One Man, One Vote

Romans Pancs and Tridib Sharma*

January 18, 2023

Abstract

In the United States, electoral districts must be equipopulous. This requirement is known as the one man one vote doctrine. We propose welfare-based justifications for this requirement under the economic view, according to which voters care about the policy, and under the political view, according to which voters care about representation. Both justifications assume that the districter is partisan. If the districter is benevolent, one man one vote is harmless under the economic view but may reduce voter welfare under the political view by as much as the reduction from K to $\sqrt{K}$ districts would.

Keywords: one man one vote, electoral districting, gerrymandering

JEL Classification Numbers: D72, D71

1 Introduction

In the United States, each representative, a member of the state legislature, is elected by the voters who live in the electoral district where he runs for the office. District maps are redrawn decennially, after each census, and must satisfy the requirement that districts remain equipopulous as constituents migrate, are born, and die. The requirement of equipopulous districts—the one man one vote (1M1V) doctrine—was codified in a series of U.S. Supreme Court cases in the 1960s.

Neither the Founding Fathers nor the Supreme Court Justices who interpreted the Founding Fathers articulated the exact reasons for 1M1V. While the Equal Protection Clause of the 14th

*Pancs (rpancs@gmail.com) and Sharma (sharma@itam.mx) are both at ITAM. We thank Asen Kochov for the conversations that have sharpened the focus of the paper, and the seminar audiences at USC and UCSD, and Nageeb Ali for their feedback. This paper admits no more footnotes.

Amendment to the United States Constitution clearly inspired the Court, the precise logical chain that has led the Court to conclude that the Equal Protection Clause demanded 1M1V is unknown. In the words of the constitutional scholar Jacobsohn (1977), "the reasoning, whereby a newly legitimated principle was interpreted retroactively and designated unchanging, may leave those who value a reason-oriented jurisprudence disappointed." The present paper proposes the missing logical chain that culminates in the optimality of 1M1V. In doing so, the paper answers in the affirmative the question about whether the Supreme Court's rulings on 1M1V can be rationalized as features of an optimal mechanism, a solution to a problem of constitutional design. With its descriptive assumptions exposed and its normative principles explicit, the mechanism offers an opportunity to reassess the normative appeal of 1M1V.

While our question is of political nature, our analytical framework is economic, based on the idea of optimal delegation (Holmstrom, 1977, 1984). In our model, a regulator delegates the drawing of district maps to a districter. The motivation for the delegation is that the districter (typically, the political party that controls the legislature) knows how voters and their ideologies are distributed in space at the time of districting, whereas the regulator (the 1960s Supreme Court) is denied such clairvoyance and is unable to enforce contracts that are complete and contingent on future realizations of voter ideologies. Motivated by this asymmetry, the model assumes that the regulator is only capable of constraining the sizes of the districts that the districter may subsequently draw. Throughout, the number of districts is exogenously fixed (by the considerations examined by Stigler, 1976, and surveyed by Santo and Maux, 2022, and orthogonal to the forces in our model), and all districts are single-member.

The districter and the regulator have conflicting objectives. The districter is partisan. The regulator is benevolent; he cares about voter welfare. We show that if voter welfare derives from voters' concern about the policy chosen by the elected legislature (the economic view), then the regulator optimally restricts district sizes to satisfy 1M1V. He does so because 1M1V prevents the districter from inducing a policy that would be too extreme from voters' perspective.

If, by contrast, voter welfare derives from each voter's concern about how well his elected representative represents him (the political view), then the exact 1M1V is suboptimal. Instead, the regulator imposes the approximate 1M1V:

$$\frac{(\text{\# people in the largest district})}{(\text{\# people in the smallest district})} \leq \frac{(\text{\#districts}) + 1}{(\text{\#districts}) - 1}, \tag{1}$$

where the right-hand side of the inequality is pretty close to one. For example, with twenty-one districts, the right-hand side of (1) is $(21+1)\ /\ (21-1) = 1.1$, meaning that it is not merely permissible but desirable to give the districter the freedom to violate exact 1M1V by ten percent. This freedom helps exploit the partial alignment between the regulator's and the districter's conflicting objectives.

Both the ten-percent rule suggested by the example above and the rule's exact operationalization by inequality (1) are the law. In *Chapman v. Meier* (1975), the Supreme Court has ruled that 1M1V can be violated by at most ten percent and verbally formulated the "maximum population deviation" criterion that is equivalent to (1).

It is unclear whether, when deciding on *Chapman v. Meier*, the Court perceived the ten-percent rule as a bug or a feature—as a necessary allowance in the face of a districting technology incapable of precise districting or as a desideratum. In our model, the ten-percent rule is a feature. The rule emerges as optimal under the political view when in (1) we set $(\text{\#districts}) = 21$, which happens to be the number of senate districts in Alaska. In the other states, the number of districts is greater, and (1) prescribes a smaller departure from 1M1V.

