Representation in District-Based Elections*

Yunus C. Aybas† Oguzhan Celebi‡ Surabhi Dutt†

July 26, 2026

Abstract

District-based elections must reconcile two objectives: representing each district's electorate and matching the statewide delegation to overall political support. We study the design of seat allocation rules when these objectives conflict. The optimal rule is characterized by an endogenous cutoff that compares the district-level and statewide representational effects of awarding a party an additional seat. In a linear benchmark, the optimal rules form a family ranging from first-past-the-post to proportional representation as the weight on statewide representation increases. Along this path, it becomes increasingly difficult for an already overrepresented party to gain further seats through gerrymandering. The characterization yields a measure of first-past-the-post robustness to statewide representation concerns and provides a unified framework for comparing it with established electoral diagnostics. Applying these measures to U.S. House elections from 2002 to 2024, we find that maps drawn by single-party legislatures generally perform worse across our measures than those produced by other institutions, and Republican-legislature maps fare worse on average than Democratic-legislature maps. The efficiency gap, a widely used wasted-vote measure of partisan gerrymandering, is more strongly associated with struck or litigated maps than the other measures, yet it considerably amplifies the Republican–Democratic asymmetry.

*We are grateful to Peter Buisseret, Steve Callander, Andres Carvajal, Hulya Eraslan, SoRelle Wyckoff Gaynor, Matthew Jackson, Murat Koksalan, Justin Levitt, Scott Page, Maggie Penn, Doron Ravid, Nicholas Stephanopoulos and seminar participants at Texas A&M, UC Davis and Rochester for helpful discussions and comments.

†Texas A&M University, Department of Economics.

‡University of Michigan, Ross School of Business.

1 Introduction

In a representative democracy, elections translate votes into representation. In district-based legislatures, representation is determined seat by seat: each district elects one representative, and these representatives collectively determine the partisan composition of the delegation. This structure, central to U.S. House and state legislative elections, creates two representational objects that can come into conflict: who represents each district and how the statewide delegation is divided across parties.

The two familiar benchmark election rules illustrate the tension. First-past-the-post assigns every seat to the local district winner, but it can produce a delegation far from proportional to statewide support. Proportional representation matches the delegation's seat totals to statewide vote shares, but within a fixed district map it may require assigning some districts to a party that lost locally.[1] No seat allocation rule, that is, no mapping from district vote profiles to district winners, can in general preserve every district winner and implement a proportional statewide delegation.

This paper studies how district seats should be allocated when both district-level and statewide representation matter. We study this problem within the institutional constraints of single-member district elections, holding fixed the district boundaries, one representative per district, and the resulting distribution of district vote shares. The question is then how to allocate those district seats according to two objectives: how well each district's seat represents its local electorate and how closely the partisan composition of the statewide delegation reflects statewide political support.

The district component assigns a local cost to each district-seat assignment, based on the vote profile in that district. The statewide component assigns a delegation-level cost to the party seat total, relative to a statewide target based on the statewide vote shares. A *relative weight* sets the balance between the two components, and an aggregation function combines them into a single objective, which we call total misrepresentation. The framework is flexible: the district component can encode different local representation criteria, and the statewide target can encode different views about how statewide support should translate into seats.

The main institutional departure (from first-past-the-post) is that, once statewide representation receives positive weight, district outcomes in the optimal seat allocation are

[1] Throughout, proportional representation refers to a delegation-level target within a fixed district-based system: party seat totals are chosen as they would be under proportional representation, and district seats are then assigned to implement those totals. Related institutional alternatives include party-list PR (Israel, the Netherlands), mixed-member proportional systems (Germany, New Zealand), parallel mixed systems (Japan, Mexico), and compensatory or leveling-seat systems (Sweden, Norway, Denmark). Our analysis takes a different approach: each district continues to receive one representative, and we study rules that map the fixed district vote profile into the set of district winners.

no longer independent. The assignment of a marginal district can depend on the party's statewide support, its current seat total, and outcomes in other districts. These cross-district spillovers are the mechanism through which a district-based rule can correct statewide under- or over-representation.

Our formulation admits a simple organizing principle. Conditional on a party receiving S seats, the misrepresentation-minimizing assignment gives that party its S strongest districts. We call this the top-S allocation. The allocation problem therefore reduces to choosing a seat total along the path between two benchmarks: the district-only outcome, which minimizes district-level cost alone, and the statewide-only outcome, which minimizes delegation-level cost alone. When these outcomes differ, the party whose district-only seat total lies below its statewide-only seat total is the party underrepresented in the district-only allocation; we label this party A.

Our first main result characterizes the adjacent step along this path. Ordering districts from strongest to weakest support for party A, moving from S to $S+1$ gives party A one additional seat and changes only the marginal district $S+1$. As long as S is below party A's statewide-only seat total, this move raises district-level cost and lowers delegation-level cost. Theorem 1 shows that the move is improving when the weighted delegation-level gain is large enough to justify the local cost of assigning the marginal district to party A, and expresses this comparison as an endogenous cutoff for the marginal district.

To move from adjacent comparisons to global comparisons, we impose diminishing marginal rate of substitution in the aggregation function between district-level and delegation-level misrepresentation. Intuitively, additional district-level cost becomes harder to justify as district-level misrepresentation grows, while reductions in delegation-level cost become less valuable as the allocation approaches the statewide-only outcome. Theorem 2 shows that, under this condition, the adjacent comparisons generated by the same cutoff formula are ordered, so they identify the optimal seat total along the top-S path.

The cutoff condition has a single crossing structure in seat total and the relative weight. Starting from the district-only outcome, seats are awarded to party A until the next marginal district no longer clears the cutoff. This ordering also gives the main comparative static. Proposition 1 shows that increasing the relative weight on statewide representation moves the underrepresented party's optimal seat total monotonically from the district-only outcome toward the statewide-only outcome. As this weight rises, the cutoff admits progressively weaker marginal districts until the statewide-only benchmark is reached.

We then specialize the framework to a linear model that puts the two components on a common population scale, using the population of one district as the unit. District-level misrepresentation is the total mass of voters whose own district is represented by their nonpreferred party. Statewide misrepresentation is the absolute gap between the population represented by a party's seats and the population supporting that party statewide. The objective is a weighted sum of the two. We show that, even though the feasible set of allocations is discrete and nonconvex, varying the weight traces the Pareto frontier of district and statewide misrepresentation. We also give the linear model a voter-level foundation: when voters value being represented by their preferred party in their own district and value their preferred party's representation in the statewide delegation, aggregating voter payoffs yields exactly the linear model.

In the linear model, the cutoff characterization becomes an explicit family of optimal electoral rules indexed by the relative weight on statewide representation. Theorem 3 characterizes, for each weight, the rule that maps every vote profile into its optimal, misrepresentation-minimizing seat allocation. At one endpoint, when only district-level representation matters, first-past-the-post is optimal: the cutoff is $1/2$, and each district is assigned to its local majority winner. At the other endpoint, when only statewide representation matters, the optimal rule selects the integer seat total closest to proportionality and assigns those seats to the parties' strongest districts. For the party that is underrepresented under first-past-the-post relative to proportional representation, increasing the relative weight on statewide representation lowers the cutoff from the first-past-the-post threshold of $1/2$, making additional districts easier to win until the proportional seat total is reached. Seat totals change only when this moving cutoff crosses an observed district vote share, so the theorem gives a smooth interpolation in the tradeoff parameter and a stepwise path in seats, with first-past-the-post and proportional representation as the two canonical endpoints of the same optimal family.

We next study properties of the linear optimal family. The first concerns the geography of support. Holding statewide support fixed, greater concentration lowers district-level cost for any top-S allocation because the party's strongest districts contain more of its voters. We ask when this ordering is preserved by an endogenous seat-allocation rule, that is, if one vote profile is more concentrated than another with the same statewide support, when does it generate lower misrepresentation? Using majorization order to quantify concentration, we show that, for a given relative weight governing the tradeoff between district-level and statewide representation, this monotonicity holds within the optimal family if and only if the rule is proportional representation or the rule is calibrated to that same relative weight to match the designer's preferences.

Second, we characterize the endpoints of the linear optimal family. A rule is strongly district-monotone if increasing a party's support district by district cannot cause it to lose a seat it previously held. A rule is gerrymandering-proof if, holding the statewide vote shares fixed, moving votes across districts cannot change the party seat total. Within the linear optimal family, first-past-the-post is the unique strongly district-monotone rule, and proportional representation is the unique gerrymandering-proof rule. Thus the two canonical endpoints can also be pinned down by distinct robustness requirements, in addition to their preference-based characterization through the relative weight. Interior rules deliberately accept cross-district spillovers to balance district-level responsiveness against statewide proportionality.

This characterization leaves open how vulnerable the nonproportional rules are to gerrymandering. We therefore study vulnerability continuously. This question is directly relevant to modern redistricting disputes. Standard reforms constrain how maps may distribute voters across districts. Our analysis asks whether the allocation rule itself can reduce the electoral return to the same map manipulation, holding the mapmaking technology fixed.

Holding statewide vote shares fixed, we consider reshufflings of votes across districts that would give an already overrepresented party additional seats. We show that, within the optimal family, increasing the weight on statewide representation shrinks the set of reshufflings that can achieve this. Thus, even away from the proportional endpoint, greater concern for statewide representation makes gerrymandering harder. This identifies a rule-based lever for limiting the electoral returns to gerrymandering, complementary to standard redistricting constraints and map diagnostics: moving away from first-past-the-post by increasing the weight on statewide representation can reduce the scope for gerrymandering even when district boundaries are held fixed.

The framework maps common electoral diagnostics into choices of statewide targets, normalizations, and counterfactual vote profiles. Under the proportional benchmark, this yields two measures: the first-switch weight, which captures the robustness of the observed first-past-the-post allocation to statewide representation concerns, and the signed proportional seat gap, which records the direction and magnitude of its departure from proportionality. Comparing them with the efficiency gap, whose statewide target awards the statewide winner a seat bonus relative to proportional representation, shows that the measures can diverge sharply: an exactly proportional first-past-the-post outcome can have a large efficiency gap, while an outcome with zero efficiency gap can become suboptimal after an arbitrarily small weight is placed on statewide representation. The same framework also encompasses standard disproportionality indices and counterfactual symmetry measures.

We then bring these measures to U.S. House elections from 2002 to 2024, combining realized district-level returns with information on the institutions that produced each congressional map. We find that the maps that were struck down or litigated are much more concentrated in the upper tail of the efficiency-gap distribution than in the corresponding tails of the proportional seat gap or first-switch weight. This concentration makes it especially important to understand what the efficiency gap captures and how its winner's-bonus benchmark differs from alternative measures. Comparing map-drawing authorities yields three further patterns. First, maps enacted by single-party legislatures generally produce less robust first-past-the-post allocations and feature larger departures from proportionality than maps drawn under split control, by nonpartisan authorities, or by courts. Second, among partisan legislatures, Republican-legislature maps exhibit larger average party-favorable deviations than Democratic-legislature maps. Third, the efficiency gap substantially amplifies this Republican–Democratic asymmetry: the difference is moderate under the proportional seat gap and first-switch weight, but much larger under the efficiency gap. Together, these findings show that the choice of representational benchmark materially shapes the empirical evaluation of electoral maps.

Related literature. The paper contributes to the political-economy literature on electoral institutions. A large body of work studies how majoritarian and proportional systems affect party competition, political selection, public spending, accountability, and polarization (Cox, 1997; Milesi-Ferretti, Perotti, and Rostagno, 2002; Persson and Tabellini, 2005). Our contribution is complementary. Rather than comparing majoritarian and proportional systems as institutional categories, we study the rule design problem inside district-based elections and characterize the family of allocations that optimally trade off local and statewide representation. Carey and Hix (2011) ask when electoral systems can combine proportional representation with accountability and local linkage, identifying a favorable region of an empirically estimated representation–accountability frontier across electoral institutions; we instead derive the complete local–statewide representation frontier within a fixed district-based institution.

Related work on electoral allocation studies the joint distribution of seats across parties and territorial districts. Fair-majority and biproportional methods first determine aggregate party seat totals and then assign those party seats across districts subject to party and territorial constraints (Balinski, 2008; Gaffke and Pukelsheim, 2008). Antweiler (2019) is especially close to our conditional assignment problem: taking proportional party totals as given, he assigns party seats across constituencies to maximize local representation. Our framework endogenizes the party seat total, choosing it jointly with the district assignments as part of the tradeoff between district-level and statewide loss. Jeong (2024) studies party-list compensation as a way to reconcile local accountability with proportionality, a design margin orthogonal to our focus on optimally trading off district and statewide representation within a fixed set of district seats.

The paper also contributes to the measurement of electoral representation. Disproportionality indices summarize departures of seat shares from vote shares (Loosemore and Hanby, 1971; Gallagher, 1991), while partisan-bias and seats–votes measures evaluate partisan symmetry, often at counterfactual vote profiles (King and Browning, 1987; Gelman and King, 1994). We represent these measures through their statewide target, loss function, and, where relevant, counterfactual vote profile. This makes transparent why different measures can evaluate the same election differently.

The efficiency gap is a central example. Building on the relative-wasted-votes approach of McGhee (2014), Stephanopoulos and McGhee (2015) propose it as a measure of partisan gerrymandering. The measure compares the parties' statewide totals of losing votes and surplus winning votes. Under equal turnout, a zero efficiency gap requires a party's seat share to equal twice its statewide vote share minus one half, rather than its proportional share. Chambers, Miller, and Sobel (2017) emphasize this winner's-bonus target and show that recalibrating the implicit value of a seat instead yields the proportional vote–seat gap. We place this distinction inside our framework by treating the two measures as evaluations of the same first-past-the-post seat total against different statewide targets. Relatedly, Warrington (2018, 2019) develop and compare measures that use different features of the district vote distribution, illustrating why alternative diagnostics can evaluate the same election differently.

Our contribution is to place this benchmark inside a general allocation framework. Proportionality and the efficiency gap correspond to different statewide targets, and our model studies how each target interacts with district-level representation in determining both the party seat total and the districts receiving those seats. This target-based perspective also guides our empirical comparison of electoral outcomes across mapmaking institutions. Barton and Eguia (2024) evaluate the 2021–22 congressional maps by projecting earlier statewide elections onto enacted and simulated district plans and decompose partisan advantage under alternative fairness ideals into components attributable to the ideal, political geography, legal constraints, and mapmaker discretion. They find that partisan-controlled maps generally exhibit larger discretionary advantages than commission- or court-drawn maps, a pattern echoed by our finding that partisan-legislature maps generally fare worse across our fairness measures in realized congressional elections from 2002 to 2024. Complementarily, Gaynor and Hayes (2026) find that congressional plans drawn by partisan institutions, particularly Republican-controlled legislatures, shift vulnerable constituencies more than plans drawn by commissions or under split control.

The paper also relates to work on political geography and partisan redistricting. Existing models study how geography and strategic district boundaries generate vote–seat distortions under a fixed district-winner rule (Coate and Knight, 2007; Friedman and Holden, 2008; Chen and Rodden, 2013; Kolotilin and Wolitzky, 2026; Bouton, Genicot, Castanheira, and Stashko, 2023). We instead hold the map and induced vote profile fixed and make the allocation rule the object of design. Redistricting changes the distribution of votes across districts under a fixed rule; our framework changes the rule applied to a fixed distribution.

Finally, the paper connects to axiomatic electoral design and revealed preference. Social-choice and apportionment theories characterize rules through properties such as responsiveness, neutrality, and consistency (May, 1952; Balinski and Young, 2001). We instead begin with an explicit representational objective and then characterize the endpoint rules within the resulting optimal family.

2 The Model

We study a two-party election between parties A and B that determines the composition of a statewide delegation consisting of $N \geq 3$ single-member district seats.

Districts have equal population, and we take the distribution of electoral support across districts as given. We denote party A's vote share in district $d \in \{1, \dots, N\}$ with $p_d \in [0, 1]$, and write $\mathbf{p} := (p_1, \dots, p_N)$ for the voting profile. Since there are only two parties, party B's vote share in district d is $1 - p_d$. We write $\rho := \sum_{d=1}^{N} p_d$ for party A's statewide support measured in seat units. The corresponding statewide vote share is $\bar{p} := \rho/N$.

A *seat-allocation rule* assigns the N district seats as a function of the voting profile $\mathbf{p}$. Formally, for each profile $\mathbf{p}$, the rule R selects a subset $R(\mathbf{p}) \subseteq \{1, \dots, N\}$, where $d \in R(\mathbf{p})$ means that party A wins district d and party B wins the complement. Equivalently, an allocation can be represented by an indicator vector $\mathbf{x} := (x_1, \dots, x_N)$, where $x_d = 1$ if party A wins district d and $x_d = 0$ otherwise. For example, under first-past-the-post, denoted R^{FPTP}, party A wins exactly those districts in which it receives a majority: $R^{\text{FPTP}}(\mathbf{p}) = \{d : p_d > 1/2\}$. For a generic allocation $\mathbf{x}$, write $S(\mathbf{x}) := \sum_{d=1}^{N} x_d$ for the number of seats awarded to party A.

Throughout the text, we relabel districts so that $p_1 \geq p_2 \geq \cdots \geq p_N$.[2] We will frequently refer to the monotonic allocation that assigns party A a fixed number of seats in the districts where it is strongest. We call this the top-S allocation. Formally, for any seat total S, let $\mathbf{x}^S$ denote the top-S allocation defined by $x_d^S = 1$ if $d \leq S$, and $x_d^S = 0$ if $d > S$.

Our goal is to evaluate an electoral outcome by how accurately it translates votes into political representation. In district-based elections, two objectives are natural, yet often in tension. First, *within each district*, the winning party should represent as many local voters as possible. Second, *at the statewide level*, the delegation's party balance should align with party support in the electorate as a whole.

[2] We assume distinct district vote shares and no district-level electoral tie, that is, $p_i \neq p_j$ for $i \neq j$ and $p_i \neq 1/2$ for all i, only for notational convenience.

District-level representation. The district-level component captures the idea that each district seat represents its own local electorate, and we define local misrepresentation as the mismatch between that assignment and the voters in that district.

We impose that any admissible measure of district-level representation must be *anonymous*: two districts with the same party A vote share and the same assignment must generate the same local loss. For any district with party A vote share p, we denote the loss from assigning it to party A by $\delta(1, p)$, and the loss from assigning it to party B by $\delta(0, p)$. We write

$$\Delta(p) := \delta(1, p) - \delta(0, p)$$

for the district switch margin, the change in local loss from assigning the district to party A rather than party B.

We also assume additive separability across districts. The goal of this component is to capture the local mismatch between a district's vote share and the party assigned to represent it. That mismatch should not depend on assignments made elsewhere. Otherwise, the loss assigned to one district could vary because of outcomes in other districts, thereby importing statewide considerations into a component intended to be local.

The two assumptions above are especially natural when district-level misrepresentation is interpreted as a function of the voters whose own district seat is held by their nonpreferred party: districts are exchangeable, and voter-level losses accumulate across districts.[3,4] Under these two assumptions, total district-level misrepresentation is

$$\mathrm{Dist}(\mathbf{x}, \mathbf{p}) := \sum_{d=1}^{N} \delta(x_d, p_d),$$

and we define the district-optimal seat total by

$$S^{\mathrm{Dist}} := \min \left\{ S(\mathbf{x}) : \mathbf{x} \in \operatorname*{arg\,min}_{\mathbf{y} \in \{0,1\}^N} \mathrm{Dist}(\mathbf{y}, \mathbf{p}) \right\}.$$

Thus S^{Dist} is the smallest party A seat total attained by any allocation that minimizes

[3]These assumptions may be less appropriate when the representational value of a district seat depends on district identity or on assignments elsewhere, as with cross-district spillovers, delegation-level bargaining, committee assignments, minority-opportunity districts, or other complementarities across seats. Such effects lie outside the district-level component considered here. The statewide composition of the delegation is captured separately by the statewide loss.

[4]Although it is natural, we do not require district losses to be party neutral as our results do not rely on it. When *party neutrality* is imposed, the loss from assigning a district to a party depends only on that party's vote share in the district, not on the party's name. In that case, the loss function takes the one-dimensional form $\delta(p) := \delta(1, p)$, and $\delta(1 - p) := \delta(0, p)$.

district-level misrepresentation.

Finally, we impose the ordinal restriction of *district responsiveness*: as the party A vote share in a district rises, assigning the district to party A should not become less representative relative to assigning it to party B.[5] Equivalently, additional support for party A may strengthen the case for assigning the district to party A, or leave that comparison unchanged, but it cannot strengthen the case for assigning the district to party B. Formally, the district switch margin $\Delta(p)$ is weakly decreasing in p. This is the standard *single crossing* condition.

Statewide representation. The statewide component formalizes the second dimension of representation. Even if each district seat is assigned in a way that best fits its own local electorate, the resulting delegation may still give a party too many or too few seats relative to its statewide support. The statewide delegation has a party composition, and we define statewide misrepresentation as the mismatch between that composition and party support in the electorate as a whole.

To capture this, we assume that there is a *target* seat count $T(\rho) \in [0, N]$ for party A that depends on its statewide vote total ρ, and that statewide misrepresentation depends on party A's seat total $S(\mathbf{x})$ and its distance from this target $T(\rho)$.[6] A natural target is statewide proportionality, which is the special case $T(\rho) = \rho$.

Formally, let $Q : [0, N] \to \mathbb{R}_+$ be strictly increasing and weakly convex, with $Q(0) = 0$, and define

$$\mathrm{State}(\mathbf{x}, \mathbf{p}) := Q\big(|S(\mathbf{x}) - T(\rho)|\big).$$

Because statewide misrepresentation depends on an allocation only through its seat total, we also write $\mathrm{State}(S, \rho) := Q(|S - T(\rho)|)$ for the statewide loss associated with seat total S and statewide vote total ρ. This form defines statewide misrepresentation as a failure of the delegation to align with a function of the statewide popular vote, measured by $|S(\mathbf{x}) - T(\rho)|$. Since Q is strictly increasing, larger gaps are more costly. And since Q is weakly convex, the marginal cost of distance from the target is weakly increasing.[7] We define

$$S^{\mathrm{State}} := \min \operatorname*{arg\,min}_{S \in \{0, \ldots, N\}} \mathrm{State}(S, \rho)$$

[5]We call this district responsiveness, because it is the loss-function analog of the positive responsiveness condition in the celebrated Theorem of May (1952). In a binary choice, the axiom states that increasing support for an alternative should not make that alternative less favored.

[6]Because there are only two parties and N seats, party A's seat total pins down the full partisan composition of the delegation: if party A receives S seats, party B receives $N - S$. We therefore express the statewide component in terms of party A's seat total and statewide support.

[7]Again, although it's natural to assume party neutrality in the main text, the results hold without this assumption. Imposing it implies that the target satisfies $T(N - \rho) = N - T(\rho)$.

as the smallest statewide-optimal integer seat total. Since Q is strictly increasing, S^{State} is the smallest integer seat total closest to the target $T(\rho)$.

Total misrepresentation. The two components measure different failures of representation. Improving one component can require worsening the other. We evaluate allocations by aggregating the two losses via $F(\cdot,\cdot)$. For $w \in [0,1]$, define

$$M(\mathbf{x}, \mathbf{p}; w) := F\Big((1-w)\mathrm{Dist}(\mathbf{x}, \mathbf{p}),\ w\mathrm{State}(\mathbf{x}, \mathbf{p})\Big),$$

where $F : \mathbb{R}_+^2 \to \mathbb{R}$ is differentiable and increasing in both arguments. Since both arguments are misrepresentation terms, lower values of M are preferred.

The weight w determines the relative scale of the two losses, while F determines how the planner substitutes between them once they are placed on that scale. This substitution is summarized locally by the marginal rate of substitution (MRS), $F_2(y_1, y_2)/F_1(y_1, y_2)$. This is the compensating increase in district-level misrepresentation that leaves total misrepresentation unchanged after a one-unit reduction in statewide misrepresentation, evaluated at the loss pair (y_1, y_2). We write $\phi := w/(1-w)$ for the relative weight on statewide misrepresentation.

We say that F has diminishing MRS if F is continuously differentiable with positive partial derivatives $F_1 > 0$, $F_2 > 0$, and the ratio $F_2(y_1, y_2)/F_1(y_1, y_2)$ is weakly decreasing in y_1 and weakly increasing in y_2. Under this condition, the substitution rate between the two losses falls as district-level misrepresentation grows and rises as statewide misrepresentation grows. Equivalently, additional district-level loss becomes harder to compensate when district-level misrepresentation is already high, while reductions in statewide misrepresentation become more valuable when statewide misrepresentation is high.

A useful benchmark is the linear aggregation, which adds the weighted district and statewide losses: $F(y_1, y_2) = y_1 + y_2$, so that $M(\mathbf{x}, \mathbf{p}; w) = (1-w)\mathrm{Dist}(\mathbf{x}, \mathbf{p}) + w\mathrm{State}(\mathbf{x}, \mathbf{p})$. For this aggregator, $F_1 = F_2 = 1$. Hence the planner's pointwise marginal rate of substitution is constant and equal to one.[8]

Axiomatization. The specification can also be derived from primitive requirements on how district and statewide losses should behave. These requirements lead to an additive district-loss component and a statewide-loss component based on distance from a target seat count. Appendix B gives the formal characterization.

[8] The effective willingness to exchange district loss for statewide improvement is determined jointly by the product of relative weight ϕ and the marginal rate of substitution generated by F, which in this case results in ϕ.

3 Misrepresentation-Minimizing Rules

Given a fixed vote profile $\mathbf{p}$, an allocation $\mathbf{x}$ *minimizes misrepresentation* at weight w if

$$\mathbf{x} \in \arg\min_{\mathbf{y}\in\{0,1\}^N} M(\mathbf{y}, \mathbf{p}; w).$$

A rule $R(\cdot)$ minimizes misrepresentation at weight w if for every profile $\mathbf{p}$, the induced allocation minimizes $M(\cdot, \mathbf{p}; w)$.[9]

The solution has three steps. First, holding the seat total fixed, district responsiveness implies that Party A should receive its strongest districts. Second, this conditional result reduces the allocation problem to a one-dimensional choice over seat totals, where optimality can be studied through adjacent-seat comparisons. Third, single crossing orders the adjacent-seat comparisons and yields a cutoff implementation.

Conditional assignment. We begin with a simple conditional assignment observation. Holding fixed the number of seats awarded to party A, district responsiveness implies that those seats can be assigned to the districts where party A is strongest.

Lemma 1. *Top-S allocation $\mathbf{x}^S$ minimizes district-level misrepresentation among all allocations that award party A exactly S seats.*

The logic is best explained with a simple seat-swap argument. Suppose an allocation with S seats for party A gives one of those seats to district j but not to district i, where $p_i \geq p_j$. Consider swapping the assignments. By district responsiveness, the relative loss from assigning a district to party A rather than party B is weakly decreasing in p, so this expression is weakly negative. Thus the swap cannot increase district-level misrepresentation. Repeating this argument eliminates every inversion, leaving an allocation in which party A receives exactly the S districts where it is strongest, namely the top-S allocation.

Reduction to seat totals. For seat totals and any ordered profile $\mathbf{p}$, we write $\mathrm{Dist}(S, \mathbf{p}) := \mathrm{Dist}(\mathbf{x}^S, \mathbf{p})$. By Lemma 1, this is the minimum district loss among allocations that award party A exactly S seats.

The statewide term depends on an allocation only through the seat total $S(\mathbf{x})$, while Lemma 1 identifies an optimal assignment of districts conditional on any fixed seat total. This reduces the allocation problem to a one-dimensional choice over seat totals: choose S, then implement it by a top-S allocation.

[9] When the set of minimizers is not single-valued, one may fix any deterministic selection. None of the structural results depends on how ties among minimizers are resolved.

Lemma 2. *Every optimal seat total has an optimal top-S implementation. Moreover, the optimal seat total is attained by some S between S^{Dist} and S^{State}.*

To see the intuition behind this, suppose that $S^{\text{Dist}} \leq S^{\text{State}}$. If $S < S^{\text{Dist}}$, then raising the seat total to S^{Dist} weakly improves the district component and also moves the delegation weakly closer to the statewide target. If $S > S^{\text{State}}$, then lowering the seat total to S^{State} weakly improves the statewide component and, because this remains weakly above the district optimum, single crossing in δ implies that this does not worsen the district component. Thus any seat total outside $[S^{\text{Dist}}, S^{\text{State}}]$ is dominated by an endpoint of the interval. Since F is increasing in both loss arguments, an optimum can be chosen only where improving one component may require worsening the other: between the district-optimal and statewide-optimal seat totals.

Throughout the remainder of the main text, unless stated otherwise, we state the results for the case $S^{\text{Dist}} \leq S^{\text{State}}$. This is the case in which the district-optimal allocation gives party A weakly fewer seats than the statewide-optimal seat total. If $S^{\text{Dist}} > S^{\text{State}}$, the same arguments apply after reversing party labels.[10] If $S^{\text{Dist}} = S^{\text{State}}$, the district and statewide components select the same seat total, so the tradeoff is degenerate.

District margin. For each S, the district loss of the top-S allocation is

$$\text{Dist}(S, \mathbf{p}) = \sum_{d \leq S} \delta(1, p_d) + \sum_{d > S} \delta(0, p_d).$$

The district margin compares two adjacent points on $\text{Dist}(S, \mathbf{p})$. Moving from S to $S+1$ keeps the top-S districts assigned to party A and changes only district $S+1$, which is switched from party B to party A. Using the district switch margin defined above, the marginal district cost of increasing party A's seat total from S to $S+1$ can be written as

$$\Delta\text{Dist}(S, \mathbf{p}) := \text{Dist}(S+1, \mathbf{p}) - \text{Dist}(S, \mathbf{p}) = \Delta(p_{S+1}).$$

District responsiveness says that $p \mapsto \Delta(p)$ is weakly decreasing. Since the shares are ordered, p_{S+1} is weakly decreasing in S. Consequently, the sequence $S \mapsto \Delta\text{Dist}(S, \mathbf{p})$ is weakly increasing and therefore $\text{Dist}(S, \mathbf{p})$ is discretely convex.

Statewide margin. Consider increasing party A's seat total from S to $S+1$. Along the range $S = S^{\text{Dist}}, \ldots, S^{\text{State}} - 1$, the move raises district-level misrepresentation and lowers

[10] This is a notational reduction rather than an additional party-neutrality assumption. If party neutrality is not imposed, the relabeled problem uses the correspondingly relabeled district and statewide primitives.

statewide misrepresentation. The statewide improvement is

$$\Delta\mathrm{State}(S,\rho) := \mathrm{State}(S,\rho) - \mathrm{State}(S+1,\rho) > 0.$$

Because $S \mapsto \mathrm{State}(S,\rho)$ is discretely convex and decreases as S moves toward S^{State} from the left, the sequence $S \mapsto \Delta\mathrm{State}(S,\rho)$ is weakly decreasing on this range. Intuitively, each additional seat moves the delegation closer to the statewide target, but the remaining statewide gain becomes weakly smaller as the target is approached.

