Electoral Maldistricting

Andrei Gomberg, Romans Pancs, Tridib Sharma*

November 18, 2020

Abstract

We introduce a framework to examine, both theoretically and empirically, electoral maldistricting. Maldistricting is defined as districting in pursuit of a partisan objective at the expense of voter welfare. Analysis is performed on the set of implementable (via some district map) legislatures, which we characterize both geometrically (via majorization) and in information theoretic terms. Drawing on data from the 2008 presidential election and the 2010 census–based districts, we compute our index for 42 U.S. states and find that observed districting predominantly favors Republicans over Democrats. In three case studies, our index aligns with courts' purported motivations for requesting redistricting.

Keywords: maldistricting, electoral districting, gerrymandering

JEL Classification Numbers: D72, D82

1 Introduction

Many legislative chambers around the world are elected based on geographic districts whose boundaries are regularly redrawn in order to account for population changes over time. In the case of the United States, such redrawing is required after each decennial census, both for the House of

*Gomberg (gomberg@itam.mx), Pancs (rpancs@gmail.com), and Sharma (sharma@itam.mx) are all at ITAM. We thank Matthew Elliot and Ben Golub for urging us to set up a model that we could solve, and conference participants at the Southern Economic Association 89th Annual Meeting, the Sixth Latin American Workshop in Law and Economics, the 2020 Canadian Economic Theory Conference, and the International Seminar Series on Social Choice for comments. For their incisive feedback, we also thank Emir Kamenica, Joel Sobel, Shaun McRae, Thomas Wiseman, and Michael Peress. Dave Bradlee of *Dave's Redistricting* has generously released the electoral and demographic data that we use. Carlos Aburto has provided excellent research assistance.

Representatives and for the state legislatures, and must deliver equipopulous districts.[1] Over the years, this regular redistricting has sparked numerous controversies due to alleged malpractice by the political actors involved in it. Indeed, because voter ideology varies across space, the choice of district boundaries affects the ideological composition of the elected legislature. Naturally, the problem has been a subject of many studies in law, political science, and economics.[2]

In this paper, we propose an index for the intent to maldistrict. The index takes a map of electoral districts and associates with it a relative likelihood that this map has been drawn with the intent to advance partisan interests (maldistricting) rather than to maximize voter welfare (well-districting). In economics, the recovery of intent (e.g., preferences from consumer choices) is a classical problem. In criminal law, intent is a prerequisite to guilt according to the *mens rea* ("guilty mind") liability doctrine. In the context of electoral districting, intent as a prerequisite for illegal partisan districting was suggested in the plurality opinion in *Davis v. Bandemer,* 478 U.S. 109 (1986). While partisan gerrymandering remains a nonjusticiable political question at the federal level (*Rucho v. Common Cause*, U.S., 2019), it is has been declared unconstitutional at the state level in Pennsylvania (*League of Women Voters of Pennsylvania v. Commonwealth of Pennsylvania*, 2018) and in North Carolina (*Common Cause v. Lewis*, 2019), and, perhaps, would be deemed justiciable in more states if a principled way to determine intent existed.

Our examination of intent is conducted in a model in which a districter partitions a set of locations, each comprised of known numbers of Republican and Democrat voters, into electoral districts, each of which elects a representative, who goes on to sit in a legislature. The legislature votes on a policy. Maldistricting is an act of optimization: maximization or minimization (depending on the districter's partisanship) of the policy.[3] The novelty of our approach is in understanding this act in relation to another act of optimization: the maximization of voter welfare. Welfare

[1] In the U.S., the requirement to equalize district populations dates back to the Supreme Court's decisions in *Baker v. Carr*, 369 U.S. 186 (1962), according to which redistricting is justiciable, and to the consequent decisions in *Wesberry v. Sanders*, 376 U.S. 1 (1964), and *Reynolds v. Sims*, 377 U.S. 533 (1964), which require equipopulous districts for the U.S. Congress and for the state legislatures, respectively.

[2] For a survey of the political science literature on the consequences of the "reapportionment revolution" started by *Baker v. Carr* see Cox and Katz (2002). Much research has been devoted to the topic of biased district drawing (commonly referred to as gerrymandering); an important early paper is Owen and Grofman (1988). Much recent work has concentrated on formulating criteria for identifying gerrymandering (Grofman and King, 2007; Chen and Rodden, 2013; Stephanopoulos and McGhee, 2015; Cervas and BernardGrofman, 2020, among others). We survey the closely related literature in Section 6.

[3] We also discuss how to model alternative expressions of partisanship, such as seat maximization and the protection of incumbents.

maximization strikes a balance between each voter's concerns for the policy and for representation, with the latter captured by the ideology of the district representative whom the voter elects. While the payoff-relevance of the policy is standard in political economics, the payoff-relevance of representation has a long pedigree in political science (Chamberlin and Courant, 1983; Monroe, 1995) and is implicit in many a legal opinion on electoral districting and in legislation (e.g., the 1965 Voting Rights Act).

The index of maldistricting is defined as (a monotone increasing function of) the distance of an observed district map from the closest welfare-maximizing district map relative to the closest maldistricted district map. Welfare-maximizing maps are generated by the social welfare functions with varying weights assigned to the policy and representation concerns. Maldistricted maps include those favored by the Republican and by the Democrat districters. The appropriate notion of distance between an observed district map and an intended one (either welfare-maximizing or maldistricted) is the measure of voters who would have to move for the observed map to induce the same composition of the legislature that the intended map would have induced. This distance is easy to compute as a Manhattan distance (the L_1 distance) between two vectors that represent the corresponding legislatures and is motivated by voter migration.

We refrain from advocating a specific threshold for the index to condemn a legislature as maldistricted. Instead, we envisage (and perform) comparisons across legislatures and over time.

We emphasize that our index quantifies districter's intent, not the damage done to voter welfare. The associated welfare loss can be readily assessed using the tools that we develop but is not our focus.

District shapes play no role in our analysis.[4] While in most U.S. states the law mandates that electoral districts be connected sets, this is a weak restriction because disconnected sets can be approximated by connected ones. In particular, Sherstyuk (1998, Proposition 4) identifies the conditions under which "[t]here exists, generically, a connected map with the value of the objective function close to the one of an optimal map." A fortiori, we also neglect "compactness," which stands for a "degree of convexity" (operationalized by Chambers and Miller, 2010, and Fryer and Holden, 2011, among others) and for the absence of elongation. Compactness is not particularly

[4]Therefore, we eschew the term "gerrymandering," which is loaded with geometric connotations (its etymology references "salamander"), in favor of "maldistricting."

restrictive in commonly studied districting settings. For instance, the districter who represents the majority party can partition the map into convex districts so that his party is in the majority in *each* of these districts (Soberon, 2017, Theorem 2). Vickrey (1961) is a classical illustration of the impotence of shape restrictions, while Alexeev and Mixon (2018) is a recent exemplar. In our setting, we show that even mechanically drawn compact district maps exhibit a partisan bias. In neglecting district shapes, we conform to the literature that we build on.

We position our paper in the economics literature here and discuss additional related papers in Section 6. Conceptually, our paper inverts Friedman and Holden (2008) and Coate and Knight (2007). Friedman and Holden ask: How to district in order to meet a certain partisan objective? Coate and Knight ask: How to district to maximize a certain social welfare function? We ask: Given the observed district map, what is the likely maximand?

Our model is chosen to be tractable and to depend only on the parameters that have direct counterparts in the data, comprised, precinct by precinct, of the number of voters who reveal themselves as Republicans by voting for a Republican president and the number of voters who reveal themselves as Democrats by voting for a Democratic president. Our setting is closest to Gul and Pesendorfer's (2010) (whose is a special case of Friedman and Holden's 2008). If one labels Gul and Pesendorfer's voters as locations, then the unit of attention becomes the same in their model and ours. The difference between the two models is that each of Gul and Pesendorfer's district representatives has an exogenous ideology, Republican or Democrat, and is elected with some probability, whereas in our model, this probability is replaced with the ideology of the elected representative. The models are similar enough for the sets of all feasible districting outcomes to be described by analogous majorization conditions.[5]

In our model, feasible district outcomes can be summarized by the set of implementable legislatures, which we must first characterize in order to then maximize voter welfare or the partisan objective function, as appropriate. This set is formally related to the set of posterior beliefs that can be induced in a model of Bayesian persuasion (Kamenica and Gentzkow, 2011; Kamenica, 2019). In particular, the mean voter theorem (Hinich, 1977) puts the ideology of a district representative in correspondence to the mean of a posterior belief in a model of Bayesian persuasion—which is

[5]Other modeling differences include Gul and Pesendorfer's two competing districters and uncountably many districts versus our one districter and finitely many districts.

the case analyzed by Gentzkow and Kamenica (2016). Majorization theory plays a role both in our characterization and, as pointed out by Kleiner, Moldovanu and Strack (2020), in Bayesian persuasion. The contemporaneous work of Kolotilin and Wolitzky (2020) exploits the parallels between Bayesian persuasion and electoral districting in order to generalize the models of Friedman and Holden (2008) and Gul and Pesendorfer (2010).

The paper is organized as follows. Section 2 describes how a legislature is formed and how a policy is chosen. Section 3 offers multiple characterizations of the set of implementable legislatures, each of which can be induced by some district map, and reports a comparative statics result. One of the characterizations is information-theoretic: a legislature is implementable if and only if it is less informative about voter ideology than a certain extreme legislature. The comparative statics result then says that more informative geography enlarges the set of implementable legislatures. Section 4 characterizes the sets of maldistricted and well-districted legislatures, which are inputs into the index of maldistricting that we propose. Section 5 empirically validates our index by showing that the value of the index fell in North Carolina, Texas, and Virginia after the courts of law suspected foul play and called for redistricting. This section also shows that both the observed legislatures and the "natural" ones (induced by district maps with regularly shaped districts, which we also construct) tend to favor Republicans over Democrats, and explains why. Section 6 concludes by clarifying the relationship between our model and the existing literature; we also discuss literature as we develop our theory. Proofs of facts and propositions that are missing from the main text are in Appendix A.

2 A Model of An Electoral System Based on Geographic Districts

The model is concerned with the decision problem of a districter who, by choosing a district map, affects the composition of a legislature and the policy that the legislature adopts. The mapping from a district map to the legislature is motivated by the unmodeled behavior of voters, candidates, and the elected representatives.

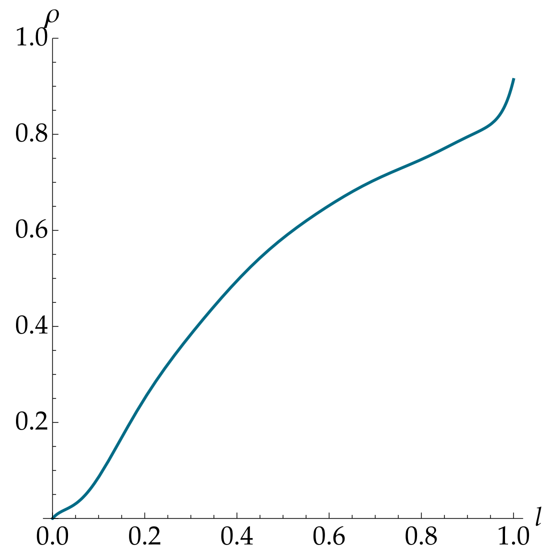


Figure 1: **An affiliation function.**

2.1 District Maps and Legislatures

A unit measure of voters are distributed uniformly on the unit square $\mathcal{L} \times [0,1]$, where $\mathcal{L} \equiv [0,1]$ is the set of geographic locations, with a typical location denoted by l. Each voter is of an ideology $x \in \mathcal{X} \equiv \{0,1\}$. The proportion of voters with the ideology $x = 1$ at a location l is $\rho(l)$, where the **(ideological) affiliation function** ρ is nondecreasing.[6,7] The mean ideology in the population is denoted by $R \equiv \int_{\mathcal{L}} \rho(l)\,\mathrm{d}l$. Figure 1 illustrates an affiliation function.[8]

The set of electoral districts is $\mathcal{K} \equiv \{1,2,\ldots,K\}$, with a typical element k, where K is odd. A function $g : \mathcal{L} \to \mathcal{K}$ maps locations into districts. Its inverse is $g^{-1}(k) \equiv \{l \in \mathcal{L} \mid g(l) = k\}$ for all $k \in \mathcal{K}$. The function g is a **district map** if, for every district $k \in \mathcal{K}$, $\int_{g^{-1}(k)} \mathrm{d}l = \frac{1}{K}$, ensuring that all districts are equipopulous. Let G be the set of all district maps.

A **legislature** is a collection $\mathbf{r} \equiv (r_1, r_2, \ldots, r_K)$, in $[0,1]^K$, of ideologies of district representatives. We assign no significance to district labels and, therefore, without loss of generality, restrict attention to legislatures that are **ordered**: $r_1 \le r_2 \le \ldots \le r_K$. A legislature $\mathbf{r}$ is **implemented by a district map** g if it is the collection of induced district means: for every district $k \in \mathcal{K}$, $r_k = K \int_{g^{-1}(k)} \rho(l)\,\mathrm{d}l$. That is, by assumption, each elected representative's ideology, in $[0,1]$,

[6]Definitions (and later vectors) are in bold.

[7]Friedman and Holden (2008) assume that the districter observes a noisy signal about each voter's ideology but not the ideology itself. In our model, voter's location is such as signal.

[8]The depicted affiliation function corresponds to Georgia. Section 5 explains its construction.

matches the mean voter ideology in his district. A **legislature is implementable** if some district map implements it.

Given a legislature, the ideology of the median legislator is called the **policy**: $p \equiv r_M$, where $M \equiv (K+1)/2$ is the index of the median-ideology district. A **policy is implementable** if it is associated with some implementable legislature.

An agency, called the **districter**, knows the model and chooses a district map g from the set G.

2.2 The Economic Content of the Model's Assumptions

Geography

The model encapsulates geography in the assumption that a districter cannot cherrypick individual voters on the basis of their ideology. Instead, voters come in chunks called "locations," which are indivisible geographic units to which voters are tied. One can interpret a location as the smallest geographic unit for which electoral data are available.

Confining locations to the unit interval and ordering them in the ascending order of voter ideology entails no loss of generality. Locations scattered in a space that is, say, two-dimensional can be ordered and mapped onto a unit interval.[9] Such a mapping destroys information about the geographic contiguity of locations. We have no use for this information in our analysis because we do not require each district to be a connected set.

Mixing Means and Medians

Our model implicitly assumes the mean-voter theorem at the district level and the median-voter theorem at the legislature level. Because there is a continuum of voters in each district, the pivotal-voter model would not discipline their behavior. Instead, we can assume that each voter tends to vote for the candidate whose ideology is closest to his.[10] In particular, each voter is concerned about the ideology of the winning candidate and is affected by each candidate's charisma, which is independent of the candidate's ideology. The candidates know voters' ideological preferences

[9]The assertion is easy to visualize when locations are finitely many. With uncountably many locations, (i) there exists a bijective map between a unit square and a unit interval, and (ii) while not every measurable function $f : \mathcal{L} \to [0,1]$ can be "sorted" to yield a nondecreasing affiliation function ρ, ρ can be interpreted as a weakly increasing rearrangement of f (Hardy and Littlewood, 1930, Part III, §8; Marshall, Olkin and Arnold, 2011, Definition 1.D.1).

[10]Such voting is called expressive, is a common modeling assumption (Fiorina, 1976, 1996), and is supported empirically (Pons and Tricaud, 2018).

but treat charisma as a random variable. When candidates' uncertainty about voter preferences is sufficiently large, the mean-voter theorem (Hinich, 1977; Duggan, 2017, Theorems 7, 10, and 11) predicts that, in a two-candidate contest, both candidates run on the ideologies that equal the mean ideology of voters.[11] As the uncertainty about voter preferences vanishes (as we believe it does in the legislature, whose members we treat, politically speaking, as known quantities), the candidates' equilibrium ideologies converge to the median instead (Banks and Duggan, 2005; Duggan and Jackson, 2005), consistent with our median-voter theorem assumption for the legislature.

Equipopulous Districts

The legal requirement that districts be equipopulous is substantive. Without it, "anything goes" in the sense of Proposition 1.

Proposition 1. *Suppose that the affiliation function is continuous and satisfies $\rho(0) = 0$ and $\rho(1) = 1$, and that districts need not be equipopulous. Then, any policy p in $(0, 1)$ is implementable.*

Proof. For any p in $(0, 1)$, let $[a, b]$ with $0 < a < b < 1$ be a set of locations whose mean ideology is p:

$$\frac{1}{b-a}\int_a^b \rho(l)\,\mathrm{d}l = p. \tag{1}$$

Such a set exists because p is interior and ρ is continuous. Partition $[0, a)$ and $(b, 1]$ arbitrarily into $M - 1$ nonempty districts each. The district $[a, b]$ is the median district. The constructed partition of $[0, 1]$ is a district map that implements the policy p. ■

The conclusion of Proposition 1 no longer holds when districts are equipopulous, as the characterization in Section 3 reveals.

[11] To illustrate, suppose that a district-k voter i of ideology $x_i \in \mathcal{X}$ (recall that $\mathcal{X} \equiv \{0, 1\}$) votes for candidate Alice of ideology $r_{Alice} \in [0, 1]$ over candidate Bob of ideology $r_{Bob} \in [0, 1]$ if and only if $-(r_{Alice} - x_i)^2 + \varepsilon_i > -(r_{Bob} - x_i)^2$, where ε_i is voter i's privately observed idiosyncratic preference for Alice over Bob. If each ε_i is distributed uniformly on $[-1, 1]$, then the number of votes that Alice amasses in a district k can be verified to be

$$\frac{1}{2} + \frac{1}{2}\int_{g^{-1}(k)} \left((r_{Bob} - \rho(l))^2 - (r_{Alice} - \rho(l))^2 \right) \mathrm{d}l,$$

which is maximized at $r_{Alice} = K \int_{g^{-1}(k)} \rho(l)\,\mathrm{d}l$, the district mean.

