium characterized in proposition 2, with* $\underline{x} < \bar{x}$. *There exists* $\bar{\kappa} > 0$ *such that, for every* $\kappa \in [0, \bar{\kappa})$, *the game with idiosyncratic shocks admits* $(\bar{x}, \underline{x})$ *and* $(\underline{x}, \bar{x})$ *as strict pure-strategy equilibria.*

Moreover, there exists a neighborhood $\mathcal{N}$ *of* $(\bar{x}, \underline{x})$ *such that, for every* $(x_A, x_B) \in \mathcal{N}$ *with* $x_A > x_B$,

$$V_A^\kappa(x_A, x_B) = C_\kappa(\mathbf{s}) + \nu(0) + \sum_{i=1}^{N} F(\Delta_i(x_A, x_B))\Big(\nu(\bar{s}_i) - \nu(\bar{s}_{i+1})\Big),$$

for a constant $C_\kappa(\mathbf{s})$ *independent of* (x_A, x_B). *Hence*

$$\frac{\partial V_A^\kappa(x_A, x_B)}{\partial x_A} = \frac{1}{\phi}\sum_{i=1}^{N}(x_i - x_A)\Big(\nu(\bar{s}_i) - \nu(\bar{s}_{i+1})\Big),$$

and symmetrically, for (x_A, x_B) *in a neighborhood of* $(\underline{x}, \bar{x})$ *with* $x_A < x_B$,

$$\frac{\partial V_A^\kappa(x_A, x_B)}{\partial x_A} = \frac{1}{\phi}\sum_{i=1}^{N}(x_i - x_A)\Big(\nu(\underline{s}_i) - \nu(\underline{s}_{i-1})\Big).$$

Consequently, the local first-order conditions coincide with those in proposition 2.

Proof. Let $x^\star = (\bar{x}, \underline{x})$, and note that

$$\Delta_{i+1}(x^\star) - \Delta_i(x^\star) = 2(\bar{x} - \underline{x})(x_{i+1} - x_i) > 0$$

for each $i = 1, \ldots, N-1$, while $-\phi < \Delta_1(x^\star) < \Delta_N(x^\star) < \phi$ by the support assumption used in section 4. Hence,

$$m := \min\Big\{\phi - \Delta_N(x^\star),\ \Delta_1(x^\star) + \phi,\ \min_{i<N}\Big(\Delta_{i+1}(x^\star) - \Delta_i(x^\star)\Big)\Big\} > 0.$$

Fix $\bar{\kappa}_1 < m/4$. By continuity of $(\Delta_i)_{i\in I}$, for each $\kappa < \bar{\kappa}_1$ there exists a neighborhood $\mathcal{N}$ of $x^\star$ such that, for every $(x_A, x_B) \in \mathcal{N}$ with $x_A > x_B$,

$$-\phi < \Delta_1 - \kappa < \Delta_1 + \kappa < \cdots < \Delta_N - \kappa < \Delta_N + \kappa < \phi. \tag{14}$$

On this neighborhood, as ϵ increases, voter groups peel off one at a time. Therefore,

$$S_A^\kappa(\epsilon; x_A, x_B) = \begin{cases} 1, & \epsilon \in [-\phi, \Delta_1 - \kappa], \\ \bar{s}_{i+1} + s_i \frac{\Delta_i + \kappa - \epsilon}{2\kappa}, & \epsilon \in [\Delta_i - \kappa, \Delta_i + \kappa],\ i = 1, \ldots, N, \\ \bar{s}_{i+1}, & \epsilon \in [\Delta_i + \kappa, \Delta_{i+1} - \kappa],\ i = 1, \ldots, N-1, \\ 0, & \epsilon \in [\Delta_N + \kappa, \phi]. \end{cases}$$

Integrating and performing the change of variables $u = (\Delta_i + \kappa - \epsilon)/(2\kappa)$ on each ramp yields

$$V_A^\kappa(x_A, x_B) = C_\kappa(\mathbf{s}) + \nu(0) + \sum_{i=1}^N F(\Delta_i(x_A, x_B))\Big(\nu(\bar{s}_i) - \nu(\bar{s}_{i+1})\Big),$$

where $C_\kappa(\mathbf{s})$ depends only on $(\mathbf{s}, \kappa, \nu, \phi)$. Differentiating and using

$$\frac{\partial \Delta_i(x_A, x_B)}{\partial x_A} = 2(x_i - x_A)$$

gives the displayed derivative formula. Since these weights sum to $\nu(1) - \nu(0) = 1$, the corresponding second derivative is $-1/\phi < 0$, so the local best response is unique. The case $x_A < x_B$ is symmetric and yields the formula with $(\underline{s}_i)_{i \le N}$.

Since proposition 2 implies that $x^\star = (\bar{x}, \underline{x})$ is a strict equilibrium of the baseline game on the compact action set $[x_1, x_N]^2$, there exist neighborhoods $U_A \ni \bar{x}$, $U_B \ni \underline{x}$ and $\delta > 0$ such that, after shrinking U_A and U_B if necessary, $U_A \times U_B \subseteq \mathcal{N}$ and

$$V_A^0(\bar{x}, \underline{x}) \ge V_A^0(x_A, \underline{x}) + 3\delta \quad \text{for all } x_A \notin U_A,$$

$$V_B^0(\bar{x}, \underline{x}) \ge V_B^0(\bar{x}, x_B) + 3\delta \quad \text{for all } x_B \notin U_B.$$

Finally, for every $(x_A, x_B) \in [x_1, x_N]^2$,

$$|V_p^\kappa(x_A, x_B) - V_p^0(x_A, x_B)| \leq \frac{N\kappa}{\phi},$$

because, for each type i, $G_\kappa(\Delta_i - \epsilon)$ differs from $\mathbb{1}\{\Delta_i \geq \epsilon\}$ only when $|\Delta_i - \epsilon| \leq \kappa$, a set of ϵ-measure at most 2κ, and $\nu(1) - \nu(0) = 1$. Choosing $\bar{\kappa}_2 = \delta\phi/N$, the above strict inequalities continue to hold under V_p^κ for every $\kappa < \bar{\kappa}_2$. Inside $U_A \times U_B$, the first-order conditions are unchanged, so $\bar{x}$ and $\underline{x}$ remain the unique local best responses to one another. Hence $(\bar{x}, \underline{x})$ is a strict pure-strategy equilibrium of the perturbed game for every $\kappa < \bar{\kappa} := \min\{\bar{\kappa}_1, \bar{\kappa}_2\}$. The symmetric profile $(\underline{x}, \bar{x})$ is handled analogously. ∎

Remark 14 *For general voter distributions, the non-overlap condition used in the proof of proposition A.1 is a sufficient device ensuring that the perturbed first-order conditions coincide exactly with those of proposition 2; proposition 3 survives exactly; and proposition 4 survives exactly whenever the polarization comparison is generated by changes in bliss points alone. The non-overlap condition is not generally necessary. For instance, in the case of two equally sized voter types, overlap remains tractable: one obtains an implicit one-dimensional characterization of the symmetric divergent equilibrium, and the polarization comparative static of proposition 4 continues to hold along that branch.*