Under the top-S representation, both components of the problem live on the same one-dimensional seat-count domain: a seat total S determines the district loss $\mathrm{Dist}(S,\mathbf{p})$ and the statewide loss $\mathrm{State}(S,\rho)$, so we can write the full objective directly as a function of S:

$$M(S,\mathbf{p};w) := M(\mathbf{x}^S,\mathbf{p};w) = F\big((1-w)\mathrm{Dist}(S,\mathbf{p}), w\mathrm{State}(S,\rho)\big).$$

Marginal rate of substitution. The discrete move from S to $S+1$ traces a line segment in the two-dimensional space of district and statewide losses. The relevant tradeoff is therefore the average marginal rate of substitution along that segment. Define the loss along the segment connecting S and $S+1$, parameterized by $\nu \in [0,1]$:

$$z_S(\nu;\mathbf{p},w) := \Big((1-w)(\mathrm{Dist}(S,\mathbf{p}) + \nu\Delta\mathrm{Dist}(S,\mathbf{p})), w(\mathrm{State}(S,\rho) - \nu\Delta\mathrm{State}(S,\rho))\Big),$$

Define the average marginal rate of substitution along this line segment:

$$\lambda^F(S,\mathbf{p};w) := \frac{\int_0^1 F_2(z_S(\nu;\mathbf{p},w))\,d\nu}{\int_0^1 F_1(z_S(\nu;\mathbf{p},w))\,d\nu}.$$

When the profile and weight are fixed, we suppress them and write $z_S(\nu)$ and $\lambda^F(S)$. The linear case illustrates the definition. If $F(y_1,y_2) = y_1+y_2$, then $F_1(z_S(\nu;\mathbf{p},w)) = F_2(z_S(\nu;\mathbf{p},w)) = 1$ at every point of the segment. Hence both integrals equal one, and $\lambda^F(S,\mathbf{p};w) = 1$.

Since the optimal implementation of any seat total is top-S allocation, the resulting allocation is a cutoff allocation: party A receives all districts above the cutoff and no districts below it. Using the relative weight on statewide misrepresentation $\phi = w/(1-w)$, we define the optimal cutoff function:

$$\tau(S,\mathbf{p};w) := \Delta^{-1}\left(\phi \cdot \Delta\mathrm{State}(S,\rho) \cdot \lambda^F(S,\mathbf{p};w)\right).$$

When the profile and weight are fixed, we write simply $\tau(S)$.

Theorem 1. *For each $S \in \{S^{\text{Dist}}, \dots, S^{\text{State}} - 1\}$, the move from S to $S+1$ is weakly improving if and only if the next district in the ordered list clears the cutoff:*

$$M(S+1, \mathbf{p}; w) \leq M(S, \mathbf{p}; w) \quad \iff \quad p_{S+1} \geq \tau(S, \mathbf{p}; w).$$

Intuition for Theorem 1. For each seat total S, the top-S allocation generates the loss pair $\big(\text{Dist}(S, \mathbf{p}), \text{State}(S, \rho)\big)$. Moving from S to $S+1$ moves along this frontier. It raises district-level misrepresentation by $\Delta\text{Dist}(S, \mathbf{p})$, because district $S+1$ is switched from party B to party A, and it lowers statewide misrepresentation by $\Delta\text{State}(S, \rho)$.

To capture the tradeoff on the frontier, we define the ratio:

$$\kappa(S, \mathbf{p}; w) := \frac{\Delta\text{Dist}(S, \mathbf{p})}{\phi \cdot \Delta\text{State}(S, \rho)}.$$

This is the scaled marginal rate of transformation along the misrepresentation possibilities frontier when the seat total increases from S to $S+1$. Since $\Delta\text{Dist}(S, \mathbf{p})$ is weakly increasing and $\Delta\text{State}(S, \rho)$ is weakly decreasing on this range, the sequence $S \mapsto \kappa(S, \mathbf{p}; w)$ is weakly increasing. When the profile and weight are fixed, we also write $\kappa(S)$.

The discrete first order condition behind Theorem 1 compares this frontier cost with the planner's average marginal rate of substitution. For each $S \in \{S^{\text{Dist}}, \dots, S^{\text{State}} - 1\}$,

$$M(S+1, \mathbf{p}; w) \leq M(S, \mathbf{p}; w) \quad \iff \quad \kappa(S, \mathbf{p}; w) \leq \lambda^F(S, \mathbf{p}; w). \tag{1}$$

The condition in (1) compares the cost of moving along the frontier with the planner's willingness to make that move. $\kappa(S, \mathbf{p}; w)$ is the scaled cost of the adjacent move: it measures how much additional district-level misrepresentation must be accepted per weighted unit of statewide improvement. $\lambda^F(S, \mathbf{p}; w)$ is the planner's average marginal rate of substitution along the same move. Thus moving from S to $S+1$ is weakly improving when the planner's willingness to trade district-level loss for statewide improvement is at least as large as the frontier cost.

Using $\Delta\text{Dist}(S, \mathbf{p}) = \Delta(p_{S+1})$, condition (1) can be written as

$$\Delta(p_{S+1}) \leq \phi \cdot \Delta\text{State}(S, \rho) \cdot \lambda^F(S, \mathbf{p}; w).$$

This is the cutoff inequality in Theorem 1. The left-hand side is the district cost of awarding the next district to party A. The right-hand side is the statewide value of doing so: the statewide gain from the additional seat, scaled by the relative weight ϕ and by the planner's substitution rate. Since this expression is the argument of Δ^{-1} in $\tau(S, \mathbf{p}; w)$, a larger statewide value lowers the cutoff. The rule is then willing to award party A the next seat in a weaker district.[11]

Optimal allocation rule. Theorem 1 gives the local cutoff test for a single adjacent move. With diminishing MRS, these local tests are ordered across seat totals, so they aggregate into a global cutoff rule.

Theorem 2. *Suppose that F has diminishing MRS. Then there is an optimal cutoff allocation: party A receives exactly those districts whose ordered vote shares satisfy $p_S \geq \tau(S-1, \mathbf{p}; w)$.*

Theorem 2 turns the adjacent comparison into a full allocation rule. Diminishing MRS makes the cutoff tests single crossing, so once a marginal district fails the test, weaker districts fail it as well. The next step is to relate the statewide weight w to the seat total selected by the optimal allocation. The frontier representation above is useful for this purpose because each adjacent move changes both coordinates of misrepresentation: it raises district-level loss while lowering statewide loss. Diminishing MRS gives these adjacent comparisons a single crossing structure, which lets us describe the weights that support each seat total.

Supporting weights. For each $S \in \{S^{\text{Dist}}, \dots, S^{\text{State}}\}$, define $\mathcal{W}_S(\mathbf{p})$ as the set of weights $w \in [0,1]$ for which S is an optimal seat total:

$$\mathcal{W}_S(\mathbf{p}) := \left\{ w \in [0,1] : S \in \operatorname*{arg\,min}_{s \in \{0,\dots,N\}} M(s, \mathbf{p}; w) \right\}.$$

Proposition 1 shows that these supporting-weight sets are intervals, and that they are ordered in the same direction as the seat totals.

Proposition 1. *If F has diminishing MRS, then each $\mathcal{W}_S(\mathbf{p})$ is a nonempty closed interval. Moreover, these intervals are ordered by seat total: writing $\mathcal{W}_S(\mathbf{p}) = [\ell_S, u_S]$, we have*

$$0 = \ell_{S^{\text{Dist}}} \leq \ell_{S^{\text{Dist}}+1} \leq \cdots \leq \ell_{S^{\text{State}}}, \quad \textit{and} \quad u_{S^{\text{Dist}}} \leq u_{S^{\text{Dist}}+1} \leq \cdots \leq u_{S^{\text{State}}} = 1.$$

The result also formalizes the monotone path from the district-optimal seat total to the statewide-optimal seat total. At low ϕ, the objective places little value on reducing statewide misrepresentation, so the optimal allocation stays near S^{Dist}. As ϕ rises, the statewide gain from adding seats for party A receives more weight, and the rule becomes willing to admit districts in which party A has weaker support. Diminishing MRS gives the path its direction: once a higher seat total is justified, further increases in ϕ do not push the optimum back

[11] If the district switch margin Δ is only weakly decreasing, the same statement holds with the corresponding generalized inverse and the maintained tie-breaking convention.

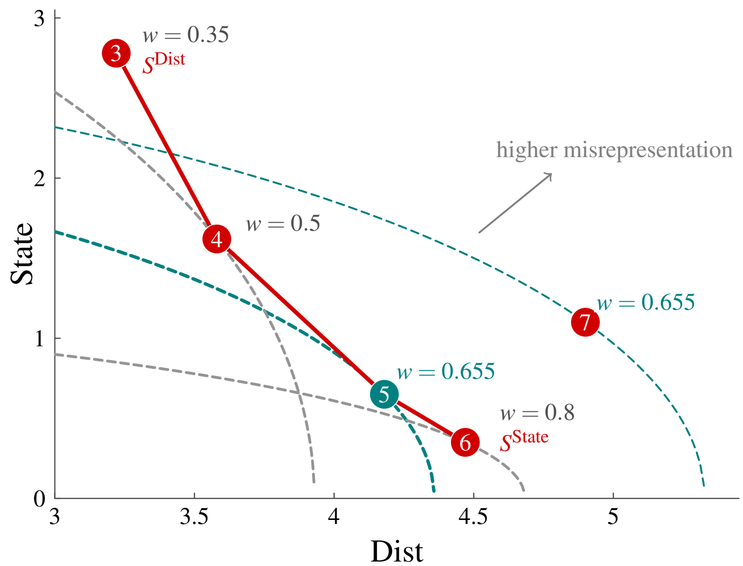


Figure 1: Red circles mark Pareto-frontier top-S allocations, and the red path connects them. Dashed curves are iso-misrepresentation level sets. The teal point and lower teal curve highlight the allocation selected at $w = 0.655$, while the upper teal curve shows a higher-misrepresentation level set passing through a Pareto-inefficient top-S allocation.

toward fewer seats. Thus the optimal seat total moves weakly upward with the relative weight on statewide representation.

Illustration. Figure 1 illustrates the optimality condition in (1) and the comparative static in Proposition 1. For a fixed vote profile, the red points are the feasible frontier points generated by Pareto optimal top-S allocations, and the red path connects them for visual representation.[12] Moving southeast along the frontier awards the underrepresented party A more seats: district-level misrepresentation rises, while statewide misrepresentation falls.

The optimality condition compares the frontier cost of the next southeast move, $\kappa(S, \mathbf{p}; w)$, with the planner's willingness to trade statewide for district-level loss, $\lambda^F(S, \mathbf{p}; w)$. This is a discrete tangency condition: the frontier's marginal rate of transformation is compared with the planner's marginal rate of substitution. Graphically, the selected frontier point is the top-S allocation supported by the lowest attainable iso-misrepresentation level set at that value of w. With diminishing MRS, supporting w intervals are ordered: increasing w makes statewide loss more important and moves the supported seat total weakly southeast toward the statewide target.

[12] Because seat totals are discrete, the frontier consists of the displayed top-S points, and the connecting line is only a visual representation. So an iso-misrepresentation curve may cross it between neighboring seat totals without affecting the discrete optimality comparison. Allocations assigning fewer than 3 seats or more than 8 seats to party A are not plotted because their misrepresentation pairs lie outside the displayed axes.

Bonus vote. The cutoff representation can also be interpreted as a vote-bonus. Let p^0 denote the district-only indifference threshold, defined by $\Delta(p^0) = 0$. Lowering the cutoff from p^0 to $\tau(S, \mathbf{p}; w)$ is equivalent to adding a uniform statewide bonus $b(S, \mathbf{p}; w) := p^0 - \tau(S, \mathbf{p}; w)$ to party A's vote share in every district and then applying the district-only threshold. That is, $p_d \geq \tau(S, \mathbf{p}; w)$ if and only if $p_d + b(S, \mathbf{p}; w) \geq p^0$. When party A is short of the statewide target, the rule can therefore be read as crediting party A with the same statewide adjustment in every district, with the size of the adjustment determined by the statewide gain from adding a seat and the planner's substitution rate.[13]

4 The Linear Benchmark

This section specializes the model to a linear benchmark. District-level misrepresentation is the mass of voters whose own district is represented by their nonpreferred party, statewide misrepresentation is the vote-seat gap, and the planner aggregates the two losses linearly. We first define this benchmark. We then give two reasons to focus on it: varying the weight on statewide misrepresentation in the linear model traces the full Pareto frontier of district and statewide losses, and the same objective admits a direct voter-loss foundation. We then analyze the linear model and derive the cutoff rules that implement the optimum and show how the rule moves from first-past-the-post toward proportional representation as the relative weight on statewide representation rises.

We define district-level misrepresentation explicitly as the mass of voters whose own district is represented by their nonpreferred party. Thus, for a district with party A vote share p, we set $\delta(1, p) = 1 - p$, and $\delta(0, p) = p$. For any allocation $\mathbf{x}$, district-level misrepresentation becomes:

$$\text{Dist}(\mathbf{x}, \mathbf{p}) = \sum_{d=1}^{N} \Big[x_d(1 - p_d) + (1 - x_d)p_d \Big].$$

We also explicitly define statewide misrepresentation as the vote-seat gap. That is, we set the statewide target equal to party A's proportional seat total, $T(\rho) = \rho$, and use the linear loss function $Q(z) = z$ to obtain:

$$\text{State}(\mathbf{x}, \mathbf{p}) = |S(\mathbf{x}) - \rho|.$$

Note that both components lie on the same scale of $[0, N]$.[14] Dist is the normalized mass

[13]Unlike first-past-the-post, the rule gives votes in locally safe districts a representational role through the statewide component. A vote that cannot change its own district winner can still affect the statewide target, or the cutoff applied to other districts, creating a participation channel through statewide representation.

[14]For Dist, the value is 0 when every district is assigned to a party supported by all voters in that district,

of voters locally represented by their nonpreferred party, while State is the number of seats by which the delegation departs from proportionality.

Finally, we specialize to the linear aggregator: $F(y_1, y_2) = y_1 + y_2$. For $w \in [0, 1]$, the misrepresentation objective becomes:

$$M(\mathbf{x}, \mathbf{p}; w) := (1 - w)\text{Dist}(\mathbf{x}, \mathbf{p}) + w|S(\mathbf{x}) - \rho|.$$

The parameter w is the weight on statewide misrepresentation. At $w = 0$, the objective is purely district-level and awards each district to its local majority. At $w = 1$, only statewide misrepresentation matters, so the rule selects an integer seat total closest to proportionality.

4.1 Pareto frontier

The linear objective selects, for each weight, an allocation that minimizes a weighted sum of district-level and statewide misrepresentation. We now ask which feasible tradeoffs between these two losses can be justified in this way. For a fixed vote-share profile $\mathbf{p}$, each allocation generates a loss pair $\big(\text{Dist}(\mathbf{x}, \mathbf{p}), \text{State}(S(\mathbf{x}), \rho)\big)$. An allocation $\mathbf{x}$ is *Pareto efficient* if no alternative allocation weakly reduces both district-level and statewide misrepresentation, with at least one strict reduction.

In multiobjective problems, weighted sums identify *supported* Pareto points: points on the lower convex envelope of the feasible loss set. In finite, nonconvex problems, some Pareto-efficient points can be *unsupported*; they are efficient, but no linear weight makes them optimal. The next proposition shows that the linear benchmark supports every Pareto-efficient point in this problem.

Proposition 2. *An allocation $\widehat{\mathbf{x}}$ is Pareto efficient if and only if there exists a weight $w \in [0, 1)$ such that*

$$\widehat{\mathbf{x}} \in \underset{\mathbf{x} \in \{0,1\}^N}{\arg\min} \left\{(1 - w)\text{Dist}(\mathbf{x}, \mathbf{p}) + w|S(\mathbf{x}) - \rho|\right\}.$$

Moreover, $\widehat{\mathbf{x}}$ is a top-S allocation.

The intuition for Proposition 2 is as follows. The conditional-assignment result simplifies the efficient-set problem. If an allocation has seat total S but is not a top-S allocation, then replacing it with the top-S allocation leaves statewide misrepresentation unchanged

---

and N when every district is assigned to a party supported by no voters in that district. For State, the value is 0 when the seat total equals party A's proportional seat total, $S(\mathbf{x}) = \rho$, and N when one party receives all seats despite receiving no statewide support.

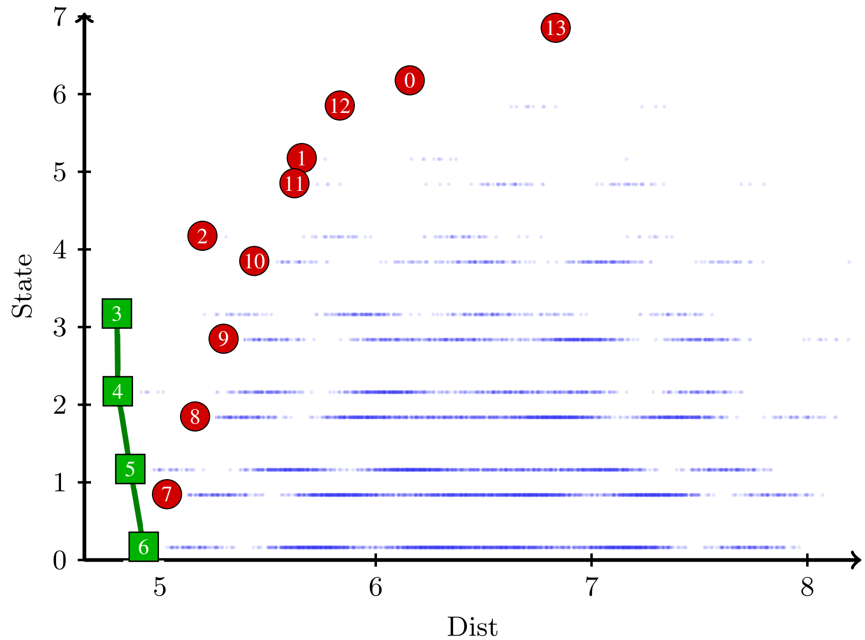


Figure 2: Feasible misrepresentation pairs (Dist, State) for the profile in the right panel of Figure 3. Red points are the top-S allocations. Green squares trace the misrepresentation-minimizing allocation as w varies. The number inside each top-S allocation denotes the corresponding seat total S. Blue points are all other feasible allocations.

and weakly lowers district-level misrepresentation. Hence every Pareto-efficient allocation must be conditionally district-cost minimizing at its seat total. Among top-S allocations, any seat total outside the Pareto frontier is weakly dominated in both dimensions by another top-S' allocation. In the linear benchmark, the remaining top-S frontier has ordered adjacent tradeoffs, so every Pareto-efficient frontier point is supported by some linear weight.

The relative weight $\phi = w/(1-w)$ has the revealed-preference interpretation used in Section 7. It is the amount of additional district-level misrepresentation the rule is willing to accept in order to reduce statewide misrepresentation by one unit. Equivalently, if an observed allocation is rationalized by normalized weight w, then ϕ is the implicit price it places on moving the delegation one seat closer to proportionality.

Figure 2 visualizes the Pareto frontier of feasible misrepresentation pairs (Dist, State) via the green squares. As w increases, the misrepresentation-minimizing allocation moves along this frontier by jumping between successive green points. At any switching weight between two consecutive green squares, the indifference condition implies that the connecting segment has slope $-1/\phi$, i.e. the weighted marginal rate of substitution between State and Dist at the switch.

Welfare Theorems. Proposition 2 relates to the familiar *Welfare Theorems*. The *First*

Welfare Theorem analog is that any allocation minimizing a strictly positive linear combination of district-level and statewide misrepresentation is Pareto efficient. If another allocation reduced both losses, the linear objective would also be lower.

The *Second Welfare Theorem* analog is the converse. Even though the feasible set of allocations is discrete and hence nonconvex, every Pareto-efficient point has a supporting *price* $\phi \geq 0$, so varying ϕ traces the whole efficient frontier. This supporting-price logic is not specific to the voter-loss or vote-seat-gap normalization used in the linear benchmark. It is a consequence of the top-S structure. Appendix C unpacks this frontier result via supporting-prices, and then records their implications. In particular, it proves Proposition 2 for the general Dist$(\cdot)$ and State$(\cdot)$ introduced in Section 2.

The supporting-price interpretation also yields a law of demand for statewide misrepresentation. For a fixed vote profile, a higher ϕ weakly lowers statewide loss at the selected price-optimal seat total by moving that total toward $T(\rho)$. Along this path, district loss weakly rises because statewide correction is obtained through locally weaker districts. Changes in the vote profile also admit a Slutsky-style decomposition: a compensated distributional effect, which changes the marginal district-cost schedule $\Delta\text{Dist}(S, \mathbf{p})$ holding $T(\rho)$ fixed, and a target effect, which changes $T(\rho)$ itself. Appendix C also formalizes these implications.

4.2 A Voter-Loss Foundation

The frontier result shows that the linear family is rich enough to capture all efficient tradeoffs. We next give the same objective a voter-level interpretation. Voters value representation along two dimensions. First, they value local representation: they prefer their own district seat to be held by their preferred party. Second, they value statewide representation: they prefer their preferred party to hold more seats in the statewide delegation. We express these preferences in loss form. Voters incur a local loss when their own district is represented by their nonpreferred party and a statewide loss from any shortfall between their party's statewide representation and its statewide support.[15]

We write $(z)_+$ to mean $\max\{z, 0\}$. For a voter who supports party A in district d, define loss under allocation $\mathbf{x}$ by

$$\ell_d^A(\mathbf{x}, \mathbf{p}; w) = (1-w)(1-x_d) + w\frac{(\rho - S(\mathbf{x}))_+}{\rho}.$$

[15]Appendix D gives an equivalent utility formulation in which voter payoffs are increasing in local and statewide representation, rather than decreasing in nonrepresentation and underrepresentation. Aggregating voter utility yields the same linear objective.

For a voter who supports party B in district d, define loss under allocation $\mathbf{x}$ by

$$\ell_d^B(\mathbf{x}, \mathbf{p}; w) = (1-w)x_d + w\frac{(S(\mathbf{x}) - \rho)_+}{N - \rho}.$$

The first term in each expression captures local representation. It reflects the loss a voter experiences when her own district is represented by a member of the other party: an A-voter incurs this loss when district d is assigned to party B, and a B-voter incurs it when district d is assigned to party A. For an A-voter, the statewide loss term measures the shortfall between party A's seat total and its statewide support. The loss is maximal when party A receives no seats, decreases linearly as $S(\mathbf{x})$ rises, and reaches zero once party A's seat total equals its statewide support ρ. In loss form, this term is $(\rho - S(\mathbf{x}))^+/\rho$. For a B-voter, the analogous shortfall is between party B's seat total $N - S(\mathbf{x})$ and its statewide support $N - \rho$, which gives the term $(S(\mathbf{x}) - \rho)^+/(N - \rho)$.

The planner minimizes aggregate voter loss,

$$L(\mathbf{x}, \mathbf{p}; w) := \sum_{d=1}^{N} \left[\underbrace{p_d \ell_d^A(\mathbf{x}, \mathbf{p}; w)}_{\substack{\text{total loss of party } A \text{ voters} \\ \text{in district } d}} + \underbrace{(1-p_d)\ell_d^B(\mathbf{x}, \mathbf{p}; w)}_{\substack{\text{total loss of party } B \text{ voters} \\ \text{in district } d}} \right].$$

Proposition 3. *For any allocation* $\mathbf{x}$*, aggregate voter loss is*

$$L(\mathbf{x}, \mathbf{p}; w) = (1-w)\text{Dist}(\mathbf{x}, \mathbf{p}) + w|S(\mathbf{x}) - \rho|.$$

Therefore, minimizing aggregate voter loss is equivalent to minimizing the linear misrepresentation objective.

4.3 The Optimal Election Rule

We now fully characterize the optimal election rules in the linear model. To do so, we first formalize the class of cutoff rules used in Theorem 2: rules that assign seats to Party A exactly in the districts where its vote share clears a threshold.

Cutoff rules. A rule R is a cutoff rule if, for every profile $\mathbf{p}$, there is a threshold $t(\mathbf{p}) \in [0, 1]$ such that party A wins the districts whose vote shares are above the cutoff:

$$R(\mathbf{p}) = \{d : p_d \geq t(\mathbf{p})\}.$$

Two canonical rules are cutoff rules. In the two-party setting, first-past-the-post, denoted

R^{FPTP}, is the district-by-district plurality rule; equivalently, $R^{\text{FPTP}}(\mathbf{p}) = \{d : p_d \geq 1/2\}$.

Proportional representation, denoted R^{PR}, is the district-based analog of party-list proportional representation.[16] It chooses the integer seat total closest to party A's proportional entitlement, $S^{PR}(\mathbf{p}) := \min \arg\min_{S\in\{0,\dots,N\}} |S - \rho|$, and assigns those seats to party A's strongest districts: $R^{PR}(\mathbf{p}) = \{d : d \leq S^{PR}(\mathbf{p})\}$. Equivalently, R^{PR} is a cutoff rule with profile-dependent threshold

$$t(\mathbf{p}) := \begin{cases} 1, & S^{PR}(\mathbf{p}) = 0, \\ p_{S^{PR}(\mathbf{p})}, & S^{PR}(\mathbf{p}) \in \{1, \dots, N\}. \end{cases}$$

Note that the proportional seat total depends on the fractional part of ρ. Let $\underline{\rho} := \lfloor \rho \rfloor$ denote the integer part of statewide support. Then $S^{PR}(\mathbf{p}) = \underline{\rho}$ when $\rho - \underline{\rho} \leq 1/2$, while $S^{PR}(\mathbf{p}) = \underline{\rho} + 1$ when $\rho - \underline{\rho} > 1/2$. In the first case, proportional representation rounds down; in the second, it rounds up.

Optimal rule. The linear environment makes the general cutoff theorem explicit. Under the linear district loss, $\Delta(p) = 1 - 2p$, so $\Delta^{-1}(y) = (1 - y)/2$. Since $F(y_1, y_2) = y_1 + y_2$, the average marginal rate of substitution is $\lambda^F(S, \mathbf{p}; w) = 1$. Thus, whenever an additional seat reduces statewide loss by one unit, Theorem 1 gives the cutoff

$$\Delta(p) = \phi \quad \Longleftrightarrow \quad p = \frac{1 - \phi}{2}.$$

This is the basic linear cutoff: ϕ is the relative value of a one-seat statewide improvement, and $(1 - \phi)/2$ is the district vote share at which that statewide gain exactly offsets the district-level cost of awarding the seat to party A.

Under our maintained convention $S^{\text{Dist}} \leq S^{\text{State}}$, the optimal path starts at the district-optimal seat total at $\phi = 0$ and as ϕ increases, follows this basic cutoff until party A reaches $\underline{\rho}$ seats. Let $\phi_{\underline{\rho}}$ denote the relative weight at which the top-$\underline{\rho}$ allocation is first implemented.[17] Thus, for $0 \leq \phi < \phi_{\underline{\rho}}$, the optimal cutoff is $(1 - \phi)/2$.

Once $\underline{\rho}$ seats are reached, the cutoff no longer follows the linear path $(1 - \phi)/2$. Consider first the round-down case, $\rho - \underline{\rho} \leq 1/2$. Then the top-$\underline{\rho}$ allocation is already proportional representation. Awarding party A one additional seat would increase statewide misrepresentation, since ρ is weakly closer to $\underline{\rho}$ than to $\underline{\rho} + 1$, and it would also increase district-level

[16] In standard party-list proportional representation, seats are first allocated to parties in proportion to their vote shares and then assigned to candidates from the party lists. In our district-based setting, R^{PR} matches that party-level seat total and assigns party A's seats (and therefore, party B's seats) to the districts in which its vote share is highest.

[17] Formally, $\phi_{\underline{\rho}} := 0$ if $S^{\text{Dist}} = \underline{\rho}$, and $\phi_{\underline{\rho}} := 1 - 2p_{\underline{\rho}}$ if $S^{\text{Dist}} < \underline{\rho}$.

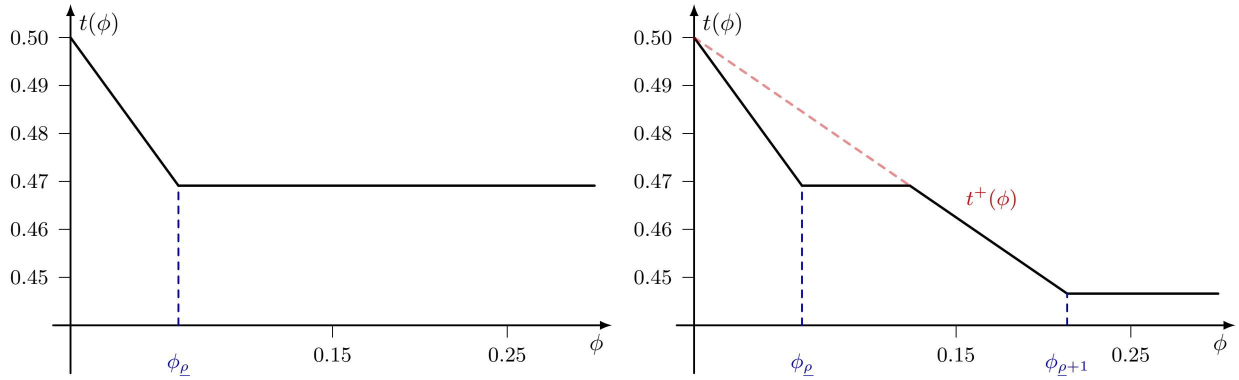


Figure 3: Optimal cutoff rule $t(\phi)$ as the relative weight on statewide misrepresentation ϕ varies. Left: the actual 2018 North Carolina vote profile; proportionality rounds down. Right: a perturbed version; proportionality rounds up.

misrepresentation, since the next district was lost by party A under first-past-the-post. Hence, once proportional representation is reached, the optimal cutoff remains constant at $p_{\underline{\rho}}$.

The left panel of Figure 3 illustrates this case. At $\phi = 0$, the rule coincides with first-past-the-post. As ϕ rises, the cutoff falls along $(1-\phi)/2$, bringing in the strongest districts that party A lost under first-past-the-post, until ϕ reaches $\phi_{\underline{\rho}}$. After that point, the cutoff stays fixed and continues to implement proportional representation.

Now consider the round-up case, $\rho - \underline{\rho} > 1/2$, so proportional representation requires one additional seat beyond $\underline{\rho}$. The statewide gain from this final seat is smaller than the gain from the earlier seats: moving from $\underline{\rho}$ to $\underline{\rho}+1$ reduces the statewide gap from $\rho - \underline{\rho}$ to $\underline{\rho} + 1 - \rho$, a gain of $2(\rho - \underline{\rho}) - 1$. Thus the cutoff associated with the rounding seat is

$$t^+(\phi) := \frac{1 - \phi\big(2(\rho - \underline{\rho}) - 1\big)}{2}$$

which is larger than $\frac{1-\phi}{2}$, and converges to it as $\rho - \underline{\rho}$ goes to 1. Let $\phi_{\underline{\rho}+1}$ denote the relative weight at which the top-($\underline{\rho}+1$) allocation is first implemented.[18] The cutoff $t^+(\phi)$, however, need not be the operative cutoff immediately after the top-$\underline{\rho}$ allocation is reached. It prices the fractional correction from $\underline{\rho}$ to $\underline{\rho}+1$, while the $\underline{\rho}$-th seat was justified by a full-unit improvement, so $t^+(\phi)$ may initially lie above $p_{\underline{\rho}}$. The marginal cutoff is therefore $\min\{p_{\underline{\rho}}, t^+(\phi)\}$.

The right panel of Figure 3 illustrates this case. The cutoff first falls along $(1-\phi)/2$, as in the round-down case, until the top-$\underline{\rho}$ allocation is reached. It then follows the rounding-seat logic: initially the cutoff may remain flat at $p_{\underline{\rho}}$, and once $t^+(\phi) \le p_{\underline{\rho}}$, it declines with $t^+(\phi)$

[18]Formally, $\phi_{\underline{\rho}+1} := \frac{1-2p_{\underline{\rho}+1}}{2(\rho-\underline{\rho})-1}$.

until reaching $p_{\underline{\rho}+1}$. As in the first case, once the proportional seat total is implemented, the cutoff remains constant.