3 Implementable Legislatures

3.1 Characterization

Proposition 2 characterizes the polytope of implementable legislatures twice: by its vertices and by linear inequalities. The vertex characterization highlights some special legislatures, such as those that maximize welfare or deliver extreme policies. The inequalities characterization suggests an information-theoretic reinterpretation.

Both characterizations rely on the notion of an extreme legislature. A legislature $\mathbf{r}^e \equiv (r_1^e, r_2^e, \ldots, r_K^e)$ is **extreme** if $r_k^e = K \int_{\frac{k-1}{K}}^{\frac{k}{K}} \rho(l)\, \mathrm{d}l$ for each $k \in \mathcal{K}$. The extreme legislature is induced by the district map in which each district contains ideologically adjacent locations. Informally, the extreme legislature delivers the representatives who are maximally ideologically "spread out."

Proposition 2. *The following are equivalent.*

1. *$\mathcal{P}$ is a set of implementable legislatures.*

2. *$\mathcal{P}$ is the convex polytope X comprised of the legislatures majorized by the extreme legislature:*[12]

$$X \equiv \left\{ \mathbf{r} \in [0,1]^K \middle| \begin{array}{l} (i)\ (\forall m \in \mathcal{K}) \sum_{k=1}^m r_k^e \leq \sum_{k=1}^m r_k \\ (ii)\ \sum_{k \in \mathcal{K}} r_k^e = \sum_{k \in \mathcal{K}} r_k \\ (iii)\ r_1 \leq r_2 \leq \ldots \leq r_K \end{array} \right\}. \tag{2}$$

3. *$\mathcal{P}$ is the convex polytope each vertex $(r_1^v, r_2^v, \ldots, r_K^v)$ of which is induced by a partition of the extreme legislature $\mathbf{r}^e$ into intervals; each interval $\left(r_k^e, r_{k+1}^e, \ldots, r_{k+m}^e\right)$ in the partition induces $r_k^v = r_{k+1}^v = \ldots = r_{k+m}^v = \frac{1}{m+1} \sum_{i=k}^{k+m} r_i^e$.*

Proof. The equivalence between parts 2 and 3 is a result in majorization proved by Hoffman (1969, Theorems 1 and 2) and reported by Marshall, Olkin and Arnold (2011, Proposition 2.G.3.a, p. 58). A direct proof is given in Appendix A. The proof of the equivalence between parts 1 and 2 is also in Appendix A. ■

[12] An ordered vector $\mathbf{r}$ is **majorized** by an ordered vector $\mathbf{r}^e$ if conditions (i) and (ii) in (2) hold (Marshall, Olkin and Arnold, 2011, Definition 1.A.1). With a continuum of districts, the counterpart of the majorization inequality (i) is Gul and Pesendorfer's (2010) segregation constraint.

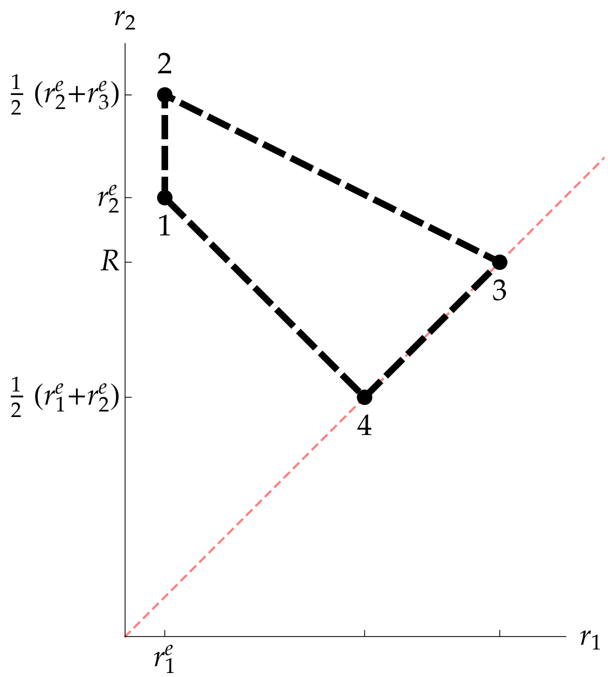


Figure 2: **The projection of the convex polytope of implementable legislatures when $K = 3$.** The projection has $2^{3-1} = 4$ vertices and is bounded by bold dashed line segments. The suppressed coordinate is $r_3 = 3R - r_1 - r_2$. Because ordered, all implementable legislatures lie above the 45-degree line.

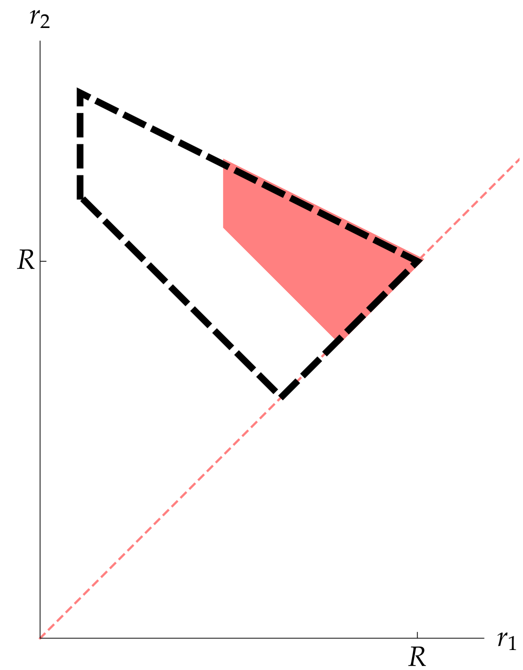


Figure 3: **More legislatures become implementable as geography becomes more informative about ideology.** The shaded polytope expands to become the polytope with the dashed boundary.

Part 3 of Proposition 2 immediately implies that the set of implementable legislatures is nonempty. The vertex $(R, R, \ldots, R)$ is always implementable.

Figure 2 illustrates the polytope of implementable legislatures for the affiliation function in Figure 1 assuming $K = 3$ districts. Vertex 1, obtained from the finest partition $\{\{r_1^e\}, \{r_2^e\}, \{r_3^e\}\}$, is the extreme legislature $\mathbf{r}^e$. Vertex 2, obtained from the partition $\{\{r_1^e\}, \{r_2^e, r_3^e\}\}$, is $\left(r_1^e, \frac{1}{2}(r_2^e + r_3^e), \frac{1}{2}(r_2^e + r_3^e)\right)$. Vertex 3, obtained from the coarsest partition $\{\{r_1^e, r_2^e, r_3^e\}\}$, is (R, R, R). Vertex 4, obtained from the partition $\{\{r_1^e, r_2^e\}, \{r_3^e\}\}$, is $\left(\frac{1}{2}(r_1^e + r_2^e), \frac{1}{2}(r_1^e + r_2^e), r_3^e\right)$.

The polytope of implementable legislatures has up to 2^{K-1} vertices, a handful of which we invest with economic significance later in the paper. To count the vertices, consider how they are constructed in part 3 of Proposition 2. There are K elements in $\mathbf{r}^e$ and, so, $K - 1$ "gaps" between adjacent elements. Each of these gaps can be turned On or Off, with On denoting the boundary between two adjacent intervals in a partition. There are 2^{K-1} On–Off constellations in total. When

ρ is strictly increasing, the induced 2^{K-1} partitions correspond to 2^{K-1} district vertices. When ρ is constant, all partitions induce the same vertex $(R, R, \ldots, R)$.

3.2 An Information Theoretic Interpretation

The equivalence between parts 1 and 2 of Proposition 2 can be restated in information-theoretic terms: A legislature is implementable if and only if it is less informative about voter ideology than the extreme legislature is. A legislature is a signal. Imagine an alien who lands in Georgia and meets a voter drawn uniformly at random. The voter proceeds to divulge the district where he votes, thereby revealing to the alien (who knows the model) something about his ideology. The informativeness of this revelation depends on how well the districts segregate voters of disparate ideologies or, equivalently, on how ideologically dispersed the induced legislature is.

Formally, let $(\Omega, \mathcal{B}, \mathbb{P})$ be the probability space. Here, $\Omega \equiv \mathcal{X} \times \mathcal{L}$ (the set of voter ideology–location pairs); $\mathcal{B} \equiv 2^{\mathcal{X}} \times B$, where B is the Borel σ-algebra on the set $\mathcal{L}$; and, for any interval $I \subset \mathcal{L}$, $\mathbb{P}(\mathcal{X} \times I) \equiv \int_I \mathrm{d}l$ and $\mathbb{P}(\{1\} \times I) \equiv \int_I \rho(l)\,\mathrm{d}l$. Define the random variables $\tilde{x} : \Omega \to \mathcal{X}$, a randomly drawn voter's ideology, and $\tilde{k} : \Omega \to \mathcal{K}$, a signal about $\tilde{x}$.[13] This signal is **monotone** if

$$k < k' \implies \mathbb{E}\left[\tilde{x} \mid \tilde{k} = k\right] \leq \mathbb{E}\left[\tilde{x} \mid \tilde{k} = k'\right] \qquad \text{for all } k, k' \in \mathcal{K} \tag{3}$$

and **uniform** if all realizations are equiprobable, $\Pr\left\{\tilde{k} = k\right\} = \frac{1}{K}$ for all $k \in \mathcal{K}$.[14] For any two monotone uniform signals $\tilde{k}$ and $\tilde{k}'$, the signal $\tilde{k}'$ is **(weakly) more informative** than the signal $\tilde{k}$ if

$$\mathbb{E}\left[\tilde{x} \mid \tilde{k}' \leq k\right] \leq \mathbb{E}\left[\tilde{x} \mid \tilde{k} \leq k\right] \qquad \text{for all } k \in \mathcal{K}. \tag{4}$$

Under the assumptions made in this paragraph, condition (4) captures informativeness in Blackwell's (1951; 1953) sense.[15]

[13] To avoid ambiguities, a tilde distinguishes a random variable from its realization.

[14] Monotonicity is a normalization; uniformity is not.

[15] Because $\tilde{x}$ is binary, the convex stochastic dominance order on the distributions of posterior beliefs about $\tilde{x}$ (which is equivalent to the Blackwell order on the signals that generate these distributions) is equivalent to the mean-preserving spread order on the distributions of posterior expectations of $\tilde{x}$. Courtault, Crettez and Hayek (2006, Proposition 2) characterize the mean-preserving spread order for the discrete case as a condition on c.d.f.s (textbooks typically report the continuous case); here, this condition is equivalent to (4).

The realization of $\tilde{k}$ announces the district of residence of a voter drawn at random. This announcement induces a posterior belief about the voter ideology $\tilde{x}$. The support of the induced belief is the legislature $\mathbf{r}$ with $r_k = \mathbb{E}\left[\tilde{x} \mid \tilde{k} = k\right]$ for each $k \in \mathcal{K}$. When $\tilde{k}$ is monotone, the induced legislature is ordered. When $\tilde{k}$ is uniform, Bayes plausibility requires that $\frac{1}{K}\sum_{k\in\mathcal{K}} r_k = R$. The converse is also true: any ordered legislature that satisfies $\frac{1}{K}\sum_{k\in\mathcal{K}} r_k = R$ is induced by some signal $\tilde{k}$ (Kamenica and Gentzkow, 2011, Proposition 1), which is monotone and uniform.

To restate the equivalence of parts 1 and 2 in Proposition 2, for any ordered legislature $\mathbf{r}$ with $\frac{1}{K}\sum_{k\in\mathcal{K}} r_k = R$, let $\tilde{k}$ be a monotone uniform signal that induces it. For the extreme legislature $\mathbf{r}^e$, this signal is $\tilde{k}^e \equiv \sum_{k=1}^{K} 1_{\left\{\frac{k-1}{K} \leq \tilde{l} \leq \frac{k}{K}\right\}} k$, where $\tilde{l}$ is a random voter's location. Then, the inequality in condition (i) of equation (2) can be equivalently rewritten as an instance of (4):

$$\begin{aligned}
\sum_{k=1}^{m} r_k^e \leq \sum_{k=1}^{m} r_k \iff & \sum_{k=1}^{m} \mathbb{E}\left[\tilde{x} \mid \tilde{k}^e = k\right] \leq \sum_{k=1}^{m} \mathbb{E}\left[\tilde{x} \mid \tilde{k} = k\right] \\
\iff & \sum_{k=1}^{m} \mathbb{E}\left[\tilde{x} \mid \tilde{k}^e = k\right] \frac{1}{K} \leq \sum_{k=1}^{m} \mathbb{E}\left[\tilde{x} \mid \tilde{k} = k\right] \frac{1}{K} \\
\iff & \mathbb{E}\left[\tilde{x} \mid \tilde{k}^e \leq m\right] \leq \mathbb{E}\left[\tilde{x} \mid \tilde{k} \leq m\right],
\end{aligned} \tag{5}$$

meaning that $\tilde{k}^e$ is more informative than $\tilde{k}$ or, as we equivalently say, the legislature induced by $\tilde{k}^e$ is more informative than the one induced by $\tilde{k}$. Corollary 1 follows from (5) and Proposition 2.

Corollary 1. *An ordered legislature $\mathbf{r}$ that satisfies $\frac{1}{K}\sum_{k\in\mathcal{K}} r_k = R$ is implementable if and only if it is less informative than the extreme legislature.*

3.3 Comparative Statics

Proposition 3 shows that, as geography becomes more informative about ideology, the set of implementable legislatures expands. The geography captured by an affiliation function ρ' is **(weakly) more informative** than the geography captured by an affiliation function ρ if ρ' majorizes ρ:

$$\int_0^1 \rho'(l)\,\mathrm{d}l = \int_0^1 \rho(l)\,\mathrm{d}l \qquad \Longrightarrow \qquad (\forall s \in \mathcal{L}) \int_0^s \rho'(l)\,\mathrm{d}l \leq \int_0^s \rho(l)\,\mathrm{d}l. \tag{6}$$

Condition (6) parallels (4) and likewise captures Blackwell's informativeness, here of a voter location about his ideology.[16] Geography is **uninformative** when the affiliation function is flat (i.e., $\rho(l) = R$ for all $l \in \mathcal{L}$). Geography is **perfectly informative** when the affiliation function is an indicator function (i.e., $\rho(l) = \mathbf{1}_{\{l>1-R\}}$). The uninformative and the perfectly informative geographies are, respectively, minimal and maximal according to the informativeness order defined above.

Proposition 3. *More (in the inclusion sense) legislatures are implementable the more informative geography is about ideology. When geography is uninformative, the only implementable legislature is $(R, R, \ldots, R)$. When geography is perfectly informative, the set of implementable legislatures is maximal.*

Proof. Take any two affiliation functions ρ and ρ' that satisfy $\int_0^1 \rho'(l)\,\mathrm{d}l = \int_0^1 \rho(l)\,\mathrm{d}l$ and (6). Let $\mathcal{P}$ and $\mathcal{P}'$ be the sets of implementable legislatures for ρ and ρ', respectively. To show that $\mathcal{P} \subset \mathcal{P}'$, take an arbitrary legislature $\mathbf{r} \in \mathcal{P}$. For any $n \in \mathcal{K} \backslash K$,

$$\sum_{k=1}^{n} r_k \geq K \int_0^{\frac{n}{K}} \rho(l)\,\mathrm{d}l \geq K \int_0^{\frac{n}{K}} \rho'(l)\,\mathrm{d}l,$$

where the first inequality follows from $\mathbf{r} \in \mathcal{P}$, and the second inequality follows from (6). Taken together, the two inequalities imply $\mathbf{r} \in \mathcal{P}'$, thereby establishing $\mathcal{P} \subset \mathcal{P}'$.

The argument for $\mathcal{P} \subset \mathcal{P}'$ above implies that the minimal and the maximal sets of implementable legislatures are induced, respectively, by the minimally and the maximally informative geographies. These set are found by the direct substitution of the corresponding affiliation functions into (2). ■

In Figure 3, the polytope of implementable legislatures expands as the affiliation function of Figure 1 becomes more informative.[17] Because both affiliation functions share a mean, the corresponding polytopes share the vertex (R, R, R).

[16] When ρ and ρ' are invertible, (6) is equivalent to $\int_0^{\phi} \rho^{-1}(s)\,\mathrm{d}s \leq \int_0^{\phi} (\rho')^{-1}(s)\,\mathrm{d}s$ for all $\phi \in [0,1]$, which can be ascertained graphically. The latter collection of inequalities implicates the c.d.f.s $\rho^{-1}(s) = \Pr\{\rho(\tilde{l}) \leq s\}$ and $(\rho')^{-1}(s) = \Pr\{\rho'(\tilde{l}) \leq s\}$ and says that the random variable $\rho'(\tilde{l})$ is a mean preserving spread of the random variable $\rho(\tilde{l})$. Because $\tilde{x}$ is binary, the mean-preserving spread order on the distributions of posterior expectations of $\tilde{x}$ is equivalent to the Blackwell order on the underlying signals (as in Footnote 15), thereby implying that the signal $\rho'(\tilde{l})$ is more Blackwell informative about $\tilde{x}$ than the signal $\rho(\tilde{l})$ is.

[17] This less informative affiliation function is Arizona's. The more informative one is Georgia's.

4 An Index of Maldistricting

Our index of maldistricting is inspired by the relative likelihood test for a pair of hypotheses. One hypothesis is that the districter intends a well-districted legislature. The alternative hypothesis is that the districter intends a maldistricted legislature. The closer the observed legislature is to the closest maldistricted legislature, relative to the closest well-districted one, the more maldistricted the observed legislature is declared to be, and the higher the value of the index is.