Critical to the derived optimality of 1M1V, exact or approximate, is the conflict between the regulator's and the districter's objectives. If, instead, the districter shares the regulator's objective to maximize voter welfare, then, under the political view, 1M1V can hurt voter welfare as much as the counterfactual reduction in the number of districts from K down to $\sqrt{K}$ would. That is, the best map with a hundred equipopulous districts may end up delivering as little welfare as a map with ten optimally-sized districts. Fewer districts means districts that are larger and more ideologically diverse, which, under the political view, is bad for voter welfare. Under the economic view, the imposition of 1M1V on the districter who shares the regulator's objective inflicts no harm.

The paper closest in spirit to ours is Barberà and Jackson's (2006). Their central problem is dual (in an informal sense) to the problem of optimal districting. They ask: Given a map of districts that may differ in size and composition, what is the best way to aggregate district representatives' votes in the legislature? As a warm-up exercise, Barberà and Jackson observe that "if districts are small, of similar size, and of similar degrees of heterogeneity, then weighting each representative's vote equally" is best, where "best" is for voter welfare. The converse statement that if representatives' votes are weighted equally, then districts should be of similar sizes (or possess similar degrees of heterogeneity, for that matter) does not hold in our setting. That is, under the political view, welfare maximization by the districter rejects 1M1V (Proposition 2). We instead ask: Given representatives' votes are weighted equally, in what situations is it best to require all districts to be of the same size? The pertinent situations, identified in Proposition 1, involve a regulator who is as concerned about voter welfare as Barberà and Jackson are but faces a partisan districter with a conflicting objective. Our formal set up differs from Barberà and Jackson's and is closest to Gomberg, Pancs and Sharma's (2022).

Another problem that is dual (again, in an informal sense) to ours is Koriyama, Macé, Treibich and Laslier's (2013) apportionment problem. They ask: How many representatives should each district elect when district sizes are different and fixed? Koriyama, Macé, Treibich and Laslier find that, when the goal is to maximize voter welfare, smaller districts should have disproportionately many representatives. By this logic, one should not expect 1M1V to emerge from welfare maximization alone, consistent with out findings. Balinski and Young (1982) investigate the apportionment problem from the axiomatic perspective with the emphasis on the integer problem, which has no counterpart in our setting.

Existing literature on electoral districting maintains the descriptively accurate assumption that all districts are equipopulous (1M1V) instead of and focuses on the questions other than the normative rationale for 1M1V. For instance, Coate and Knight (2007) examine electoral districting from the perspective of welfare maximization. Owen and Grofman (1988), Friedman and Holden (2008), and Gul and Pesendorfer (2010) model a partisan districter who maximizes seats. Gilligan and Matsusaka's (2006) partisan districter, like ours, maximizes or minimizes the policy.

In the remainder, Section 2 introduces the model, whose assumptions are assessed in Section 3. Section 4 reveals the main result: the optimality of 1M1V, exact or approximate. Sections 5 and 6 examine two extensions. Section 5 quantifies the harm from the imposition of 1M1V on a benevolent districter. Section 6 shows that the main result remains intact even if voter preferences are not assumed to be quadratic. Section 7 concludes. Omitted proofs are in the Appendix.

2 The Setting

The **districter** draws a district map so as to influence the policy subsequently chosen by the elected legislature. The **regulator** constrains the districter's choices so as to maximize voter welfare. The equilibrium concept is subgame perfection.

District Maps

The set of geographic locations is $\mathcal{L} \equiv [0,1]$, with a typical element l. Each location houses a continuum of voters of measure one. An ideology of each voter is binary: either 0 or 1. The proportion of voters with ideology 1 at a location l is denoted by $\rho(l)$. We call $\rho : \mathcal{L} \to [0,1]$ the **affiliation function** and assume that it is strictly increasing, continuous, and satisfies $\rho(0) = 0$ and $\rho(1) = 1$. The mean ideology across all locations is $R \equiv \int_{\mathcal{L}} \rho(l)\, \mathrm{d}l$.

A **district map** is a partition $\mathbf{g} \equiv (g_1, g_2, \ldots, g_K)$ of locations $\mathcal{L}$ into an odd number K of nonnull electoral districts, where K is exogenously given and satisfies $K \geq 3$. The set of all districts is $\mathcal{K} \equiv \{1, \ldots, K\}$, with a typical element k. A map $\mathbf{g} \equiv (g_1, g_2, \ldots, g_K)$ has a **district size distribution** $(|g_1|, |g_2|, \ldots, |g_K|)$ associated with it. The set of all district size distributions is $\operatorname{int}(\Delta^{K-1})$, the interior of a simplex, with a typical element denoted by $\mathbf{s} \equiv (s_1, \ldots, s_K)$. Abusing notation slightly, we shall sometimes write $g(l) \equiv \{k \in \mathcal{K} \mid l \in g_k\}$ to denote the district into which district map $\mathbf{g}$ places location l; the function $g : \mathcal{L} \to \mathcal{K}$ is a **districting function**.