The preceding discussion yields the optimal cutoff family: at each relative weight, the cutoff is determined by the marginal tradeoff along the path. For $\phi \geq 0$, define $t_\phi(\mathbf{p})$ by

$$\text{(i)} \quad t_\phi(\mathbf{p}) = \frac{1-\phi}{2}, \quad \text{if} \quad 0 \leq \phi < \phi_{\underline{\rho}};$$

$$\text{(ii)} \quad \rho - \underline{\rho} \leq \frac{1}{2} \quad \Longrightarrow \quad t_\phi(\mathbf{p}) = p_{\underline{\rho}} = \frac{1-\phi_{\underline{\rho}}}{2}, \quad \phi_{\underline{\rho}} \leq \phi;$$

$$\text{(iii)} \quad \rho - \underline{\rho} > \frac{1}{2} \quad \Longrightarrow \quad t_\phi(\mathbf{p}) = \begin{cases} \min\{p_{\underline{\rho}}, t^+(\phi)\}, & \phi_{\underline{\rho}} \leq \phi < \phi_{\underline{\rho}+1}, \\ p_{\underline{\rho}+1}, & \phi_{\underline{\rho}+1} \leq \phi, \end{cases}$$

The corresponding optimal cutoff rule is

$$R_\phi(\mathbf{p}) := \{d : p_d \geq t_\phi(\mathbf{p})\}.$$

Theorem 3. *Suppose $S^{\text{Dist}} \leq \underline{\rho}$. The family $\{R_\phi(\mathbf{p})\}_{\phi\in[0,\infty)}$ implements the optimal allocation under vote profile $\mathbf{p}$ and relative weight ϕ.*

The theorem gives an explicit cutoff implementation of the optimal rule. It also reveals the underlying seat-total path. Proposition 1 shows generally that, under diminishing MRS, the weights supporting each seat total form ordered intervals. In the linear benchmark, this interval structure reduces to adjacent switch weights: except at the switch points, the optimal seat total is unique, so the weight space is partitioned into regions that select successive seat totals from S^{Dist} to S^{State}. The next corollary records this step-function representation.

Corollary 1. *There exist weights $0 =: w_{S^{\text{Dist}}} < w_{S^{\text{Dist}}+1} \leq \cdots \leq w_{S^{\text{State}}} \leq 1$ such that:*

(i) *for each $S \in \{S^{\text{Dist}}, \ldots, S^{\text{State}} - 1\}$, if $w \in (w_S, w_{S+1})$, then the optimal seat total is unique and equals S;*

(ii) *if $w > w_{S^{\text{State}}}$, then the optimal seat total is unique and equals S^{State};*

(iii) *at each switch weight $w = w_{S+1}$, both S and $S + 1$ minimize the linear objective.*

Thus, for an underrepresented party, increasing the statewide weight weakly raises its optimal seat total, moving the allocation monotonically from the district-optimal outcome toward the statewide-optimal one. The movement is discrete because seats change only when the cutoff crosses an observed district vote share, as illustrated in Figure 4.

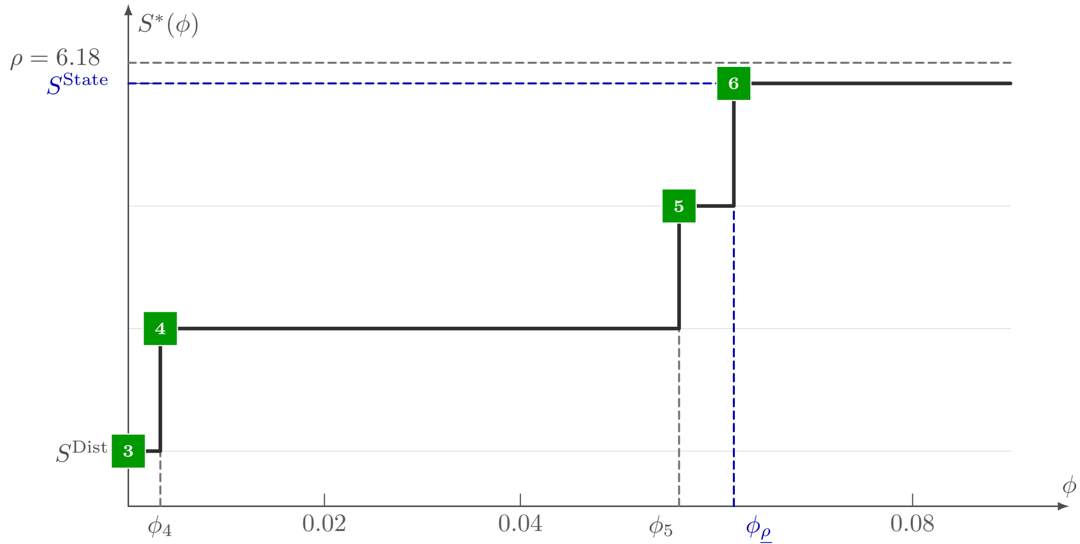


Figure 4: Optimal seat total $S^*(\phi)$ for the profile in the right panel of Figure 3.

5 Analysis of the Optimal Family

Theorem 3 characterizes the optimal rules in the linear benchmark. We collect these rules in the family

$$\mathcal{R} := \{R_\lambda : \lambda \in [0, 1]\},$$

where R_λ is the cutoff rule that minimizes misrepresentation at weight λ. We use λ to index the rule itself, reserving w for the designer's preferences over district-level and statewide misrepresentation, or equivalently ϕ for the induced relative weight, in the comparisons below. This family runs from first-past-the-post at $\lambda = 0$ to proportional representation at $\lambda = 1$, and it traces the optimal tradeoffs between district-level and statewide misrepresentation in the linear benchmark. Accordingly, $\mathcal{R}$ is the central family for the analysis that follows: at any vote-share profile, any allocation chosen by an alternative rule is either selected by some $R_\lambda \in \mathcal{R}$, or is Pareto dominated by one.

5.1 Concentration of Support and Majorization Monotonicity

We begin by studying how the rules in $\mathcal{R}$ respond when the same statewide support for party A is distributed more or less unevenly across districts. Holding the statewide vote share fixed, differences across profiles come from the district-loss term. Under a top-S allocation, the relevant statistic is the cumulative support $\sum_{d \le S} p_d$ in party A's strongest districts, which makes majorization the natural order for comparing profiles.

Formally, given profiles $\mathbf{p}$ and $\mathbf{q}$ with the same mean, say that $\mathbf{p}$ *majorizes* $\mathbf{q}$ if, after ordering both profiles in decreasing order, $\sum_{d=1}^{k} p_d \geq \sum_{d=1}^{k} q_d$ for every $k = 1, \dots, N$. Thus $\mathbf{p}$ is more concentrated than $\mathbf{q}$ if party A has weakly more cumulative support in its strongest districts at every cutoff.

The next lemma shows that misrepresentation satisfies a natural monotonicity property with respect to majorization for a fixed seat count S.

Lemma 3. *If $\mathbf{p}$ majorizes $\mathbf{q}$ and $\bar{p} = \bar{q}$, then $M(S, \mathbf{p}; w) \leq M(S, \mathbf{q}; w)$ for every S and every $w \in [0, 1]$.*

Intuitively, majorization increases Party A's cumulative support in its strongest districts, so for any fixed S, the top-S allocation places seats where fewer voters are represented by their non-preferred party.

We now move from fixing the seat total S to evaluating cutoff rules, which choose the seat total endogenously as a function of the vote-share profile. Write $S_R(\mathbf{r})$ for the seat total selected by a rule R at profile $\mathbf{r}$. The misrepresentation induced by R at evaluation weight w is $M(R, \mathbf{r}; w) := M(S_R(\mathbf{r}), \mathbf{r}; w)$. It turns out that the ordering in Lemma 3 is preserved when the seat total is not fixed but instead is chosen optimally.

Proposition 4. *If $\mathbf{p}$ majorizes $\mathbf{q}$ and $\bar{p} = \bar{q}$, then $M(R_w, \mathbf{p}; w) \leq M(R_w, \mathbf{q}; w)$.*

This conclusion need not extend to rules that keep the district winning threshold fixed. For example, under first-past-the-post the threshold is $\frac{1}{2}$, so the induced seat total can vary across profiles even when the statewide mean $\bar{p}$ is unchanged. When $w > 0$, these discrete seat swings can raise statewide distortion enough to offset the within-district improvement implied by majorization.[19]

This motivates the next question: when do the majorization comparative statics continue to hold once a *rule* chooses the seat total endogenously?

Majorization monotonicity. A rule R satisfies *majorization monotonicity at w* if, for all profiles $\mathbf{p}, \mathbf{q}$ with the same mean and with $\mathbf{p}$ majorizing $\mathbf{q}$,

$$M(R, \mathbf{p}; w) \leq M(R, \mathbf{q}; w). \tag{MM}$$

We have already shown majorization monotonicity when the seat total is fixed and when the seat total is chosen optimally for the same weight used to evaluate misrepresentation. The fixed-seat result implies the property for proportional representation because majorization comparisons hold statewide support fixed, and proportional representation fixes the party

[19] Appendix G.1 gives a four-district example, with a graphical illustration, in which a mean-preserving majorization lowers district-level misrepresentation under first-past-the-post but increases the first-past-the-post seat total, raising statewide misrepresentation enough to increase total misrepresentation.

seat total as a function of that support. The next theorem shows that these are the only such cases.

Theorem 4. *$R_\lambda \in \mathcal{R}$ satisfies majorization monotonicity at w if and only if either $\lambda = w$ or $\lambda = 1$.*

Majorization makes Party A's support more concentrated while holding the statewide mean fixed. By Lemma 3, for any fixed seat total S, this concentration can only strengthen the top-S districts and therefore weakly improves the district component of the objective. Thus any failure of majorization monotonicity must come from how the rule chooses the statewide seat total. When $\lambda = w$, the rule selects S by minimizing the same weighted criterion we use to evaluate outcomes, so greater concentration cannot induce a switch to a worse S. When $\lambda \neq w$, the rule is minimizing a different tradeoff.

We prove Theorem 4 by first showing that one can find a more concentrated profile at which R_λ becomes indifferent between different seat totals, but that indifference does not hold under w. Thus, R_λ picks a seat total that is strictly worse at the more concentrated profile, and choosing the two profiles close enough to minimize the change in district component, we obtain a contradiction to (MM). Hence majorization monotonicity uniquely pins down $\lambda = w$.

5.2 Characterization of Misrepresentation-Minimizing Rules

We next characterize two canonical members of the optimal family $\mathcal{R}$. The district endpoint R_0 is first-past-the-post: each seat is assigned to the party with the higher vote share in that district. The statewide endpoint R_1 is the adaptation of proportional representation to our district-based setting: it selects the integer seat total closest to party A's statewide support ρ. We show that these two endpoints are fully characterized within $\mathcal{R}$ by two natural properties: strong monotonicity characterizes R_0, and gerrymandering-proofness characterizes R_1.

A monotonicity characterization of first-past-the-post. A rule satisfies *strong monotonicity* if, for all profiles $\mathbf{p}, \mathbf{p}'$ with $\mathbf{p}' \geq \mathbf{p}$ componentwise, $R(\mathbf{p}) \subseteq R(\mathbf{p}')$. Strong monotonicity says that raising party A's support district by district cannot cause it to lose any district it previously won.

Proposition 5. *$R_\lambda \in \mathcal{R}$ satisfies strong monotonicity if and only if $\lambda = 0$. In that case, it coincides with first-past-the-post.*

The intuition follows from the structure of the objective. When $\lambda = 0$ the criterion is district-separable and each district is awarded to its local majority, whereas any $\lambda > 0$ couples districts through the statewide term and necessarily induces cross-district reallocation of seats on some profiles, violating monotonicity. For each $\lambda > 0$, we construct a pair of profiles with the following property: increasing Party A's support in one district leaves the optimal seat total at $S = 1$, but changes which district is top-ranked, so the unique awarded seat shifts to the improved district. As a result, a district can lose its seat even though its own vote share is unchanged (and is still above $\frac{1}{2}$), which is ruled out under the fixed-threshold rule of R_0.

A gerrymandering-proof characterization of proportionality. A rule is *gerrymandering–proof* if no party can redistrict their aggregate votes across districts and increase their seat share. Define the set profiles where Party A obtains a vote share of m:

$$\mathcal{P}(m) := \{\mathbf{q} \in [0,1]^N : \bar{q} = m\}.$$

If $\mathbf{q} \in \mathcal{P}(\bar{p})$, $\mathbf{q}$ is a redistricting of support across districts relative to $\mathbf{p}$. Gerrymandering is possible if this redistricting changes the seat total in favor of some party.[20] Requiring that no such redistricting can increase a party's seat total implies that whenever $\mathbf{q} \in \mathcal{P}(\bar{p})$, we must have $|R(\mathbf{p})| \geq |R(\mathbf{q})|$. Applying the same logic after swapping the roles of $\mathbf{p}$ and $\mathbf{q}$ yields the reverse inequality, so the seat totals must coincide whenever $\bar{p} = \bar{q}$, leading to the following definition: a rule is *gerrymandering-proof* if $|R(\mathbf{p})| = |R(\mathbf{q})|$ for every $\mathbf{p}$ and $\mathbf{q} \in \mathcal{P}(\bar{p})$.

Note that for any $\lambda < 1$, the objective trades off district-level and statewide distortions, so the optimal seat total can depend on how support is distributed across districts even when the statewide mean is fixed. Only at $\lambda = 1$ does the seat total depend only on the statewide vote total. This gives the following result.

Proposition 6. *$R_\lambda \in \mathcal{R}$ satisfies gerrymandering-proofness if and only if $\lambda = 1$. In that case, it coincides with the proportional rule.*

A district-rule trilemma. The endpoint characterizations imply a trilemma within the misrepresentation-minimizing family. No rule in this family can combine strong monotonicity, gerrymandering-proofness, and a nondegenerate concern for both district-level and statewide representation. At $\lambda = 0$, the rule is first-past-the-post: raising party A's support district by district cannot take away a seat it already won, but the seat total can change under

[20] We use redistricting in an abstract sense. A redistricting profile is any reassignment of fixed voters to districts that preserves voters' party allegiances and the statewide vote total, so it includes changes that could be generated by redrawing district boundaries. The feasible set $\mathcal{P}(\bar{p})$ deliberately does not impose geographic, legal, or administrative constraints. This abstraction isolates how the allocation rule responds to the distribution of partisan support across districts, rather than modeling the mapmaking constraints that determine which redistributions are feasible in a particular state.

mean-preserving redistricting of votes. At $\lambda = 1$, the rule is proportional: the seat total is invariant to such redistrictings, but district outcomes can depend on votes cast elsewhere. For $\lambda \in (0,1)$, the rule keeps both objectives in the criterion and therefore gives up the global version of both properties. Moving from first-past-the-post to proportionality is therefore not only a change in formulas; it is a movement between two incompatible notions of electoral principles.

5.3 Gerrymandering Vulnerability

Gerrymandering-proofness is a binary property: a rule either makes the seat total invariant to mean-preserving redistrictings, or it does not. We now measure vulnerability to gerrymandering more finely by asking how much room a rule leaves for such redistrictings to change the outcome. Fix party A's statewide vote share and consider the redistributions of the same statewide support that can give party A additional seats.

For a baseline profile $\mathbf{p}$, an integer $k \in \{0, \ldots, N\}$, and $\lambda \in [0,1]$, define

$$\mathcal{G}_k(\mathbf{p}; \lambda) := \{\mathbf{q} \in \mathcal{P}(\bar{p}) : |R_\lambda(\mathbf{q})| \geq k\}.$$

Thus $\mathcal{G}_k(\mathbf{p}; \lambda)$ is the set of mean-preserving redistrictings of $\mathbf{p}$ that give party A at least k seats under R_λ.

Attainable seat gains. Fix a baseline index $\lambda_0 \in [0,1)$ such that party A is overrepresented at the baseline profile: $|R_{\lambda_0}(\mathbf{p})| > \rho$. Also fix an integer $k > |R_{\lambda_0}(\mathbf{p})|$. Thus k is a seat total strictly above the number of seats party A already receives at the baseline. The next result compares, across $\lambda \in [\lambda_0, 1]$, the set of mean-preserving redistrictings that can deliver at least k seats.

Proposition 7. *For every λ', λ with $1 \geq \lambda' > \lambda \geq \lambda_0$, $\mathcal{G}_k(\mathbf{p}; \lambda') \subseteq \mathcal{G}_k(\mathbf{p}; \lambda)$. Moreover, if $\mathcal{G}_k(\mathbf{p}; \lambda) \neq \emptyset$ and $\lambda' > \lambda$, then $\mathcal{G}_k(\mathbf{p}; \lambda') \subset \mathcal{G}_k(\mathbf{p}; \lambda)$.*

The proposition states that the attainable set is nested along the optimal family. When party A is already overrepresented, reaching k seats moves the delegation still farther from the statewide target. A higher value of λ selects a rule with greater concern for statewide proportionality, so success under a higher-λ rule implies success under every lower-λ rule, while some redistrictings that succeed at lower λ may fail once λ rises. Thus increasing λ makes gerrymandering harder in a set-theoretic sense: it reduces the collection of maps on which the additional seats are attainable.

Figure 5 illustrates the same nesting on the relative rule-index scale $\phi_\lambda = \lambda/(1-\lambda)$. The gray polygon is the set of sorted mean-preserving redistrictings, while each colored

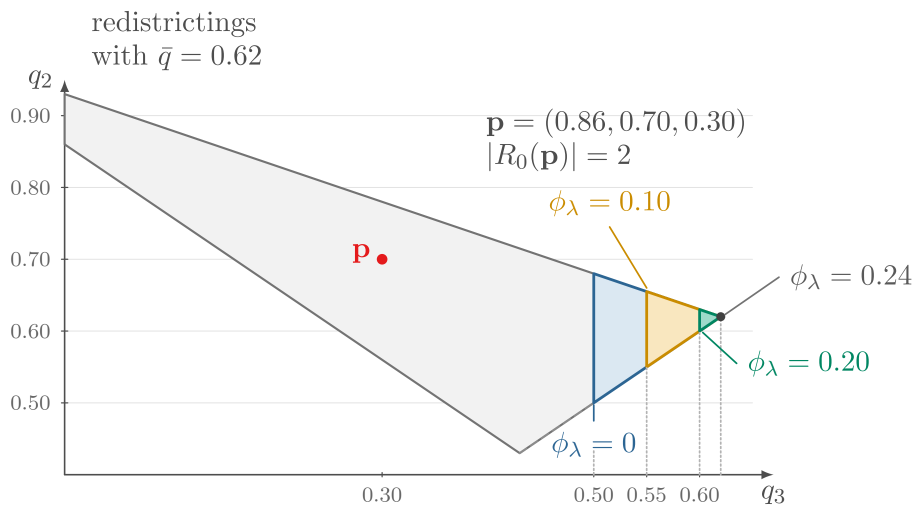


Figure 5: Gray area denotes the set of redistrictings $\mathbf{q}$ with mean $\bar{q} = \bar{p} = 0.62$. On the relative rule-index scale $\phi_\lambda = \lambda/(1-\lambda)$, Party A gets three seats iff $q_3 \geq (1+\phi_\lambda)/2$, denoted by the colored areas at different ϕ_λ values. Thus the feasible gerrymandering sets that allocate 3 seats to party A, $\mathcal{G}_3(\mathbf{p};\lambda)$, shrink as λ rises, and collapse to $(0.62, 0.62, 0.62)$ at $\phi_\lambda = 0.24$.

triangle imposes the additional cutoff needed to give party A all three seats. As ϕ_λ rises, the cutoff moves right and the set shrinks; at $\phi_\lambda = 0.24$, only the equalized profile remains. In Proposition 8, we characterize the point at which this shrinking set becomes empty: above a simple rule-index threshold, no mean-preserving redistricting can deliver the target k seats.

A cost version of this result follows by assigning each mean-preserving redistricting a cost based on its distance from the baseline profile. Appendix E defines this gerrymandering cost and shows that increasing λ weakly raises the minimal cost of obtaining additional seats for an already overrepresented party.

A capacity bound. The next result gives an upper bound on how much overrepresentation can be generated by redistributing the statewide support.

Proposition 8. *Consider a voting profile* $\mathbf{p}$ *with statewide support* $\rho > 0$*, and a seat target* k *such that* $k \geq 1 + \rho$*. If* $\lambda > 1 - \frac{k}{2\rho}$ *then* $\mathcal{G}_k(\mathbf{p};\lambda) = \emptyset$.

The intuition is that seat gains must be supported by enough votes in the strongest districts. When party A is already overrepresented, increasing λ raises the cutoff for awarding it another seat. To make k seats attainable, the k^{th} strongest district in some mean-preserving redistricting must clear the cutoff $\frac{1}{2(1-\lambda)}$. But then the top k districts together require at least $\frac{k}{2(1-\lambda)}$ units of statewide support. Once this exceeds the available support ρ, no redistricting of the same statewide vote can deliver k seats. If $k > 2\rho$, the target is already beyond this capacity even under first-past-the-post.

6 Measures of Representation and Fairness

The previous section characterizes a family of rules that trade off district-level and statewide representation in the linear benchmark. We now use the same primitives to organize measures of partisan advantage. We focus on three measures—first-past-the-post robustness, the efficiency gap, and the signed proportional seat gap—and then relate standard disproportionality and counterfactual symmetry measures to the same framework.

6.1 First Switch Weight

Given a vote-share profile $\mathbf{p}$, recall that S^{Dist} denotes the seat total produced by first-past-the-post. We call $w_{S^{\text{Dist}}+1}(\mathbf{p})$ the *first switch weight*. It is the largest weight on statewide proportionality for which the first-past-the-post seat total S^{Dist} remains optimal:

$$w_{S^{\text{Dist}}+1}(\mathbf{p}) := \sup\{w \in [0,1] : S^{\text{Dist}} \in \arg\min_{S} M(S, \mathbf{p}; w)\}.$$

The first switch weight converts the district-statewide tradeoff into a robustness measure.[21] A high value means that the first-past-the-post allocation remains optimal even when the objective puts substantial weight on statewide proportionality. A low value means that the outcome depends on giving near-exclusive priority to district-level representation: even a small increase in the statewide component moves the optimum to a different seat total.

6.2 Efficiency Gap

The efficiency gap is widely used to diagnose partisan gerrymandering, and it is often invoked in debates about electoral representativeness more broadly. This prominence is empirically relevant in our data: Section 7.3 shows that maps struck down or litigated in court are concentrated much more strongly in the upper tail of the efficiency-gap distribution than in the corresponding tails of the alternative measures.

The efficiency gap is defined using the concept of *wasted votes*. Under first-past-the-post, all votes cast for the losing party in a district are wasted, because they do not elect a representative. For the winning party, votes above the one-half threshold are wasted; these votes are surplus to the minimum needed to win.[22] Let $W_A(\mathbf{p})$ and $W_B(\mathbf{p})$ denote the

[21] This value can be also be written as $w_{S^{\text{Dist}}+1}(\mathbf{p}) = \Delta\text{Dist}(S^{\text{Dist}}, \mathbf{p})/[\Delta\text{Dist}(S^{\text{Dist}}, \mathbf{p}) + \Delta\text{State}(S^{\text{Dist}}, \rho)]$, using the adjacent-seat comparison.

[22] To see the construction, suppose party A wins a district with vote share $1/2 + x$, so party B's vote share is $1/2 - x$. Then party B's wasted votes are all of its votes, $1/2 - x$, while party A's wasted votes are only its surplus above the winning threshold, x.

total wasted vote shares of parties A and B across N districts under the first-past-the-post allocation. The signed efficiency gap for party A is

$$EG_A(\mathbf{p}) := \frac{W_B(\mathbf{p}) - W_A(\mathbf{p})}{N}.$$

The rationale for the measure is that differences in wasted votes summarize the packing-and-cracking logic of gerrymandering: packing produces surplus votes in lopsided wins, whereas cracking produces votes cast in losing districts. Positive values mean that party B wastes more votes than party A, so party A is advantaged; the absolute value records the size of the imbalance.

Proposition 9. *At profile* $\mathbf{p}$*,*

$$|EG_A(\mathbf{p})| = \frac{1}{N}\left|S^{\text{Dist}} - \left(2\rho - \frac{N}{2}\right)\right|.$$

Hence the absolute efficiency gap is the statewide-misrepresentation term with linear cost $Q(z) = z$ *and target* $T^{EG}(\rho) = 2\rho - N/2$*.*

Proposition 9 shows that although the efficiency gap is defined from wasted votes computed district by district, its aggregate content is entirely captured by a statewide target. Once the first-past-the-post seat total S^{Dist} and statewide support ρ are fixed, the measure is simply the distance between S^{Dist} and the target $T^{EG}(\rho) = 2\rho - N/2$. This illustrates the flexibility of our framework: different measures of representation can often be embedded by changing the district or statewide targets while keeping the allocation problem fixed.

Replacing the proportional statewide term with the efficiency-gap target leaves the top-S and cutoff logic unchanged. What changes is the endpoint. As the statewide weight rises, the rule moves toward $T^{EG}(\rho)$, rather than toward proportionality.

Proportionality versus the Efficiency Gap. We define party A's *proportional seat gap* under first-past-the-post as

$$\mathrm{PG}_A(\mathbf{p}) := S^{\text{Dist}} - \rho.$$

Note that its absolute value is exactly statewide misrepresentation at the first-past-the-post allocation; $|\mathrm{PG}_A(\mathbf{p})| = \mathrm{State}(S^{\text{Dist}}, \rho)$. Thus, the proportional seat gap is simply the signed counterpart of our statewide-misrepresentation measure.

As the efficiency gap evaluates the same first-past-the-post seat total against the target $T^{EG}(\rho) = 2\rho - N/2 = \rho + (\rho - N/2)$, the proportional seat gap is its proportional-target analogue. The targets coincide only when $\rho = N/2$; otherwise, the efficiency-gap target shifts away from proportionality in favor of the statewide winner. Thus a zero efficiency gap generally incorporates a winner's bonus rather than proportional representation.[23] Two examples make the distinction between efficiency gap and our statewide measure transparent.

Example 1. Consider the ten-district profile

$$\mathbf{p} = (0.51, 0.51, 0.51, 0.21, 0.21, 0.21, 0.21, 0.21, 0.21, 0.21).$$

Here $\rho = 3$, so proportionality gives $S^{\text{State}} = 3$. First-past-the-post also gives $S^{\text{Dist}} = 3$. Hence first-past-the-post is exactly proportional in seat totals. But $EG_A = \frac{3}{10} - 2\frac{3}{10} + \frac{1}{2} = 0.2$, a large value relative to the empirical distribution of observed U.S. House maps.[24] The efficiency gap records an imbalance in wasted votes even though statewide misrepresentation is zero. In fact, the seat total that sets the efficiency gap to zero is $S_A = 1$.

Example 2. Consider a twelve-district profile with

$$\mathbf{p} = (0.75, 0.75, 0.75, 0.75, 0.75, 0.75, 0.75, 0.75, 0.75, 0.75, 0.75, 0.75).$$

Then $\rho = 9$, so proportionality gives $S^{\text{State}} = 9$. Under first-past-the-post, party A wins every district, so $S^{\text{Dist}} = 12$. Thus party A receives all seats with only 75% of the vote. Yet $EG_A = \frac{12}{12} - 2\frac{9}{12} + \frac{1}{2} = 0$. The efficiency gap classifies this outcome as unbiased because the winner's bonus exactly balances wasted votes, even though the seat allocation is far from proportional.

Separation of the two measures. The two examples show more than a difference in scale. The efficiency gap, proportional seat gap, and first-switch weight are not ordinally equivalent: the same election can be extreme under one benchmark and standard under another. These examples are not pathological cases. In Appendix F, we prove two general separation results. First, for every feasible seat total, there are profiles for which first-past-the-post and proportionality select that same seat total and for which the absolute efficiency gap is as large as possible among all profiles with that common seat total. Thus agreement between first-past-the-post and proportionality does not imply a small efficiency gap: when the statewide vote share is far from one half, an exactly proportional seat total can still create a large wasted-vote imbalance.

Second, efficiency-gap zero does not imply robustness of first-past-the-post to statewide representation concerns. There exist voting profiles with $EG_A = 0$ for which the interval of

[23] See Chambers et al. (2017) for an early criticism of the efficiency gap based on its winner's-bonus benchmark, and Stephanopoulos and McGhee (2018) for a reply defending its equal-wasted-votes principle and implied responsiveness.

[24] In the empirical distribution shown in Figure 9, an efficiency gap of 0.2 lies in the far right tail, well beyond the conventional 7% threshold.

statewide weights over which first-past-the-post remains optimal can be made arbitrarily small. The reason is that the efficiency gap depends only on statewide support and the first-past-the-post seat total, while robustness depends on the pivotal district that would switch first. A profile can balance wasted votes exactly and still award a nearly tied district to a party that is already overrepresented, so an arbitrarily small positive weight on statewide proportionality can overturn the first-past-the-post outcome.

6.3 Other Measures

Disproportionality indices. The statewide component in Section 2 nests standard two-party disproportionality indices as special normalizations of the proportional target. For example, the Loosemore and Hanby (1971) index is

$$\frac{1}{2}\left(\left|\frac{S}{N}-\frac{\rho}{N}\right|+\left|\left(1-\frac{S}{N}\right)-\left(1-\frac{\rho}{N}\right)\right|\right)=\left|\frac{S}{N}-\frac{\rho}{N}\right|.$$

The Gallagher (1991) least-squares index gives the same expression in the two-party case, up to the usual percentage convention. Thus both are proportional-target versions of our statewide loss, with $T(\rho)=\rho$ and $Q(z)=z/N$. If one works with the squared least-squares loss before taking the square root, the corresponding normalization is $Q(z)=(z/N)^2$.

Counterfactual symmetry measures. Partisan bias and seats–votes summaries are also nested in our framework, but they evaluate the statewide component at counterfactual profiles rather than at the observed profile. Partisan bias holds the map fixed and applies a uniform swing until statewide support is tied. For an observed profile $\mathbf{p}$, let $\mathbf{p}^{1/2}:=\mathbf{p}+\left(\frac{1}{2}-\bar{p}\right)\mathbf{1}$ denote this half-vote counterfactual, whenever it lies in $[0,1]^N$. Partisan bias asks whether first-past-the-post gives party A more or fewer than half the seats at $\mathbf{p}^{1/2}$ (Gelman and King, 1994; Grofman and King, 2007). Thus, using the already-defined first-past-the-post rule R_0,

$$PB_A(\mathbf{p}):=\frac{|R_0(\mathbf{p}^{1/2})|}{N}-\frac{1}{2}.$$

With the proportional statewide loss $\mathrm{State}(S,\rho)=|S-\rho|$, this gives

$$N|PB_A(\mathbf{p})|=\left||R_0(\mathbf{p}^{1/2})|-\frac{N}{2}\right|=\mathrm{State}\left(|R_0(\mathbf{p}^{1/2})|,\frac{N}{2}\right).$$

Appendix F gives the formal details. Seats–votes curves and partisan-Gini summaries extend the same idea by applying the statewide component over a family of counterfactual swing profiles $\{\mathbf{p}(t)\}$ and aggregating the resulting deviations.

7 Empirical Analysis

The theoretical analysis above characterizes the main trade-off between district-level and statewide representation. We now bring that framework to U.S. House elections. For each state and election year, we observe the district vote-share profile $\mathbf{p}$. We proceed in three steps. We first analyze first-switch weights in the 2024 election. We then examine where state-years governed by struck or litigated maps fall in the distributions of the three measures. Finally, we compare the measures across map-drawing authorities, including partisan legislatures, courts, and split-control or nonpartisan bodies.

7.1 Data and construction

We use U.S. House general-election returns from MIT Election Data and Science Lab (2024) at the candidate-district level, covering candidate names, party labels, and vote totals. We treat Republicans as Party A and Democrats as Party B and construct district-level two-party vote shares p_d for each state-year.