4.1 The Index as a Relative Distance

The construction of our index is motivated by the fictitious sequence of events:

1. The districter draws a district map g, intending either a well-districted or a maldistricted legislature.

2. Voters move randomly.

3. Voters vote, and a legislature is observed.

In other words, the observed legislature equals the intended legislature plus "noise" due to unanticipated voter migration. More migration is assumed to be less likely.[18] The **distance** $d(\mathbf{r}, \mathbf{r}')$ between two legislatures $\mathbf{r}$ and $\mathbf{r}'$ is defined as the minimal (over g in G) measure of voters who must change districts for the district map g to induce $\mathbf{r}$ instead of $\mathbf{r}'$.

Fact 1. *The described distance between any two ordered legislatures* $\mathbf{r}$ *and* $\mathbf{r}'$ *that share a mean is the Manhattan distance*

$$d(\mathbf{r}, \mathbf{r}') = \sum_{k \in \mathcal{K}} |r_k - r'_k| \,. \tag{7}$$

Let $\mathbf{r}^{data}$ be an observed legislature. Let $\mathcal{W} \subset [0,1]^K$ and $\mathcal{M} \subset [0,1]^K$ be the disjoint sets of the legislatures that are, respectively, well-districted and maldistricted. Then, the **index of**

[18] In our model, the districter neglects voter migration, implicitly assumed to be sufficiently small to justify neglect. This neglect is not very costly. The districter's payoff is continuous in the measure of the voters who move because each winning candidate's ideology is continuous in the ideological composition of his district. The payoff continuity in our model is in contrast to the models that restrict the candidate ideology to be binary (e.g., Republican or Democrat). In those models, the districter who aims to win a district by, say, 0.1% of votes must plan for the possibility that the district can flip if its ideological composition changes ever so slightly.

maldistricting ι is defined as

$$\iota \equiv \frac{1}{1+\Lambda}, \qquad \text{where} \quad \Lambda \equiv \frac{\inf_{\mathbf{r}\in\mathcal{M}} d\left(\mathbf{r}^{data}, \mathbf{r}\right)}{\inf_{\mathbf{r}\in\mathcal{W}} d\left(\mathbf{r}^{data}, \mathbf{r}\right)} \tag{8}$$

and where $d\left(\mathbf{r}^{data}, \mathbf{r}\right)$ is the distance in (7). The index has its values in $[0, 1]$, with higher values corresponding to more maldistricting. The rest of this section populates the sets $\mathcal{M}$ and $\mathcal{W}$.

4.2 Maldistricted Legislatures

We entertain three motives for maldistricting:

1. A partisan districter seeks an extremal (i.e., maximal or minimal) policy.
2. A partisan districter seeks to maximize the number of representatives elected from his own party.
3. A bipartisan districter seeks to cement the positions of incumbents.

A legislature is **maldistricted**—is in the set $\mathcal{M}$ in (8)—if it is inspired by one of the maldistricting motives above and none of the well-districting motives introduced in the following subsection. The maldistricting motive we emphasize theoretically and (in Section 5) empirically is motive 1.

Policy Extremization

Recall that $M \equiv \frac{K+1}{2}$ is the index of the median-ideology district.

Proposition 4. *The minimal and the maximal implementable policies are, respectively,*

$$\underline{p} \equiv \frac{K}{M} \int_0^{\frac{M}{K}} \rho\left(l\right) dl \qquad \text{and} \qquad \bar{p} \equiv \frac{K}{M} \int_{1-\frac{M}{K}}^{1} \rho\left(l\right) dl.$$

Either extreme policy above is induced by packing the fraction M/K of the appropriately ideologically extreme (i.e., minimal for $\underline{p}$, maximal for $\bar{p}$) locations into M equal-ideology districts, with the remaining districts formed arbitrarily.[19] *Moreover, $\underline{p}$ weakly decreases and $\bar{p}$ weakly increases as K increases or as locations become more informative about ideology.*

[19]One can show that the distance from any legislature $\mathbf{r}$ to the closest legislature that induces $\underline{p}$ is $\sum_{k=1}^{M}\left|r_k - \underline{p}\right| + \sum_{k=M+1}^{K}\left(r_k^e - r_k\right)$ and to the closest legislature that induces $\bar{p}$ is $\sum_{k=1}^{M-1}\left(r_k - r_k^e\right) + \sum_{k=M}^{K}\left|r_k - \bar{p}\right|$.

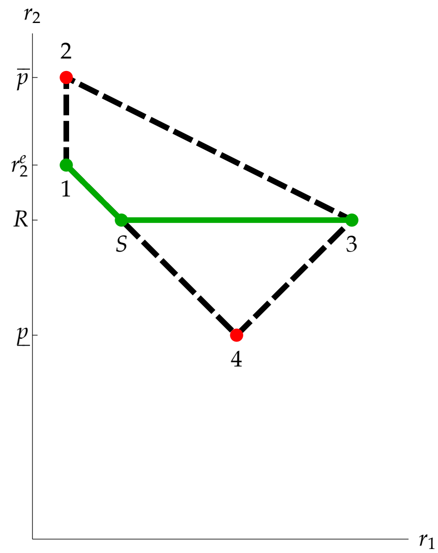


Figure 4: **Maldistricted and well-districted legislatures.** The maximal-policy and the minimal-policy legislatures are attained at the vertices 2 and 4, respectively. The set of socially optimal legislatures is the union of the line segments that connect the vertex 1 to the point S and then to the vertex 3.

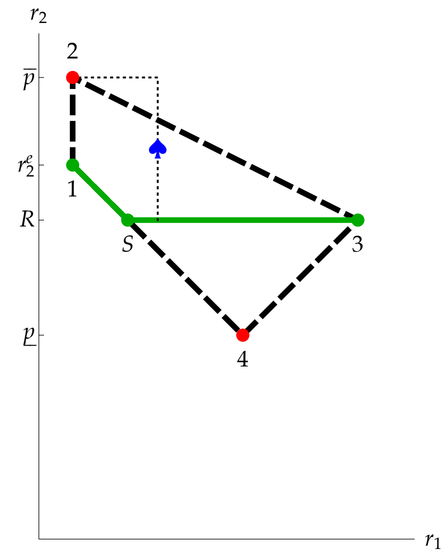


Figure 5: **The distance between legislatures.** The observed legislature is ♠. The dotted vertical line segment that emanates from ♠ and drops onto the line segment $S3$ is the distance from the observed legislature to the set of socially optimal legislatures. The crooked two-piece piecewise-linear dotted path from ♠ to the vertex 2 is the distance from the observed legislature to the set of maldistricted legislatures (which comprises the vertices 2 and 4). The maldistricting index is monotone in the ratio of the two distances.

For the affiliation function in Figure 1 and $K = 3$ districts, the legislatures that deliver the maximal and the minimal policies are the vertices 2 and 4, respectively, in Figure 4. In general, the extreme legislatures of Proposition 4 are the vertices in part 3 of Proposition 2 that are extreme in the dimension M, corresponding to the median district. The ability to form more districts and to better target voters on the basis of their ideology helps the districter pursue any objective, including policy extremization, as the "moreover" part of the proposition confirms.[20] The "moreover" part is a corollary to Proposition 3.

[20] In the language of Gilligan and Matsusaka (2006)—and consistent with their findings—the "moreover" part of Proposition 4 can be read to say that the "policy bias" is increasing in the number of districts, where the policy bias can be defined as either $\left|\underline{p} - R\right|$ or $|\bar{p} - R|$, the distance between an extreme policy and some fixed "fair" policy, such as R, the mean ideology in the population.

Seat Maximization

Our model has no political parties. Nevertheless, as a matter of interpretation, one can label a district-k representative a Republican if $r_k \geq \alpha$ and a Democrat if $r_k < \alpha$, for some $\alpha \in [0,1]$. Among the natural choices for α are $\alpha = \frac{1}{2}$ and $\alpha = \rho\left(\frac{1}{2}\right)$. The former honors symmetry but is otherwise arbitrary. The latter is motivated by (unmodeled) party competition, assumed to lead to the political platforms that would induce half of the locations to vote Republican and half to vote Democrat.

A Republican partisan districter is assumed to maximize the number of Republican representatives. This number, denoted by N_R, is $N_R = 0$ if $r_K^e < \alpha$. If $r_K^e \geq \alpha$, then

$$N_R = \max\left\{N \in \mathcal{K} \mid \frac{K}{N}\int_{\frac{K-N}{K}}^{1} \rho\left(l\right) \mathrm{d}l \geq \alpha\right\},$$

which is the maximal number of districts that can be constructed without the mean ideology in any of them dipping below α.[21]

To uniquely assign ideology to each district, one can apply a refinement that shares its motivation with our distance measure. The refinement assumes that the districter strives to maximize the ideology of the Republican representative with the lowest ideology, thereby minimizing the (unmodeled) probability that the Republican district flips to become Democratic if voters move. The lowest ideology of a Republican district is maximized by setting the ideology in each Republican districts to $\bar{r} = \frac{K}{N_R}\int_{\frac{K-N_R}{K}}^{1} \rho\left(l\right) \mathrm{d}l$. The refinement further assumes that the districter *maximizes* the probability that a Democratic district flips by maximizing the ideology of the highest ideology Democratic district and then, lexicographically, maximizing the probability that the second-highest ideology Democratic district flips, and so on. In the end, the partisan districter who is

[21] Gilligan and Matsusaka's (1999) partisan districter differs from ours in that theirs can reach inside locations to cherry-pick voters by ideology. Cherry-picking is a special case of our model when $\rho\left(l\right) = \mathbf{1}_{\{l \geq 1-R\}}$. With this ρ and with $\alpha = \frac{1}{2}$, the fraction of Republican districts (ignoring indivisibilities) is $N_R/K = \min\{2R, 1\}$, which is also the corresponding fraction in Gilligan and Matsusaka's model when the number of voters approaches infinity.

Republican aims for the legislature[22]

$$(r_1^e, \ldots, r_{K-N_R}^e, \underbrace{\bar{r}, \ldots, \bar{r}}_{N_R \text{ times}}). \tag{9}$$

The legislature favored by an analogously motivated Democratic districter is defined similarly.

Suppose that $\alpha = \frac{1}{2}$. Then, for the affiliation function of Figure 1 with $K = 3$, the legislature that is districted for Republican partisanship and is described by (9) corresponds to the vertex 2 in Figure 4. At that vertex, $r_1 \approx 0.27 < 0.5$ and $r_2 = r_3 \approx 0.60 > 0.5$, and, so, two representatives are Republicans. The legislature that is districted for Democratic partisanship corresponds to the vertex 3. At that vertex, $r_1 = r_2 = r_3 \approx 0.49$; all three representatives are Democrats.

Incumbency

A pro-incumbency districter is assumed to minimize the (unmodeled) probability that any district flips parties if voters move.[23] He does so by maximizing the ideological distance of each representative from α (the threshold from the discussion of partisanship above) subject to target numbers N_R and $N_D \equiv K - N_R$ of, respectively, Republican representatives and Democratic representatives.[24] The result of this maximization is the legislature

$$(\underbrace{\underline{r}, \ldots, \underline{r}}_{N_D \text{ times}}, \underbrace{\bar{r}, \ldots, \bar{r}}_{N_R \text{ times}}), \tag{10}$$

where $\underline{r}$, $\bar{r}$, N_R, and N_D satisfy

$$\underline{r} \equiv \int_0^{\frac{N_D}{K}} \rho(l)\, \mathrm{d}l \leq \alpha \qquad \text{and} \qquad \bar{r} \equiv \int_{\frac{K-N_R}{K}}^1 \rho(l)\, \mathrm{d}l \geq \alpha.$$

[22]Owen and Grofman's (1988) districter similarly expects the ideological composition of districts to change over time. To minimize the probability that Republican districts flip, their Republican districter seeks ideological uniformity across the districts he intends to win, just as our districter does. In addition, Owen and Grofman predict ideological uniformity of Democratic districts, whereas our districter lends Democratic districts distinct ideologies.

[23]The role of the incumbency motive in electoral districting has been asserted by Tufte (1973) and challenged by Ferejohn (1977).

[24]Owen and Grofman (1988) emphasize similarities between partisan and pro-incumbency districting. By contrasting the legislatures in (10) and (9), we highlight potential differences.

Suppose that $\alpha = \frac{1}{2}$. Then, for the affiliation function in Figure 1 and $K = 3$ districts, the vertex 3 in Figure 4 is pro-incumbent if $N_R = 0$, the vertex 4 is pro-incumbent if $N_R = 1$, and the vertex 2 is pro-incumbent if $N_R = 2$.

4.3 Well-Districted Legislatures

A legislature is well-districted—is in the set $\mathcal{W}$ in (8)—if it is induced by a district map that maximizes voter welfare, which balances the considerations of policy and representation, both of which are motivated by Powell (2000, pp. 7-9). Not to commit to an arbitrary resolution of the trade-off between the two considerations, we treat a class of social welfare functions as admissible.

Socially Optimal Legislatures

A **socially optimal district map** $g : \mathcal{L} \to \mathcal{K}$ maximizes the utilitarian social welfare function

$$-\gamma \underbrace{\int_0^1 \left[\rho(l)(1-p)^2 + (1-\rho(l))\, p^2\right] \mathrm{d}l}_{\text{aggregate disutility from a policy } p} - (1-\gamma) \underbrace{\int_0^1 \left[\rho(l)\left(1 - r_{g(l)}\right)^2 + (1-\rho(l))\, r_{g(l)}^2\right] \mathrm{d}l}_{\text{aggregate disutility from misrepresentation}} \tag{11}$$

for some $\gamma \in [0,1]$, where γ is the weight that every voter at any location l attaches to his utility from the policy p relative to his utility from the ideology-$r_{g(l)}$ representative in the district that contains the location l.[25] By not committing to a particular value of γ, we recognize a class of district maps as socially optimal.

Lemma 1 shows that a legislature summarizes all the welfare-relevant aspects of the district map that induces it.

Lemma 1. *The value of the social welfare function in (11) depends on a district map only through the legislature* $\mathbf{r}$ *that this district map induces. This legislature in turn enters the social welfare function through the quadratic term*

$$W(\mathbf{r}) \equiv \gamma\left(2Rr_M - r_M^2\right) + \frac{1-\gamma}{K} \sum_{k \in \mathcal{K}} r_k^2. \tag{12}$$

[25]The disutility from the district representative must be quadratic for the mean-voter theorem to imply, as we maintain, that the ideology of the elected representative is the district mean.

Proof. The proof proceeds by straightforward rearrangement of (11) and is omitted. ■

Lemma 1 motivates the definition: a **socially optimal legislature** maximizes (12) over implementable legislatures for some relative weight γ. A legislature is socially optimal if and only if it is induced by a socially optimal district map.

For the affiliation function in Figure 1 and $K = 3$ districts, the set of socially optimal legislatures in Figure 4 is the two-piece piecewise-linear path through the vertex 1, the point S, and the vertex 3. As γ rises from 0 to 1, the socially optimal legislature moves from the extreme legislature at the vertex 1 to the point S, at which the induced policy is R. When $\gamma = 1$, any legislature on the line segment that connects the point S and the vertex 3 induces the policy R and, therefore, is socially optimal. Figure 5 illustrates the construction of the maldistricting index in (8) when the set $\mathcal{W}$ comprises socially optimal legislatures, and the set $\mathcal{M}$ comprises the two legislatures that extremize the policy.[26]

With five or more districts, the polytope of implementable legislatures cannot be profitably visualized. An individual legislature can be drawn, though. An ordered legislature is a staircase in which the height of each step equals the ideology of the corresponding district representative. Proposition 5 shows that the extreme legislature—the "steepest" staircase that one can construct—is socially optimal when $\gamma = 0$. When $\gamma = 1$, optimality requires setting the policy to the population mean, which can be accomplished, in particular, by making each district to be a "representative sample" of all locations; the associated "staircase" is flat. When $\gamma \in (0, 1)$, the socially optimal legislature flattens the steepest staircase by replacing some of its steps with a horizontal segment. The case of $\gamma \in (0, 1)$ invokes Assumption 1.

Assumption 1 (Strict monotonicity)**.** *The affiliation function ρ is strictly increasing.*

Assumption 1 is not required for the conclusion of Proposition 5 but simplifies the number of cases that one must consider to prove part 3.

Proposition 5. *Let $\mathbf{r}^*$ denote a socially optimal legislature, which exists. Let $p^* \equiv r^*_M$ denote the corresponding optimal policy.*

[26] Because the r_3 axis points directly at the reader, the component of the Manhattan distance along this axis is invisible.

1. *When $\gamma = 1$ (i.e., only the disutility from policy affects welfare), any implementable legislature $\mathbf{r}^*$ with $p^* = R$ is socially optimal.*[27]

2. *When $\gamma = 0$ (i.e., only the utility from representation affects welfare) or when $\gamma \in (0,1)$ and $R = r^e_M$, the extreme legislature is uniquely socially optimal: $\mathbf{r}^* = \mathbf{r}^e$.*

3. *Suppose that Assumption 1 holds. Then, when $\gamma \in (0,1)$ and $R \neq r^e_M$, any socially optimal legislature is flattened, that is, of the form*

$$\mathbf{r}^* = \begin{cases} (r^e_1, \ldots, r^e_{M-1}, \underbrace{p^*, \ldots, p^*}_{q \text{ times}}, b, r^e_{M+q+1}, \ldots, r^e_K) & \text{if } R > r^e_M \\ (r^e_1, \ldots, r^e_{M-q-1}, b, \underbrace{p^*, \ldots, p^*}_{q \text{ times}}, r^e_{M+1}, \ldots, r^e_K) & \text{if } R < r^e_M, \end{cases} \tag{13}$$

where q, p^*, b are all chosen optimally. Moreover, as γ increases, q increases weakly and p^* can be selected (from the set of optimal policies, a singleton for almost all γ) to move monotonically away from r^e_M (which is the optimal policy when $\gamma = 0$) and towards R (which is the optimal policy when $\gamma = 1$). Always, p^* lies between r^e_M and R.

Proof. Because the set X of implementable legislatures defined in (2) is compact, and the social welfare function (12) is continuous, an optimal legislature exists; $\max_{\mathbf{r} \in X} W(\mathbf{r})$ has a solution for every $\gamma \in [0,1]$.

When $\gamma = 1$, the maximand (12), now $2Rr_M - r^2_M$, is maximized at $p^* = R$.

When $\gamma = 0$, the maximand (12) is strictly Schur-convex (Marshall, Olkin and Arnold, 2011, 3.C.1.a.(i)). Because any implementable legislature is majorized by $\mathbf{r}^e$, strict Schur-convexity implies that (12) is uniquely maximized by $\mathbf{r}^* = \mathbf{r}^e$ (Marshall, Olkin and Arnold, 2011, Definition 3.A.1).[28]

[27] One can show that the distance from any legislature $\mathbf{r}$ to the closest legislature that is socially optimal for $\gamma = 1$ is $2\sum_{k=M-q+1}^{M} \max\{0, r_k - R\}$ with $q \equiv \min\left\{i \geq 0 \mid \frac{1}{i+1}\sum_{k=M-i}^{M} r_k \leq R\right\}$ whenever $r_M > R$, and is $2\sum_{k=M}^{M+q-1} \max\{0, R - r_k\}$ with $q \equiv \min\left\{i \geq 0 \mid \frac{1}{i+1}\sum_{k=M}^{M+i} r_k \geq R\right\}$ whenever $r_M \leq R$. Here, q is the number of districts (in the closest optimal legislature) whose mean ideology is R.

[28] For an alternative, information theoretic argument, note that, when $\gamma = 0$, the social welfare function (12) is $\sum_{k\in\mathcal{K}} r^2_k/K$ and is the value function (sans a constant) for the decision problem $\min_{a\in[0,1]} \mathbb{E}\left[-(a - \tilde{x})^2 \mid \tilde{k}\right]$, where—following the interpretation in Section 3.2—$\tilde{x}$ is a random voter's ideology and $\tilde{k}$ is a legislature formulated as a signal. Because any decision maker prefers the most informative signal, the value function is maximized at the most informative legislature, which is $\mathbf{r}^e$, the extreme legislature.

When $\gamma \in (0,1)$, by the argument in the paragraph above, $\mathbf{r}^e$ continues to uniquely maximize the component of (12) that multiplies $(1-\gamma)$. In addition, when $R = r^e_M$, $\mathbf{r}^e$ maximizes the component of (12) that multiplies γ. As a result, when $\gamma \in (0,1)$ and $R = r^e_M$, $\mathbf{r}^e$ uniquely maximizes (12).

The (involved) proof of part 3 is in Appendix A. ■

Proposition 5 suggests a sniff test for the districter's failure to maximize welfare: if the observed policy falls outside the interval bounded by r^e_M and R, then welfare is not maximized, no matter the γ.

5 Empirical Analysis

This section interprets electoral data through the prism of the model developed above. To preview:

1. In most states, observed electoral districting favors Republicans over Democrats.
2. In most states, an impartial districting benchmark would also favor Republicans over Democrats.
3. In North Carolina, Texas, and Virginia, the index of maldistricting fell after courts of law requested redistricting.

Point 3 above provides external validation for our index of maldistricting.

Demographic and Electoral Data and Natural Legislatures

The unit of observation is an electoral precinct in the year 2008; precincts are the counterparts of the model's locations. We examine precincts from 42 states: Alabama (AL), Arkansas (AR), Arizona (AZ), California (CA), Colorado (CO), Connecticut (CT), Delaware (DE), Florida (FL), Georgia (GA), Hawaii (HI), Iowa (IA), Idaho (ID), Illinois (IL), Indiana (IN), Kansas (KS), Louisiana (LA), Massachusetts (MA), Michigan (MI), Minnesota (MN), Missouri (MO), Mississippi (MS), Montana (MT), North Carolina (NC), North Dakota (ND), Nebraska (NE), New Jersey (NJ), New Mexico (NM), Nevada (NV), New York (NY), Ohio (OH), Oklahoma (OK), Oregon (OR), Pennsylvania (PA), Rhode Island (RI), South Carolina (SC), Tennessee (TN), Texas (TX), Utah (UT), Virginia (VA), Washington (WA), Wisconsin (WI), and Wyoming (WY).[29] We use the 2010 census–based districts that are drawn to elect members of the lower house of each state's legislature (usually called the House of Representatives).[30] Each district elects one representative except for Arizona, Idaho, New Jersey, North Dakota, and Washington, whose districts elect a pair. We treat each pair as one, thus equating the number of representatives to the number of districts.[31]

An ideology at a precinct is the ratio of votes cast in the 2008 presidential election for the Republican candidate to the total number of votes cast for either the Republican candidate (McCain) or the Democratic candidate (Obama). The affiliation function ρ for each state is constructed by setting $\rho(l)$ to be the smallest ideology such that at least the fraction l of the state's population live in a precinct with an ideology that does not exceed $\rho(l)$. This construction assumes (counterfactually) that the same fraction of the population turn out to vote in every precinct; it is without loss of generality to further assume that this fraction is 100%.[32]

Dave's Redistricting (http://gardow.com/davebradlee/redistricting/) contains all the electoral and population data we need except for the partition of precincts into districts. To partition, we superimpose the map of districts on the map of precincts. Each precinct is allocated to the district that contains its "center" as defined by QGIS's plugin *realcentroid* (https://github.com/zsiki/realcentroid); the center is guaranteed to lie within the boundaries of the precinct, even for nonconvex precincts.[33] The relevant maps are available online from a variety of official sources (www.census.gov, catalog.data.gov, www.rigis.org, www.ncleg.gov, data.imap.maryland.gov, www.legis.iowa.gov, and azredistricting.org).

Observed districts induce observed legislatures and observed policies by applying the definitions of Section 2.

We define the **natural legislature** as the legislature induced by the natural district map, which is constructed following the shortest splitline algorithm, proposed by RangeVoting.org and described on their website. First, we identify each precinct with its "center" and then apply the

[29]For California, Montana, Oregon, and Rhode Island, we replace precincts with census block groups. California precincts are unusually large; census block groups are smaller. For Montana, Oregon, and Rhode Island, precinct data are incomplete.

[30]Nebraska's legislature is unique in that it is unicameral.

[31]In a suitable stochastic voting setting, m winners in an $(m+1)$-candidate race will all tend to be of the ideology that equals the district mean. Therefore, we treat $m = 2$ district representatives as one.

[32]We drop the precincts (0.05% by population) where no one voted for McCain or Obama.

[33]This procedure allocates factually incorrectly the precincts that in practice are split between multiple districts; North Carolina is a special offender.

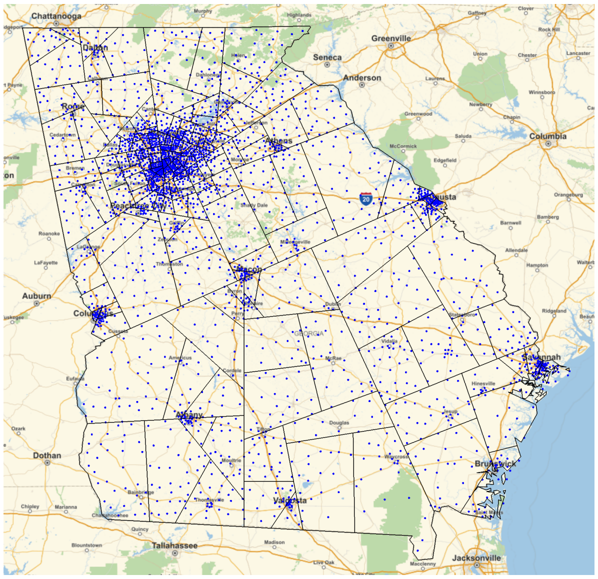


Figure 6: **Natural districts for Georgia.** Precinct "centers" (dots) are partitioned into natural districts (polygons except for state boundaries).

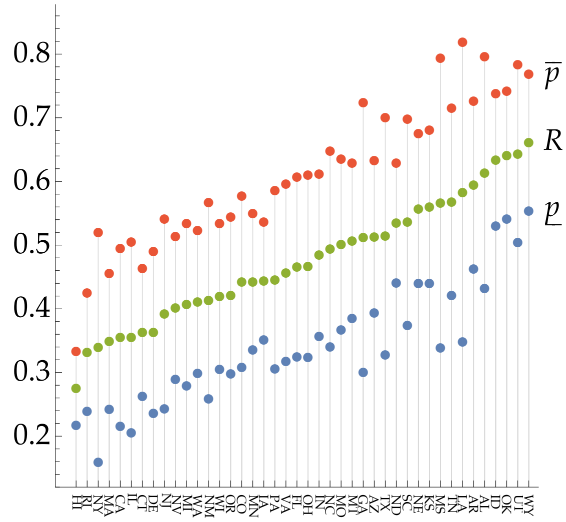


Figure 7: **Policy bands.** The minimal and the maximal implementable policies, $\underline{p}$ and $\bar{p}$, straddle the mean ideology in the state, R.

shortest splitline algorithm to allocate each precinct to a district. By construction, each natural district is a convex and "not too elongated" polygon (modulo state boundaries). Figure 6 depicts the natural district map for Georgia.[34]

Maldistricting (Mostly) Favors Republicans

For each state in our data set, Figure 7 plots the mean ideology in that state straddled by the minimal and the maximal implementable policies, as reported in Proposition 4. The mean is always implementable. The range of implementable policies in Figure 7 is substantially larger for Georgia than for Arizona, even though mean ideologies are about the same. Figure 8 indicates the reason: geography in Georgia is more informative about voter ideology than in Arizona. The conclusion of Proposition 3 applies.

Figure 9 plots the policies implied by the district maps observed in the data. These policies exceed state means in most states. The means are benchmark policies that correspond to populating districts by picking precincts uniformly at random. The discrepancy between the observed

[34] All maps are in Appendix B.

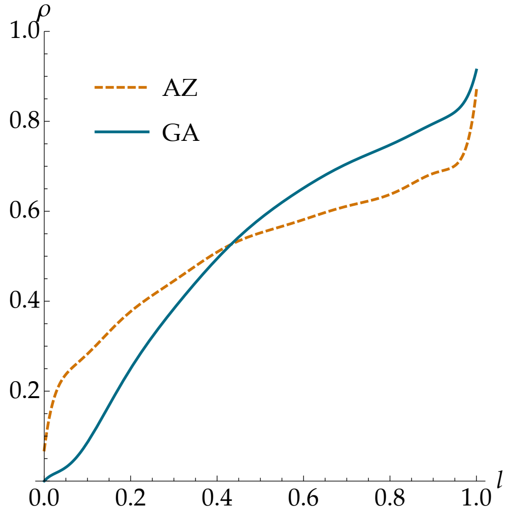


Figure 8: **Affiliation functions ranked by informativeness.** Geography in Georgia is more informative about voter ideology than in Arizona. For instance, at $l = 0.1$ or $l = 0.9$, there is more uncertainty about voter ideology (i.e., $\rho(l)$ is farther away from $\frac{1}{2}$) in Arizona than in Georgia.

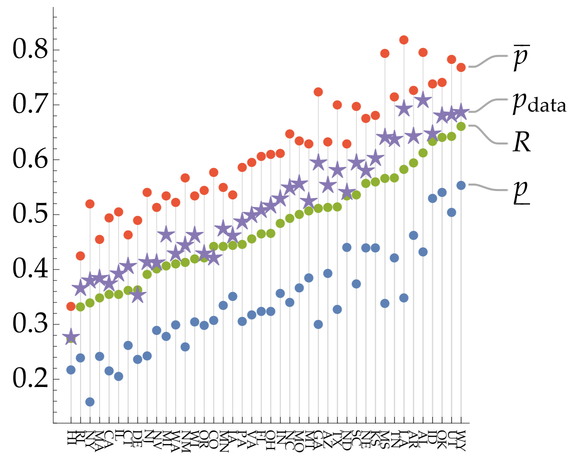
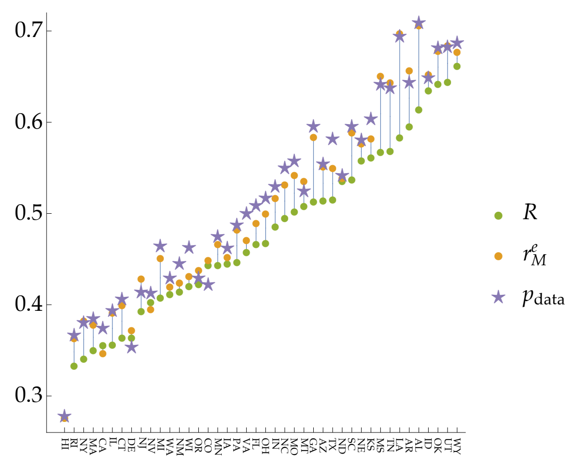


Figure 9: **Policies implied by the data.** The stars, labelled p_{data}, denote the policies calculated from the district maps in the data. The rest is as in Figure 7.

policies and state means need not betray maldistricting and may arise as a result of welfare maximization. To check whether observed policies are consistent with welfare maximization, Figure 10 appeals to Proposition A.7 to plot, for each state, the interval of the policies consistent with optimality: either $[r^e_M, R]$ or $[R, r^e_M]$, whichever is nonempty. The policies observed in the data are mostly outside these intervals, suggesting alternative motives for the districters.

While Figure 10 is suggestive of maldistricting, it fails to provide a measure of its extent. The figure also discards all the information contained in the observed legislature apart from one number: the policy. Figure 11 rectifies these shortcomings by reporting the values of the maldistricting index (8). The index is constructed by letting the set $\mathcal{M}$ of maldistricted legislatures to be the legislatures that extremize the policy, and by letting the set $\mathcal{W}$ of well-districted legislatures to be the legislatures that maximize of the social welfare function (11) for some γ. In the figure, the



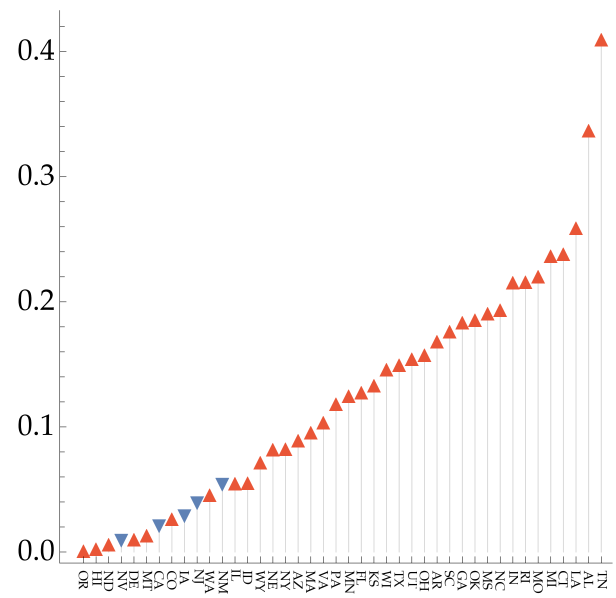


Figure 10: **Policies consistent with welfare maximization.** Vertical dumbbells are intervals $[r^e_M, R]$, which contain policies consistent with welfare maximization; r^e_M is the policy at the legislature that minimizes the disutility from misrepresentation. The rest is as in Figure 9 except the vertical axis has been magnified, and the minimal and the maximal policies have been dropped.

Figure 11: **Observed values of the maldistricting index.** Maldistricting extremizes the policy; well-districting maximizes a social welfare function. A triangle points upwards if the closest maldistricted legislature favors Republicans and points downwards if it favors Democrats.

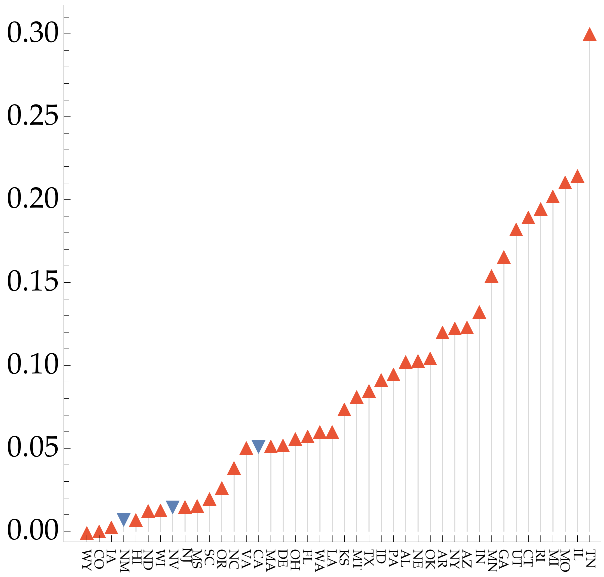


Figure 12: **Values of the maldistricting index for natural legislatures**. A triangle points upwards if the closest maldistricted legislature favors Republicans and points downwards if it favors Democrats.