Our panel is the set of state-years with at least eight congressional districts over 2002–2024, the period for which the enacted maps are coded from Levitt (2026) and a same-map presidential reference is available.[25] Levitt (2026) also provides map level information for elections from 2002 onward. This includes the date of the map's passing, the time period it was active for, the authority that drew the final map, the political tilt of the map drawing authority and whether the map was superseded, active, struck down by courts or continues to be litigated as of May 2026.[26]

To align the returns with the two-party framework we apply a set of preprocessing rules. We (i) consolidate fusion candidates; (ii) remove invalid ballot entries; and (iii) flag elections in which either major party is unrepresented or the third-party vote is large, and proxy their vote shares using other elections in the same district. We implement this preprocessing separately for each district-year as follows.

If one candidate appears under multiple party lines and one of those is a major party (Democrat or Republican), we combine that candidate's votes across lines and assign the total to their major party. Votes for candidates belonging to neither major party are combined

[25]The eight-district restriction keeps attention on delegations large enough for the seat-conversion tradeoff to be nontrivial; the model assumes equal-population districts, a close approximation for U.S. House seats within a state, though turnout and eligibility differ across districts.

[26]We use "struck or litigated" as a descriptive marker of legal scrutiny. Such scrutiny is plausibly related to representational concerns, since maps that attract legal challenges are unlikely to be a random subset. However, the cases in our sample involve different legal claims, standards, remedies, and procedural postures, so the classification should not be interpreted as a common judicial finding of partisan unfairness or as definitive evidence of a map's representativeness.

and assigned to *third party*. This yields a candidate-level returns dataset.

We then rank candidates by votes and use a classification rule to decide whether the observed congressional returns can be used directly or must be proxied. The goal is to use observed congressional returns whenever they provide a credible measure of underlying two-party support, since these races preserve the district-specific information needed for the model. If the top two finishers are one Democrat and one Republican and the third-party share is at most 15%, the race is used as is: invalid ballots, write-ins, and third-party votes are discarded, and the Democratic and Republican votes are normalized to sum to one. Every other race is flagged for proxy, either because a major party is absent from the top two, or because the third-party share exceeds 15% and the raw two-party share is materially distorted.[27]

If a flagged race lacks one major-party candidate, we proxy the missing two-party vote share using contested elections in the same district under the same congressional map. Specifically, for the party that ran uncontested, we use its highest observed two-party vote share in a contested race from that district-map combination.[28] If no such congressional proxy is available, we use the nearest presidential election in the same district under the same enacted map.[29] Where neither proxy is available, we drop the state-year.

When the flagged race is unreliable as a measure of two party preferences due to third party candidates holding a large vote share, we use the presidential proxy as our primary one and the congressional district-map proxy when that fails.[30]

As a robustness check, we reverse the proxy order, using the presidential proxy first and the district-map proxy as the secondary source. The main empirical patterns are unchanged under this alternative specification.

7.2 First Switch Weight in the 2024 Election

We now apply the first-switch-weight measure introduced in Section 6.1 to the 2024 U.S. House elections. Figure 6 reports the first switch weight for each state in 2024. Each bar gives the largest weight on statewide proportionality for which the observed first-past-the-post seat

[27] 12% of the races in our sample require proxy for the election results.

[28] Uncontested races create a selection problem because the missing party's absence is itself informative about the district's partisan balance. Using the uncontested party's strongest contested performance within the same district-map combination preserves that information while anchoring the proxy to an observed two-party race in the same electoral geography. Because the same-map proxy is restricted to elections held under the same enacted district plan, it normally reaches no farther than one decennial redistricting cycle, and in most cases uses an adjacent election.

[29] Presidential votes by congressional districts were sourced from https://sites.google.com/view/presidentialbycongressionaldis/.

[30] Three state-years in which neither a presidential proxy nor a within-district contested share could be constructed are dropped. State-years in which an independent candidate won a seat are excluded entirely.

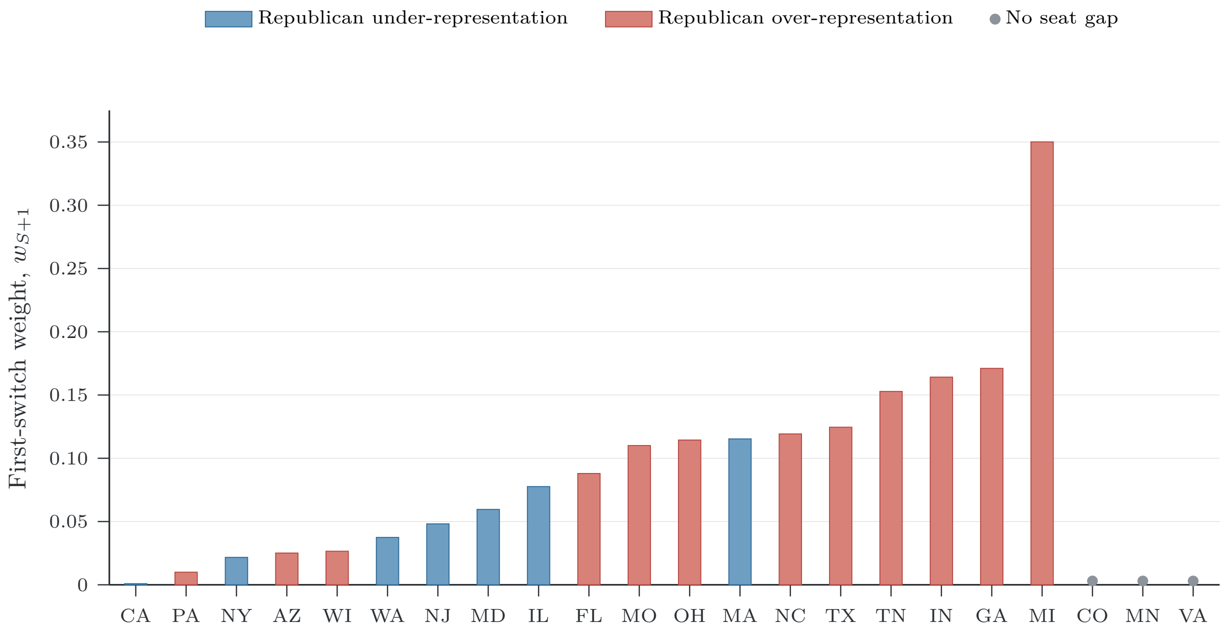


Figure 6: First-switch weights by state, 2024. Gray markers denote no seat gap.

total remains optimal. Low values indicate fragile first-past-the-post outcomes: only a small statewide component is needed to move the misrepresentation-minimizing rule to a different seat total.

The ordering also shows that robustness is distinct from the direction of partisan advantage. Low first switch weights appear in states with Republican under-representation, such as California and New York, and in states with Republican over-representation, such as Pennsylvania, Arizona, and Wisconsin. The common feature is the pivotal seat: it is the narrowest win of the already overrepresented party, and the closer that win is to the cutoff, the less statewide weight is needed to displace the first-past-the-post allocation. Appendix H.1 reports the corresponding proportional switch weights, at which the proportional seat total becomes optimal.

Texas and California. The contrast most relevant to current redistricting debate is Texas versus California, the two largest delegations and the focus of renewed attention to the interaction between maps and rules.[31] In 2024, Republicans are under-represented in

[31] In 2025 Texas adopted a new congressional map widely described as engineered to add several Republican-leaning seats; a three-judge panel initially enjoined it, but the Supreme Court stayed the injunction, leaving the map eligible for 2026 while litigation proceeds (Kruzel and Chung, 2025). California responded with Proposition 50, framed as a counter to Texas's mid-cycle redistricting, which the Supreme Court declined to block for 2026 (Chung, 2026). The 2024 cross-section in Figure 6 therefore predates both new maps and is the natural baseline against which to read their effects.

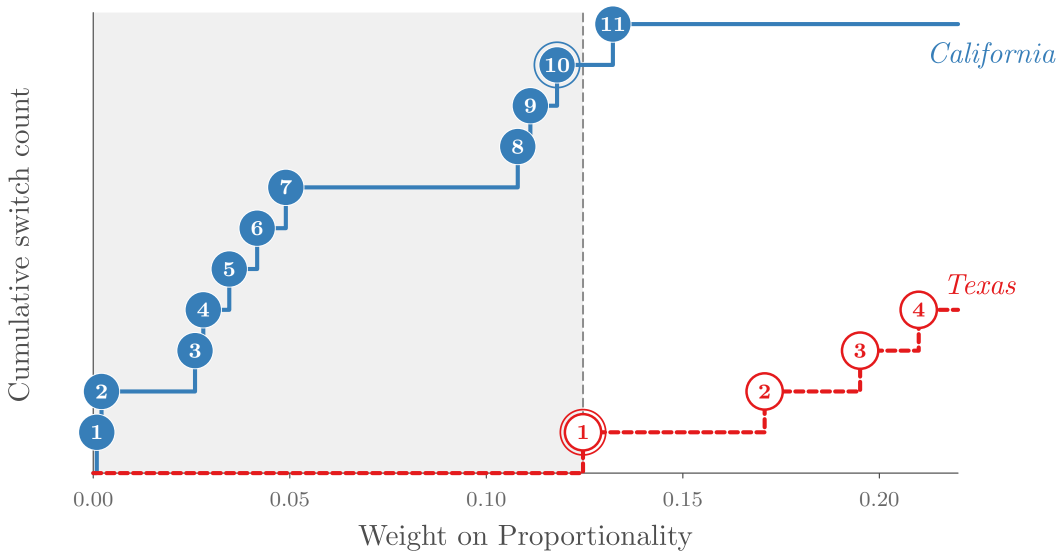


Figure 7: Cumulative seat switches in California and Texas in 2024 as the weight on proportionality increases. California reaches its tenth switch at $w_{\text{CA},10} = 0.1180$, before Texas reaches its first at $w_{\text{TX},1} = 0.1246$.

California and over-represented in Texas. California's first switch weight is only a fraction of Texas's, 0.0009 versus 0.1246, so the California seat allocation stops being optimal once the objective places even a very small weight on statewide proportionality. Figure 7 shows that this is not only a first-seat difference: California would switch ten seats to Republicans before Texas would switch its first seat to Democrats. The Texas allocation is therefore more robust over the relevant range, while California's seat total moves quickly once statewide proportionality enters the objective. This makes the Texas–California pair a natural baseline for tracking how Proposition 50 and the 2025 Texas map reshape the underlying vote profiles in 2026.

7.3 Comparing measures on enacted maps

We now compare the efficiency gap, the first switch weight, and the proportional seat gap on enacted maps. These measures summarize different features of the same vote-seat profile: the efficiency gap incorporates the winner's-bonus target described in Section 6; the proportional seat gap measures over- or under-representation relative to proportionality; and the first switch weight measures the robustness of the observed first-past-the-post outcome.

Appendix H.3 gives the detailed decomposition. The empirical question here is how these distinctions appear in observed maps, especially among maps that were struck down or litigated.

Figures 8 and 9 show that maps struck down or litigated are concentrated in the high-

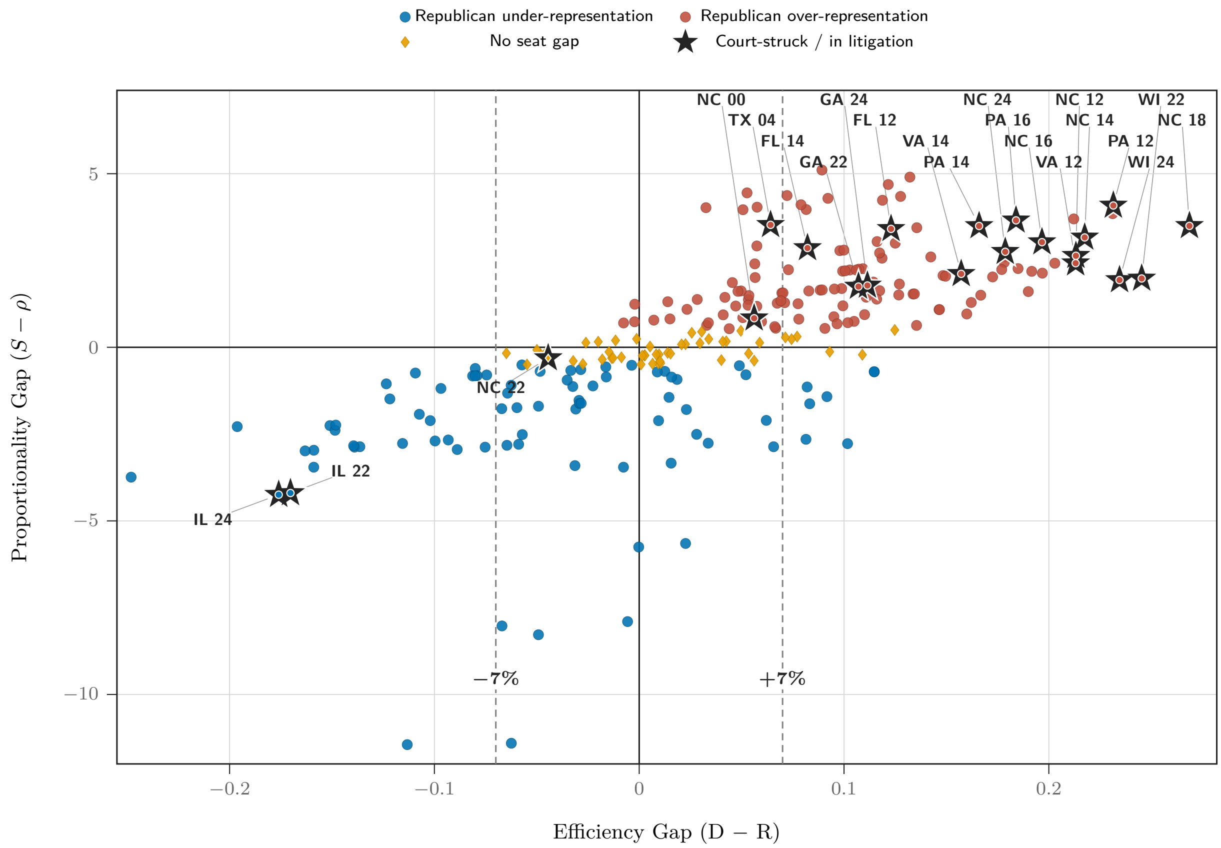


Figure 8: Proportional seat gap $S^{\text{Dist}} - \rho$ against the efficiency gap. Positive values indicate Republican over-representation, and negative values indicate Republican under-representation. Stars mark struck or litigated maps; dashed lines mark the conventional ±7% efficiency-gap thresholds.

efficiency-gap region. This pattern is notable because the efficiency gap is already a prominent measure in legal and public debates over partisan gerrymandering, and our analysis shows that this salience corresponds to a distinctive empirical pattern.

This association is much weaker for the proportional seat gap and the first switch weight, even though those measures capture highly intuitive representational concerns. Many maps with large proportional distortions or low first switch weights were not litigated, while some litigated maps are not especially extreme on those dimensions. Table 2 in Appendix H.3 reports the percentile rank of each struck or litigated map in our sample for all three measures. This contrast is clear in the percentile ranks of struck or litigated maps. Struck or litigated maps are disproportionately drawn from the extreme upper tail of the efficiency-gap distribution: twelve exceed the 90th percentile, compared with one for the first switch weight and three for the proportional seat gap; eight exceed the 95th percentile, compared with one for the first switch weight and none for the proportional seat gap.

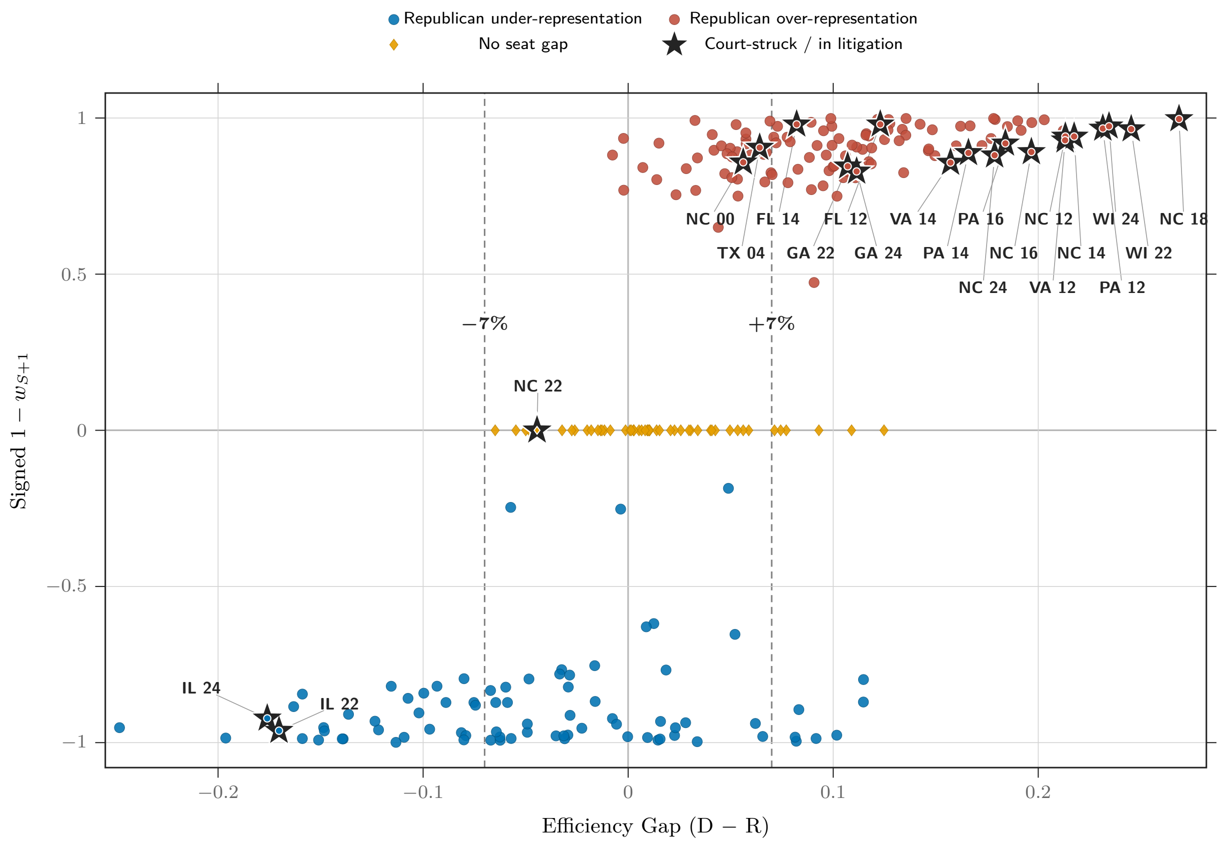


Figure 9: Signed inverse first-switch weight $1 - w_{S^{\text{Dist}}+1}$ against the efficiency gap. Positive values indicate Republican over-representation, negative values under-representation. Yellow points denote no proportional seat gap. Stars mark struck or litigated maps; dashed lines mark the conventional $\pm 7\%$ efficiency-gap thresholds.

This contrast is what makes the decomposition useful. The efficiency gap evaluates maps against a benchmark distinct from both proportionality and robustness, and that benchmark appears to have been particularly influential in litigation. Comparing it with the proportional seat gap and the first switch weight therefore clarifies what is being captured when maps are described as extreme.

7.4 Misrepresentation by map-drawing authority

The preceding subsection showed that the efficiency gap is more closely associated with struck or litigated maps than either the proportional seat gap or the first switch weight. We now ask whether the three measures also differ systematically by the institution that draws the map. This comparison helps clarify whether the efficiency gap, in particular, is more sensitive to some map-drawing contexts than the other measures.

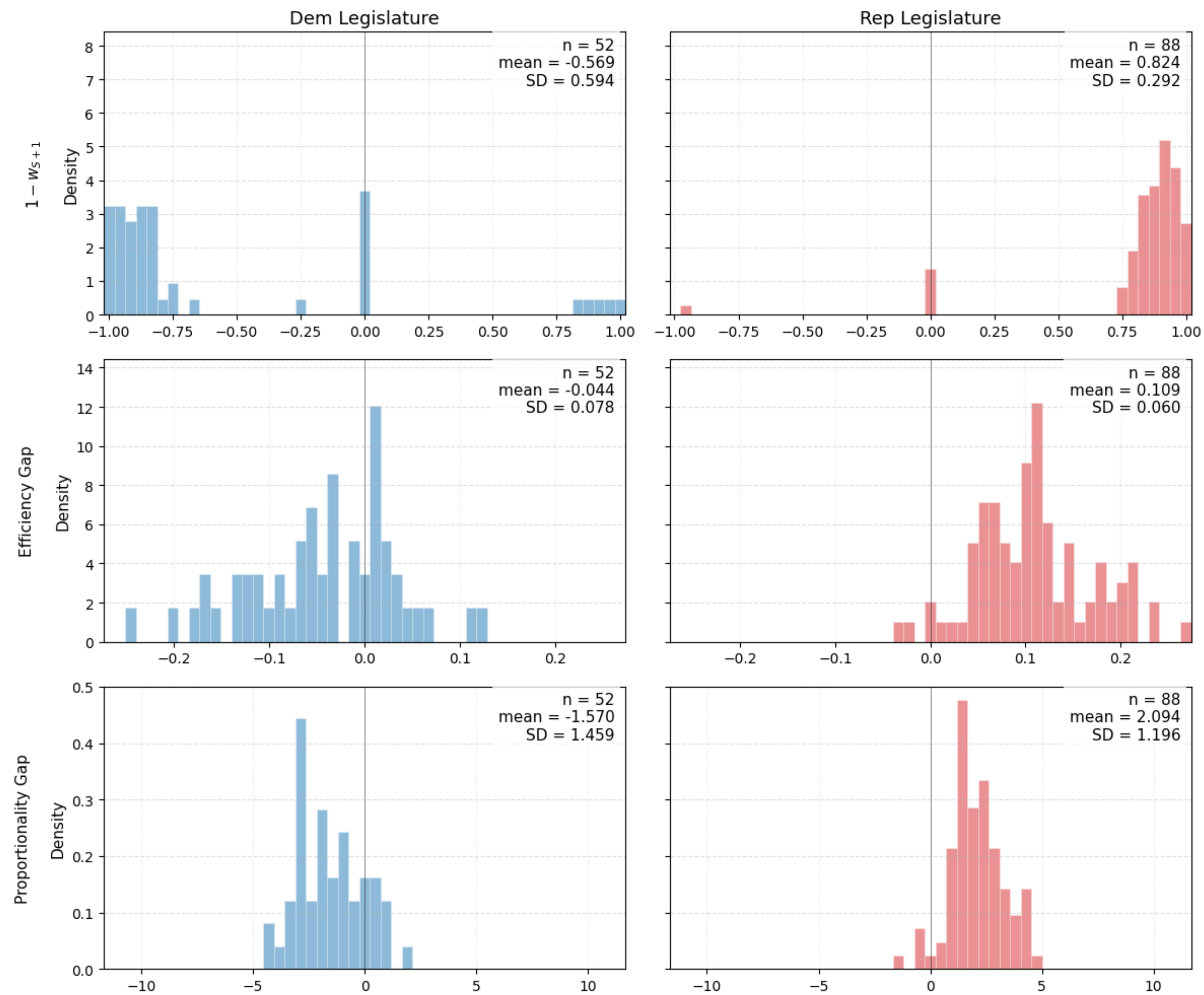


Figure 10: Distribution of inverse first switch weight $1 - w_{S^{\text{Dist}}+1}$, the efficiency gap, and the proportional seat gap $S^{\text{Dist}} - \rho$ by partisan legislatures. Positive values are Republican-favorable throughout. Group means and standard deviations are reported in each panel.

Figures 10 and 11 group state-years by the authority responsible for the congressional map: Democratic legislature, Republican legislature, split-control or nonpartisan authority, and court. For each group, the figures plot the distribution of three signed measures: inverse first switch weight, $1 - w_{S^{\text{Dist}}+1}$; the efficiency gap; and the proportional seat gap. Positive values are Republican-favorable throughout.[32]

Three patterns stand out. First, comparing Figures 10 and 11 shows that maps enacted by single-party legislatures are farther from zero than maps drawn under split control, by nonpartisan authorities, or by courts. Table 1 reports this comparison using standardized

[32]For the first measure, using $1 - w_{S^{\text{Dist}}+1}$ means that larger values correspond to less robust first-past-the-post outcomes, aligning its direction with the efficiency gap and proportional seat gap.

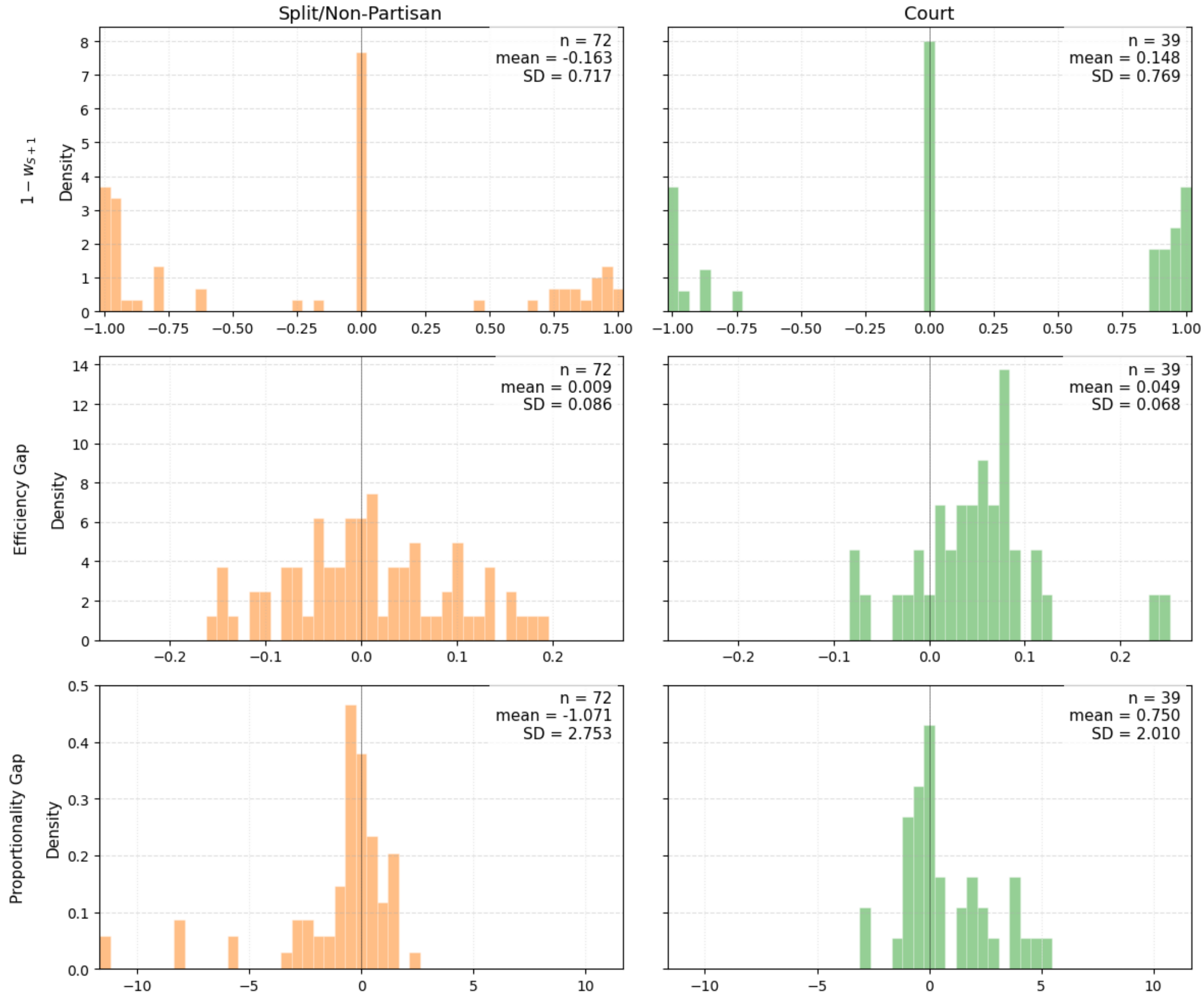


Figure 11: Distribution of the signed inverse switch weight $1 - w_{S^{\text{Dist}}+1}$, the efficiency gap and the proportional seat gap by split/Non-partisan legislatures and courts. Positive values are Republican-favorable on the three measures. Group means and standard deviations are reported in each panel.

means, defined as the group mean divided by the within-group standard deviation. On this scale, Democratic- and Republican-legislature maps have larger absolute values than split/nonpartisan and court-drawn maps. The main exception is court-drawn maps, whose efficiency-gap mean is slightly farther from zero than the Democratic-legislature mean. This exception is consistent with the third finding below. In this sample, the efficiency gap makes Republican-favorable deviations appear larger, and Democratic-favorable deviations smaller, relative to the proportional seat gap and the first switch weight.

Second, among maps drawn by partisan legislatures, Republican-legislature maps show larger average party-favorable deviations than Democratic-legislature maps. Using the

Map-drawing authority	Mean	SD	n	Mean/SD
Panel A: Inverse first switch weight				
Democratic legislature	−0.569	0.594	52	−0.958
Republican legislature	0.824	0.292	88	2.822
Split/nonpartisan	−0.163	0.717	72	−0.227
Court	0.148	0.769	39	0.192
Panel B: Efficiency gap				
Democratic legislature	−0.044	0.078	52	−0.564
Republican legislature	0.109	0.060	88	1.817
Split/nonpartisan	0.009	0.086	72	0.105
Court	0.049	0.068	39	0.721
Panel C: Proportional seat gap				
Democratic legislature	−1.570	1.459	52	−1.076
Republican legislature	2.094	1.196	88	1.751
Split/nonpartisan	−1.071	2.753	72	−0.389
Court	0.750	2.010	39	0.373

Table 1: Representation measures by map-drawing authority. Positive values are Republican-favorable; Mean/SD reports the group mean divided by the within-group standard deviation.

absolute value of the group means, the Republican-legislature mean is about 45% larger for inverse first switch weight, about 33% larger for the proportional seat gap, and about 148% larger for the efficiency gap.[33]

These ratios point to a third pattern: the efficiency gap amplifies the Republican–Democratic asymmetry among partisan legislatures. Under the proportional seat gap and first switch weight measures, Republican-legislature maps are more favorable to Republicans than Democratic-legislature maps are to Democrats, but the difference is moderate. Under the efficiency gap, which incorporates the winner's-bonus benchmark, the Republican-legislature mean is much farther from the Democratic-legislature mean. Thus the efficiency gap does not simply restate the same partisan-legislature contrast on another scale; it makes that contrast substantially larger.

This matters because the efficiency gap is both widely used and, as shown above, closely aligned with litigated and struck maps. Heavy reliance on it can therefore shape which maps appear most problematic in public debate and litigation. The main source of divergence is the winner's-bonus benchmark, a feature that is less transparent than the proportional seat gap or first switch weight. When these reasonable measures point in different directions,

[33]These comparisons use the ratio of absolute group means, $(|\bar{m}_R|/|\bar{m}_D| - 1) \times 100$, where $\bar{m}_R$ and $\bar{m}_D$ are the Republican- and Democratic-legislature means for the relevant measure. This statistic compares the average party-favorable shift across the two partisan-legislature groups.

the efficiency gap may tilt interpretation toward the patterns emphasized by its benchmark. Understanding when that tilt is substantively informative, and when it reflects the metric itself, is an important question for further research.

8 Conclusion

District-based elections ask a single rule to perform two tasks: represent each district and determine the partisan composition of the statewide delegation. This paper studies the conflict between these objectives while holding the district map and the induced vote profile fixed. Once a party's seat total is chosen, its seats should be assigned to the districts where its support is strongest. The allocation problem therefore reduces to deciding how far to move from the district-by-district outcome toward the statewide target.