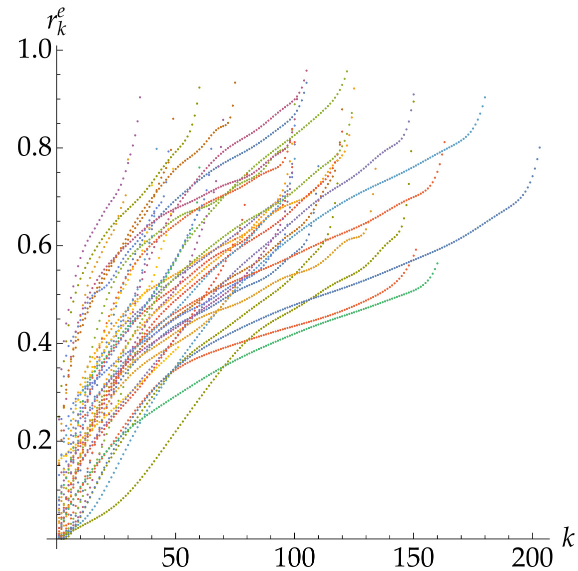


Figure 13: **Extreme legislatures for each state.** The legislatures exhibit "concave tendencies," thereby implying higher ideological variation across Democratic-leaning districts than Republican-leaning ones.

least maldistricted state is Oregon, and the most maldistricted one is Tennessee.[35] Most states are maldistricted to favor Republicans.

The Republican Bias is (Mostly) Shared by Natural Legislatures

Why are most states maldistricted to favor Republicans? Possibly, because Republican maldistricting may be easier to conceal. Natural district maps (defined above and illustrated in Figure 6) are politically easy to advocate. Deviations from natural districts' "compactness" (convexity and the lack of elongation) are condemned by both the court of law and the court of public opinion. If natural legislatures are biased towards Republicans, then so may be the observed ones.

Figure 12 confirms that most (39 out of 42) natural legislatures indeed favor Republicans. The comparison of Figures 11 and 12 reveals that all states with a Republican-biased natural legislature retain this bias in the observed legislature. By contrast, among the states with a Democratic-biased

[35] Unseen in Figure 11 is the fact that, in every state, the well-districted legislature that is the closest to the observed one is a socially optimal one for $\gamma = 1$. This empirical regularity owes to the multiplicity and the flexibility of socially optimal legislatures for $\gamma = 1$, contrasted to the almost uniqueness of the socially optimal legislatures for each $\gamma < 1$ (Proposition 5).

natural legislature, only those in which both chambers are controlled by Democrats manage to retain this bias in the observed legislature. The control of legislatures matters because these are typically state legislatures who draw district boundaries.

Why do natural legislatures favor Republicans? Leaving extensive theorizing on this subject to future research, we convey intuition in a fictional example.

Example 1 (Ideological Dispersion and a Republican Bias)**.** Let $K = 3$. Suppose that the tendency of ideologically similar precincts to cluster together is maximally strong, equating the natural legislature to the extreme one: $\mathbf{r}^{natural} = \mathbf{r}^{e}$. Proposition 4 and its attendant Footnote 19 can then be verified to imply that $\mathbf{r}^{e}$ is closer to a policy-maximizing legislature than to a policy-minimizing one if and only if $r_2^e - r_1^e > r_3^e - r_2^e$, when the extreme legislature $\mathbf{r}^{e}$ is "concave." $\triangle$

Example 1 suggests that the natural legislature favors Republicans when

1. similar-ideology precincts cluster together in space, causing the natural legislature to resemble the extreme one, and

2. the extreme legislature exhibits "concave tendencies."

Both conditions are approximated for most states in our data set. The empirical validity of condition 1 owes to the distribution of voter ideologies in space. Ideologically similar voters tend to live near each other. By construction, natural districts tend to bring together into the same district the voters who live nearby. The empirical validity of condition 2 is corroborated by Figure 13. The "concave tendencies" in the figure owe to the higher ideological variation across Democratic-leaning districts than Republican-leaning ones. This comparative variation can be explained by residential segregation by race, as Example 2 illustrates by specializing Example 1.

Example 2 (Ideological Dispersion and Racial Segregation)**.** Let $K = 3$. Label the locations in $\left[0, \frac{1}{2}\right]$ as urban and in $\left(\frac{1}{2}, 1\right]$ as exurban. Half of the voters are Republicans, and half are Democrats: $R = \frac{1}{2}$. Two thirds of Republicans prefer exurban locations. Two thirds of Democrats prefer urban locations. Assume that all urban lovers "mix uniformly" on $\left[0, \frac{1}{2}\right]$, where $\rho\left(l\right) = \frac{1}{3}$. Assume that all suburban lovers mix uniformly on $\left(\frac{1}{2}, 1\right]$, where $\rho\left(l\right) = \frac{2}{3}$. The extreme legislature $\mathbf{r}^{e} = \left(\frac{1}{3}, \frac{1}{2}, \frac{2}{3}\right)$ is equidistant from the policy-maximizing legislature $\left(\frac{1}{3}, \frac{7}{12}, \frac{7}{12}\right)$ and from the policy minimizing legislature $\left(\frac{5}{12}, \frac{5}{12}, \frac{2}{3}\right)$.

Now assume that African Americans comprise one sixth of the population and are all Democrats and urban lovers. Assuming residential segregation by race, all African Americans live in urban locations on $\left[0, \frac{1}{6}\right]$, where $\rho(l) = 0$. The remaining urban lovers, Democrats and Republicans, mix uniformly on $\left(\frac{1}{6}, \frac{1}{2}\right)$, where $\rho(l) = \frac{1}{2}$. All suburban lovers mix uniformly on $\left[\frac{1}{2}, 1\right]$, where $\rho(l) = \frac{2}{3}$. The extreme legislature $\mathbf{r}^e = \left(\frac{1}{4}, \frac{7}{12}, \frac{2}{3}\right)$ is closer to the policy-maximizing legislature $\left(\frac{1}{4}, \frac{5}{8}, \frac{5}{8}\right)$ than to the policy-minimizing legislature $\left(\frac{5}{12}, \frac{5}{12}, \frac{2}{3}\right)$. $\triangle$

In Example 2, all Republicans mix with Democrats, but not all Democrats mix with Republicans. As a result, there is more variety in ideology across Democratic precincts than there is across Republican ones.

Our finding that natural districts tend to favor Republicans is consistent with the findings of Chen and Rodden (2013), who document correlation between geographic concentration of Democratic voters and electoral bias in favor of Republicans.

The Index of Maldistricting Echoes Courts' Interpretations of District Maps in NC, TX, and VA

In North Carolina, Texas, and Virginia, the 2010 census–based district maps, drawn by a Republican-dominated legislature in 2011, were all struck down by federal courts for being an "impermissible racial gerrymander." Consistent with the courts' motivation, the maps that eventually replaced the 2011 ones are less maldistricted according to our index.[36,37] The path towards the final 2010 census–based map is different in each of the three states and is signposted in Figure 14.

In North Carolina, the halfhearted attempt by the state legislature in 2017 to obey federal courts' ruling against the 2011 map was struck down by the same courts, on the same grounds of racial gerrymander. In 2018, the US Supreme Court proposed an alternative map. It, too, was struck down by courts for being an "impermissible partisan gerrymander"—impermissible according to the newly interpreted state law, even though not according to the federal law. Finally, in 2019, the state legislature proposed a new map, to comply with state courts' requirements.

[36]The maps are at https://www.ncleg.gov/Redistricting for North Carolina, at https://data.capitol.texas.gov/organization/tlc for Texas, and at https://catalog.data.gov/dataset/tiger-line-shapefile-2017-state-virginia-current-state-legislative-district-sld-lower-chamber-s and https://redistricting.dls.virginia.gov/CensusDownload.aspx for Virginia (all accessed on 4 June 2020).

[37]We thank Thomas Wiseman for suggesting this exercise.

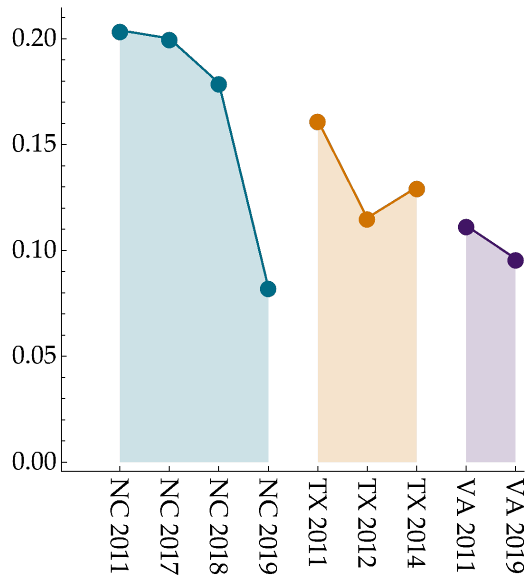


Figure 14: **The evolution of the index of maldistricting in NC, TX and VA.**

Texas's story resembles North Carolina's but has a twist. The alternative to the 2011 map was proposed in 2012 by the federal courts. The US Supreme Court struck down this map because it ignored state legislature's preferences. The legislature itself then produced the 2014 map, thereby generating an uptick in the maldistricting index in Figure 14.

The legislature-drawn map that replaced the 2011 map in Virginia was so effective at rectifying Republicans' impermissible gerrymander that it elected a Democratic majority in 2019.

6 Remarks on the Literature

The literature on electoral districting falls into three strands: papers that focus on bizarre shapes of electoral districts as indirect evidence of foul play by the districters, papers that use the votes–seats curve to assess the unfairness of electoral outcomes as indirect evidence of foul play by districters, and papers that directly model districter's motives.

The first strand includes Chambers and Miller (2010), who review prior art and propose a compelling measure of district convexity. Fryer and Holden (2011) propose a related measure, amenable to principled aggregation to assess the bizarreness of a district map as a whole. Ely (2019) identifies harmful nonconvexities through equilibrium analysis. The problem with relying on district geometry in order to infer districter's intentions is that these intentions cannot be reliably read off geometry alone. One must consider the ideological composition of districts. Vickrey's (1961) example makes the point. In his example, a district map that caters to voters' local representation concerns and a distinct map that maximizes the number of seats won by a party share the same set of convex district shapes. Our measure of maldistricting differs from measures of bizarreness in two ways: We neglect district geometry and instead focus on ideology. We are also agnostic about individual districts and can only assess maldistricting of the entire district map.

In the second strand of the literature, the votes–seats curve $s : [0,1] \to [0,1]$ assigns to any share v of votes that a party obtains in an election a share $s(v)$ of seats that that the party wins (Erikson, 1972; Tufte, 1973). An electoral system is deemed unbiased if $s\left(\frac{1}{2}\right) = \frac{1}{2}$, with $\left|s\left(\frac{1}{2}\right) - \frac{1}{2}\right|$ being the measure of bias.[38] Our model does not endorse this measure. When $v = \frac{1}{2}$ is interpreted as the mean ideology $R = \frac{1}{2}$, and $s\left(\frac{1}{2}\right)$ is interpreted as the share of districts with $r_k > \frac{1}{2}$ (i.e., won by Republicans), the bias $\left|s\left(\frac{1}{2}\right) - \frac{1}{2}\right|$ does not align with our index of maldistricting, if only because the former is discontinuous in the legislature while the latter is continuous.[39] Even if $s\left(\frac{1}{2}\right)$ is replaced with the policy p, the bias $\left|p - \frac{1}{2}\right|$ fails to align with our index. When voters care about representation (i.e., when $\gamma < 1$), the welfare-maximizing legislature induces zero maldistricting while $p \neq R$. We share with the votes–seats literature the neglect of the shapes of electoral districts.

Ensemble sampling (reviewed by Ellenberg, 2020) extends the votes–seats curve approach by, first, computing an empirical distribution for the values of $s(v) - v$ generated by "randomly" sampling from all admissible district maps and then assessing how much of an outlier the observed value of $s(v) - v$ is. This approach implicitly ascribes normative significance to both the votes–seat curve and the "random" map, neither of which has basis in law, which does not pro-

[38]Gelman and King (1994) and Katz, King and Rosenblatt (2020) provide rich applications of this model. Gov. Arnold Schwarzenegger is among the activists who endorse $\left|s\left(\frac{1}{2}\right) - \frac{1}{2}\right|$ as a measure of electoral unfairness (https://youtu.be/E2EnuHsRJd4, accessed 17 June 2020).

[39]Nor does maldistricting align with the efficiency gap (proposed by Stephanopoulos and McGhee, 2015, critiqued by Chambers, Miller and Sobel, 2017, and recognized by the Supreme Court in *Gill v Whitford*, 585 US, 2018), which coincides with the bias when $R = \frac{1}{2}$.

	Party Welfare	Voter Welfare
Policy	Gilligan and Matsusaka (2006)	Coate and Knight (2007)
Representation	Owen and Grofman (1988) Gilligan and Matsusaka (1999) Friedman and Holden (2008)	Chamberlin and Courant (1983) Monroe (1995)

Table 1: **Related literature.**

hibit the toolkit of partisan districting, only some uses of this toolkit.[40] Our approach ascribes normative (though not legal) significance to the set $\mathcal{W}$ of well-districted legislatures and, instead of generating an empirical distribution as ensemble sampling would have it, implicitly assumes a theoretical one, whose salient features are captured by the Manhattan distance from $\mathcal{W}$.

Our paper contributes to the third strand of the literature, which explicitly models the motives of the districter. This strand is further subdivided depending on whether the districter is concerned with party welfare or voter welfare, and whether this welfare is derived from policy or representation. Table 1 collates some of the foundational papers according to the aforementioned taxonomy. Our model accommodates districter motives in each cell of the table.

Coate and Knight (2007) focus on a districter who maximizes voters' utility from policy, which corresponds to $\gamma = 1$ in our model. The first-best has a linear relationship between votes and seats, which can be interpreted to require $p = R$ in our model. When independent voters are few, the first-best is implementable; it always is in our model (which has no counterpart for independents). Our principal difference from Coate and Knight (2007) is the assumption that, instead of being constrained by party platforms, each candidate responds to the ideological composition of the district in which he runs; a Democratic candidate in the Deep South is quite different from a Democratic candidate in the northeast. Following Chamberlin and Courant (1983) and Monroe (1995), we also acknowledge that voter welfare may depend on the characteristics of the legislature

[40]In *Rucho v. Common Cause* (2019), Justice Roberts writes: 'Partisan gerrymandering claims rest on an instinct that groups with a certain level of political support should enjoy a commensurate level of political power and influence. Such claims invariably sound in a desire for proportional representation, but the Constitution does not require proportional representation, and federal courts are neither equipped nor authorized to apportion political power as a matter of fairness. It is not even clear what fairness looks like in this con- text. It may mean achieving a greater number of competitive districts by undoing packing and cracking so that supporters of the dis- advantaged party have a better shot at electing their preferred candidates. But it could mean engaging in cracking and packing to ensure each party its "appropriate" share of "safe" seats. Or perhaps it should be measured by adherence to "traditional" districting criteria. Deciding among those different visions of fairness poses basic questions that are political, not legal. There are no legal standards discernible in the Constitution for making such judgments.'

that are not summarized by the policy that the legislature adopts. For tractability, we focus on one such characteristic: the voter's utility from being ideologically near to his district representative.

We adopt the utilitarian approach to social welfare by assuming that well-districting maximizes the welfare of people (voters) not of institutions (political parties). Maldistricting, in turn, is concerned with party welfare, either derived from the policy or from representation as captured by the number of seats won. Gilligan and Matsusaka (2006) focus on policy extremization and, when the number of voters is "large" (we have a continuum), derive comparative statics consistent with ours. Owen and Grofman (1988) show that seat maximization involves packing (concentrating) and cracking (dispersing) supporters of the opposing party. Our districter goes through the same stylized motions, whether he maximizes the seats or extremizes the policy. Friedman and Holden (2008) and Gul and Pesendorfer (2010) resemble Owen and Grofman but accommodate a much richer structure of uncertainty.

7 Concluding Remarks

The index of maldistricting we propose in this paper is modular. The user specifies a set of maldistricted legislatures ($\mathcal{M}$) and a set of well-districted ones ($\mathcal{W}$), as well as an economically motivated notion of a distance between legislatures (d). These modules are then combined into an index, a function that takes the observed legislature as an input. Our contribution consists in proposing a setting in which the described modular design is meaningful and in making particular choices for the modules: the maldistricted and well-districted sets motivated by the maximization of party and voter welfare, respectively, and a distance between the observed and the intended legislatures motivated by voter migration. In future work, both theoretical and empirical, alternative choices of the modules may prove profitable.[41]

The tractability of our approach owes to the focus on the environments in which the mean-voter theorem is a good description of voting outcomes at the district level. The promise of such environments lies in the synergies between the problems of electoral districting and information design.

[41] Alternative notions of distance would call for reassessment of the assertions in Footnotes 19 and 27.

A Omitted Proofs (for publication)

A.1 The Proof of the Equivalence between Parts 1 and 2 of Proposition 2

To show the equivalence between parts 1 and 2, we shall show that $\mathcal{P} = X$, where $\mathcal{P}$ is the set of implementable legislatures, and X is defined in (2). That X is a polytope is immediate.

We first show that $\mathcal{P} \subset X$. If $\mathbf{r} \in \mathcal{P}$, then $\mathbf{r}$ satisfies properties (ii) and (iii) in (2) by virtue of being a (ordered) legislature. For property (i), we must equivalently show that, for any $m \in \mathcal{K}$,

$$\sum_{k=1}^{m} \int_{g^{-1}(k)} \rho(l)\, \mathrm{d}l \geq \int_{0}^{\frac{m}{K}} \rho(l)\, \mathrm{d}l.$$

Due to the monotonicity of ρ, the right hand-side of the inequality above equals

$$\min_{g' \in G} \sum_{k=1}^{m} \int_{g'^{-1}(k)} \rho(l)\, \mathrm{d}l.$$

The sought inequality follows. The inclusion $P \subset X$ follows.

We now show that $X \subset P$. If $\mathbf{r} \in X$, then $\mathbf{r}$ is ordered by property (iii) in (2). It remains to show that $\mathbf{r}$ is implementable by some g. The proof is recursive.

From property (i), $r_1^e \leq r_1$. From properties (i), (ii), and (iii), $r_1 \leq r_K \leq r_K^e$. Define the interval $S(a) \equiv \left[a, a + \frac{1}{K}\right) \subset \mathcal{L}$ and the associated mean $r_1(a) = K \int_{S(a)} \rho(l)\, \mathrm{d}l$, which satisfies $r_1(0) = r_1^e \leq r_1 \leq r_1\left(\frac{K-1}{K}\right) = r_K^e$. Because $r_1(a)$ is continuous in a, there exists an a^* such that $r_1(a^*) = r_1$. Set $g(l) = 1$ for all $l \in S(a^*)$.

We now reduce the initial problem by one dimension by removing the district $S(a^*)$, with its representative r_1, and constructing an extreme legislature $\mathbf{r}^{e'} \equiv \left(r_2^{e'}, r_3^{e'}, \ldots, r_K^{e'}\right)$ for the residual economy $\mathcal{L}' \equiv \mathcal{L} \backslash S(a^*)$. To do so, define, for each $l \in \mathcal{L}$,

$$\varphi(l) = \begin{cases} l & \text{if } l \leq a^* \\ 0 & \text{if } a^* < l \leq a^* + \frac{1}{K} \\ l - \frac{1}{K} & \text{if } l > a^* + \frac{1}{K}, \end{cases}$$

and define $\mathbf{r}^{e'}$ by setting, for each $k \in \mathcal{K}' \equiv \mathcal{K} \setminus \{1\}$,

$$r_k^{e'} = K \int_{l: \frac{k-2}{K} < \varphi(l) < \frac{k-1}{K}} \rho(l) \, \mathrm{d}l.$$

To show that $r_2^{e'} \leq r_2 \leq r_K^{e'}$, we use Lemma A.1.

Lemma A.1. *1.* $\frac{1}{K} \sum_{k \in \mathcal{K}'} r_k = \int_{\mathcal{L}'} \rho(l) \, dl$

2. $(\forall m \in \mathcal{K}') \sum_{k=2}^{m} r_k \geq \sum_{k=2}^{m} r_k^{e'}$.

Proof. Part 1 follows from part (ii) in (2) and by the construction of $S(a^*)$.

For part 2, note that, for any m such that $a^* \geq \frac{m-1}{K}$, $\left(r_2^{e'}, \ldots, r_m^{e'}\right) = \left(r_1^e, \ldots, r_{m-1}^e\right)$. In this case, part 2 follows from

$$\sum_{k=2}^{m} r_k^{e'} = \sum_{k=1}^{m-1} r_k^e \leq \sum_{k=1}^{m-1} r_k \leq \sum_{k=2}^{m} r_k,$$

where the first inequality follows from part (i) in (2), and the second inequality follows by part (iii) in (2).

When $a^* < \frac{m-1}{K}$, part 2 follows from

$$\begin{aligned} \sum_{k=2}^{m} r_k^{e'} &= K \int_0^{\frac{m}{K}} \rho(l) \, \mathrm{d}l - K \int_{S(a^*)} \rho(l) \, \mathrm{d}l \\ &= \sum_{k=1}^{m} r_k^e - r_1 \leq \sum_{k=2}^{m} r_k, \end{aligned}$$

where the inequality follows from property (i) in (2): $\sum_{k=1}^{m} r_k^e \leq \sum_{k=1}^{m} r_k = r_1 + \sum_{k=2}^{m} r_k$. ■

By part 2 of Lemma A.1, $r_2 \geq r_2^{e'}$. By parts 1 and 2 of Lemma A.1, given the definition of $\mathcal{K}' \equiv \mathcal{K} \setminus \{1\}$, we have $r_K \leq r_K^{e'}$. By property (iii) in (2), we have $r_2 \leq r_K$. Therefore, $r_2 \leq r_K \leq r_K^{e'}$. We can now repeat for r_2 the procedure that we have performed for r_1 and thereby recover $g^{-1}(2)$. By successively invoking Lemma A.1, the procedure is repeated until the entire district map g is recovered.

A.2 A Direct Proof of the Equivalence of Part 3 and 2 in Proposition 2

Let V be the set of all vertices described in part 3 of the proposition and let X be the implementable polytope defined in (2). By construction of V, $V \subset X$. The proof proceeds in two steps. Step 1 shows that V contains only the vertices of X. Step 2 (which contains an auxiliary lemma) shows that V does not omit any vertices of X. From steps 1 and 2, we conclude that V characterizes the polytope X, which is equivalent to $\mathcal{P}$.

Step 1: Each $\mathbf{r}^v$ in V is a vertex of X.

By contradiction, suppose that an $\mathbf{r}^v \in V$ is not a vertex. Then, there exist $\mathbf{r}'$ and $\mathbf{r}''$ in X such that $\mathbf{r}' \neq \mathbf{r}''$ and

$$\mathbf{r}^{\mathbf{v}} = \theta \mathbf{r}' + (1-\theta)\,\mathbf{r}'' \qquad \text{for some} \qquad \theta \in (0,1)\,. \tag{A.1}$$

Let $t \in \mathcal{K}$ be the first coordinate at which $\mathbf{r}'$ and $\mathbf{r}''$ differ: $r'_t \neq r''_t$. Without loss of generality, let $r'_t > r''_t$. Because of (A.1), $r'_t > r''_t$ implies $r'_t > r^v_t > r''_t$, and $r'_k = r''_k$ for all $k < t$ implies $r'_k = r^v_k = r''_k$.

Let m be the smallest integer in $\{0, 1, \ldots, K-t\}$ such that $\sum_{k=1}^{t+m} r^v_k = \sum_{k=1}^{t+m} r^e_k$. (Such an m exists because part (ii) of (2) demands that $\sum_{k=1}^{K} r^v_k = \sum_{k=1}^{K} r^e_k$.) Then, because $\mathbf{r}^v \in V$, it must be that, for any $j \in \{0, 1, \ldots, m\}$, we have $r^v_t = r^v_{t+j}$. Then, combining $r'_{t+j} \geq r'_t$ (by part (iii) of (2)) and $r'_t > r^v_t$ with $r^v_t = r^v_{t+j}$ we get $r'_{t+j} > r^v_{t+j}$ for every $j \in \{0, 1, \ldots, m\}$.

Now, (A.1) and $r'_{t+j} > r^v_{t+j}$ for all $j \in \{0, 1, \ldots, m\}$ imply $r^v_{t+j} > r''_{t+j}$ for all $j \in \{0, 1, \ldots, m\}$. Recall that $r^v_k = r''_k$ for all $k < t$. The last two sentences taken together imply $\sum_{k=1}^{t+m} r^v_k > \sum_{k=1}^{t+m} r''_k$. Because, $\sum_{k=1}^{t+m} r^v_k = \sum_{k=1}^{t+m} r^e_k$, we have $\sum_{k=1}^{t+m} r^e_k > \sum_{k=1}^{t+m} r''_k$, which contradicts part (i) of (2) for $\mathbf{r}'' \in X$.

Step 2: No $\mathbf{r}$ in $X \backslash V$ is a vertex of X.

Fix any $\mathbf{r} \in X \backslash V$. We shall construct legislatures $\mathbf{r}'$ and $\mathbf{r}''$—each in X and each a local perturbation of $\mathbf{r}$—such that $\mathbf{r} = \frac{1}{2}\mathbf{r}' + \frac{1}{2}\mathbf{r}''$, thereby demonstrating that $\mathbf{r}$ is not a vertex of the polytope X. The delicate part of the proof is in determining which components of $\mathbf{r}$ can be perturbed so that both perturbations remain in X and, in particular, satisfy property (i) in (2).

Given an $\mathbf{r} \in X$, let $\mathcal{I}_{\mathbf{r}} \equiv \{I_{\mathbf{r},1}, I_{\mathbf{r},2}, \ldots\}$ be the interval partition of the set of districts $\mathcal{K}$ into equivalence classes such that

- for any two districts k and k' in the same class, $r_k = r_{k'}$, and
- for any k in $I_{\mathbf{r},i}$ and any k' in $I_{\mathbf{r},j}$, if $i < j$ then $r_k < r_{k'}$.[A.1]

[A.1] The partition $\mathcal{I}_{\mathbf{r}}$ is unrelated to the partitions used to generate the elements of V in part 3 of the proposition's statement.

Because $\mathbf{r}$ is ordered, whenever $r_i \in I_{\mathbf{r},n}$ and $r_j \in I_{\mathbf{r},m}$, the inequality $n < m$ implies that $r_i < r_j$. Because $\mathbf{r} \notin X \backslash V$, it cannot be that, for all $I_{\mathbf{r},h} \in \mathcal{I}_{\mathbf{r}}$, the equality $\sum_{i\in I_{\mathbf{r},h}} r_i = \sum_{i\in I_{\mathbf{r},h}} r_i^e$ holds. Without loss of generality, assume that[A.2]

$$\sum_{i\in I_{\mathbf{r},1}} r_i > \sum_{i\in I_{\mathbf{r},1}} r_i^e. \tag{A.2}$$

Write $I_{\mathbf{r},1} = \{1, \ldots, t\}$ and $I_{\mathbf{r},2} = \{t+1, \ldots, t+m\}$ for some $t \geq 1$ and $m \geq 1$. We know that $I_{\mathbf{r},2}$ is nonempty; otherwise, $r_k = R$, for all $k \in \mathcal{K}$, and then $\mathbf{r} \in V$, which we have ruled out. Note that, for some $\bar{r}_1$ and all $k \in I_{\mathbf{r},1}$, we have $r_k = \bar{r}_1$, and, for some $\bar{r}_2$ and all $k \in I_{\mathbf{r},2}$, we have $r_k = \bar{r}_2$, with $\bar{r}_1 < \bar{r}_2$. The following lemma is crucial.

Lemma A.2. *For all $k < t+m$, the inequality $\sum_{i=1}^k r_i > \sum_{i=1}^k r_i^e$ holds.*

Proof. First, let $k \in I_{\mathbf{r},1}$. Because r_i^e is nondecreasing in i, $r_i^e \geq r_j^e$ for all $i \in \{k+1, \ldots, t\}$ and all $j \in \{1, \ldots, k\}$. Adding up these inequalities over j gives $kr_i^e \geq \sum_{j=1}^k r_j^e$ (for every $i \in \{k+1, \ldots, t\}$), which, in turn, can be added up over i to give

$$\begin{aligned}
\sum_{i=k+1}^{t} kr_i^e \geq (t-k)\sum_{j=1}^{k} r_j^e \iff \\
\sum_{i=1}^{t} kr_i^e \geq t\sum_{j=1}^{k} r_j^e \iff \\
\frac{k}{t}\sum_{i\in I_{\mathbf{r},1}} r_i^e \geq \sum_{i=1}^{k} r_i^e.
\end{aligned} \tag{A.3}$$

Assumption (A.2) is equivalent to the inequality $t\bar{r}_1 > \sum_{i\in I_{\mathbf{r},1}} r_i^e$. Multiplying its both sides by k/t, we get the inequality $k\bar{r}_1 > \frac{k}{t}\sum_{i\in I_{\mathbf{r},1}} r_i^e$. Since $k \in I_{\mathbf{r},1}$, this inequality is equivalent to $\sum_{i=1}^k r_i > \frac{k}{t}\sum_{i\in I_{\mathbf{r},1}} r_i^e$. Therefore, by (A.3), we have $\sum_{i=1}^k r_i > \sum_{i=1}^k r_i^e$, which is the sought inequality when $k \in I_{\mathbf{r},1}$.

If $m = 1$ (i.e., $I_{\mathbf{r},2}$ is a singleton), the lemma's conclusion follows.

Assume that $m > 1$. Let $k = (t+j) \in I_{\mathbf{r},2}$ with $j < m$ (recall that $I_{\mathbf{r},2} \equiv \{t+1, \ldots, t+m\}$). Because $\mathbf{r} \in X$, we have $\sum_{i=1}^k r_i \geq \sum_{i=1}^k r_i^e$ and $\sum_{i=1}^{k+1} r_i \geq \sum_{i=1}^{k+1} r_i^e$. By contradiction, suppose that $\sum_{i=1}^k r_i = \sum_{i=1}^k r_i^e$. The contradiction hypothesis combined with $\sum_{i=1}^{k+1} r_i \geq \sum_{i=1}^{k+1} r_i^e$ implies $r_{k+1} \geq r_{k+1}^e$ or, equivalently, $\bar{r}_2 \geq r_{k+1}^e$. We can now rewrite the contradiction hypothesis as

[A.2] If $\sum_{i\in I_{\mathbf{r},1}} r_i = \sum_{i\in I_{\mathbf{r},1}} r_i^e$, then the argument given below for $I_{\mathbf{r},1}$ is adapted for the first $I_{r,h}$ for which $\sum_{i\in I_{\mathbf{r},h}} r_i > \sum_{i\in I_{\mathbf{r},h}} r_i^e$.

$\sum_{i \in I_{\mathbf{r},1}} r_i + j\bar{r}_2 = \sum_{i \in I_{\mathbf{r},1}} r_i^e + \sum_{i=t+1}^{k} r_i^e$, which, by $\bar{r}_2 \geq r_{k+1}^e$, implies $\sum_{i \in I_{\mathbf{r},1}} r_i + j r_{k+1}^e \leq \sum_{i \in I_{\mathbf{r},1}} r_i^e + \sum_{i=t+1}^{k} r_i^e$. Therefore, because $\mathbf{r}^e$ is ordered, $\sum_{i \in I_{\mathbf{r},1}} r_i \leq \sum_{i \in I_{\mathbf{r},1}} r_i^e$. But this inequality contradicts assumption (A.2). Therefore, it must be that $\sum_{i=1}^{k} r_i > \sum_{i=1}^{k} r_i^e$ for $k = (t+j) \in I_{\mathbf{r},2}$ with $j < m$. The lemma's conclusion follows. ■

We now express $\mathbf{r}$ as $\frac{1}{2}\mathbf{r}' + \frac{1}{2}\mathbf{r}''$ for appropriately chosen legislatures $\mathbf{r}'$ and $\mathbf{r}''$ in X. For all $k > t+m$, set $r_k' = r_k'' = r_k$. Take some ε. For every $k \in I_{\mathbf{r},1}$, set $r_k' = r_k - \varepsilon$ and $r_k'' = r_k + \varepsilon$. For every $k \in I_{\mathbf{r},2}$, set $r_k' = r_k + \frac{t}{m}\varepsilon$ and $r_k'' = r_k - \frac{t}{m}\varepsilon$. The legislatures $\mathbf{r}'$ and $\mathbf{r}''$ satisfy property (ii) in (2) by construction. By the inequality in Lemma A.2, there exists a sufficiently small $\varepsilon > 0$ such that $\mathbf{r}'$ and $\mathbf{r}''$ each satisfies not only property (iii) in (2) but also property (i) in (2). Hence, $\mathbf{r}'$ and $\mathbf{r}''$ are in X. By $\mathbf{r} = \frac{1}{2}\mathbf{r}' + \frac{1}{2}\mathbf{r}''$, $\mathbf{r}$ is not a vertex of the polytope X.

A.3 Proof of Fact 1

When two legislatures are ordered, the total number of ideology-1 voters who must move to a differently labelled district in order to transform one legislature into the other equals the total variation distance.[A.3] The total number of voters who move is twice this number because ideology-1 voters displace an equal number of ideology-0 voters. Because the number of districts is finite, twice the total variation distance equals the Manhattan distance (Levin, Peres and Wilmer, 2017, Proposition 4.2).

It remains to confirm that the right way to minimize the number of voters who must move into a differently labelled district is to order both legislatures in the ascending order. Indeed, take two legislatures, $\mathbf{r}'$ and $\mathbf{r}$. Without loss of generality, relabel the districts so that $\mathbf{r}'$ is ordered. By contradiction, suppose that the total variation distance between $\mathbf{r}'$ and $\mathbf{r}$ is minimized only if $\mathbf{r}$ is unordered, that is, only if, for some districts k and k' with $k' > k$, $r_{k'} < r_k$. But, in this case, switching the districts k and k' either makes no difference to the total variation distance between $\mathbf{r}'$ and $\mathbf{r}$ or strictly decreases it, which is a contradiction.

[A.3] The total variation distance (defined in Wikipedia) is a standard distance for probability measures.

A.4 The Proof of Proposition 4

The extreme legislatures correspond to the vertices of the implementable polytope (part 3 of Proposition 2) that have the minimal and the maximal components in the dimension M. For the "moreover" part of the proposition, note that $\partial \underline{p}/\partial K \leq 0$ and $\partial \bar{p}/\partial K \geq 0$ because ρ is non-decreasing. The comparative statics with respect to ρ follows immediately by the definition of second-order stochastic dominance and by inspection of the expressions for $\underline{p}$ and $\bar{p}$.

A.5 The Proof of Part 3 of Proposition 5

In the proof, $\mathcal{I}_{\mathbf{r}}$ is the interval partition of $\mathcal{K}$ for a legislature $\mathbf{r}$ as defined in Step 2 of the proof of Proposition 2 (Section A.2, preceding the statement of Lemma A.2).

The proof proceeds in a series of lemmas. Lemmas A.3 and A.4 prove two inequalities used in subsequent lemmas. Lemma A.5 shows that a welfare maximizing legislature belongs to a certain class. Lemma A.6 shrinks that class to the class in the proposition's statement. Lemma A.7 establishes the "moreover" part. By specifying the exact values of p^* as γ varies, Lemma A.7 does more than required.