In the linear benchmark, this decision has a direct interpretation. District-level misrepresentation counts voters represented locally by their nonpreferred party, while statewide misrepresentation measures the delegation's departure from proportionality. Their relative weight generates a single family of optimal rules. First-past-the-post is optimal when only district-level representation matters, proportional representation is optimal when only statewide representation matters, and increasing the statewide weight moves the allocation monotonically between them. The same movement also reduces the scope for an already overrepresented party to gain further seats through redistricting.

This perspective clarifies why familiar measures can disagree. The proportional seat gap and first-switch weight evaluate outcomes relative to proportionality, whereas the efficiency gap uses a benchmark that gives the statewide winner a seat bonus. In U.S. House elections from 2002 to 2024, this distinction is empirically important. Struck or litigated maps are especially concentrated among elections with large efficiency gaps. Maps enacted by single-party legislatures also tend to be less robust and less proportional than maps produced under split control, by nonpartisan authorities, or by courts. Republican-legislature maps show larger average party-favorable deviations than Democratic-legislature maps, and the efficiency gap makes this asymmetry considerably more pronounced.

These differences illustrate a central implication of our analysis: the representational consequences of a district map depend not only on how it distributes political support, but also on the rule used to translate district votes into seats. Most analyses focus on one of these margins, either characterizing electoral rules or studying redistricting under a fixed rule. Our approach separates them. The map determines the distribution of votes across districts; the allocation rule determines how those district votes become a statewide delegation. This makes it natural to analyze maps and electoral rules jointly, asking how the map shapes the operation of the rule and how the rule changes the representational consequences of the map.

References

Antweiler, W., 2019. Electoral economics: Maximizing local representation under proportionality. Economics Letters 182, 109–113.

Balinski, M., 2008. Fair majority voting (or how to eliminate gerrymandering). The American Mathematical Monthly 115, 97–113.

Balinski, M. L., Young, H. P., 2001. Fair Representation: Meeting the Ideal of One Man, One Vote. Brookings Institution Press, Washington, D.C., second ed.

Barton, J. T., Eguia, J. X., 2024. A decomposition of partisan advantage in electoral district maps. Electoral Studies 92, 102871.

Bouton, L., Genicot, G., Castanheira, M., Stashko, A. L., 2023. Pack-crack-pack: Gerrymandering with differential turnout. Tech. rep., National Bureau of Economic Research.

Carey, J. M., Hix, S., 2011. The electoral sweet spot: Low-magnitude proportional electoral systems. American Journal of Political Science 55, 383–397.

Chambers, C. P., Miller, A. D., Sobel, J., 2017. Flaws in the efficiency gap. JL & Pol. 33, 1.

Chen, J., Rodden, J., 2013. Unintentional gerrymandering: Political geography and electoral bias in legislatures. Quarterly Journal of Political Science 8, 239–269.

Chung, A., 2026. Us supreme court allows pro-democratic california voting map. Reuters, accessed 2026-02-08.

Coate, S., Knight, B., 2007. Socially optimal districting: A theoretical and empirical exploration. The Quarterly Journal of Economics 122, 1409–1471.

Cox, G. W., 1997. Making votes count: strategic coordination in the world's electoral systems. cambridge university Press.

Friedman, J. N., Holden, R. T., 2008. Optimal gerrymandering: sometimes pack, but never crack. American Economic Review 98, 113–144.

Gaffke, N., Pukelsheim, F., 2008. Divisor methods for proportional representation systems: An optimization approach to vector and matrix apportionment problems. Mathematical Social Sciences 56, 166–184.

Gallagher, M., 1991. Proportionality, disproportionality and electoral systems. Electoral Studies 10, 33–51.

Gaynor, S. W., Hayes, S. D., 2026. New lines, new districts, new representation: Institutional impacts of congressional redistricting. Political Research Quarterly 79, 370–386.

Gelman, A., King, G., 1994. A unified method of evaluating electoral systems and redistricting plans. American Journal of Political Science 38, 514–554.

Grofman, B., King, G., 2007. The future of partisan symmetry as a judicial test for partisan gerrymandering after lulac v. perry. Election Law Journal 6, 2–35.

Jeong, B.-h., 2024. The cost of proportional representations in electoral system design. Economic Theory Bulletin 12, 47–56.

King, G., Browning, R. X., 1987. Democratic representation and partisan bias in congressional elections. American Political Science Review 81, 1251–1273.

Kolotilin, A., Wolitzky, A., 2026. The economics of partisan gerrymandering. Econometrica 94, 71–103.

Kruzel, J., Chung, A., 2025. Supreme court revives pro-republican texas voting map sought by trump. Reuters, accessed 2026-02-08.

Levitt, J., 2026. All about redistricting. https://redistricting.lls.edu/, loyola Law School. Accessed: 2026-06-27.

Loosemore, J., Hanby, V. J., 1971. The theoretical limits of maximum distortion: Some analytic expressions for electoral systems. British Journal of Political Science 1, 467–477.

May, K. O., 1952. A set of independent necessary and sufficient conditions for simple majority decision. Econometrica 20, 680–684.

McGhee, E., 2014. Measuring partisan bias in single-member district electoral systems. Legislative Studies Quarterly 39, 55–85.

Milesi-Ferretti, G. M., Perotti, R., Rostagno, M., 2002. Electoral systems and public spending. The quarterly journal of economics 117, 609–657.

MIT Election Data and Science Lab, 2024. U.S. House 1976–2024.

Persson, T., Tabellini, G., 2005. The economic effects of constitutions. MIT press.

Stephanopoulos, N. O., McGhee, E. M., 2015. Partisan gerrymandering and the efficiency gap. University of Chicago Law Review 82, 831.

Stephanopoulos, N. O., McGhee, E. M., 2018. The measure of a metric: The debate over quantifying partisan gerrymandering. Stan. L. Rev. 70, 1503.

Warrington, G. S., 2018. Quantifying gerrymandering using the vote distribution. Election Law Journal: Rules, Politics, and Policy 17, 39–57.

Warrington, G. S., 2019. A comparison of partisan-gerrymandering measures. Election Law Journal: Rules, Politics, and Policy 18, 262–281.

A Proofs

This appendix collects the proofs of the main results. Throughout, the districts are ordered so that $p_1 \geq p_2 \geq \cdots \geq p_N$. Whenever a proof considers another profile, its districts are also ordered in decreasing party-A vote share.

Reduction to Seat Totals

Proof of Lemma 1. Fix any allocation $\mathbf{x}$ with $S(\mathbf{x}) = S$. If $\mathbf{x}$ is not a top-S allocation, then there exist districts $i < j$ such that $x_i = 0$, $x_j = 1$, and $p_i \geq p_j$. Let $\mathbf{x}'$ be the allocation obtained by swapping the party-A assignment from district j to district i: $x'_i = 1$, $x'_j = 0$, and $x'_d = x_d$ for all $d \notin \{i, j\}$. The seat total is unchanged but the change in district loss from this swap is

$$\begin{aligned}\mathrm{Dist}(\mathbf{x}', \mathbf{p}) - \mathrm{Dist}(\mathbf{x}, \mathbf{p}) &= \delta(1, p_i) + \delta(0, p_j) - \delta(0, p_i) - \delta(1, p_j) \\ &= \big[\delta(1, p_i) - \delta(0, p_i)\big] - \big[\delta(1, p_j) - \delta(0, p_j)\big].\end{aligned}$$

By single crossing, this expression is weakly negative. Thus the swap does not increase district loss. Repeating this exchange finitely many times removes all inversions and yields the top-S allocation with weakly lower district loss. □

Proof of Lemma 2. Fix an allocation $\mathbf{x}$ and let $S = S(\mathbf{x})$. The statewide term depends only on the seat total, so $\mathrm{State}(S(\mathbf{x}), \rho) = \mathrm{State}(S, \rho)$. By Lemma 1, the top-$S$ allocation weakly lowers district loss: $\mathrm{Dist}(S, \mathbf{p}) \leq \mathrm{Dist}(\mathbf{x}, \mathbf{p})$. Since F is increasing,

$$M(S, \mathbf{p}; w) \leq M(\mathbf{x}, \mathbf{p}; w).$$

Thus every allocation is weakly dominated, in objective value, by the top-S allocation with the same seat total. Since every top-S allocation is feasible, the allocation problem and the seat-total problem have the same value, and every optimal seat total has an optimal top-S implementation.

It remains to prove the range restriction. Suppose $S^{\text{Dist}} \leq S^{\text{State}}$. If $S < S^{\text{Dist}}$, then $\text{Dist}(S^{\text{Dist}}, \mathbf{p}) \leq \text{Dist}(S, \mathbf{p})$ because S^{Dist} is a district-loss minimizer. Also, since $S < S^{\text{Dist}} \leq S^{\text{State}}$, moving from S to S^{Dist} moves weakly toward the statewide optimum, so $\text{State}(S^{\text{Dist}}, \rho) \leq \text{State}(S, \rho)$. Monotonicity of F implies

$$M(S^{\text{Dist}}, \mathbf{p}; w) \leq M(S, \mathbf{p}; w).$$

Now suppose $S > S^{\text{State}}$. Since S^{State} minimizes statewide loss, $\text{State}(S^{\text{State}}, \rho) \leq \text{State}(S, \rho)$. Moreover, $S^{\text{State}} \geq S^{\text{Dist}}$, and district loss is weakly nondecreasing for seat totals weakly above S^{Dist}. Hence $\text{Dist}(S^{\text{State}}, \mathbf{p}) \leq \text{Dist}(S, \mathbf{p})$. Monotonicity of F implies

$$M(S^{\text{State}}, \mathbf{p}; w) \leq M(S, \mathbf{p}; w).$$

Therefore an optimal seat total is attained between S^{Dist} and S^{State}. □

The General Cutoff Rule and Supporting Weights

Before proving Theorems 1 and 2, we first provide a useful lemma.

Lemma 4. *Suppose $S^{\text{Dist}} < S^{\text{State}}$. For every $w \in (0,1)$ and every $S \in \{S^{\text{Dist}}, \ldots, S^{\text{State}} - 1\}$:*

(i) *The condition below holds:*

$$M(S+1, \mathbf{p}; w) \leq M(S, \mathbf{p}; w) \iff \kappa(S, \mathbf{p}; w) \leq \lambda^F(S, \mathbf{p}; w); \tag{1}$$

(ii) *$S \mapsto \kappa(S, \mathbf{p}; w)$ is weakly increasing;*

(iii) *If F has diminishing MRS, then $S \mapsto \lambda^F(S, \mathbf{p}; w)$ is weakly decreasing. Hence the first order conditions stated in (1) are single crossing in S, and the minimizing seat totals in $\{S^{\text{Dist}}, \ldots, S^{\text{State}}\}$ form an interval of consecutive integers.*

(iv) *If F has diminishing MRS, then for each fixed $S \in \{S^{\text{Dist}}, \ldots, S^{\text{State}} - 1\}$, the first order condition is single crossing in w. That is, for any $0 \leq w \leq w' \leq 1$,*

$$M(S+1, \mathbf{p}; w) - M(S, \mathbf{p}; w) \leq 0 \implies M(S+1, \mathbf{p}; w') - M(S, \mathbf{p}; w') \leq 0.$$

Moreover, if both weights are interior and the move strictly raises district loss, then the move becomes strictly improving at every higher weight.

Proof of Lemma 4. Part (i). By construction, $z_S(0) = \big((1-w)\mathrm{Dist}(S, \mathbf{p}), w\mathrm{State}(S, \rho)\big)$, and $z_S(1) = \big((1-w)\mathrm{Dist}(S+1, \mathbf{p}), w\mathrm{State}(S+1, \rho)\big)$. The derivative of the affine path is $z_S'(\nu) = \big((1-w)\Delta\mathrm{Dist}(S, \mathbf{p}), -w\Delta\mathrm{State}(S, \rho)\big)$. Integrating the derivative of $F(z_S(\nu))$ along this path gives

$$\begin{aligned} M(S+1, \mathbf{p}; w) - M(S, \mathbf{p}; w) &= F(z_S(1)) - F(z_S(0)) \\ &= \int_0^1 \left[F_1(z_S(\nu))(1-w)\Delta\mathrm{Dist}(S, \mathbf{p}) - F_2(z_S(\nu))w\Delta\mathrm{State}(S, \rho)\right] d\nu \\ &= w\Delta\mathrm{State}(S, \rho)\left[\kappa(S, \mathbf{p}; w)\int_0^1 F_1(z_S(\nu))\, d\nu - \int_0^1 F_2(z_S(\nu))\, d\nu\right]. \end{aligned}$$

Since $w\Delta\mathrm{State}(S, \rho) > 0$ and $F_1 > 0$, $M(S+1, \mathbf{p}; w) \le M(S, \mathbf{p}; w)$ if and only if

$$\kappa(S, \mathbf{p}; w) \le \frac{\int_0^1 F_2(z_S(\nu))\, d\nu}{\int_0^1 F_1(z_S(\nu))\, d\nu} = \lambda^F(S, \mathbf{p}; w).$$

This proves (i).

Part (ii). On the stated range, $S \mapsto \Delta\mathrm{Dist}(S, \mathbf{p})$ is weakly increasing and nonnegative, while $S \mapsto \Delta\mathrm{State}(S, \rho)$ is positive and weakly decreasing. Therefore $S \mapsto \kappa(S, \mathbf{p}; w) = \frac{\Delta\mathrm{Dist}(S,\mathbf{p})}{\phi\Delta\mathrm{State}(S,\rho)}$ is weakly increasing. This proves (ii).

Part (iii). Now suppose F has diminishing MRS. Fix $S \in \{S^{\mathrm{Dist}}, \dots, S^{\mathrm{State}} - 2\}$ and take any $\nu, \mu \in [0, 1]$. Every point on the segment starting from $S+1$ has weakly larger district loss term and weakly smaller statewide loss term than every point on the segment starting from S. Diminishing MRS therefore implies $\frac{F_2(z_{S+1}(\mu))}{F_1(z_{S+1}(\mu))} \le \frac{F_2(z_S(\nu))}{F_1(z_S(\nu))}$.

Writing,

$$\lambda^F(S, \mathbf{p}; w) = \frac{\int_0^1 \frac{F_2(z_S(\nu))}{F_1(z_S(\nu))} F_1(z_S(\nu))\, d\nu}{\int_0^1 F_1(z_S(\nu))\, d\nu},$$

it follows that $\lambda^F(S, \mathbf{p}; w)$ is an F_1-weighted average of the local MRS values on the segment from S to $S+1$, since $F_1 > 0$. The pointwise comparison across adjacent segments, $\frac{F_2(z_{S+1}(\mu))}{F_1(z_{S+1}(\mu))} \le \frac{F_2(z_S(\nu))}{F_1(z_S(\nu))}$, therefore gives $\lambda^F(S+1, \mathbf{p}; w) \le \lambda^F(S, \mathbf{p}; w)$.

Combining this with (ii), we conclude that $S \mapsto \kappa(S, \mathbf{p}; w) - \lambda^F(S, \mathbf{p}; w)$ is weakly increasing. By the first order condition (1), the adjacent move $S \to S+1$ is weakly improving when this difference is weakly negative. Hence once an adjacent move fails to be weakly improving, every later adjacent move also fails to be weakly improving. The reduced objective therefore decreases, possibly has a flat segment, and then increases. Its minimizers form an interval of consecutive integers. This proves (iii).

Part (iv). Fix $S \in \{S^{\text{Dist}}, \dots, S^{\text{State}} - 1\}$. On this range, $\text{Dist}(S+1, \mathbf{p}) \geq \text{Dist}(S, \mathbf{p})$, and $\text{State}(S+1, \rho) < \text{State}(S, \rho)$.

If $\text{Dist}(S+1, \mathbf{p}) = \text{Dist}(S, \mathbf{p})$, then for every $w > 0$, strict monotonicity of F in its second argument gives $M(S+1, \mathbf{p}; w) < M(S, \mathbf{p}; w)$, while at $w = 0$ the two values are equal. Hence the first order condition is single crossing in w in this case.

Now suppose $\text{Dist}(S+1, \mathbf{p}) > \text{Dist}(S, \mathbf{p})$. Take $0 < w < w' < 1$ and assume

$$M(S+1, \mathbf{p}; w) \leq M(S, \mathbf{p}; w).$$

For $y_1 \in [\text{Dist}(S, \mathbf{p}), \text{Dist}(S+1, \mathbf{p})]$, define the level set $y_2(y_1)$ by

$$F((1-w)y_1, wy_2(y_1)) = F\big((1-w)\text{Dist}(S, \mathbf{p}), w\text{State}(S, \rho)\big). \tag{2}$$

We first show that $y_2(y_1)$ is well defined for each such y_1. At $y_2 = \text{State}(S+1, \rho)$, the left-hand side is weakly below the right-hand side of the defining equation: this follows from $y_1 \leq \text{Dist}(S+1, \mathbf{p})$, monotonicity of F in its first argument, and the premise $M(S+1, \mathbf{p}; w) \leq M(S, \mathbf{p}; w)$. At $y_2 = \text{State}(S, \rho)$, the left-hand side is weakly above the right-hand side because $y_1 \geq \text{Dist}(S, \mathbf{p})$. Continuity and strict monotonicity of F in its second argument therefore give a unique solution $y_2(y_1) \in [\text{State}(S+1, \rho), \text{State}(S, \rho)]$.

Differentiating the level-set equation (2) gives

$$y_2'(y_1) = -\frac{1-w}{w} \frac{F_1((1-w)y_1, wy_2(y_1))}{F_2((1-w)y_1, wy_2(y_1))}.$$

Now evaluate the same level set at the higher weight w'. Its derivative with respect to y_1 is

$$\begin{aligned}
&\frac{d}{dy_1} F((1-w')y_1, w'y_2(y_1)) \\
&= F_1((1-w')y_1, w'y_2(y_1)) \\
&\quad \times \left[(1-w') - w' \frac{1-w}{w} \frac{F_1((1-w)y_1, wy_2(y_1))}{F_2((1-w)y_1, wy_2(y_1))} \frac{F_2((1-w')y_1, w'y_2(y_1))}{F_1((1-w')y_1, w'y_2(y_1))}\right].
\end{aligned}$$

Because $w' > w$, moving from weight w to weight w' lowers the scaled district loss coordinate from $(1-w)y_1$ to $(1-w')y_1$ and raises the scaled statewide loss coordinate from $wy_2(y_1)$ to

$w'y_2(y_1)$. By diminishing MRS, this change weakly raises the local ratio F_2/F_1. Therefore

$$\frac{F_2((1-w')y_1, w'y_2(y_1))}{F_1((1-w')y_1, w'y_2(y_1))} \geq \frac{F_2((1-w)y_1, wy_2(y_1))}{F_1((1-w)y_1, wy_2(y_1))}.$$

Together with $\frac{w'}{1-w'} > \frac{w}{1-w}$, this makes the bracketed term strictly negative, and we conclude that $\frac{d}{dy_1}F((1-w')y_1, w'y_2(y_1)) < 0$. This implies that

$$F((1-w')\text{Dist}(S+1,\mathbf{p}), w'y_2(\text{Dist}(S+1,\mathbf{p}))) < F((1-w')\text{Dist}(S,\mathbf{p}), w'y_2(\text{Dist}(S,\mathbf{p}))).$$

By definition in (2), $y_2(\text{Dist}(S,\mathbf{p})) = \text{State}(S,\rho)$. So, the right-hand side is $M(S,\mathbf{p};w')$. Since $y_2(\text{Dist}(S+1,\mathbf{p})) \geq \text{State}(S+1,\rho)$, monotonicity of F gives

$$M(S+1,\mathbf{p};w') \leq F((1-w')\text{Dist}(S+1,\mathbf{p}), w'y_2(\text{Dist}(S+1,\mathbf{p}))) < M(S,\mathbf{p};w').$$

Thus the move from S to $S+1$ is strictly improving at w'.

We finally consider the endpoints. The endpoint $w'=1$ follows directly from

$$M(S+1,\mathbf{p};1) = F(0,\text{State}(S+1,\rho)) < F(0,\text{State}(S,\rho)) = M(S,\mathbf{p};1).$$

Finally, if $w=0$ and $\text{Dist}(S+1,\mathbf{p}) > \text{Dist}(S,\mathbf{p})$, then $M(S+1,\mathbf{p};0) > M(S,\mathbf{p};0)$, so the premise of the implication cannot hold. This proves (iv). □

Proof of Theorem 1. Take $w \in (0,1)$ and $S \in \{S^{\text{Dist}}, \ldots, S^{\text{State}}-1\}$. By Lemma 4,

$$M(S+1,\mathbf{p};w) \leq M(S,\mathbf{p};w) \iff \kappa(S,\mathbf{p};w) \leq \lambda^F(S,\mathbf{p};w).$$

Substituting the definition of $\kappa(S,\mathbf{p};w)$, the RHS becomes $\Delta\text{Dist}(S,\mathbf{p}) \leq \phi \cdot \Delta\text{State}(S,\rho) \cdot \lambda^F(S,\mathbf{p};w)$. Also substituting $\Delta\text{Dist}(S,\mathbf{p}) = \Delta(p_{S+1})$, gives the condition

$$M(S+1,\mathbf{p};w) \leq M(S,\mathbf{p};w) \iff \Delta(p_{S+1}) \leq \phi \cdot \Delta\text{State}(S,\rho) \cdot \lambda^F(S,\mathbf{p};w).$$

If $p \mapsto \Delta(p)$ is strictly decreasing, the RHS is equivalent to

$$p_{S+1} \geq \Delta^{-1}\left(\phi \cdot \Delta\text{State}(S,\rho) \cdot \lambda^F(S,\mathbf{p};w)\right) = \tau(S,\mathbf{p};w).$$

This proves the adjacent cutoff condition.

If the district switch margin Δ is only weakly decreasing, the same argument uses the generalized inverse $\Delta_g^{-1}(c) := \inf\{p \in [0,1] : \Delta(p) \leq c\}$, together with the maintained tie-breaking convention. □

Proof of Theorem 2. If $S^{\text{Dist}} = S^{\text{State}}$, Lemma 2 implies that the top-S^{Dist} allocation is optimal. Hence suppose $S^{\text{Dist}} < S^{\text{State}}$.

By Lemma 4, the first order conditions are single crossing in S, and the minimizing seat totals in $\{S^{\text{Dist}}, \dots, S^{\text{State}}\}$ form an interval of consecutive integers. Let S^* be the largest minimizing seat total in this interval. By Lemma 2, the top-S^* allocation is an optimal implementation.

Since S^* is the largest minimizer and the adjacent signs are single crossing, every adjacent move before S^* is weakly improving: $M(S+1, \mathbf{p}; w) \le M(S, \mathbf{p}; w)$ for all $S = S^{\text{Dist}}, \dots, S^*-1$. By Theorem 1, this is equivalent to $p_{S+1} \ge \tau(S, \mathbf{p}; w)$ for all $S = S^{\text{Dist}}, \dots, S^* - 1$.

Again by the single crossing property every adjacent move after S^* is not strictly improving: $M(S+1, \mathbf{p}; w) > M(S, \mathbf{p}; w)$ for all $S = S^*, \dots, S^{\text{State}}-1$ and by Theorem 1, this is equivalent to the failure of the first order condition: $p_{S+1} < \tau(S, \mathbf{p}; w)$ for all $S = S^*, \dots, S^{\text{State}} - 1$.

With a weakly decreasing district switch margin, the same argument applies using the generalized inverse and the maintained tie-breaking convention.

□

Proof of Proposition 1. By Lemma 2, for every $w \in [0,1]$ at least one minimizer lies in $\{S^{\text{Dist}}, \dots, S^{\text{State}}\}$. Hence it is enough to characterize optimality on this range. If $S^{\text{Dist}} = S^{\text{State}}$, then S^{Dist} is optimal for every $w \in [0,1]$, so $\mathcal{W}_{S^{\text{Dist}}}(\mathbf{p}) = [0,1]$ and the result is immediate.

Suppose therefore that $S^{\text{Dist}} < S^{\text{State}}$. For each $S \in \{S^{\text{Dist}}, \dots, S^{\text{State}} - 1\}$, write

$$\Gamma_S(w) := M(S+1, \mathbf{p}; w) - M(S, \mathbf{p}; w).$$

Thus the adjacent move from S to $S+1$ is weakly improving when $\Gamma_S(w) \le 0$. Define

$$\mathcal{U}_S := \{w \in [0,1] : \Gamma_S(w) \le 0\}, \qquad \mathcal{L}_S := \{w \in [0,1] : \Gamma_S(w) \ge 0\}.$$

By Lemma 4(iv), the first order condition is single crossing in w. Therefore $\mathcal{U}_S$ is an upper interval and $\mathcal{L}_S$ is a lower interval. Continuity of $w \mapsto M(\cdot, \cdot; w)$ implies that $\mathcal{U}_S$ and $\mathcal{L}_S$ are closed subsets of $[0,1]$.

Moreover, $\Gamma_S(0) \ge 0$, because at $w = 0$ only district loss matters and district loss is weakly increasing on the stated range. Also, $\Gamma_S(1) < 0$, because at $w = 1$ only statewide loss matters and $\text{State}(S+1, \rho) < \text{State}(S, \rho)$. Hence there exist $0 \le \ell_{S+1} \le u_S \le 1$ such that $\mathcal{U}_S = [\ell_{S+1}, 1]$, and $\mathcal{L}_S = [0, u_S]$.

Next we convert adjacent signs into support sets. By Lemma 4(iii), for each fixed w, the first order conditions are single crossing in S. Therefore a seat total is optimal when all adjacent moves below it are weakly improving and all adjacent moves above it are weakly non-improving. Therefore, by definition $\mathcal{W}_{S^{\text{Dist}}}(\mathbf{p}) = \mathcal{L}_{S^{\text{Dist}}}$, and $\mathcal{W}_{S^{\text{State}}}(\mathbf{p}) = \mathcal{U}_{S^{\text{State}}-1}$, and, for every $S \in \{S^{\text{Dist}}+1, \ldots, S^{\text{State}}-1\}$, $\mathcal{W}_S(\mathbf{p}) = \mathcal{U}_{S-1} \cap \mathcal{L}_S$.

Using the interval forms above, this can be stated as

$$\begin{aligned}
\mathcal{W}_{S^{\text{Dist}}}(\mathbf{p}) &= [0, u_{S^{\text{Dist}}}], \\
\mathcal{W}_{S^{\text{State}}}(\mathbf{p}) &= [\ell_{S^{\text{State}}}, 1], \\
\mathcal{W}_S(\mathbf{p}) &= [\ell_S, u_S] \quad \text{for } S \in \{S^{\text{Dist}}+1, \ldots, S^{\text{State}}-1\}.
\end{aligned}$$

It remains to show that the interior intervals are nonempty. Suppose, toward a contradiction, that $\ell_S > u_S$ for some $S \in \{S^{\text{Dist}}+1, \ldots, S^{\text{State}}-1\}$. Choose any $w \in (u_S, \ell_S)$. Then $w \notin \mathcal{U}_{S-1}$, so $\Gamma_{S-1}(w) > 0$, and $w \notin \mathcal{L}_S$, so $\Gamma_S(w) < 0$. This positive-then-negative pattern is ruled out by single crossing in S. Hence $\ell_S \le u_S$, and every support set is a nonempty closed interval.

Finally, single crossing in S orders the adjacent comparison sets, that is, for every $S \in \{S^{\text{Dist}}, \ldots, S^{\text{State}}-2\}$, $\mathcal{U}_{S+1} \subseteq \mathcal{U}_S$, and $\mathcal{L}_S \subseteq \mathcal{L}_{S+1}$. If $w \in \mathcal{U}_{S+1}$, then $\Gamma_{S+1}(w) \le 0$, so single crossing gives $\Gamma_S(w) \le 0$, and therefore $w \in \mathcal{U}_S$. Similarly, if $w \in \mathcal{L}_S$, then $\Gamma_S(w) \ge 0$, so single crossing gives $\Gamma_{S+1}(w) \ge 0$, and therefore $w \in \mathcal{L}_{S+1}$. Since $\mathcal{U}_S = [\ell_{S+1}, 1]$, $\mathcal{U}_{S+1} = [\ell_{S+2}, 1]$, $\mathcal{L}_S = [0, u_S]$, and $\mathcal{L}_{S+1} = [0, u_{S+1}]$, these inclusions imply $\ell_{S+1} \le \ell_{S+2}$ and $u_S \le u_{S+1}$.

Writing $\mathcal{W}_S(\mathbf{p}) = [\ell_S, u_S]$, the formulas above give $\ell_{S^{\text{Dist}}} = 0$ and $u_{S^{\text{State}}} = 1$. Therefore

$$0 = \ell_{S^{\text{Dist}}} \le \ell_{S^{\text{Dist}}+1} \le \cdots \le \ell_{S^{\text{State}}}, \quad \text{and} \quad u_{S^{\text{Dist}}} \le u_{S^{\text{Dist}}+1} \le \cdots \le u_{S^{\text{State}}} = 1.$$

This proves the claimed ordering of the support intervals. $\square$

Linear Benchmark

Proof of Proposition 2. Fix a profile $\mathbf{p}$. First, note that only top-S allocations can be Pareto efficient or minimize the linear objective. If an allocation with seat total S assigns party A to district j but not to district i, with $p_i > p_j$, then swapping the two assignments changes district loss by

$$\big[(1-p_i) + p_j\big] - \big[p_i + (1-p_j)\big] = 2(p_j - p_i) < 0.$$

where the strict inequality holds as $p_i \ne p_j$ for $i \ne j$. Repeating this swap eliminates every inversion. Thus, for each seat total S, the top-S allocation is the unique district-loss minimizer among allocations with that seat total. Since statewide loss is constant within a seat-total class, if $\mathbf{x}$ is not a top-$S(\mathbf{x})$ allocation, it cannot be Pareto efficient or minimize the linear objective for any $w < 1$.

Write $m_w(S) := (1-w)\mathrm{Dist}(S, \mathbf{p}) + w|S - \rho|$. Suppose first that $\widehat{\mathbf{x}}$ minimizes the linear objective for some $w \in [0,1)$. If $w = 0$, then $\widehat{\mathbf{x}}$ is the unique district-loss minimizer because there are no district-level ties. Hence no other allocation can weakly lower district loss and strictly improve at least one coordinate. If $w \in (0,1)$, any allocation that weakly lowered both Dist and State, with at least one strict reduction, would strictly lower the weighted objective, contradicting optimality. Thus $\widehat{\mathbf{x}}$ is Pareto efficient.

Conversely, suppose $\widehat{\mathbf{x}}$ is Pareto efficient. By the first paragraph, $\widehat{\mathbf{x}}$ is the top-S^* allocation for $S^* = S(\widehat{\mathbf{x}})$. By the same dominance argument used in Lemma 2, S^* lies between S^{Dist} and S^{State} in the maintained case $S^{\mathrm{Dist}} \leq S^{\mathrm{State}}$.

If $S^{\mathrm{Dist}} = S^{\mathrm{State}}$, this common seat total minimizes both components, so the top-S^* allocation minimizes the linear objective for every $w \in [0,1)$. Now suppose $S^{\mathrm{Dist}} < S^{\mathrm{State}}$. For $S = S^{\mathrm{Dist}}, \ldots, S^{\mathrm{State}} - 1$, define

$$w_{S+1} := \frac{\Delta\mathrm{Dist}(S, \mathbf{p})}{\Delta\mathrm{Dist}(S, \mathbf{p}) + \Delta\mathrm{State}(S, \rho)}, \qquad w_{S^{\mathrm{Dist}}} := 0.$$

Along this path, $\Delta\mathrm{Dist}(S, \mathbf{p}) \geq 0$ is weakly increasing and $\Delta\mathrm{State}(S, \rho) > 0$ is weakly decreasing, so these critical weights are ordered and all are strictly below one. Moreover,

$$m_w(S+1) - m_w(S) = (1-w)\Delta\mathrm{Dist}(S, \mathbf{p}) - w\Delta\mathrm{State}(S, \rho),$$

which is weakly negative if and only if $w \geq w_{S+1}$.