Lemma A.3. *Take an implementable legislature $\mathbf{r} \in X$ with its associated partition $\mathcal{I}_{\mathbf{r}}$. For every element $I_{\mathbf{r},h} = \{i, \ldots, i+j\}$ of $\mathcal{I}_{\mathbf{r}}$ with $j \geq 1$, $0 \leq q < j$ implies $\sum_{k=1}^{i+q} r_k > \sum_{k=1}^{i+q} r_k^e$.*

Proof. For the objects in the lemma's statement, $\mathbf{r} \in X$ implies $\sum_{k=1}^{i+q} r_k \geq \sum_{k=1}^{i+q} r_k^e$ for all q with $0 \leq q < j$. By way of contradiction, suppose that, for some such q, $\sum_{k=1}^{i+q} r_k = \sum_{k=1}^{i+q} r_k^e$.

If $i + q = 1$, then the contradiction hypothesis implies $r_1 = r_1^e$.

If $i + q \geq 2$, then $\mathbf{r} \in X$ implies $\sum_{k=1}^{i+q-1} r_k \geq \sum_{k=1}^{i+q-1} r_k^e$, which, combined with the contradiction hypothesis, implies $r_{i+q}^e \geq r_{i+q}$.

For any $i + q$, $\mathbf{r} \in X$ implies $\sum_{k=1}^{i+q+1} r_k \geq \sum_{k=1}^{i+q+1} r_k^e$, which, combined with the contradiction hypothesis, implies $r_{i+q+1} \geq r_{i+q+1}^e$.

Then, $r_{i+q+1} \geq r_{i+q+1}^e > r_{i+q}^e \geq r_{i+q}$, where the strict inequality is implied by Assumption 1. The conclusion $r_{i+q+1} > r_{i+q}$, however, contradicts assumption that $i + q$ and $i + q + 1$ are both in $I_{\mathbf{r},h}$. As a result, the contradiction hypothesis must be wrong, and the lemma's conclusion follows. ∎

Lemma A.4. *Let* $\mathbf{r}$ *and* $\mathbf{r}'$ *be two legislatures in* X*. Suppose that* $\mathbf{r}$ *and* $\mathbf{r}'$ *differ only in* $r'_i = r_i - \varepsilon$ *and* $r'_j = r_j + \varepsilon$ *for some* i *and* j *in* $\mathcal{K}\setminus\{M\}$ *and some* $\varepsilon > 0$*. Suppose that* $\mathbf{r}_M = \mathbf{r}'_M$ *(i.e., the the* ε*-perturbation above leaves the policy unchanged). Then,* $i < j$ *and* $\gamma < 1$ *imply* $W(\mathbf{r}') > W(\mathbf{r})$*, where* W *is defined in (12).*

Proof. Note that

$$\begin{aligned} W(\mathbf{r}') &= \gamma\left(2Rr_M - r_M^2\right) + \frac{1-\gamma}{K}\left(\sum_{k\in\mathcal{K}\setminus\{i,j\}} r_k^2 + (r_i-\varepsilon)^2 + \left(r_j+\varepsilon\right)^2\right) \\ &= W(\mathbf{r}) + \frac{2(1-\gamma)\,\varepsilon\left(r_j - r_i + \varepsilon\right)}{K} > W(\mathbf{r}), \end{aligned}$$

where the inequality follows from $r_i \leq r_j$ (by $i < j$), $\varepsilon > 0$, and $\gamma < 1$. ■

Lemma A.5. *Let* $\mathbf{r}^*$ *be a socially optimal legislature for* $\gamma \in (0,1)$ *with the associated partition* $\mathcal{I}_{\mathbf{r}^*} = \{I_{\mathbf{r}^*,1}, I_{\mathbf{r}^*,2}, \ldots\}$*. Let* $I_{\mathbf{r}^*,p}$ *be the element of the partition that contains the median district:* $M \in I_{\mathbf{r}^*,p} = \{\underline{k}, \underline{k}+1, \ldots, M, \ldots, \bar{k}\}$*, where* $\underline{k} \in \{1,2,\ldots,M\}$ *and* $\bar{k} \in \{M, M+1, \ldots, K\}$*. Then, all the remaining partition elements in* $\mathcal{I}_{\mathbf{r}^*}\setminus I_{\mathbf{r}^*,p}$ *are singletons. Moreover,* $r_k^* = r_k^e$ *for each* $k \notin I_{\mathbf{r}^*,p} \cup \{\underline{k}-1, \bar{k}+1\}$*.*

Proof. We first show that every element of $\mathcal{I}_{\mathbf{r}^*}\setminus I_{\mathbf{r}^*,p}$ is a singleton. By way of contradiction, let $I_{\mathbf{r}^*,p'}$ be some nonsingleton element of $\mathcal{I}_{\mathbf{r}^*}\setminus I_{\mathbf{r}^*,p}$, with cardinality $c \equiv \left|I_{\mathbf{r}^*,p'}\right| > 1$. Suppose that $r_k^* = \bar{r}$ for all $k \in I_{\mathbf{r}^*,p'}$. Now perturb $\mathbf{r}^*$ by an ε to obtain

$$\mathbf{r}' \equiv \mathbf{r}^* + \left(0, \ldots, 0, \underbrace{-\varepsilon, \ldots, -\varepsilon, (c-1)\,\varepsilon}_{\text{at the positions in } I_{\mathbf{r}^*,p'}}, 0, \ldots, 0\right).$$

By Lemma A.3, there exists an $\varepsilon > 0$ small enough such that $\mathbf{r}' \in X$.[A.4] Then, $c > 1$ and $\varepsilon > 0$ imply $W(\mathbf{r}') > W(\mathbf{r}^*)$ by Lemma A.4, thereby contradicting the optimality of $\mathbf{r}^*$. Hence, every element of $\mathcal{I}_{\mathbf{r}^*}\setminus I_{\mathbf{r}^*,p}$ must be a singleton.

If $\underline{k} \leq 2$ and $\bar{k} \geq K-1$, then the lemma is vacuously true.

[A.4]If it were to be the case that $p' < K$, then by construction of $I_{\mathbf{r}^*}$, we have $\bar{r} < r_k^*$ for every $k \in I_{\mathbf{r}^*,p'+1}$. Hence, the construction of such an ε is feasible.

Suppose that $\underline{k} \geq 3$. We now show that, for all $k \in \{1, 2, \ldots, \underline{k} - 2\}$, $r_k^* = r_k^e$. Pick the smallest $l \in \{1, 2, \ldots, \underline{k} - 2\}$ such that $r_l^* > r_l^e$. If such an l does not exist, then the desired conclusion obtains. Suppose that such an l exists. Perturb $\mathbf{r}^*$ by an ε to obtain an $\mathbf{r}'$ that coincides with $\mathbf{r}^*$ except for $r_l' = r_l^* - \varepsilon$ and $r_{l+1}' = r_{l+1}^* + \varepsilon$. Because $r_l^e < r_l^* < r_{l+1}^* < r_{l+2}^*$, there exists an $\varepsilon > 0$ small enough such that $\mathbf{r}' \in X$. Then, $\varepsilon > 0$ implies $W(\mathbf{r}') > W(\mathbf{r}^*)$ by Lemma A.4, thereby contradicting the optimality of $\mathbf{r}^*$. Hence, every element of $\mathcal{I}_{\mathbf{r}^*} \backslash I_{\mathbf{r}^*,p}$ must be a singleton. Hence, for all $k \in \{1, 2, \ldots, \underline{k} - 2\}$, $r_k^* = r_k^e$.

An analogous argument shows that $r_k^* = r_k^e$ for all $k \in \{\bar{k} + 2, \bar{k} + 3, \ldots, K\}$. ∎

Lemma A.6. *Let $\mathbf{r}^*$ be a socially optimal legislature for $\gamma \in (0, 1)$ with the associated partition $\mathcal{I}_{\mathbf{r}^*} = \{I_{\mathbf{r}^*,1}, I_{\mathbf{r}^*,2}, \ldots\}$. Let $I_{\mathbf{r}^*,p}$ be the element of the partition that contains the median district: $M \in I_{\mathbf{r}^*,p} = \{\underline{k}, \underline{k} + 1, \ldots, M, \ldots, \bar{k}\}$, where $\underline{k} \in \{1, 2, \ldots, M\}$ and $\bar{k} \in \{M, M + 1, \ldots, K\}$. Then, $R > r_M^e$ implies $\underline{k} = M$, and $R < r_M^e$ implies $\bar{k} = M$.*

Proof. We prove the lemma for the case of $R > r_M^e$. The case of $R < r_M^e$ is analogous.

Assume that $R > r_M^e$. By way of contradiction, suppose that $\underline{k} < M$. There are two cases to consider.

Case 1. $M < \bar{k}$.

The assumption $M < \bar{k}$ coupled with $\underline{k} < M$ implies $\underline{k} < \bar{k}$. Perturb $\mathbf{r}^*$ by an ε to obtain an $\mathbf{r}'$ that coincides with $\mathbf{r}^*$ except for $r_{\underline{k}}' = r_{\underline{k}}^* - \varepsilon$ and $r_{\bar{k}}' = r_{\bar{k}}^* + \varepsilon$. There exists an $\varepsilon > 0$ small enough such that $\mathbf{r}' \in X$. Indeed, there exists an $\varepsilon > 0$ such that $\mathbf{r}'$ satisfies condition (i) in (2) by Lemma A.3. Condition (ii) in (2) holds by construction of the perturbation, for any ε. There exists an $\varepsilon > 0$ such that $\mathbf{r}'$ satisfies condition (iii) in (2) ($\mathbf{r}'$ is ordered) because $\mathbf{r}^*$ is ordered, and because $\underline{k}$ and $\bar{k}$ are the endpoints of an element of the partition $\mathcal{I}_{\mathbf{r}^*}$. Then, $\varepsilon > 0$ implies $W(\mathbf{r}') > W(\mathbf{r}^*)$ by Lemma A.4, thereby contradicting the optimality of $\mathbf{r}^*$.

Case 2. $\bar{k} = M$.

We consider two special cases:[A.5]

[A.5] The case $r_{M+1}^* > r_{M+1}^e$ (equivalently, $r_{\bar{k}+1}^* > r_{\bar{k}+1}^e$) cannot occur by Lemma A.5.

Case i. $r^*_{M+1} < r^e_{M+1}$

Perturb $\mathbf{r}^*$ by an ε to obtain an $\mathbf{r}'$ that coincides with $\mathbf{r}^*$ except for $r'_{\underline{k}} = r^*_{\underline{k}} - \varepsilon$ and $r'_{\bar{k}+1} = r^*_{\bar{k}+1} + \varepsilon$. There exists an $\varepsilon > 0$ small enough such that $\mathbf{r}' \in X$. In particular, there exists an $\varepsilon > 0$ such that $\mathbf{r}'$ is ordered because $\mathbf{r}^*$ is ordered, $\underline{k}$ is an endpoint of an element of the partition $\mathcal{I}_{\mathbf{r}^*}$, and $r^*_{\bar{k}+1} < r^*_{\bar{k}+2}$ by Lemma A.5. Then, $\varepsilon > 0$ implies $W(\mathbf{r}') > W(\mathbf{r}^*)$ by Lemma A.4, thereby contradicting the optimality of $\mathbf{r}^*$.

Case ii. $r^*_{M+1} = r^e_{M+1}$

Perturb $\mathbf{r}^*$ by an ε to obtain an $\mathbf{r}'$ that coincides with $\mathbf{r}^*$ except for $r'_{\underline{k}} = r^*_{\underline{k}} - \varepsilon$ and $r'_M = p^* + \varepsilon$.

By hypothesis, $\bar{k} = M$ and $r^*_{M+1} = r^e_{M+1}$. By Lemma A.5, $r^*_{\underline{k}-1} + \sum_{k=\underline{k}}^{\bar{k}} p^* = \sum_{k=\underline{k}-1}^{\bar{k}} r^e_k$. Then, $\underline{k} < \bar{k}$ and Lemma A.3 imply $r^e_M > p^*$. Since $R > r^e_M$ by lemma's hypothesis, it follows that $R > p^*$. Now note that

$$\left.\frac{\mathrm{d}W(\mathbf{r}')}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = 2\gamma(R-p^*) + \frac{2(1-\gamma)\left(p^* - r^*_{\underline{k}}\right)}{K} > 0,$$

where the inequality follows because $\gamma > 0$, $R > p^*$, and $p^* = r^*_{\underline{k}}$.

There exists an $\varepsilon > 0$ small enough such that $\mathbf{r}' \in X$. In particular, there exists an $\varepsilon > 0$ such that $\mathbf{r}'$ is ordered because $\mathbf{r}^*$ is ordered, and because $\underline{k}$ and M are the endpoints of an element of the partition $\mathcal{I}_{\mathbf{r}^*}$.

Therefore, there exists an $\varepsilon > 0$ small enough such that both $\mathbf{r}' \in X$ and $W(\mathbf{r}') > W(\mathbf{r}^*)$, thereby contradicting the optimality of $\mathbf{r}^*$.

Because all possible cases end up in contradiction, it must be that $\underline{k} = M$. ■

Lemma A.7 calls for additional notation. Define

$$z_i \equiv \begin{cases} \frac{1}{i+1}\sum_{k=M}^{M+i} r^e_k & \text{if } R \geq r^e_M \\ \frac{1}{i+1}\sum_{k=M-i}^{M} r^e_k & \text{if R<}r^e_M \end{cases}, \qquad i \in \{0,1,\ldots,K-M\}. \tag{A.4}$$

For every $i \in \{1, 2, \ldots, Q-1\}$, define

$$\gamma_i \equiv \left(1 + K \frac{2R - (z_i + z_{i-1})}{i\left(r^e_{M+i} - z_{i-1}\right)}\right)^{-1}. \tag{A.5}$$

Finally, define Q by the inclusion $R \in (z_{Q-1}, z_Q]$,[A.6] and then define

$$\gamma^* \equiv \left(1 + \frac{K}{Q(Q+1)} \frac{R - z_{Q-1}}{z_Q - z_{Q-1}}\right)^{-1}. \tag{A.6}$$

Lemma A.7. *Let q, b, and p^* be the objects introduced in part 3 of Proposition 5. Suppose that Assumption 1 holds. Then, $\gamma_0 \equiv 0 < \gamma_1 < \gamma_2 < \ldots < \gamma_{Q-1} < \gamma_Q \equiv \gamma^*$, and*

1. *if $\gamma \in (\gamma_i, \gamma_{i+1}]$ for some $i \in \{0, 1, \ldots, Q-1\}$, then $q = i$, $p^* = z_i$, and $b = r^e_{M+q}$*
2. *if $\gamma > \gamma^*$, then $q = Q$, $p^* = R + \frac{R - z_Q}{\frac{\gamma}{1-\gamma}\frac{K}{Q(Q+1)} - 1}$, and $b = p^* + Q(z_Q - p^*)$.*

Thus, in particular, p^* is between r^e_M and R. Moreover, as γ increases, q weakly increases, and p^* moves monotonically away from r^e_M and towards R.[A.7]

Proof. Assume that $R > r^e_M$, in which case z_i is strictly increasing in i. The case of $R < r^e_M$ can be treated analogously.

Step 1: $p^* \leq R$.

By way of contradiction, suppose that $p^* > R$. For further contradiction, suppose that p^* solves $\max_{x \in (z_{i-1}, z_i]} W_i(x)$ for some $i \in \{Q+1, Q+2, \ldots, K-M\}$, where

$$W_i(x) \equiv \gamma\left(2Rx - x^2\right) + \frac{1-\gamma}{K}\left(\sum_{k=1}^{M-1}\left(r^e_k\right)^2 + ix^2 + \left(\sum_{k=M}^{M+i} r^e_k - ix\right)^2 + \sum_{k=M+i+1}^{K}\left(r^e_k\right)^2\right). \tag{A.7}$$

But

$$W_i'(p^*) \equiv \left.\frac{\partial W_i(x)}{\partial x}\right|_{x=p^*} = 2\left[\gamma(R - p^*) + \frac{1-\gamma}{K} i(i+1)(p^* - z_i)\right] < 0$$

because $\gamma > 0$, $R < p^*$ (by $i > Q$), and $p^* \leq z_i$. The inequality $W_i'(p^*) < 0$ contradicts $p^* \in \arg\max_{x \in (z_{i-1}, z_i]} W_i(x)$. Conclude that $p^* \leq z_Q$. If $R = z_Q$, then $p^* \leq R$ follows.

[A.6] By convention, the order in which the endpoints that define an interval are specified is irrelevant. That is, $(z_{Q-1}, z_Q]$ and $(z_Q, z_{Q-1}]$ would refer to the same interval.

[A.7] In the limit when $K \to \infty$, p^* is a continuous function of γ if ρ is differentiable. In this sense, when districts are "many," the discontinuity of the policy p^* in the parameter γ is not economically significant.

Assume that $R < z_Q$. Then, the contradiction hypothesis $p^* > R$ combined with the conclusion $p^* \leq z_Q$ requires p^* to solve $\max\limits_{x \in \left(R, z_Q\right]} W_Q(x)$. But then $W_Q'(p^*) < 0$ because $\gamma > 0$, $R < p^*$, and $p^* \leq z_Q$. The inequality $W_Q'(x) < 0$ contradicts $p^* \in \arg\max_{x \in \left(R, z_Q\right]} W_i(x)$. Conclude that the hypothesis $p^* > R$ is false.

Step 2: $p^* \in \{z_0, z_1, \ldots, z_{Q-2}\} \cup [z_{Q-1}, R]$.

We show that $p^* \notin (z_{i-1}, z_i)$ *for* $i \in \{1, 2, \ldots, Q-1\}$, which, combined with Step 1, establishes $p^* \in \{z_0, z_1, \ldots, z_{Q-2}\} \cup [z_{Q-1}, R]$.

By way of contradiction, suppose that p^* solves $\max_{x \in (z_{i-1}, z_i)} W_i(x)$ for some $i \in \{1, 2, \ldots, Q-1\}$. Then, $\gamma > 0$ and $R > x$ imply that

$$W_i'(x) = 2\left[\gamma(R-x) + \frac{1-\gamma}{K} i(i+1)(x - z_i)\right] > 0$$

when x is close enough to z_i. Furthermore,

$$W_i''(x) \equiv \frac{\partial^2 W_i(x)}{\partial x^2} = 2\left(\frac{1-\gamma}{K} i(i+1) - \gamma\right)$$

is a constant. The last two sentences imply that W_i cannot be maximized in (z_{i-1}, z_i). Step 1 then implies that $p^* \in \{z_0, z_1, \ldots, z_{Q-2}\} \cup [z_{Q-1}, R]$.

Step 3: The monotonicity of p^* when in $[z_{Q-1}, R]$.

For this step, assume that $p^* \in \arg\max_{x \in \left[z_{Q-1}, R\right]} W_Q(x)$, where W_Q is (A.7) with $i = Q$:

$$W_Q(x) = \gamma\left(2Rx - x^2\right) + \frac{1-\gamma}{K}\left(\sum_{k=1}^{M-1}(r_k^e)^2 + Qx^2 + \left(\sum_{k=M}^{M+Q} r_k^e - Qx\right)^2 + \sum_{k=M+Q+1}^{K}(r_k^e)^2\right).$$

By Topkis's monotone selection theorem (Topkis, 1998), the supermodularity condition

$$\frac{\partial^2 W_Q(x)}{\partial x \partial \gamma} \equiv 2\left(R - x + \frac{z_Q - x}{\kappa}\right) \geq 0, \qquad \text{where} \quad \kappa \equiv \frac{K}{Q(Q+1)},$$

implies that the optimal-policy correspondence $\arg\max_{x \in \left[z_{Q-1}, R\right]} W_Q(x)$ has a selection that is nondecreasing in γ in $(0, 1)$.

Recall from (A.6) that $\gamma^* \equiv \left(1 + \kappa \left(R - z_{Q-1}\right) / \left(z_Q - z_{Q-1}\right)\right)^{-1}$. Define $\gamma^{**} \equiv (1+\kappa)^{-1}$. Note that $\gamma^* > \gamma^{**}$ when $R < z_Q$, and $\gamma^* = \gamma^{**}$ when $R = z_Q$.

Case 1. $R < z_Q$.

Assume that $\gamma > \gamma^*$. Then, $\gamma^* > \gamma^{**}$ implies $\gamma > \gamma^{**}$, which guarantees that the maximand W_Q is strictly concave:

$$W_Q''(x) = \frac{2\left(\gamma^{**} - \gamma\right)}{1 - \gamma^{**}} < 0.$$

Therefore, the first-order condition

$$W_Q'(x) = 2\left[\gamma\left(R - x\right) + \frac{1-\gamma}{\kappa}\left(x - z_Q\right)\right] = 0$$

characterizes the optimal policy:

$$p^* = R - \frac{z_Q - R}{\frac{\gamma\kappa}{1-\gamma} - 1}, \tag{A.8}$$

which satisfies $p^* > z_{Q-1}$ (by $\gamma > \gamma^*$) and $p^* < R$ (by $\gamma > \gamma^{**}$ and $R < z_Q$).

Now assume that $\gamma \leq \gamma^*$. Because the uniquely optimal policy in (A.8) converges to z_{Q-1} as $\gamma \downarrow \gamma^*$, Topkis's monotone selection conclusion above implies that, when γ is reduced to satisfy $\gamma \leq \gamma^*$, the policy $p^* = z_{Q-1}$ solves $\max_{x \in [z_{Q-1}, R]} W_Q(x)$; the problem's value is $W_{Q-1}\left(z_{Q-1}\right)$.

Case 2. $R = z_Q$.

Assume that $\gamma > \gamma^*$. Then, $\gamma^* = \gamma^{**}$ (which holds by $R = z_Q$) implies $\gamma > \gamma^{**}$, which guarantees that the maximand W_Q is strictly concave. Therefore, the first-order condition (A.8) implies $p^* = R$.

Now assume that $\gamma \leq \gamma^*$. The equality $\gamma^* = \gamma^{**}$ implies that W_Q is weakly convex, and, as a result, either $p^* = R = z_Q$ or $p^* = z_{Q-1}$ (maybe both) solves $\max_{x \in [z_{Q-1}, R]} W_Q(x)$. One can verify that, when $R = z_Q$, $W_{Q-1}\left(z_{Q-1}\right) \geq W_Q(R)$ if and only if $\gamma \leq \gamma^{**}$. As a result, $\gamma \leq \gamma^* = \gamma^{**}$ implies that the policy $p^* = z_{Q-1}$ solves $\max_{x \in [z_{Q-1}, R]} W_Q(x)$; the problem's value is $W_{Q-1}\left(z_{Q-1}\right)$.

Step 4: The monotonicity of p^* when in $\{z_0, z_1, \ldots, z_{Q-1}\}$.

For any $i \in \{1, 2, \ldots, Q-1\}$,

$$\frac{\partial \left(W_i\left(z_i\right) - W_{i-1}\left(z_{i-1}\right)\right)}{\partial \gamma} = \left(z_i - z_{i-1}\right)\left[2\left(R - \frac{z_{i-1} + z_i}{2}\right) + \frac{1}{K} i \left(r^e_{M+i} - z_{i-1}\right)\right] > 0,$$

where the inequality follows from $z_{i-1} < z_i < R$ and $z_{i-1} < r^e_{M+i}$.

For every $i \in \{1, 2, \ldots, Q-1\}$, the value of γ at which $W_i\left(z_i\right) = W_{i-1}\left(z_{i-1}\right)$ is γ_i given in (A.5). Therefore, $W_i\left(z_i\right) \geq W_{i-1}\left(z_{i-1}\right)$ if and only if $\gamma \geq \gamma_i$.

We show that $\gamma_i < \gamma_{i+1}$ by showing that $(i+1)\left(r^e_{M+i+1} - z_i\right) > i\left(r^e_{M+i} - z_{i-1}\right)$ and that $2R - \left(z_{i+1} + z_i\right) < 2R - \left(z_i + z_{i-1}\right)$. The latter inequality follows from $z_{i-1} < z_{i+1}$. The former one is equivalent to

$$(i+1)\, r^e_{M+i+1} - \sum_{k=M}^{M+i} r^e_k > i r^e_{M+i} - \sum_{k=M}^{M+i-1} r^e_k \quad \Longleftrightarrow \quad r^e_{M+i+1} > r^e_{M+i},$$

where $r^e_{M+i+1} > r^e_{M+i}$ follows by Assumption 1.

For further reference, $\gamma_{Q-1} < \gamma^*$. Indeed,

$$\begin{aligned}
\gamma^* &\equiv \left(1 + \frac{K}{Q\left(Q+1\right)} \frac{R - z_{Q-1}}{z_Q - z_{Q-1}}\right)^{-1} \\
&= \left(1 + \frac{K\left(R - z_{Q-1}\right)}{Q\left(r^e_{M+Q} - z_{Q-1}\right)}\right)^{-1} \\
&> \left(1 + \frac{K\left(2R - z_{Q-1} - z_{Q-2}\right)}{\left(Q-1\right)\left(r^e_{M+Q-1} - z_{Q-2}\right)}\right)^{-1} \equiv \gamma_{Q-1},
\end{aligned}$$

where the identity is definitional, the equality follows by rearranging using the definitions of z_Q and z_{Q-1}, and the inequality follows from $R - z_{Q-1} < \left(R - z_{Q-1}\right) + \left(R - z_{Q-2}\right)$ and $Q\left(r^e_{M+Q} - z_{Q-1}\right) > \left(Q-1\right)\left(r^e_{M+Q-1} - z_{Q-2}\right)$.

To summarize, $\gamma_0 \equiv 0 < \gamma_1 < \gamma_2 < \ldots < \gamma_{Q-1} < \gamma_Q \equiv \gamma^*$.

Assume that $p^* \in \arg\max_{x \in \left\{z_0, z_1, \ldots, z_{Q-1}\right\}} W_{i:x=z_i}\left(z_i\right)$. If $\gamma \in \left(\gamma_i, \gamma_{i+1}\right)$ for some $i \in \{0, 1, \ldots, Q-1\}$, then $\gamma < \gamma_{i+1} < \ldots < \gamma_{Q-1} < \gamma_Q$ implies $W_i\left(z_i\right) > W_{i+1}\left(z_{i+1}\right) > \ldots > W_{Q-1}\left(z_{Q-1}\right)$, and

$\gamma > \gamma_i > \ldots > \gamma_1 > \gamma_0$ implies $W_i(z_i) > W_{i-1}(z_{i-1}) > W_{i-2}(z_{i-2}) > \ldots > W_0(z_0)$. As a result, $p^* = z_i$ is uniquely optimal. If $\gamma = \gamma_i$ for some $i \in \{1, 2, \ldots, Q-1\}$, then $p^* = z_{i-1}$ and $p^* = z_i$ are both optimal. That is, the optimal-policy correspondence $\arg\max_{x \in \{z_0, z_1, \ldots, z_{Q-1}\}} W_{i:x=z_i}(x)$ has a selection that is nondecreasing in γ on $(0, \gamma^*)$.

Step 5: The global monotonicity of p^*

Steps 1 and 2 imply that $p^* \in \{z_0, z_1, \ldots, z_{Q-1}\} \cup [z_{Q-1}, R]$.

If $\gamma \in (\gamma_i, \gamma_{i+1}]$ for some $i \in \{0, 1, \ldots, Q-1\}$ (recall that $\gamma_0 \equiv 0$ and $\gamma_Q \equiv \gamma^*$), Step 3 implies that the value of the social welfare function (SWF) when p^* is restricted to $[z_{Q-1}, R]$ is the value of the SWT when $p^* = z_{Q-1}$, which is the choice also available in the problem considered in Step 4. Step 4 implies that the SWF achieves its highest value on $\{z_0, z_1, \ldots, z_{Q-1}\}$ when $p^* = z_i$, which is also the optimal policy in the unconstrained SWF maximization problem. Moreover, $q = i$.

If $\gamma > \gamma^*$, Step 3 implies that the maximal value of the SWF when p^* is restricted to $[z_{Q-1}, R]$ exceeds the value of the SWF when $p^* = z_{Q-1}$. Step 4 implies that SWF achieves its highest value on $\{z_0, z_1, \ldots, z_{Q-1}\}$ when $p^* = z_{Q-1}$. As a result, $\gamma > \gamma^*$ implies the optimality of choosing the p^* that satisfies (A.8). Moreover, $q = Q$.

Note that, as γ increases, both q and the described selection from the optimal-policy correspondence (singleton-valued for almost all γ) increase weakly. ■

References

Alexeev, Boris and Dustin G. Mixon, "Partisan gerrymandering with geographically compact districts," *Journal of Applied Probability*, 2018, *55* (4), 1046–1059.

Banks, Jeffrey S. and John Duggan, "Probabilistic Voting in the Spatial Model of Elections: The Theory of Office-Motivated Candidates," in David Austen-Smith and John Duggan, eds., *Social Choice and Strategic Decisions: Essays in Honor of Jeffrey S. Banks*, 2005.

Blackwell, David, "Comparisons of Experiments," *Proceedings of the Second Berkeley Symposium in Mathematical Statistics*, 1951, pp. 93–102.

—, "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 1953, *24* (2), 265–272.

Cervas, Jonathan R. and BernardGrofman, "Tools for identifying partisan gerrymandering with an application to congressional districting in Pennsylvania," *Political Georgraphy*, 2020, *76.*

Chamberlin, John R. and Paul N. Courant, "Representative Deliberations and Representative Decisions: Proportional Representation and the Borda Rule," *The American Political Science Review*, 1983, *77* (3), 718–733.

Chambers, Christopher P., Alan D. Miller, and Joel Sobel, "Flaws in the Efficiency Gap," *Journal of Law and Politics*, 2017, *33* (1), 1–33.

_ and _ , "A Measure of Bizarreness," *Quarterly Journal of Political Science*, 2010, *5* (1), 27–44.

Chen, Jowei and Jonathan Rodden, "Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures," *Quarterly Journal of Political Science*, 2013, *8*, 239–269.

Coate, Stephen and Brian Knight, "Socially Optimal Districting: A Theoretical and Empirical Exploration," *Quarterly Journal of Economics*, 2007, *122* (4), 1409–1471.

Courtault, Jean-Michel, Bertrand Crettez, and Naïla Hayek, "Characterization of Stochastic Dominance for Discrete Random Variable," 2006.

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

Duggan, John, "A Survey of Equilibrium Analysis in Spatial Models of Elections," 2017.

_ and Matthew O. Jackson, "Mixed Strategy Equilibrium and Deep Covering in Multidimensional Electoral Competition," 2005.

Ellenberg, Jordan S., "Geometry, Inference, Complexity, and Democracy," 2020.

Ely, Jeff, "A Cake-Cutting Solution to Gerrymandering," 2019.

Erikson, Robert S., "Malapportionment, Gerrymandering, and Party Fortunes in Congressional Elections," *American Political Science Review*, 1972, *66* (4), 1234–1245.

Ferejohn, John A., "On the Decline of Competition in Congressional Elections," *The American Political Science Review*, 1977, *71* (1), 166–176.

Fiorina, Morris P., "The Voting Decision: Instrumental and Expressive Aspects," *The Journal of Politics*, 1976, *38* (2), 390–413.

— , "Voting Behavior," in Dennis C. Mueller, ed., *Perspectives on Public Choice: A Handbook*, Cambridge University Press, 1996, pp. 391–414.

Friedman, John N. and Richard T. Holden, "Optimal Gerrymandering: Sometimes Pack, but Never Crack," *The American Economic Review*, 2008, *98* (1), 113–144.

Fryer, Roland G. and Richard Holden, "Measuring the Compactness of Political Districting Plans," *The Journal of Law and Economics*, 2011, *54* (3), 493–535.

Gelman, Andrew and Gary King, "A Unified Method of Evaluating Electoral Systems and Redistricting Plans," *American Journal of Political Science*, 1994, *38*, 514–554.

Gentzkow, Matthew and Emir Kamenica, "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review: Papers and Proceedings*, 2016, *106* (5), 597–601.

Gilligan, Thomas and John Matsusaka, "Structural constraints on partisan bias under the efficient gerrymander," *Public Choice*, 1999, *100*, 65–84.

Gilligan, Thomas W. and John G. Matsusaka, "Public choice principles of redistricting," *Public Choice*, 2006, *129*, 381–398.

Grofman, Bernard and Gary King, "The future of partisan symmetry as a judicial test for partisan gerrymandering after LULAC v. Perry," *Election Law Journal*, 2007.

Gul, Faruk and Wolfgang Pesendorfer, "Strategic Redistricting," *American Economic Review*, 2010, *100* (4), 1616–1641.

Hardy, G. H. and J. E. Littlewood, "A maximal theorem with function-theoretic applications," *Acta Mathematica*, 1930, *54*, 81–116.

Hinich, Melvin J., "Equilibrium in Spacial Voting: The Median Voter Result is an Artifact," *Journal of Economic Theory*, 1977, *16* (2), 208–219.

Hoffman, Alan J., "A Special Class of Doubly Stochastic Matrices," *Aequationes Mathematicae*, 1969, pp. 319–326.

Kamenica, Emir, "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 2019, *11*, 249–272.

— **and Matthew Gentzkow**, "Bayesian Persuasion," *American Economic Review*, 2011, *101* (6), 2590–2615.

Katz, Jonathan N., Gary King, and Elizabeth Rosenblatt, "Theoretical Foundations and Empirical Evaluations of Partisan Fairness in District-Based Democracies," *American Political Science Review*, 2020, *114* (1), 164–178.

Kleiner, Andy, Benny Moldovanu, and Philipp Strack, "Extreme Points and Majorization: Economic Applications," 2020.

Kolotilin, Anton and Alexander Wolitzky, "The Economics of Partisan Gerrymandering," 2020.

Levin, David A., Yuval Peres, and Elizabeth L. Wilmer, *Markov Chains and Mixing Times*, 2nd ed., American Mathematical Society, 2017.

Marshall, Albert W., Ingram Olkin, and Barry C. Arnold, *Inequalities: Theory of Majorization and Its Applications* Springer Series in Statistics, 2nd ed., Springer, 2011.

Monroe, Burt L., "Fully proportional representation," *The American Political Science Review*, 1995, *89* (4), 925–940.

Owen, Guillermo and Bernard Grofman, "Optimal partisan gerrymandering," *Political Geography Quarterly*, 1988, *7*, 5–22.

Pons, Vincent and Clémence Tricaud, "Expressive Voting and Its Cost: Evidence From Runoffs With Two or Three Candidates," *Econometrica*, 2018, *86* (5), 1621–1649.

Powell, G. Bingham, *Elections as Instruments of Democracy: Majoritarian and Proportional Visions*, Yale University Press, 2000.

Sherstyuk, Katerina, "How to gerrymander: A formal analysis," *Public Choice*, 1998, *95*, 27–49.

Soberon, Pablo, "Gerrymandering, Sandwiches, and Topology," *Notices of the AMS*, 2017, *64* (9).

Stephanopoulos, Nicholas O. and Eric M. McGhee, "Partisan Gerrymandering and the Efficiency Gap," *The University of Chicago Law Review*, 2015, *82* (2), 831–900.

Topkis, Donald M., *Supermodularity and Complementarity*, Princeton: Princeton University Press, 1998.

Tufte, Edward R., "The Relationship between Seats and Votes in Two-Party Systems," *The American Political Science Review*, 1973.

Vickrey, William, "On the Prevention of Gerrymandering," *Political Science Quarterly*, 1961, *76* (1), 105–110.