If $S^* < S^{\mathrm{State}}$, choose any $w \in [w_{S^*}, w_{S^*+1}]$. The ordered critical weights imply that $m_w(S)$ weakly decreases up to S^* and weakly increases after S^*, so S^* minimizes m_w. Since $w_{S^*+1} < 1$, this supporting weight satisfies $w < 1$. If $S^* = S^{\mathrm{State}}$, choose any $w \in (w_{S^{\mathrm{State}}}, 1)$. Then all adjacent moves up to S^{State} weakly lower the objective, and Lemma 2 rules out minimizers beyond S^{State}. Hence S^* minimizes m_w for some $w < 1$. Therefore $\widehat{\mathbf{x}}$ minimizes the original allocation problem at a weight $w \in [0,1)$. $\square$

Proof of Proposition 3. The local part of aggregate voter loss is

$$\begin{aligned}(1-w)\sum_{d=1}^{N}\big[p_d(1-x_d) + (1-p_d)x_d\big] &= (1-w)\sum_{d=1}^{N}\big(x_d(1-p_d) + (1-x_d)p_d\big) \\ &= (1-w)\mathrm{Dist}(\mathbf{x}, \mathbf{p}).\end{aligned}$$

For the statewide component, use $\sum_d p_d = \rho$ and $\sum_d (1-p_d) = N - \rho$. Suppressing the argument $\mathbf{x}$ in $S(\mathbf{x})$,

$$w \sum_{d=1}^{N} \left[p_d \frac{(\rho - S)_+}{\rho} + (1 - p_d) \frac{(S - \rho)_+}{N - \rho} \right] = w\left[(\rho - S)_+ + (S - \rho)_+\right] = w|S - \rho|.$$

Therefore $L(\mathbf{x}, \mathbf{p}; w) = (1 - w)\mathrm{Dist}(\mathbf{x}, \mathbf{p}) + w|S(\mathbf{x}) - \rho|$. Minimizing aggregate voter loss is equivalent to minimizing $(1 - w)\mathrm{Dist}(\mathbf{x}, \mathbf{p}) + w|S(\mathbf{x}) - \rho|$. $\square$

Proof of Theorem 3. For finite ϕ, minimizing the linear objective is equivalent to minimizing $\mathrm{Dist}(S, \mathbf{p}) + \phi|S - \rho|$. For any adjacent move $S \to S + 1$,

$$\mathrm{Dist}(S+1, \mathbf{p}) + \phi|\rho - (S+1)| - \big[\mathrm{Dist}(S, \mathbf{p}) + \phi|\rho - S|\big] = 1 - 2p_{S+1} + \phi\big(|\rho - (S+1)| - |\rho - S|\big).$$

The aggregate term can also be stated as

$$|\rho - (S+1)| - |\rho - S| = \begin{cases} -1, & S \le \underline{\rho} - 1, \\ 1 - 2(\rho - \underline{\rho}), & S = \underline{\rho}, \\ 1, & S \ge \underline{\rho} + 1. \end{cases}$$

Since p_{S+1} is weakly decreasing in S, these adjacent differences are weakly increasing in S. Thus the reduced objective decreases as long as the adjacent difference is weakly negative and increases once the adjacent difference is weakly positive. Hence an optimal seat total is located at the first switch from weakly negative to weakly positive adjacent differences.

First suppose $0 \le \phi < \phi_{\underline{\rho}}$. By the definition of $\phi_{\underline{\rho}}$ this range is nonempty only when $S^{\mathrm{Dist}} < \underline{\rho}$, in which case $\phi_{\underline{\rho}} = 1 - 2p_{\underline{\rho}}$. Thus $\frac{1-\phi}{2} > p_{\underline{\rho}}$. The cutoff $t(\phi) = (1 - \phi)/2$ therefore implements some top-S allocation with $S \le \underline{\rho} - 1$.

For every adjacent move below $\underline{\rho}$,

$$1 - 2p_{S+1} - \phi = 2\left(\frac{1-\phi}{2} - p_{S+1}\right).$$

and the adjacent move is weakly improving for districts whose vote share clears the cutoff $t(\phi)$. By the single crossing of adjacent differences, the top-S allocation induced by this cutoff is optimal. This proves part (i).

Now suppose $\rho - \underline{\rho} \le 1/2$. We show that $\underline{\rho}$ is optimal for every $\phi \ge \phi_{\underline{\rho}}$. For every $S < \underline{\rho}$, the adjacent difference is $1 - 2p_{S+1} - \phi$.

If $S^{\mathrm{Dist}} < \underline{\rho}$, then $p_{S+1} \ge p_{\underline{\rho}}$ and $\phi \ge 1 - 2p_{\underline{\rho}}$, so this expression is weakly negative.

If $S^{\mathrm{Dist}} = \underline{\rho}$, then all moves below $\underline{\rho}$ are already district-improving, so the same expression is strictly negative for all $\phi \geq 0$. Hence the objective weakly decreases up to $\underline{\rho}$.

At $S = \underline{\rho}$, the adjacent difference is

$$1 - 2p_{\underline{\rho}+1} + \phi\big(1 - 2(\rho - \underline{\rho})\big).$$

By supposition $S^{\mathrm{Dist}} \leq \underline{\rho}$, so it must be that $p_{\underline{\rho}+1} < 1/2$ whenever this district exists. Otherwise district $\underline{\rho}+1$ would be won by party A under the district-optimal rule, contradicting $S^{\mathrm{Dist}} \leq \underline{\rho}$. So the first term of the adjacent difference, $1 - 2p_{\underline{\rho}+1}$, is positive.

Since $\rho - \underline{\rho} \leq 1/2$, the second term of the adjacent difference, $\phi\big(1 - 2(\rho - \underline{\rho})\big)$, is weakly nonnegative.

Thus the adjacent difference at $\underline{\rho}$ is positive, and all later adjacent differences are positive by single crossing. Therefore $\underline{\rho}$ is optimal and it is implemented by the cutoff $p_{\underline{\rho}}$. Note that when $S^{\mathrm{Dist}} < \underline{\rho}$, the definition gives $\phi_{\underline{\rho}} = 1 - 2p_{\underline{\rho}}$. This proves part (ii).

Finally suppose $\rho - \underline{\rho} > 1/2$. For every $S < \underline{\rho}$, the adjacent difference remains $1 - 2p_{S+1} - \phi$, which is weakly negative for all $\phi \geq \phi_{\underline{\rho}}$ by the same argument as in the previous case. Hence the objective weakly decreases up to $\underline{\rho}$, and the only remaining comparison is the move from $\underline{\rho}$ to $\underline{\rho} + 1$ with the adjacent difference

$$1 - 2p_{\underline{\rho}+1} - \phi\big(2(\rho - \underline{\rho}) - 1\big).$$

By definition,

$$\phi_{\underline{\rho}+1} = \frac{1 - 2p_{\underline{\rho}+1}}{2(\rho - \underline{\rho}) - 1}.$$

Since $\rho - \underline{\rho} > 1/2$, the denominator is positive. So the adjacent move is strictly worse when $\phi < \phi_{\underline{\rho}+1}$ and weakly improving when $\phi \geq \phi_{\underline{\rho}+1}$. Thus, if $\phi_{\underline{\rho}} \leq \phi < \phi_{\underline{\rho}+1}$, $\underline{\rho}$ is optimal.

It remains to verify that the cutoff $\min\{p_{\underline{\rho}}, t^+(\phi)\}$ implements the top-$\underline{\rho}$ allocation. To see this, note that since $\rho - \underline{\rho} > 1/2$, the denominator in $\phi_{\underline{\rho}+1}$ is positive. Thus $\phi < \phi_{\underline{\rho}+1}$ implies

$$t^+(\phi) = \frac{1 - \phi\big(2(\rho - \underline{\rho}) - 1\big)}{2} > p_{\underline{\rho}+1}.$$

Since districts are ordered and district vote shares are distinct, $p_{\underline{\rho}} > p_{\underline{\rho}+1}$. Hence both terms inside $\min\{p_{\underline{\rho}}, t^+(\phi)\}$ are greater than $p_{\underline{\rho}+1}$. By construction, $\min\{p_{\underline{\rho}}, t^+(\phi)\} \leq p_{\underline{\rho}}$. Therefore $p_{\underline{\rho}} \geq \min\{p_{\underline{\rho}}, t^+(\phi)\} > p_{\underline{\rho}+1}$. Thus this cutoff implements the top-$\underline{\rho}$ allocation.

If $\phi \geq \phi_{\underline{\rho}+1}$, the move from $\underline{\rho}$ to $\underline{\rho} + 1$ is weakly improving. If $\underline{\rho} + 1 = N$, there is no later adjacent move to check. Otherwise, the next adjacent move has difference $1 - 2p_{\underline{\rho}+2} + \phi$, which is strictly positive because $S^{\mathrm{Dist}} \leq \underline{\rho}$ implies $p_{\underline{\rho}+2} < 1/2$. By single crossing, all later adjacent differences are also weakly positive. Hence $\underline{\rho} + 1$ is optimal, and it is implemented by the cutoff $p_{\underline{\rho}+1}$. This proves part (iii).

$\square$

Proof of Corollary 1. If $S^{\text{Dist}} = S^{\text{State}}$, the statement is immediate. Suppose $S^{\text{Dist}} < S^{\text{State}}$. By the first order condition (1), specialized to the linear benchmark as in Theorem 3, an adjacent move from S to $S+1$ on the stated range is weakly profitable if and only if

$$(1-w)\Delta\text{Dist}(S,\mathbf{p}) \le w\Delta\text{State}(S,\rho).$$

For each $S \in \{S^{\text{Dist}}, \ldots, S^{\text{State}} - 1\}$, define

$$w_{S+1} := \frac{\Delta\text{Dist}(S,\mathbf{p})}{\Delta\text{Dist}(S,\mathbf{p}) + \Delta\text{State}(S,\rho)},$$

and set $w_{S^{\text{Dist}}} := 0$. On this range, $\Delta\text{Dist}(S,\mathbf{p}) \ge 0$ is weakly increasing and $\Delta\text{State}(S,\rho) > 0$ is weakly decreasing, so these critical weights are ordered:

$$0 = w_{S^{\text{Dist}}} < w_{S^{\text{Dist}}+1} \le \cdots \le w_{S^{\text{State}}} < 1.$$

By Lemma 2, it suffices to compare seat totals in $\{S^{\text{Dist}}, \ldots, S^{\text{State}}\}$. Now take $w \in (w_S, w_{S+1})$. For every $L < S$, the ordered critical weights imply that the move $L \to L+1$ strictly lowers the objective. For every $L \in \{S, \ldots, S^{\text{State}} - 1\}$, the move $L \to L+1$ strictly raises the objective. Hence the objective decreases up to S and increases after S, so S is the unique optimal seat total. If $w > w_{S^{\text{State}}}$, every move up to S^{State} lowers the objective, and Lemma 2 rules out minimizers beyond S^{State}; hence S^{State} is the unique optimal seat total.

At the switch weight $w = w_{S+1}$, the adjacent condition holds with equality, so S and $S+1$ both minimize. If several switch weights coincide, the same argument gives the corresponding block of tied minimizers. $\square$

Majorization Monotonicity

Proof of Lemma 3. Fix $w \in [0,1]$ and $S \in \{0, \ldots, N\}$. Let $\mathbf{p}$ and $\mathbf{q}$ have the same mean. For any ordered profile $\mathbf{r}$, write $\rho_{\mathbf{r}} := \sum_i r_i$. Then $\text{Dist}(S,\mathbf{r}) = S + \rho_{\mathbf{r}} - 2\sum_{i=1}^{S} r_i$. Therefore

$$\text{Dist}(S,\mathbf{p}) - \text{Dist}(S,\mathbf{q}) = -2\left(\sum_{i=1}^{S} p_i - \sum_{i=1}^{S} q_i\right).$$

If $\mathbf{p}$ majorizes $\mathbf{q}$, then $\sum_{i=1}^{S} p_i \ge \sum_{i=1}^{S} q_i$ for every S. Hence $\text{Dist}(S,\mathbf{p}) \le \text{Dist}(S,\mathbf{q})$. Since the two profiles have the same mean, their statewide support is the same. Writing $\rho_{\mathbf{q}} := \sum_i q_i$, we have $\rho_{\mathbf{q}} = \rho$ and therefore $|S - \rho| = |S - \rho_{\mathbf{q}}|$. Thus

$$M(S, \mathbf{p}; w) = (1 - w)\mathrm{Dist}(S, \mathbf{p}) + w|S - \rho| \leq (1 - w)\mathrm{Dist}(S, \mathbf{q}) + w|S - \rho_{\mathbf{q}}| = M(S, \mathbf{q}; w)$$

for every S and every $w \in [0, 1]$. $\square$

Proof of Proposition 4. Let S^* denote a minimizer of $\min_S M(S, \mathbf{q}; w)$. Then

$$\min_S M(S, \mathbf{p}; w) \leq M(S^*, \mathbf{p}; w) \leq M(S^*, \mathbf{q}; w) = \min_S M(S, \mathbf{q}; w),$$

where the second inequality is Lemma 3. Since R_w selects a minimizer at each profile,

$$M(R_w, \mathbf{p}; w) = \min_S M(S, \mathbf{p}; w) \leq \min_S M(S, \mathbf{q}; w) = M(R_w, \mathbf{q}; w).$$

$\square$

Proof of Theorem 4. **If.** Assume $\mathbf{p}$ majorizes $\mathbf{q}$ and $\bar{p} = \bar{q}$. If $\lambda = w$, then $R_\lambda = R_w$, so (MM) follows from Proposition 4. If $\lambda = 1$, then R_λ selects a statewide-optimal seat total, which depends only on the common mean. Thus $S_{R_\lambda}(\mathbf{p}) = S_{R_\lambda}(\mathbf{q}) =: S$, and (MM) follows from Lemma 3.

Only if. Suppose (MM) holds for all pairs $(\mathbf{p}, \mathbf{q})$ with $\mathbf{p}$ majorizing $\mathbf{q}$ and $\bar{p} = \bar{q}$. We show that either $\lambda = w$ or $\lambda = 1$.

Suppose, toward a contradiction, that $\lambda \in [0, 1)$ and $\lambda \neq w$. Since $\lambda < 1$, Proposition 6 gives two profiles with the same mean at which R_λ selects different seat totals. Choose the common mean m so that Nm is not equidistant from any two distinct integer seat totals. Let $\mathbf{m} := (m, \ldots, m)$. Every same-mean profile majorizes $\mathbf{m}$. Therefore, from the two profiles just described, we can choose a profile $\mathbf{p}$ such that $S_{R_\lambda}(\mathbf{p}) \neq S_{R_\lambda}(\mathbf{m})$. Set $\mathbf{q} := \mathbf{m}$ and write $\rho_{\mathbf{q}} := \sum_i q_i$. Then $\mathbf{p}$ majorizes $\mathbf{q}$, $\bar{p} = \bar{q} = m$, $\rho = \rho_{\mathbf{q}} = Nm$, and $S_{R_\lambda}(\mathbf{p}) \neq S_{R_\lambda}(\mathbf{q})$.

Define $\mathbf{r}(u) := (1 - u)\mathbf{q} + u\mathbf{p}$, for $u \in [0, 1]$. Then $\bar{r}(u) = m$ for all u. Moreover, if $u' > u$, then

$$\mathbf{r}(u) = \left(1 - \frac{u}{u'}\right)\mathbf{q} + \frac{u}{u'}\mathbf{r}(u').$$

Since $\mathbf{r}(u')$ majorizes $\mathbf{q}$ and the set of vectors majorized by a fixed profile is convex, it follows that $\mathbf{r}(u')$ majorizes $\mathbf{r}(u)$.

Claim. There exists $u^* \in [0, 1]$ and a seat total $S' \neq S_{R_\lambda}(\mathbf{q})$ such that

$$M(S', \mathbf{r}(u^*); \lambda) = M(S_{R_\lambda}(\mathbf{q}), \mathbf{r}(u^*); \lambda).$$

Proof of Claim. Define

$$u^* := \inf\{u \in [0,1] : S_{R_\lambda}(\mathbf{r}(u)) \neq S_{R_\lambda}(\mathbf{q})\}.$$

For each fixed S, the map $u \mapsto M(S, \mathbf{r}(u); \lambda)$ is continuous. For all $u < u^*$, $S_{R_\lambda}(\mathbf{r}(u)) = S_{R_\lambda}(\mathbf{q})$, so

$$M(S_{R_\lambda}(\mathbf{q}), \mathbf{r}(u); \lambda) \leq M(S, \mathbf{r}(u); \lambda) \quad \text{for every } S.$$

Letting $u \uparrow u^*$ gives $S_{R_\lambda}(\mathbf{q}) \in \arg\min_S M(S, \mathbf{r}(u^*); \lambda)$.

By definition of u^*, take a sequence $u_n \downarrow u^*$ with $u_n > u^*$ and $S_{R_\lambda}(\mathbf{r}(u_n)) \neq S_{R_\lambda}(\mathbf{q})$. Since the set of seat totals is finite, there is a subsequence, still denoted u_n, and a seat total $S' \neq S_{R_\lambda}(\mathbf{q})$ such that $S_{R_\lambda}(\mathbf{r}(u_n)) = S'$ along the subsequence. Optimality of S' at $\mathbf{r}(u_n)$ gives

$$M(S', \mathbf{r}(u_n); \lambda) \leq M(S_{R_\lambda}(\mathbf{q}), \mathbf{r}(u_n); \lambda).$$

Letting $n \to \infty$ and using continuity gives

$$M(S', \mathbf{r}(u^*); \lambda) \leq M(S_{R_\lambda}(\mathbf{q}), \mathbf{r}(u^*); \lambda).$$

Combined with the optimality of $S_{R_\lambda}(\mathbf{q})$ at u^*, this proves the claim. $\square$

Let $S_0 := S_{R_\lambda}(\mathbf{q})$. Since $\lambda < 1$, the indifference at weight λ implies

$$(1-\lambda)\big[\mathrm{Dist}(S', \mathbf{r}(u^*)) - \mathrm{Dist}(S_0, \mathbf{r}(u^*))\big] + \lambda\big[|S' - \rho| - |S_0 - \rho|\big] = 0.$$

Thus

$$\mathrm{Dist}(S', \mathbf{r}(u^*)) - \mathrm{Dist}(S_0, \mathbf{r}(u^*)) = -\frac{\lambda}{1-\lambda}\big[|S' - \rho| - |S_0 - \rho|\big].$$

Evaluating the same two seat totals at the evaluation weight w gives

$$\begin{aligned}
&M(S', \mathbf{r}(u^*); w) - M(S_0, \mathbf{r}(u^*); w) \\
&\quad = (1-w)\big[\mathrm{Dist}(S', \mathbf{r}(u^*)) - \mathrm{Dist}(S_0, \mathbf{r}(u^*))\big] + w\big[|S' - \rho| - |S_0 - \rho|\big] \\
&\quad = \frac{w-\lambda}{1-\lambda}\big[|S' - \rho| - |S_0 - \rho|\big].
\end{aligned}$$

Because ρ is not equidistant from any two distinct integer seat totals and $S' \neq S_0$, the final bracket is nonzero. Since $w \neq \lambda$, the evaluation-weight difference is nonzero.

Assume without loss of generality that $M(S', \mathbf{r}(u^*); w) > M(S_0, \mathbf{r}(u^*); w)$. By continuity, choose $u_- < u^* < u_+$ arbitrarily close to u^* such that $S_{R_\lambda}(\mathbf{r}(u_-)) = S_0$, $S_{R_\lambda}(\mathbf{r}(u_+)) = S'$, and

$$M(S', \mathbf{r}(u_+); w) > M(S_0, \mathbf{r}(u_-); w).$$

Since $u_+ > u_-$, $\mathbf{r}(u_+)$ majorizes $\mathbf{r}(u_-)$. But the displayed inequality says

$$M(R_\lambda, \mathbf{r}(u_+); w) > M(R_\lambda, \mathbf{r}(u_-); w),$$

contradicting (MM).

The opposite strict inequality is handled by taking the analogous switch in the opposite direction along a same-mean majorization segment.[34] Therefore (MM) cannot hold when $\lambda \in [0, 1)$ and $\lambda \neq w$. We conclude that majorization monotonicity requires either $\lambda = w$ or $\lambda = 1$. □

Characterization of Misrepresentation-Minimizing Rules

Proof of Proposition 5. **If.** Suppose $\lambda = 0$. Then the linear objective is purely district-level: $M(\mathbf{x}, \mathbf{p}; 0) = \mathrm{Dist}(\mathbf{x}, \mathbf{p})$. Thus the rule is first-past-the-post: $R_0(\mathbf{p}) = \{d : p_d \geq 1/2\}$.

If $\mathbf{p}' \geq \mathbf{p}$ componentwise, then $p_d \geq 1/2$ implies $p'_d \geq 1/2$, so $R_0(\mathbf{p}) \subseteq R_0(\mathbf{p}')$. Hence first-past-the-post satisfies strong monotonicity.

Only if. Now fix $\lambda \in (0, 1]$. Choose $0 < \varepsilon < \min\left\{\frac{1}{12}, \frac{\lambda}{2+4\lambda}\right\}$. Consider the profiles

$$\mathbf{p} = \left(\frac{1}{2} + \varepsilon, \frac{1}{2} - \varepsilon, 0, \ldots, 0\right) \quad \text{and} \quad \mathbf{q} = \left(\frac{1}{2} + \varepsilon, \frac{1}{2} + 2\varepsilon, 0, \ldots, 0\right).$$

Then $\mathbf{q} \geq \mathbf{p}$ componentwise.

At profile $\mathbf{p}$, the ordered shares are $p_1 = 1/2 + \varepsilon$ and $p_2 = 1/2 - \varepsilon$, and the statewide vote total is $\rho = 1$. The adjacent difference from 0 to 1 seats is $(1 - \lambda)(-2\varepsilon) + \lambda(-1) < 0$, and the adjacent difference from 1 to 2 seats is $(1 - \lambda)(2\varepsilon) + \lambda > 0$. Thus the unique optimal seat total is $S = 1$, and the unique top-ranked district is district 1. Hence $R_\lambda(\mathbf{p}) = \{1\}$.

At profile $\mathbf{q}$, the ordered shares are $q_1 = 1/2 + 2\varepsilon$ and $q_2 = 1/2 + \varepsilon$, the statewide vote total is $\rho_{\mathbf{q}} := \sum_i q_i = 1 + 3\varepsilon$, and the top-ranked district is district 2. The adjacent difference from 0 to 1 seats is $(1 - \lambda)(-4\varepsilon) + \lambda(-1) < 0$.

The adjacent difference from 1 to 2 seats is $(1 - \lambda)(-2\varepsilon) + \lambda(1 - 6\varepsilon)$. This expression equals $\lambda - (2 + 4\lambda)\varepsilon$, which is strictly positive by the choice of ε. Therefore the unique optimal seat total at $\mathbf{q}$ is again $S = 1$, but the unique top-ranked district is now district 2. Hence $R_\lambda(\mathbf{q}) = \{2\}$.

Thus $\mathbf{q} \geq \mathbf{p}$, but district 1 belongs to $R_\lambda(\mathbf{p})$ and not to $R_\lambda(\mathbf{q})$. Strong monotonicity fails

[34] The sign is controlled by placing ρ on the appropriate side of the midpoint between the two switching seat totals. If $w > \lambda$, choose the switch to move away from proportionality, which the evaluator at w penalizes more heavily; if $w < \lambda$, choose it to move toward proportionality, but the evaluator at w values that gain less than R_λ does. In either case, the seat total selected after the switch is strictly worse at weight w.

for every $\lambda > 0$. Therefore, within the linear misrepresentation-minimizing family, strong monotonicity holds if and only if $\lambda = 0$. $\square$

Proof of Proposition 6. **If.** Suppose $\lambda = 1$. Then only statewide misrepresentation matters. The selected seat total is an integer closest to ρ, using the maintained tie-breaking convention. Hence the seat total depends only on the statewide mean, not on the distribution of support across districts. The rule is gerrymandering-proof.

Only if. Suppose $\lambda < 1$. Pick $\delta \in (0, \frac{1}{4})$ and $\varepsilon > 0$ such that $\varepsilon < \min\left\{\delta(1-\lambda), \frac{1}{2} - 2\delta\right\}$. Consider the profiles

$$\mathbf{p} := \left(\underbrace{1 - \frac{\delta-\varepsilon}{N-2}, \ldots, 1 - \frac{\delta-\varepsilon}{N-2}}_{N-2}, \frac{1}{2} + \delta, 0\right) \quad \text{and} \quad \mathbf{q} = \left(\underbrace{1, \ldots, 1}_{N-2}, \frac{1}{2} - \delta, \varepsilon + \delta\right).$$

By construction, both profiles have the same statewide vote total $\rho = N - 2 + \frac{1}{2} + \varepsilon$. Moreover, $\varepsilon < \delta$ implies $1 - (\delta - \varepsilon)/(N-2) < 1$, and $\delta < 1/4$ implies

$$1 - \frac{\delta - \varepsilon}{N-2} \geq 1 - \delta > \frac{1}{2} + \delta.$$

Finally, $\varepsilon < 1/2 - 2\delta$ implies $1/2 - \delta \geq \varepsilon + \delta$. Hence both profiles are ordered.

At profile $\mathbf{p}$, the adjacent differences around $N-1$ seats are

$$M(N-1, \mathbf{p}; \lambda) - M(N-2, \mathbf{p}; \lambda) = -2(1-\lambda)\delta - 2\lambda\varepsilon < 0$$

and

$$M(N, \mathbf{p}; \lambda) - M(N-1, \mathbf{p}; \lambda) = 1 > 0.$$

Therefore $S_{R_\lambda}(\mathbf{p}) = N - 1$.

At profile $\mathbf{q}$, the adjacent differences around $N-2$ seats are

$$M(N-1, \mathbf{q}; \lambda) - M(N-2, \mathbf{q}; \lambda) = 2(1-\lambda)\delta - 2\lambda\varepsilon > 0,$$

where the inequality follows from $\varepsilon < \delta(1-\lambda)$ when $\lambda > 0$ and is immediate when $\lambda = 0$. Also,

$$M(N-2, \mathbf{q}; \lambda) - M(N-3, \mathbf{q}; \lambda) = -1 < 0.$$

Therefore $S_{R_\lambda}(\mathbf{q}) = N - 2$.

The profiles have the same statewide mean but induce different seat totals under R_λ. Hence R_λ is not gerrymandering-proof. Therefore, within the normalized linear family, gerrymandering-proofness holds if and only if $\lambda = 1$. □

Gerrymandering Vulnerability

Proof of Proposition 7. Fix a baseline profile $\mathbf{p}$ and w_0 such that $|R_{w_0}(\mathbf{p})| > \rho$. Let $k > |R_{w_0}(\mathbf{p})|$. Then $k > \rho$.

Since $|R_{w_0}(\mathbf{p})|$ is an integer and $k > |R_{w_0}(\mathbf{p})| > \rho$, we have $k - 1 \geq \rho$. Fix any same-mean profile $\mathbf{r} \in \mathcal{P}(\bar{p})$. Increasing the seat total from $k-1$ to k changes the statewide loss by

$$|\rho - k| - |\rho - (k-1)| = 1.$$

Thus

$$\Delta_k(\mathbf{r}; w) := M(k, \mathbf{r}; w) - M(k-1, \mathbf{r}; w) = (1-w)(1-2r_k) + w.$$

If $w' = 1$, then $\mathcal{G}_k(\mathbf{p}; w') = \emptyset$: the pure statewide rule cannot award an overrepresentation-side marginal seat. The inclusion is then immediate.

Suppose, therefore, that $w' < 1$. If $\mathbf{r} \in \mathcal{G}_k(\mathbf{p}; w')$, then the objective at weight w' is unimodal in the seat total. Under the maintained tie-breaking convention in favor of party A, obtaining at least k seats implies $M(k, \mathbf{r}; w') \leq M(k-1, \mathbf{r}; w')$, or $\Delta_k(\mathbf{r}; w') \leq 0$. Since

$$\Delta_k(\mathbf{r}; w) = 1 - 2(1-w)r_k \leq 1 - 2(1-w')r_k = \Delta_k(\mathbf{r}; w')$$

because $w' > w$. Hence $\Delta_k(\mathbf{r}; w) \leq 0$, so $\mathbf{r} \in \mathcal{G}_k(\mathbf{p}; w)$. Therefore $\mathcal{G}_k(\mathbf{p}; w') \subseteq \mathcal{G}_k(\mathbf{p}; w)$.

For strictness, suppose $\mathcal{G}_k(\mathbf{p}; w) \neq \emptyset$ and $w' > w$. For $w < 1$, the condition $\Delta_k(\mathbf{r}; w) \leq 0$ is equivalent to the threshold condition $r_k \geq \theta(w)$, and $\theta(w) := \frac{1}{2(1-w)}$.

The threshold is strictly increasing in w. If $w' = 1$, then $\mathcal{G}_k(\mathbf{p}; w') = \emptyset$ while $\mathcal{G}_k(\mathbf{p}; w)$ is nonempty by assumption. If $w' < 1$, choose an ordered same-mean profile whose kth order statistic lies weakly above $\theta(w)$ and strictly below $\theta(w')$: if $\theta(w) = \rho/k$, use the profile with the first k districts equal to ρ/k and the remaining districts equal to zero; otherwise choose $t \in [\theta(w), \min\{\theta(w'), \rho/k\})$, set the first k districts initially equal to t, set all later districts equal to zero, and distribute the remaining mass $\rho - kt$ among the first $k-1$ districts. This is feasible because $k - 1 \geq \rho$. The constructed profile belongs to $\mathcal{G}_k(\mathbf{p}; w)$ but not to $\mathcal{G}_k(\mathbf{p}; w')$. Hence the inclusion is strict. □

Proof of Proposition 8. If $\lambda = 1$, the rule minimizes only the statewide loss. Since $k \geq 1 + \rho$, we have $k - 1 \geq \rho$, so every seat total $S \geq k$ is farther from ρ than $k-1$. Hence no minimizer can give party A at least k seats, and $\mathcal{G}_k(\mathbf{p}; 1) = \emptyset$.

Now let $\lambda < 1$. Suppose, toward a contradiction, that $\mathcal{G}_k(\mathbf{p}; \lambda) \neq \emptyset$. Then there exists

$\mathbf{q} \in \mathcal{P}(\bar{p})$ such that $|R_\lambda(\mathbf{q})| \geq k$. Order $\mathbf{q}$ so that $q_1 \geq \cdots \geq q_N$. By the top-S reduction, the seat total selected by R_λ is an optimal seat total for the reduced problem. Since this optimal seat total is at least k, the adjacent move from $k-1$ to k must be weakly improving: $M(k, \mathbf{q}; \lambda) \leq M(k-1, \mathbf{q}; \lambda)$.

Because $k - 1 \geq \rho$, this adjacent difference is

$$M(k, \mathbf{q}; \lambda) - M(k-1, \mathbf{q}; \lambda) = (1-\lambda)(1-2q_k) + \lambda \leq 0.$$

Equivalently, $q_k \geq \frac{1}{2(1-\lambda)}$. Since the profile is ordered and $\mathbf{q} \in \mathcal{P}(\bar{p})$,

$$\rho = \sum_{d=1}^{N} q_d \geq \sum_{d=1}^{k} q_d \geq k q_k \geq \frac{k}{2(1-\lambda)}.$$

Rearranging gives $\lambda \leq 1 - \frac{k}{2\rho}$, contradicting the hypothesis. Therefore $\mathcal{G}_k(\mathbf{p}; \lambda) = \emptyset$. □

Efficiency Gap

Proof of Proposition 9. Let $x_d^F = 1$ if party A wins district d under first-past-the-post, and let $x_d^F = 0$ otherwise. Then $\sum_{d=1}^{N} x_d^F = S$.

If party A wins district d, party B wastes $1 - p_d$ votes and party A wastes $p_d - \frac{1}{2}$. Hence the district's contribution to $W_B(\mathbf{p}) - W_A(\mathbf{p})$ is $(1 - p_d) - \left(p_d - \frac{1}{2}\right) = \frac{3}{2} - 2p_d$. If party A loses district d, party B wastes $\frac{1}{2} - p_d$ votes and party A wastes p_d. Hence the contribution is $\left(\frac{1}{2} - p_d\right) - p_d = \frac{1}{2} - 2p_d$. Combining the two cases, district d's contribution is $x_d^F + \frac{1}{2} - 2p_d$. Summing across districts gives

$$W_B(\mathbf{p}) - W_A(\mathbf{p}) = \sum_{d=1}^{N} \left(x_d^F + \frac{1}{2} - 2p_d \right) = S + \frac{N}{2} - 2\rho.$$

Therefore

$$EG_A(\mathbf{p}) = \frac{W_B(\mathbf{p}) - W_A(\mathbf{p})}{N} = \frac{1}{N} \left[S - \left(2\rho - \frac{N}{2} \right) \right].$$

Taking absolute values yields

$$|EG_A(\mathbf{p})| = \frac{1}{N} \left| S - \left(2\rho - \frac{N}{2} \right) \right|.$$

Thus the absolute efficiency gap is the statewide-misrepresentation term with linear cost $Q(z) = z$ and target $T^{EG}(\rho) = 2\rho - N/2$, scaled by $1/N$. □

B Axiomatization

This appendix gives formal characterization results for the district and statewide loss primitives used in Section 2.

District loss. Consider a district-loss function $\mathrm{Dist}(\mathbf{x}, \mathbf{p})$. For a permutation π of districts, let $\pi\mathbf{p}$ and $\pi\mathbf{x}$ denote the correspondingly relabeled profile and allocation.

The district loss is *anonymous* if $\mathrm{Dist}(\pi\mathbf{x}, \pi\mathbf{p}) = \mathrm{Dist}(\mathbf{x}, \mathbf{p})$ for every permutation π.

It is *additively separable* if there are functions $\delta_d : \{0,1\} \times [0,1] \to \mathbb{R}_+$ such that $\mathrm{Dist}(\mathbf{x}, \mathbf{p}) = \sum_{d=1}^{N} \delta_d(x_d, p_d)$.

It is *district responsive* if, whenever $p_i \geq p_j$, replacing an allocation that assigns district i to party B and district j to party A with the allocation that assigns district i to party A and district j to party B weakly lowers district loss.

Proposition 10. *Suppose $N \geq 2$. Up to an additive constant independent of $\mathbf{x}$, the following are equivalent:*

(i) *There exists a common function $\delta : \{0,1\} \times [0,1] \to \mathbb{R}_+$ such that* $\mathrm{Dist}(\mathbf{x}, \mathbf{p}) = \sum_{d=1}^{N} \delta(x_d, p_d)$*, and the district switch margin $\Delta(p) := \delta(1,p) - \delta(0,p)$ is weakly decreasing in p.*

(ii) Dist *is anonymous, additively separable, and district responsive.*

Proof of Proposition 10. If (i) holds, anonymity and additive separability are immediate. To verify district responsiveness, compare an allocation that assigns district i to party B and district j to party A with the allocation that swaps those two assignments. The change in district loss from the swap is

$$\begin{aligned}&\delta(1,p_i) + \delta(0,p_j) - \delta(0,p_i) - \delta(1,p_j)\\ &\quad = \big[\delta(1,p_i) - \delta(0,p_i)\big] - \big[\delta(1,p_j) - \delta(0,p_j)\big] = \Delta(p_i) - \Delta(p_j) \leq 0,\end{aligned}$$

because $p_i \geq p_j$ and Δ is weakly decreasing.

Conversely, additive separability gives functions δ_d such that $\mathrm{Dist}(\mathbf{x}, \mathbf{p}) = \sum_{d=1}^{N} \delta_d(x_d, p_d)$. Fix districts i and j. Applying anonymity to the transposition that exchanges only i and j gives, for all $(y,q), (y',q') \in \{0,1\} \times [0,1]$,

$$\delta_i(y,q) + \delta_j(y',q') = \delta_i(y',q') + \delta_j(y,q). \iff \delta_i(y,q) - \delta_j(y,q) = \delta_i(y',q') - \delta_j(y',q').$$

Thus $\delta_i - \delta_j$ is constant over assignments and vote shares. Therefore there are constants c_d and a common function δ such that $\delta_d(y,q) = \delta(y,q) + c_d$ for all d, y, q. The term $\sum_d c_d$ is independent of the allocation and is absorbed into the allowed additive normalization.

Finally, district responsiveness applied to a two-district swap gives

$$\delta(1, p_i) - \delta(0, p_i) \leq \delta(1, p_j) - \delta(0, p_j)$$

whenever $p_i \geq p_j$, that is $\Delta(p) = \delta(1, p) - \delta(0, p)$ is weakly decreasing in p. $\square$

Neutrality. Party neutrality for district loss requires $\mathrm{Dist}(\mathbf{x}, \mathbf{p}) = \mathrm{Dist}(\mathbf{1} - \mathbf{x}, \mathbf{1} - \mathbf{p})$ for all $(\mathbf{x}, \mathbf{p})$. Under Proposition 10, and after the same allocation-independent normalization, neutrality is equivalent to $\delta(1, p) = \delta(0, 1 - p)$, and $\delta(0, p) = \delta(1, 1 - p)$. Thus, neutrality states

$$\sum_{d=1}^{N} [\delta(x_d, p_d) - \delta(1 - x_d, 1 - p_d)] = 0$$

for every independently chosen (x_d, p_d). Hence each bracketed term is constant in (x_d, p_d); after the allocation-independent normalization, that constant is zero. Taking $x_d = 1$ and $x_d = 0$ gives the two displayed identities. The converse follows by summing the identities district by district.

Statewide loss. For the statewide component, write the loss as $\mathrm{State}(S, \rho)$, where $S \in [0, N]$ is party A's seat total and $\rho \in [0, N]$ is party A's statewide support measured in seat units. As in the main text, for an allocation $\mathbf{x}$ and profile $\mathbf{p}$, we write $\mathrm{State}(\mathbf{x}, \mathbf{p}) := \mathrm{State}(S(\mathbf{x}), \rho)$.

The statewide loss is *target based* if, for every ρ, there is a target $T(\rho) \in [0, N]$ such that statewide loss depends on S only through the target gap $|S - T(\rho)|$. Given such a target, let

$$\mathcal{D}_T := \{\, |S - T(\rho)| : S \in [0, N],\ \rho \in [0, N] \,\}$$

be the attainable gap domain.

The loss is *gap-scale invariant* if pairs with the same target gap have the same loss.

It is *gap monotone* if larger attainable target gaps are weakly worse, and *strictly gap monotone* if larger attainable target gaps are strictly worse. It is *gap convex* if target-gap losses are weakly convex on every interval contained in $\mathcal{D}_T$.

Proposition 11. *After normalizing the zero-gap loss to zero, the statewide loss is target based, gap-scale invariant, gap monotone, and gap convex if and only if there is a target function $T : [0, N] \to [0, N]$ and a weakly increasing weakly convex function $Q : \mathcal{D}_T \to \mathbb{R}_+$, with $Q(0) = 0$, such that*

$$\mathrm{State}(S, \rho) = Q\big(|S - T(\rho)|\big).$$

If gap monotonicity is strict, then Q is strictly increasing on $\mathcal{D}_T$.

Proof of Proposition 11. If the representation holds, target dependence and gap-scale invariance are immediate because loss depends on (S, ρ) only through $|S - T(\rho)|$. Weak monotonicity and weak convexity of Q are gap monotonicity and gap convexity.

Conversely, target-based loss gives, for each ρ, a target $T(\rho)$ and a function Q_ρ on the gap interval attainable at ρ such that $\text{State}(S, \rho) = Q_\rho\big(|S - T(\rho)|\big)$.

Gap-scale invariance makes the value assigned to a gap independent of the statewide vote total at which that gap is attained. Therefore $Q(r) := \text{State}(S, \rho)$ for any (S, ρ) such that $r = |S - T(\rho)|$ is well-defined on $\mathcal{D}_T$. Gap monotonicity and gap convexity imply that this common Q is weakly increasing and weakly convex on the attainable gap domain. Since $T(\rho) \in [0, N]$, the zero gap belongs to the domain, and the zero-gap normalization gives $Q(0) = 0$. Without that normalization, the same representation holds up to the additive constant $Q(0)$. If larger attainable target gaps are strictly worse, the same construction gives strict monotonicity of Q on $\mathcal{D}_T$. □

As in the main text, one may take Q to be defined on all of $[0, N]$. Values outside $\mathcal{D}_T$ are never evaluated by feasible pairs, so extending Q from $\mathcal{D}_T$ to $[0, N]$ is only an off-domain convention.

Neutrality. Party neutrality for statewide loss requires $\text{State}(S, \rho) = \text{State}(N - S, N - \rho)$ for all (S, ρ). Under Proposition 11, if Q is strictly increasing on the gap domain, neutrality is equivalent to $T(N - \rho) = N - T(\rho)$.

To see necessity, neutrality implies

$$Q\big(|S - T(\rho)|\big) = Q\big(|N - S - T(N - \rho)|\big) = Q\big(|S - (N - T(N - \rho))|\big)$$

for every S. Strict monotonicity of Q implies

$$|S - T(\rho)| = |S - (N - T(N - \rho))|.$$

Two distance functions on an interval agree for every S only if their centers agree. Hence $T(\rho) = N - T(N - \rho)$, or equivalently $T(N - \rho) = N - T(\rho)$. Sufficiency follows by substituting this identity into the target-gap representation.

C Welfare Theorems and Supporting Prices

Section 4.1 defines Pareto efficiency for feasible pairs of district-level and statewide misrepresentation, and Proposition 2 gives the frontier characterization for the linear family. This appendix unpacks that result. We rewrite the frontier in supporting-price notation, record the corresponding welfare-theorem analogs, derive the revealed-weight intervals for observed seat totals, and state the elementary envelope facts for the price-indexed value function.

Fix a profile $\mathbf{p}$. The reduction already used in Proposition 2 implies that all welfare comparisons can be made on the top-S frontier, $\big(\mathrm{Dist}(S,\mathbf{p}), \mathrm{State}(S,\rho)\big)$. When we refer below to an efficient seat total, we mean a top-S frontier point that is Pareto efficient in the sense of Section 4.1.

Lemma 5 (First welfare theorem)**.** *Let $G : \mathbb{R}^2_+ \to \mathbb{R}$ be strictly increasing in both coordinates. If S minimizes $G(\mathrm{Dist}(S,\mathbf{p}), \mathrm{State}(S,\rho))$ over $S \in \{0,\dots,N\}$, then the top-S allocation is Pareto efficient.*

Proof of Lemma 5. Suppose the top-S allocation is not Pareto efficient. Then some allocation with seat total S' weakly lowers both district-level and statewide misrepresentation, with at least one strict reduction. Replacing that allocation by the top-S' allocation weakly lowers district loss and leaves statewide loss unchanged. Hence

$$\mathrm{Dist}(S',\mathbf{p}) \le \mathrm{Dist}(S,\mathbf{p}), \qquad \mathrm{State}(S',\rho) \le \mathrm{State}(S,\rho),$$

with at least one strict inequality. Since G is strictly increasing in both coordinates,

$$G(\mathrm{Dist}(S',\mathbf{p}), \mathrm{State}(S',\rho)) < G(\mathrm{Dist}(S,\mathbf{p}), \mathrm{State}(S,\rho)),$$

contradicting optimality of S. $\square$

The supporting price is denoted by $\phi \ge 0$. It is the price of one unit of statewide misrepresentation measured in units of district-level misrepresentation. The corresponding linear objective is

$$\mathrm{Dist}(S,\mathbf{p}) + \phi \mathrm{State}(S,\rho).$$

In normalized notation, $(1-w)\mathrm{Dist}(S,\mathbf{p}) + w\mathrm{State}(S,\rho)$ has the same minimizers as the price objective with $\phi = \frac{w}{1-w}$, and $w = \frac{\phi}{1+\phi}$, for $w \in (0,1)$.

Proposition 12. *Suppose $S^{\mathrm{Dist}} < S^{\mathrm{State}}$. For $\phi \ge 0$, consider* $\mathrm{Dist}(S,\mathbf{p}) + \phi\mathrm{State}(S,\rho)$*. Then:*

(i) *S^{Dist} is a minimizer if and only if $\phi \le \frac{\Delta\mathrm{Dist}(S^{\mathrm{Dist}},\mathbf{p})}{\Delta\mathrm{State}(S^{\mathrm{Dist}},\rho)}$.*

(ii) *$S \in \{S^{\mathrm{Dist}}+1,\dots,S^{\mathrm{State}}-1\}$ is a minimizer if and only if $\frac{\Delta\mathrm{Dist}(S-1,\mathbf{p})}{\Delta\mathrm{State}(S-1,\rho)} \le \phi \le \frac{\Delta\mathrm{Dist}(S,\mathbf{p})}{\Delta\mathrm{State}(S,\rho)}$.*

(iii) *S^{State} is a minimizer if and only if $\phi \ge \frac{\Delta\mathrm{Dist}(S^{\mathrm{State}}-1,\mathbf{p})}{\Delta\mathrm{State}(S^{\mathrm{State}}-1,\rho)}$.*

Proof of Proposition 12. For an adjacent move on the stated range,

$$\begin{aligned}&\big[\mathrm{Dist}(S+1,\mathbf{p})+\phi\mathrm{State}(S+1,\rho)\big]-\big[\mathrm{Dist}(S,\mathbf{p})+\phi\mathrm{State}(S,\rho)\big]\\&\quad=\Delta\mathrm{Dist}(S,\mathbf{p})-\phi\Delta\mathrm{State}(S,\rho)\\&\quad=\Delta\mathrm{State}(S,\rho)\left[\frac{\Delta\mathrm{Dist}(S,\mathbf{p})}{\Delta\mathrm{State}(S,\rho)}-\phi\right].\end{aligned}$$

Because $\Delta\mathrm{State}(S,\rho)>0$, moving from S to $S+1$ weakly lowers the objective if and only if $\frac{\Delta\mathrm{Dist}(S,\mathbf{p})}{\Delta\mathrm{State}(S,\rho)}\leq\phi$.

The ratios $\Delta\mathrm{Dist}(S,\mathbf{p})/\Delta\mathrm{State}(S,\rho)$ are weakly increasing on the stated range. Therefore the adjacent differences of $\mathrm{Dist}(S,\mathbf{p})+\phi\mathrm{State}(S,\rho)$ change sign at most once.

For S^{Dist}, minimization on $\{S^{\mathrm{Dist}},\dots,S^{\mathrm{State}}\}$ is equivalent to the first adjacent move not lowering the objective, which gives part (i). For an interior S, minimization is equivalent to the move from $S-1$ to S weakly lowering the objective and the move from S to $S+1$ not lowering it, which gives part (ii). For S^{State}, minimization is equivalent to the final move into S^{State} weakly lowering the objective, which gives part (iii). Since the adjacent differences have single crossing, these local conditions are also sufficient. Finally, Lemma 2 lets us restrict attention to $\{S^{\mathrm{Dist}},\dots,S^{\mathrm{State}}\}$, so the same conditions characterize minimizers over all $S\in\{0,\dots,N\}$. □

Proposition 13 (Second welfare theorem)**.** *Suppose $S^{\mathrm{Dist}}\leq S^{\mathrm{State}}$. Every efficient seat total can be supported by a strictly positive linear price. That is, for every efficient S, there exists $\phi>0$ such that*

$$S\in\underset{s\in\{0,\dots,N\}}{\arg\min}\{\mathrm{Dist}(s,\mathbf{p})+\phi\mathrm{State}(s,\rho)\}.$$

Proof of Proposition 13. If $S^{\mathrm{Dist}}=S^{\mathrm{State}}$, then the same seat total weakly dominates every other seat total in both coordinates. Any efficient seat total has the same coordinate pair as this common optimum, and any $\phi>0$ supports it.

Now suppose $S^{\mathrm{Dist}}<S^{\mathrm{State}}$. By Lemma 2, any efficient seat total is represented on the frontier or has the same coordinate pair as a point on that frontier. If S has the same loss pair as some $\bar{S}\in\{S^{\mathrm{Dist}},\dots,S^{\mathrm{State}}\}$, then any price supporting $\bar{S}$ also supports S. It is therefore enough to consider $S\in\{S^{\mathrm{Dist}},\dots,S^{\mathrm{State}}\}$.

If $S=S^{\mathrm{Dist}}$, Pareto efficiency rules out $\Delta\mathrm{Dist}(S^{\mathrm{Dist}},\mathbf{p})=0$, because $S^{\mathrm{Dist}}+1$ would then have the same district loss and strictly lower statewide loss. Hence $\Delta\mathrm{Dist}(S^{\mathrm{Dist}},\mathbf{p})>0$. Choose

$$\phi\in\left(0,\frac{\Delta\mathrm{Dist}(S^{\mathrm{Dist}},\mathbf{p})}{\Delta\mathrm{State}(S^{\mathrm{Dist}},\rho)}\right].$$

Proposition 12 supports S^{Dist}.

If $S \in \{S^{\text{Dist}}+1, \ldots, S^{\text{State}}-1\}$, Pareto efficiency rules out $\Delta\text{Dist}(S, \mathbf{p}) = 0$, because $S+1$ would then have the same district loss and strictly lower statewide loss. Since the adjacent slopes are weakly increasing, there exists $\phi > 0$ such that

$$\frac{\Delta\text{Dist}(S-1, \mathbf{p})}{\Delta\text{State}(S-1, \rho)} \leq \phi \leq \frac{\Delta\text{Dist}(S, \mathbf{p})}{\Delta\text{State}(S, \rho)}.$$

Proposition 12 supports S.

If $S = S^{\text{State}}$, choose any

$$\phi \geq \frac{\Delta\text{Dist}(S^{\text{State}}-1, \mathbf{p})}{\Delta\text{State}(S^{\text{State}}-1, \rho)}$$

with $\phi > 0$. Proposition 12 supports S^{State}. $\square$

We now restate Proposition 2 in supporting-price form. The first welfare theorem gives one direction: any minimizer of a strictly positive linear loss is Pareto efficient. The second welfare theorem gives the converse: because the top-S frontier has monotone adjacent slopes, every efficient top-S allocation is supported by some price $\phi > 0$. Thus the frontier-support result does not rely on the voter-loss or vote-seat-gap normalizations used in the linear benchmark. Those normalizations make Dist and State easy to interpret, but the welfare logic applies to the general primitives from Section 2.

Corollary 2. *Suppose $S^{\text{Dist}} \leq S^{\text{State}}$. Under the assumptions of Section 2, the top-S allocation is Pareto efficient in* $(\text{Dist}, \text{State})$*-space if and only if there exists $\phi > 0$ such that*

$$S \in \operatorname*{arg\,min}_{s \in \{0, \ldots, N\}} \{\text{Dist}(s, \mathbf{p}) + \phi \text{State}(s, \rho)\}.$$

Proof of Corollary 2. Suppose first that S minimizes $\text{Dist}(S', \mathbf{p}) + \phi\text{State}(S', \rho)$ over $S' \in \{0, \ldots, N\}$ for some $\phi > 0$. The objective is strictly increasing in both misrepresentation coordinates. Hence Lemma 5 implies that S is an efficient seat total. Since the top-S allocation attains the frontier point $(\text{Dist}(S, \mathbf{p}), \text{State}(S, \rho))$, the top-$S$ allocation is Pareto efficient in $(\text{Dist}, \text{State})$-space.

Conversely, suppose the top-S allocation is Pareto efficient in $(\text{Dist}, \text{State})$-space. Then S is an efficient seat total on the top-S frontier. Otherwise, some seat total S' would satisfy $\text{Dist}(S', \mathbf{p}) \leq \text{Dist}(S, \mathbf{p})$, and $\text{State}(S', \rho) \leq \text{State}(S, \rho)$, with at least one strict inequality, and the top-S' allocation would Pareto dominate the top-S allocation. By Proposition 13, there exists a price $\phi > 0$ such that

$$S \in \operatorname*{arg\,min}_{s \in \{0, \ldots, N\}} \{\text{Dist}(s, \mathbf{p}) + \phi \text{State}(s, \rho)\}.$$

This proves the claim. $\square$

Law of Demand. We can also state the voting analog of the law of demand. Increasing the relative price of statewide misrepresentation lowers the chosen amount. The cross-effect is that district-level misrepresentation weakly rises along the same path, because additional statewide accuracy must be produced by switching progressively weaker marginal districts.

Proposition 14 (Law of Demand)**.** *Suppose $S^{\text{Dist}} \leq S^{\text{State}}$, and let $S^*(\phi)$ be either the smallest or the largest optimal seat total for the price objective. Then $S^*(\phi)$ is weakly increasing in ϕ. Consequently,* $\text{State}(S^*(\phi), \rho)$ *is weakly decreasing in ϕ on the adjustment path from S^{Dist} to S^{State}.*

Proof of Proposition 14. By Proposition 12, the set of prices supporting each seat total is an interval, and these intervals are ordered by seat total. Hence either endpoint selection from the set of minimizers is weakly increasing in ϕ. On the range $S^{\text{Dist}}, \ldots, S^{\text{State}}$, increasing S moves the allocation weakly closer to the statewide optimum, so $\text{State}(S, \rho)$ is weakly decreasing. Combining these two monotonicities gives the result. $\square$

Hicksian, or compensated, representation demand. For a statewide cap $\bar{A}$, define the constrained problem

$$H(\bar{A}; \mathbf{p}) := \arg\min_{S} \{\text{Dist}(S, \mathbf{p}) : \text{State}(S, \rho) \leq \bar{A}\}.$$

This is the Hicksian version of the electoral problem: it fixes the amount of statewide misrepresentation that can be tolerated and asks for the least district-level misrepresentation needed to achieve it. The Lagrangian is

$$\text{Dist}(S, \mathbf{p}) + \phi\big(\text{State}(S, \rho) - \bar{A}\big),$$

so the same ϕ is the multiplier on the statewide constraint. If S_ϕ uniquely minimizes the price objective and $\bar{A} = \text{State}(S_\phi, \rho)$, then $S_\phi \in H(\bar{A}; \mathbf{p})$. Conversely, every efficient constrained solution is supported by some ϕ by Corollary 2. Thus the weighted objective and the constrained objective are dual descriptions of the same frontier.

Envelope theorem. Define the price-indexed value function

$$V(\phi, \mathbf{p}) := \min_{S \in \{0, \ldots, N\}} \{\text{Dist}(S, \mathbf{p}) + \phi \text{State}(S, \rho)\}.$$

This is the lower envelope of finitely many affine functions of ϕ, so it is concave and piecewise linear. Its superdifferential is

$$\partial V(\phi, \mathbf{p}) = \mathrm{co}\{\mathrm{State}(S, \rho) : S \in \arg\min_{s}[\mathrm{Dist}(s, \mathbf{p}) + \phi \mathrm{State}(s, \rho)]\}.$$

In particular, if the minimizer is unique, then $\frac{\partial V(\phi,\mathbf{p})}{\partial \phi} = \mathrm{State}(S^*(\phi), \rho)$. The derivative of the minimized objective with respect to the price of statewide misrepresentation is the chosen amount of statewide misrepresentation.

Slutsky-type decomposition. A change in the vote profile generally changes both the statewide target and the distribution of district-level margins. To separate these two channels, write

$$S^*(\mathbf{r}, \tilde{\rho}; \phi) \in \arg\min_{S}\{\mathrm{Dist}(S, \mathbf{r}) + \phi \mathrm{State}(S, \tilde{\rho})\},$$

where $\mathbf{r}$ determines district costs and $\tilde{\rho}$ determines the statewide target. For a change from $(\mathbf{p}, \rho)$ to $(\mathbf{p}', \rho')$, decompose

$$S^*(\mathbf{p}', \rho'; \phi) - S^*(\mathbf{p}, \rho; \phi) = \underbrace{S^*(\mathbf{p}', \rho; \phi) - S^*(\mathbf{p}, \rho; \phi)}_{\text{distributional, or compensated, effect}} + \underbrace{S^*(\mathbf{p}', \rho'; \phi) - S^*(\mathbf{p}', \rho; \phi)}_{\text{target effect}}.$$

The first term holds the statewide target fixed and changes only the district-cost schedule; it is the analog of a compensated substitution effect. The second term holds the district-cost schedule fixed and changes the statewide target.

D Alternative Utility Microfoundation

This appendix records an equivalent utility-based microfoundation for the linear benchmark. The main text uses a loss formulation: voters suffer disutility from local nonrepresentation and from statewide underrepresentation. The same aggregate objective can also be obtained from a utility formulation in which voters receive positive utility from representation on both dimensions, with an additional penalty when their party is underrepresented.

Assume again that $\rho \in (0, N)$. For a voter who supports party A in district d, define utility under allocation $\mathbf{x}$ by

$$u_d^A(\mathbf{x}) = (1 - w)x_d + w\left[\frac{S(\mathbf{x})}{\rho} - \frac{(\rho - S(\mathbf{x}))_+}{\rho}\right].$$

For a voter who supports party B in district d, define

$$u_d^B(\mathbf{x}) = (1-w)(1-x_d) + w\left[\frac{N-S(\mathbf{x})}{N-\rho} - \frac{(S(\mathbf{x})-\rho)_+}{N-\rho}\right].$$

The first term is local representation utility. The second term is statewide representation utility: voters value seats held by their party in the statewide delegation, measured relative to the party's statewide support, and incur an additional penalty when their party is underrepresented.

Let integrated voter utility be $U(\mathbf{x}) := \sum_{d=1}^{N}\left[p_d u_d^A(\mathbf{x}) + (1-p_d)u_d^B(\mathbf{x})\right]$. Then

$$U(\mathbf{x}) = N - \left[(1-w)\mathrm{Dist}(\mathbf{x},\mathbf{p}) + w|S(\mathbf{x})-\rho|\right].$$

Hence maximizing $U(\mathbf{x})$ is equivalent to minimizing the same linear misrepresentation objective as in the main text.

To see this, the local utility component is

$$(1-w)\sum_{d=1}^{N}\big[p_d x_d + (1-p_d)(1-x_d)\big] = (1-w)\big[N - \mathrm{Dist}(\mathbf{x},\mathbf{p})\big].$$

The positive statewide representation terms aggregate to an allocation-independent constant:

$$w\left[\rho\frac{S(\mathbf{x})}{\rho} + (N-\rho)\frac{N-S(\mathbf{x})}{N-\rho}\right] = wN.$$

The statewide underrepresentation penalties aggregate to $w\left[(\rho - S(\mathbf{x}))_+ + (S(\mathbf{x})-\rho)_+\right] = w|S(\mathbf{x})-\rho|$. Therefore

$$U(\mathbf{x}) = (1-w)\big[N-\mathrm{Dist}(\mathbf{x},\mathbf{p})\big] + wN - w|S(\mathbf{x})-\rho| = N - \left[(1-w)\mathrm{Dist}(\mathbf{x},\mathbf{p}) + w|S(\mathbf{x})-\rho|\right].$$

E Gerrymandering Costs

This appendix records the cost version of Proposition 7. Maintain the setup of that proposition: a baseline profile $\mathbf{p}$, baseline rule index λ_0, and target seat total k are fixed, with party A overrepresented at the baseline and $k > |R_{\lambda_0}(\mathbf{p})|$.

Given statewide vote share m, a *gerrymandering cost* is a continuous map $c : \mathcal{P}(m) \times \mathcal{P}(m) \to \mathbb{R}_+$ such that $c(\mathbf{p},\mathbf{p}) = 0$ and, for every distinct $\mathbf{p},\mathbf{p}' \in \mathcal{P}(m)$ and all $0 \le \theta < \theta' \le 1$,

$$c\big(\mathbf{p},(1-\theta)\mathbf{p} + \theta\mathbf{p}'\big) < c\big(\mathbf{p},(1-\theta')\mathbf{p} + \theta'\mathbf{p}'\big).$$

Thus moving farther away from the baseline profile along any mean-preserving segment is strictly more costly.

For an integer k, define the minimal cost of inducing at least k seats under the rule R_λ by

$$C(\lambda; \mathbf{p}, k) := \inf\left\{c(\mathbf{p}, \mathbf{r}) : \mathbf{r} \in \mathcal{G}_k(\mathbf{p}; \lambda)\right\},$$

with the convention $\inf \emptyset = \infty$. Equivalently, $C(\lambda; \mathbf{p}, k)$ is the least cost of moving from $\mathbf{p}$ to a mean-preserving profile that gives party A at least k seats under R_λ.

Corollary 3. *For every λ', λ with $1 \geq \lambda' > \lambda \geq \lambda_0$, $C(\lambda'; \mathbf{p}, k) \geq C(\lambda; \mathbf{p}, k)$.*

Proof of Corollary 3. By Proposition 7, $\mathcal{G}_k(\mathbf{p}; \lambda') \subseteq \mathcal{G}_k(\mathbf{p}; \lambda)$ whenever $\lambda' > \lambda \geq \lambda_0$. Since $C(\lambda; \mathbf{p}, k)$ is the infimum of $c(\mathbf{p}, \mathbf{r})$ over $\mathcal{G}_k(\mathbf{p}; \lambda)$, the infimum defining $C(\lambda'; \mathbf{p}, k)$ is taken over a weakly smaller feasible set. Therefore $C(\lambda'; \mathbf{p}, k) \geq C(\lambda; \mathbf{p}, k)$. □

The corollary is Proposition 7 written in cost terms. Since $\mathcal{G}_k(\mathbf{p}; \lambda') \subseteq \mathcal{G}_k(\mathbf{p}; \lambda)$, the minimization defining $C(\lambda'; \mathbf{p}, k)$ is taken over a smaller feasible set of maps. A smaller feasible set cannot lower the minimum cost. If the increase in λ removes the lowest-cost successful redistrictings, then the minimum cost rises strictly.

Taking $\lambda_0 = 0$ gives the first-past-the-post interpretation. If party A is overrepresented under first-past-the-post and $k > |R_0(\mathbf{p})|$, then moving away from first-past-the-post by increasing λ weakly raises the minimal cost of obtaining those additional seats. This gives the rule-design interpretation of the model. Redistricting constraints act on maps; the rule index acts on the allocation rule. Both can make gerrymandering harder, but they operate through different levers. Even holding the districting technology fixed, moving from first-past-the-post toward proportionality makes additional seats for an already overrepresented party harder to engineer, because the set of maps on which those extra seats are attainable becomes smaller.

F Measures of Representation

This appendix shows how several standard representation measures can be written inside the paper's district–statewide loss framework. Throughout, S denotes party A's seat total and $\rho = \sum_d p_d$ denotes party A's statewide support in seat units.

F.1 Efficiency Gap.

Proposition 15. *In the N-district equal-turnout linear benchmark:*

(i) Fix $s \in \{0, \dots, N\}$. Among all vote profiles such that $S^{\text{Dist}} = S^{\text{State}} = s$,

$$|EG_A| \le \overline{EG}(N, s) := \begin{cases} \frac{1}{2}, & \text{if } s \in \{0, N\}, \\ \left|\frac{1}{2} - \frac{s}{N}\right| + \frac{1}{N}, & \text{if } 1 \le s \le N-1. \end{cases}$$

(ii) For every $s \in \{0, \dots, N\}$, the bound is tight. That is, there are vote profiles with $S^{\text{Dist}} = S^{\text{State}} = s$ for which $|EG_A|$ is arbitrarily close to $\overline{EG}(N, s)$.

(iii) For every $N \ge 4$ and every $\epsilon > 0$, there exists a vote profile such that $EG_A = 0$, but $w_{S^{\text{Dist}}+1} < \epsilon$.

Proof of Proposition 15. Suppose first that $S^{\text{Dist}} = S^{\text{State}} = s$. Since $S^{\text{State}} = s$, the statewide support ρ lies within one half-seat of s, up to the maintained tie-breaking convention: $|\rho - s| \le \frac{1}{2}$.

For $1 \le s \le N-1$, write $\rho = s + r$, where $r \in [-1/2, 1/2]$. Then

$$EG_A = \frac{s}{N} - 2\frac{s+r}{N} + \frac{1}{2} = \frac{1}{2} - \frac{s}{N} - \frac{2r}{N} \implies |EG_A| \le \left|\frac{1}{2} - \frac{s}{N}\right| + \frac{1}{N}.$$

If $s = 0$, then $\rho \in [0, 1/2]$, so

$$EG_A = \frac{1}{2} - 2\frac{\rho}{N} \in \left[\frac{1}{2} - \frac{1}{N}, \frac{1}{2}\right] \implies |EG_A| \le 1/2.$$

If $s = N$, then $\rho \in [N - 1/2, N]$, so

$$EG_A = \frac{3}{2} - 2\frac{\rho}{N} \in \left[-\frac{1}{2}, -\frac{1}{2} + \frac{1}{N}\right], \implies |EG_A| \le 1/2.$$

This proves part (i).

For tightness, if $s = 0$, take $p_d = 0$ for every district. Then $S^{\text{Dist}} = S^{\text{State}} = 0$ and $EG_A = 1/2$. If $s = N$, take $p_d = 1$ for every district. Then $S^{\text{Dist}} = S^{\text{State}} = N$ and $EG_A = -1/2$.

Now let $1 \le s \le N-1$. If $s \le N/2$, take $\eta > 0$ small and set

$$p_1 = \dots = p_s = 1 - \frac{1}{2s} + \frac{\eta}{s}, \qquad p_{s+1} = \dots = p_N = 0.$$

Then $S^{\text{Dist}} = s$ and $\rho = s - 1/2 + \eta$. For small η, the proportional seat total is $S^{\text{State}} = s$, and

$$EG_A = \frac{1}{2} - \frac{s}{N} + \frac{1}{N} - \frac{2\eta}{N}.$$

Thus $|EG_A|$ can be made arbitrarily close to $|1/2 - s/N| + 1/N$.

If $s > N/2$, take $\eta > 0$ small and set

$$p_1 = \cdots = p_s = 1, \qquad p_{s+1} = \frac{1}{2} - \eta, \qquad p_{s+2} = \cdots = p_N = 0.$$

Then $S^{\text{Dist}} = s$ and $\rho = s + 1/2 - \eta$. For small η, the proportional seat total is $S^{\text{State}} = s$, and

$$EG_A = \frac{1}{2} - \frac{s}{N} - \frac{1}{N} + \frac{2\eta}{N}.$$

Thus $|EG_A|$ can again be made arbitrarily close to the stated bound. This proves the tightness claim, with exact attainment except where the maintained proportional tie-breaking convention makes the half-seat boundary select the adjacent lower seat total.

For part (iii), fix $N \geq 4$ and $\epsilon > 0$. Choose $0 < \varepsilon < 1/4$ such that $\frac{2\varepsilon}{1+2\varepsilon} < \epsilon$. Let

$$p_1 = \frac{1}{2} - \varepsilon, \qquad p_2 = \cdots = p_N = \frac{N-2}{4(N-1)} + \frac{\varepsilon}{N-1}.$$

For ε small enough, every district has party A vote share below one half, so $S^{\text{Dist}} = 0$. The statewide support is

$$\rho = \left(\frac{1}{2} - \varepsilon\right) + (N-1)\left(\frac{N-2}{4(N-1)} + \frac{\varepsilon}{N-1}\right) = \frac{N}{4} \implies EG_A = \frac{0}{N} - 2\frac{N/4}{N} + \frac{1}{2} = 0.$$

Since $N \geq 4$, adding the first seat for party A moves the seat total weakly toward proportionality. The first possible switch away from first-past-the-post assigns party A the strongest district, where $p_1 = 1/2 - \varepsilon$. The adjacent comparison is

$$M(1, \mathbf{p}; w) - M(0, \mathbf{p}; w) = (1-w)(1-2p_1) - w.$$

Substituting $p_1 = 1/2 - \varepsilon$, this becomes $2\varepsilon(1-w) - w$. The first switch occurs when this expression is weakly negative, so $w_{S^{\text{Dist}}+1} = \frac{2\varepsilon}{1+2\varepsilon} < \epsilon$.

Thus $EG_A = 0$, while first-past-the-post ceases to be optimal for an arbitrarily small positive weight on statewide proportionality. $\square$

Weighted wasted-vote targets. The efficiency gap is one member of a weighted wasted-vote family. Let $\alpha \geq 0$ be the weight on surplus winning votes where losing votes receive a fixed weight of one. If party A wins district d, the district contribution to $W_B^\alpha - W_A^\alpha$ is

$$(1 - p_d) - \alpha\left(p_d - \frac{1}{2}\right) = 1 + \frac{\alpha}{2} - (1+\alpha)p_d.$$

If party A loses district d, the contribution is

$$\alpha\left(\frac{1}{2}-p_d\right)-p_d=\frac{\alpha}{2}-(1+\alpha)p_d.$$

Summing across districts, $W_B^\alpha(\mathbf{p})-W_A^\alpha(\mathbf{p})=S^F+\frac{\alpha N}{2}-(1+\alpha)\rho$. Hence the absolute weighted wasted-vote gap is

$$\left|S^F-\left((1+\alpha)\rho-\frac{\alpha N}{2}\right)\right|.$$

It is a statewide loss with target $T^\alpha(\rho)=(1+\alpha)\rho-\frac{\alpha N}{2}$. The case $\alpha=1$ gives the efficiency-gap target $2\rho-N/2$. The case $\alpha=0$, which counts only losing votes as wasted, gives the proportional target ρ.

Linear winner's-bonus targets. Equivalently, we can define

$$T^\beta(\rho)=\rho+\beta\left(\rho-\frac{N}{2}\right),\qquad \beta\geq 0.$$

The proportional target is $\beta=0$, and the efficiency-gap target is $\beta=1$. With linear gap cost,

$$\mathrm{State}^\beta(S,\rho)=\left|S-\left[\rho+\beta\left(\rho-\frac{N}{2}\right)\right]\right|.$$

Positive β gives the statewide vote winner a seat bonus above proportionality and penalizes the statewide loser symmetrically.

F.2 Partisan Bias

Let $S^{\mathrm{FPTP}}(\mathbf{r})$ be the first-past-the-post seat total at profile $\mathbf{r}$. For any profile for which the half-vote counterfactual $\mathbf{p}^{1/2}$ lies in $[0,1]^N$,

$$S^{\mathrm{FPTP}}(\mathbf{p}^{1/2})=\sum_{d=1}^{N}\mathbf{1}\left\{p_d+\frac{1}{2}-\bar{p}\geq\frac{1}{2}\right\}=\sum_{d=1}^{N}\mathbf{1}\{p_d\geq\bar{p}\}.$$

Moreover,

$$\sum_{d=1}^{N}p_d^{1/2}=\sum_{d=1}^{N}p_d+N\left(\frac{1}{2}-\bar{p}\right)=\frac{N}{2}.$$

Thus, under the proportional statewide loss $\mathrm{State}(S,\rho)=|S-\rho|$,

$$N|PB_A(\mathbf{p})|=\left|S^{\mathrm{FPTP}}(\mathbf{p}^{1/2})-\frac{N}{2}\right|=\mathrm{State}\left(S^{\mathrm{FPTP}}(\mathbf{p}^{1/2}),\frac{N}{2}\right).$$

This identity uses the counterfactual profile $\mathbf{p}^{1/2}$, ignores the district-level component, and fixes the allocation rule to first-past-the-post. It is therefore distinct from total misrepresentation at the observed profile.

Since $S^{\text{FPTP}}(\mathbf{p}^{1/2})$ is an integer,

$$|PB_A(\mathbf{p})| = \left| \frac{S^{\text{FPTP}}(\mathbf{p}^{1/2})}{N} - \frac{1}{2} \right|.$$

Thus exact zero partisan bias is possible only when N is even. If N is odd, the smallest possible value of $|PB_A(\mathbf{p})|$ is $\frac{1}{2N}$.

Proposition 16. *Consider the linear benchmark, and suppose $N \geq 3$.*

(i) *For every $\eta > 0$, there exists a profile $\mathbf{p}$ with*

$$|PB_A(\mathbf{p})| = \begin{cases} 0, & \textit{if } N \textit{ is even,} \\ \frac{1}{2N}, & \textit{if } N \textit{ is odd,} \end{cases}$$

but $\text{State}(S^{\text{Dist}}, \rho) > \frac{N}{2} - \eta$, *and first-past-the-post ceases to be misrepresentation-minimizing at a statewide weight smaller than η.*

(ii) *There exists a profile $\mathbf{q}$ with* $\text{State}(S^{\text{Dist}}, \rho) = 0$, *but* $|PB_A(\mathbf{q})| = \frac{1}{2} - \frac{1}{N}$.

Proof of Proposition 16. For part (i), let $k = \lfloor N/2 \rfloor$. Choose $\varepsilon > 0$ small enough that $\frac{1}{2} + 2\varepsilon < 1$, $(N+k)\varepsilon < \frac{N}{2} - 1$, $(N+k)\varepsilon < \eta$, and $\frac{2\varepsilon}{1+2\varepsilon} < \min\{\eta, 1\}$. This is possible because $N \geq 3$. Consider the profile

$$\mathbf{p}^{\varepsilon} = \left(\underbrace{\frac{1}{2} + 2\varepsilon, \ldots, \frac{1}{2} + 2\varepsilon}_{k}, \underbrace{\frac{1}{2} + \varepsilon, \ldots, \frac{1}{2} + \varepsilon}_{N-k} \right).$$

Then $\bar{p}^{\varepsilon} = \frac{1}{2} + \frac{N+k}{N}\varepsilon$. Since $\frac{1}{2} + \varepsilon < \bar{p}^{\varepsilon} < \frac{1}{2} + 2\varepsilon$, total of k districts lie above the statewide mean and $S^{\text{FPTP}}\big((\mathbf{p}^{\varepsilon})^{1/2}\big) = k$. Therefore

$$|PB_A(\mathbf{p}^{\varepsilon})| = \left| \frac{k}{N} - \frac{1}{2} \right| = \begin{cases} 0, & \text{if } N \text{ is even,} \\ \frac{1}{2N}, & \text{if } N \text{ is odd.} \end{cases}$$

The odd-N value is the smallest possible value of $|PB_A|$, since $S^{\text{FPTP}}(\mathbf{p}^{1/2})$ is an integer.

At the profile $\mathbf{p}^\varepsilon$, every district has party A vote share above $1/2$, so first-past-the-post gives $S^{\text{Dist}} = N$. The proportional target is $\rho = N\bar{p}^\varepsilon = \frac{N}{2} + (N+k)\varepsilon$. Thus

$$\text{State}(S^{\text{Dist}}, \rho) = |N - \rho| = \frac{N}{2} - (N+k)\varepsilon > \frac{N}{2} - \eta.$$

It remains to show that first-past-the-post stops minimizing at a statewide weight below η. Since $\rho = \frac{N}{2} + (N+k)\varepsilon < N - 1$, moving from N seats to $N-1$ lowers statewide misrepresentation by one seat. The district switched away from party A has vote share $1/2 + \varepsilon$, so district-level misrepresentation rises by $\left(\frac{1}{2} + \varepsilon\right) - \left(\frac{1}{2} - \varepsilon\right) = 2\varepsilon$. Therefore

$$M(N-1, \mathbf{p}^\varepsilon; w) - M(N, \mathbf{p}^\varepsilon; w) = (1-w)2\varepsilon - w,$$

which is strictly negative whenever $w > \frac{2\varepsilon}{1+2\varepsilon}$. By the choice of ε, there exists $w \in \left(\frac{2\varepsilon}{1+2\varepsilon}, \min\{\eta, 1\}\right)$. For this w, first-past-the-post is not misrepresentation-minimizing, and the statewide weight is smaller than η.

For part (ii), choose $0 < \varepsilon < 1/N$, and define

$$\mathbf{q}^\varepsilon = \left(\frac{1}{2} + \varepsilon, \underbrace{\frac{\frac{1}{2} - \varepsilon}{N-1}, \ldots, \frac{\frac{1}{2} - \varepsilon}{N-1}}_{N-1}\right).$$

The statewide support is $\rho = \frac{1}{2} + \varepsilon + (N-1)\frac{\frac{1}{2}-\varepsilon}{N-1} = 1$. First-past-the-post awards party A one seat, so $S^{\text{Dist}} = 1$. Hence $\text{State}(S^{\text{Dist}}, \rho) = |1 - 1| = 0$.

The statewide mean is $\bar{q}^\varepsilon = 1/N$. The first district has vote share above $1/N$, while each remaining district has vote share below $1/N$. Thus $S^{\text{FPTP}}\big((\mathbf{q}^\varepsilon)^{1/2}\big) = 1$, and $|PB_A(\mathbf{q}^\varepsilon)| = \left|\frac{1}{N} - \frac{1}{2}\right| = \frac{1}{2} - \frac{1}{N}$. This proves the result. $\square$

G Additional Examples

G.1 First-Past-the-Post and Majorization

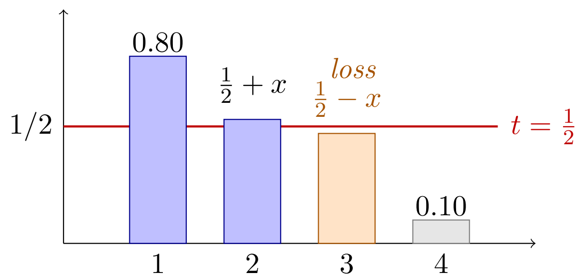
The following example illustrates why majorization monotonicity need not hold for rules with a fixed district winning threshold. Fix any evaluation weight $w > 0$. Choose $x > 0$ small enough that

$$x < \frac{1}{20} \qquad \text{and} \qquad 2x(1-w) < w.$$

Consider the four-district profile

$$\mathbf{q}^x = (0.80,\ 1/2 + x,\ 1/2 - x,\ 0.10), \qquad \rho = \sum_{d=1}^{4} q_d^x = 1.90.$$

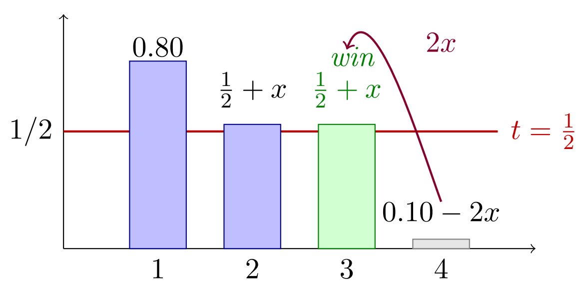
Under first-past-the-post, party A wins two districts. The seat total is therefore close to proportionality: $|2 - 1.90| = 0.10$.



Now transfer $2x$ units of party A's support from the weakest district to the marginal district:

$$\mathbf{p}^x = (0.80,\ 1/2 + x,\ 1/2 + x,\ 0.10 - 2x).$$

The profile $\mathbf{p}^x$ majorizes $\mathbf{q}^x$ and has the same statewide support. For any fixed seat total, the top-S allocation under $\mathbf{p}^x$ has weakly lower district-level misrepresentation than the top-S allocation under $\mathbf{q}^x$. Under first-past-the-post, however, the fixed threshold 1/2 changes the induced seat total: party A now wins three districts.



The district component under first-past-the-post falls by $2x$. The statewide component rises from $|2 - 1.90| = 0.10$ to $|3 - 1.90| = 1.10$, an increase of one seat. Hence, under the linear objective,

$$M(R^{\text{FPTP}}, \mathbf{p}^x; w) - M(R^{\text{FPTP}}, \mathbf{q}^x; w) = -2x(1 - w) + w.$$

By the choice of x, this expression is positive. Thus, for every $w > 0$, there is a mean-preserving majorization that improves local representation under first-past-the-post but increases total misrepresentation. The reason is that first-past-the-post keeps the district threshold fixed: concentrating support can turn a marginal district from a loss into a win, and the resulting discrete seat gain can dominate the district-level improvement.

H Empirical Appendix

H.1 Proportional Switch Weights in 2024

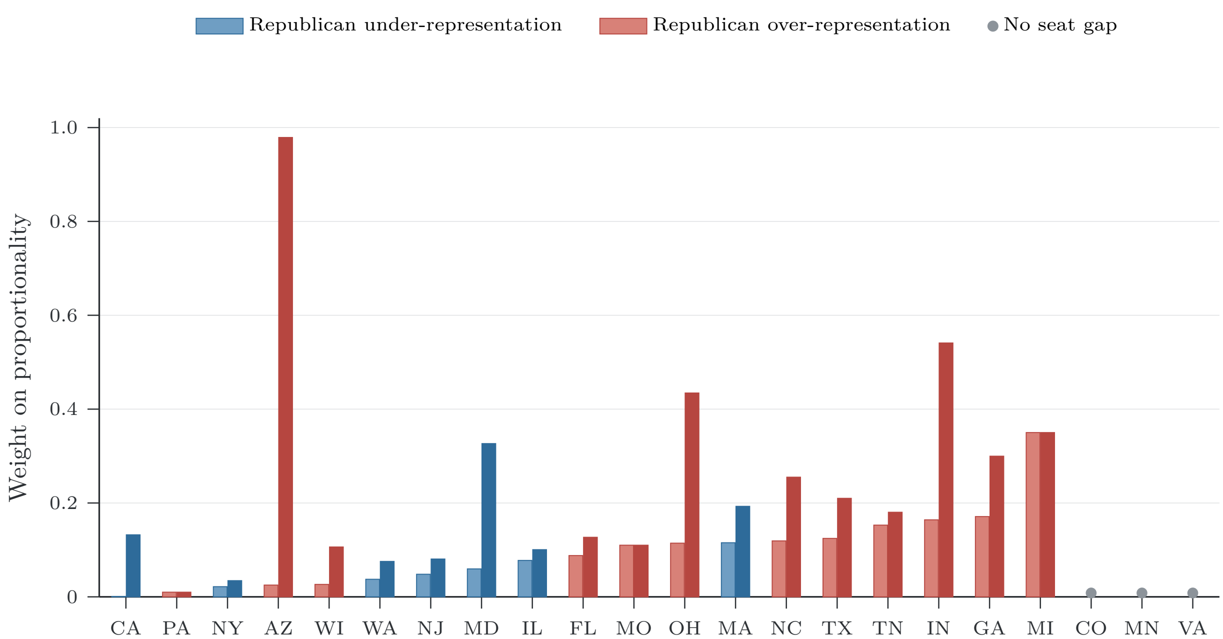


Figure 12: First-switch weight w_{S+1} and the weight required to attain proportional representation, w_{PR}, by state in 2024. Lower values indicate that less weight on proportionality is required. Gray markers denote no seat gap.

Figure 12 compares the first switch weight with the proportional switch weight, $w_{S^{\text{State}}}$, the weight at which the proportional seat total S^{State} becomes optimal. The first switch weight records when the observed first-past-the-post allocation first ceases to be optimal; the proportional switch weight records how much farther the objective must move before proportionality itself is selected.

The gap between the two weights describes the adjustment path. When the weights are close, the first move away from first-past-the-post nearly reaches proportionality, as in Michigan, Pennsylvania, and Colorado. Arizona illustrates the opposite pattern: the first switch weight is low, so the first-past-the-post allocation is fragile, but the proportional switch weight is high, so reaching proportionality requires much more weight on statewide representation.

H.2 Implied weights over time

Figure 13 plots the full sequence of complementary switch weights over 2000–2024 for four states that have been central to redistricting litigation. Starting from the first-past-the-post seat total S^{Dist}, each circle marks another switch along the optimal path toward the proportional seat total S^{State}. The black square marks the final switch. A green diamond marks a state-year with a zero seat gap, for which no switch is required.

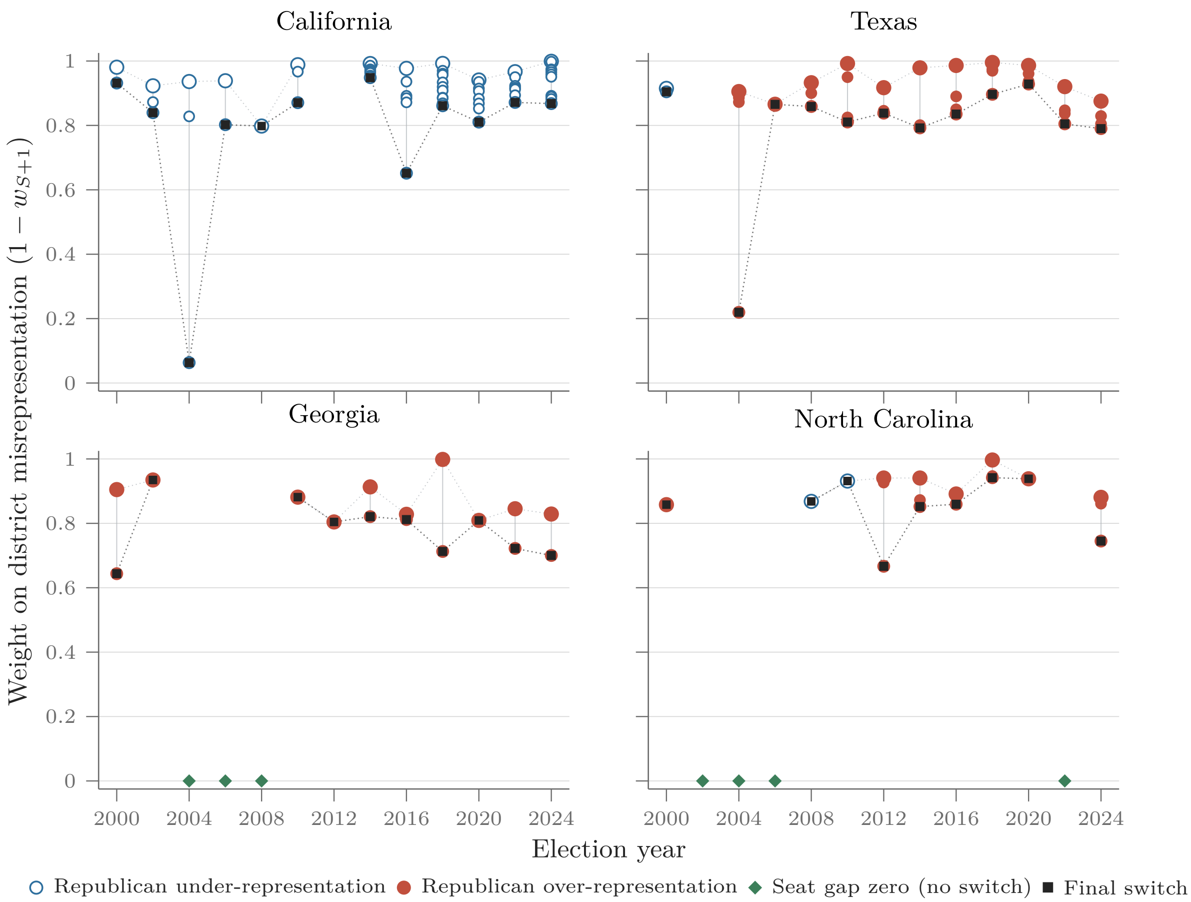


Figure 13: Complementary switch weights over time for California, Texas, Georgia, and North Carolina. Each circle is a switch along the optimal seat path from S^{Dist} to S^{State}; the black square marks the final switch. Hollow blue and filled red circles denote Republican under- and over-representation, respectively. A green diamond marks a zero seat gap and hence no switch. Missing elections break the connecting paths.

North Carolina is the clearest illustration because its map-drawing authority changes while its seat count holds at thirteen through 2020 and rises only to fourteen after the 2020 census. Under Democratic-drawn lines its delegation is exactly proportional through 2006 (seat gap zero), then the 2008 and 2010 Democratic-wave elections leave Republicans one seat short of proportional. In 2012 with the seat count unchanged at thirteen, a newly enacted Republican map flips the delegation to a three-seat Republican over-representation, holding the first switch weight low ($w_{S^{\text{Dist}}+1} = 0.06$), and it stays low through three cycles, with maps introduced for 2014 2016 and 2018 elections struck by courts. The court-drawn special-master plan returns the 2022 delegation to exact proportionality, and the legislature's 2024 map, enacted after the state court reversed course, moves it back to a three-seat Republican advantage.

Georgia shows a smaller but persistent Republican advantage that is generally robust. Proportional under Democratic maps from 2004 to 2008, its delegation runs one to two seats above proportional under every Republican map from 2012 on, and the first switch weight is moderate to high throughout ($w_{S^{\text{Dist}}+1}$ between 0.06 and 0.20 in most cycles), so the advantage rests on comfortably-won rather than narrowly-won seats. The seat gap widens from one seat to two after 2012 without the delegation becoming markedly more fragile.

Texas' 2004 mid-decade map opens a four-seat Republican advantage that persists through a decade of court-drawn interim maps and into the Republican maps of 2022 and 2024, largely unmoved by two rounds of reapportionment (thirty-two to thirty-six to thirty-eight seats). What changes is robustness: under the returning Republican legislature the first switch weight rises from about 0.08 in 2022 to 0.12 in 2024, and w_{PR} from 0.19 to 0.21, so the same four-seat advantage rests less on nearly-tied districts than it did under the court maps.

California is the outlier in size. With its 52-53 seats, Democratic advantage grows from one to three seats under the Democratic legislature of the 2000s to eight and then eleven seats under the independent commission of the 2010s and 2020s, and $1 - w_{S^{Dist}+1}$ sits near one in almost every commission-era cycle, so the advantage is large and fragile.

H.3 How measures differ

The efficiency gap and switch weight both are governed to some extent by the seat assignment. However, where the efficiency gap concerns itself with the full vote distribution, the switch weight to move away from first-past-the-post depends heavily on the most competitive races. Hence, they need not agree every time. Table 2 shows the percentile value of Efficiency Gap, the switch weight and aggregate misrepresentation for the election map struck down/ in litigation in courts. One could call these "bad maps" since they were considered unfair enough to be struck down. To illustrate the similarities and differences, we focus on some extreme examples below.

When the measures agree. Where the statewide vote is near half, the wedge is small and efficiency gap tracks the seat gap. And where the pivotal district is nearly tied, the switch weight tracks how fragile that seat gap is. North Carolina can be used to show both extremes (Figures 8 and 9). Its 2018 map records an efficiency gap of about 27%, a three-seat Republican advantage and Republicans hold the ninth seat at 50.2%, so $w_{S^{\text{Dist}}+1} = 0.003$. The statewide vote is almost exactly even ($\bar{p} = 0.5003$), the wedge is negligible, and both measures sit at the extreme for a map that was struck as a partisan gerrymander. Its 2022 court-drawn map is the opposite corner. A proportional seat assignment, an efficiency gap of -4.4% (inside the band), and no seat to flip, so the switch weight is undefined. Both measures would read this map as *"fair"*.

Disagreement I: Proportional assignment with lopsided vote profile.
There can be vote distributions such that the seat gap is zero, the switch weight is undefined and the efficiency gap equals the wedge alone (Proposition 15(i)). Illinois 2008 is the cleanest such profile. Republicans win seven of nineteen seats against a proportional target of 7.2, an exactly proportional delegation, but they hold only around 38% of the statewide vote. So the wedge is $+0.11$ and the efficiency gap is $+10.9\%$, past the threshold and Republican-favorable. In both figures Illinois 2008 sits on the seat-gap-zero line while the efficiency gap places it above the 75th percentile. The efficiency gap reads a proportional delegation as strongly skewed because its target withholds from the statewide minority the winner's bonus it would grant a majority (Proposition 9).

Disagreement II: Fragile seat assignment with lopsided vote profile.
When a map delivers a large but knife-edge advantage to the statewide majority, the seat gap and the switch weight both flag the map while the efficiency gap clears it. Texas 2010 gives Republicans a four-seat advantage resting on a pivotal district won by a small margin, 50.4% ($w_{S^{\text{Dist}}+1} = 0.008$), yet its efficiency gap is only $+3.3\%$. This is because a large Republican statewide share earns a correspondingly large winner's bonus. New York 2010 is the mirror image on the Democratic side: a three-seat Democratic edge held by a district Democrats win 50.2 to 49.8 ($w_{S^{\text{Dist}}+1} = 0.003$), and an efficiency gap of $+3.4\%$, slightly republican favoring. In both, the efficiency gap adds together the seat advantage and the majority's bonus resulting in a relatively lower value, while the seat gap records the size of the advantage and the switch weight records how thin the seats delivering it are. In Figure 9 the two maps lie well within the 7% band but are above the 90th percentile in inverse switch weight.

Disagreement III: Robust assignment with lopsided vote profile. The reverse case is a large advantage built on safe seats, where the efficiency gap flags the map but the switch weight is relatively robust. (Proposition 15(iii)). Georgia 2024 has an efficiency gap of +11.1% and a two-seat Republican advantage, but its two narrowest Republican seats are won at 61.9% and 60.3%, comfortably clear of a tie, so $w_{S^{\text{Dist}}+1} = 0.171$. The advantage is robust to a larger range of proportionality weights.[35]

Table 2: Percentile ranks (0–100, higher = more extreme) for litigated and struck congressional maps. Columns report the efficiency gap, inverse first switch weight, and proportional seat gap.

Map	Efficiency Gap	Inverse First Switch Weight	Proportional Seat Gap
GA2024	73	36	61
IL2022	**91**	76	**94**
IL2024	**92**	62	**95**
NC2024	**93**	48	78
WI2022	**99**	78	64
WI2024	**99**	81	63
GA2022	70	40	60
NC2012	**97**	68	76
NC2014	**98**	68	86
TX2004	43	56	90
VA2012	**97**	63	74
VA2014	88	42	67
FL2012	79	85	88
FL2014	57	85	81
NC2016	**96**	51	86
NC2018	**100**	**98**	90
NC2022	29	10	14
PA2012	**99**	79	**93**
PA2014	90	51	89
PA2016	**94**	61	90
Upper-tail counts among litigated or struck maps			
Above 90th percentile	12	1	3
Above 95th percentile	8	1	0

[35] Georgia's 2021 congressional map was struck on October 26, 2023 by the U.S. District Court for the Northern District of Georgia (Jones, J.) under Section 2 of the Voting Rights Act, in consolidated litigation led by *Alpha Phi Alpha Fraternity, Inc. v. Raffensperger*, for diluting Black voting strength; the General Assembly enacted the remedial map used here in December 2023, which the same court approved. Both the liability and remedial rulings are on appeal to the Eleventh Circuit (oral argument May 15, 2025), and remain pending. The challenge is thus a racial-vote-dilution claim under the Voting Rights Act, distinct from the partisan-gerrymandering standard the efficiency gap was designed for. Case materials: https://redistricting.lls.edu/case/alpha-phi-alpha-vraffensperger and https://www.aclu.org/cases/alpha-phi-alpha-fraternity-inc-v-raffensperger.