Actions

The regulator and the districter, the only two strategic players, move sequentially:

1. The regulator designates as **admissible** a subset S of the set $\operatorname{int}(\Delta^{K-1})$ of all district size distributions. The chosen S must be closed under permutation (i.e., if s' is a permutation of s, then $s \in S \implies s' \in S$), which is to say, district labels do not matter. Examples include $S = \left\{\left(\frac{1}{K}, \frac{1}{K}, \ldots, \frac{1}{K}\right)\right\}$, which corresponds to **one man one vote (1M1V)**, and $S = \operatorname{int}(\Delta^{K-1})$, which corresponds to no restrictions on district sizes.

2. The districter chooses a district map $\mathbf{g}$ with an admissible district size distribution (i.e., $(|g_1|, |g_2|, \ldots, |g_K|) \in S$).

3. The districter chosen map $\mathbf{g}$ induces a legislature $\mathbf{r}$, which, in turn, induces a policy p (as described shortly).

Payoffs

In step 3 above, a district map $\mathbf{g}$ induces a **legislature** $\mathbf{r} \equiv (r_1, r_2, \ldots, r_K)$ according to

$$r_k \equiv \frac{1}{s_k} \int_{g_k} \rho(l)\, \mathrm{d}l\,, \qquad k \in \mathcal{K}, \tag{2}$$

where r_k is interpreted as the ideology of the representative elected from a district k. This definition assumes the conclusion of the mean voter theorem (Hinich, 1977; Lindbeck and Weibull, 1987; Duggan, 2017), which holds in probabilistic voting settings. In these settings, a candidate who is uncertain about his "valence" as seen by voters aligns himself ideologically with the population mean in order to maximize the probability of winning in a two-candidate election.

The **policy** p induced by a legislature $\mathbf{r}$ is its median: $p \equiv \text{median}\{r_1, \ldots, r_K\}$. This definition assumes the conclusion of the median voter theorem (Black, 1958). Henceforth, we always relabel the districts to conform with the normalization $r_1 \leq r_2 \leq \ldots \leq r_K$, in which case the policy can be written as $p = r_M$, where $M \equiv (K+1)/2$ is the median district.

In step 2 above, the districter is either partisan (our leading formulation) or benevolent. Depending on his **partisan bias** $b \in \{min, max\}$, the **partisan** districter either minimizes ($b = min$) or maximizes ($b = max$) the policy. The **benevolent** districter maximizes voter welfare (introduced shortly). In either case, the districter knows the affiliation function ρ.

We examine two specifications of, or views about, voter welfare. The **political view** holds that each voter only cares about how well he is represented by his district representative. Given an affiliation function ρ and a district map $\mathbf{g}$, which induces a legislature $\mathbf{r}$ and has a district size distribution $\mathbf{s}$, the aggregate voter welfare from representation is

a district representative

$$W^{repr}(\rho, \mathbf{g}) \equiv -\sum_{k \in \mathcal{K}} \left[r_k\ s_k\ \left(1 - r_k\right)^2 + (1 - r_k)\ s_k\ \left(0 - r_k\right)^2 \right], \tag{3}$$

a district size

the share of ideology 1 voters

where $(1 - r_k)^2$ and $(0 - r_k)^2$ are quadratic disutilities experienced by a voter with ideologies 1 and 0, respectively, when represented by a representative with an ideology r_k. The **economic view** holds that each voter only cares about the policy, in which case the aggregate voter welfare is

$$W^{plcy}(\rho, \mathbf{g}) \equiv -\sum_{k \in \mathcal{K}} \Big[r_k \; s_k \; (1 - \overbrace{p)^2 + (1 - r_k) \; s_k \; (0 - p}^{\text{the policy}})^2 \Big], \tag{4}$$

which differs from (3) only in that a voter experiences disutility whenever his ideology departs from the policy p rather than his representative's ideology r_k.

In step 1, the regulator is benevolent: he maximizes voter welfare. When deriving the optimality of 1M1V in Proposition 1, we sometimes assume that, not knowing the realization of certain variables, the regulator plans for their worst-case realization, thereby aiming for a so-called "robust" solution (see, e.g., Ali, Haghpanah, Lin and Siegel, 2022). The variables over which the worst case must be taken for 1M1V to be optimal (exactly or approximately) vary case by case in Proposition 1 and may include the affiliation function ρ, the districter's partisan bias b, the resolution of districter's indifference among multiple district maps, and whether the economic or the political view applies. We interpret robust maximization as reflecting the regulator's ambiguity aversion in the sense of Gilboa and Schmeidler (1989); the regulator acts as if "nature" observed his choice of the admissible set S and then minimized his payoff by choosing the values of the variables that he is ambiguous about.

3 Model Assumptions in Context

Here we briefly motivate the model's assumptions. The reader who would rather judge assumptions by their implications should skip to Section 4 right now.

In its insistence on 1M1V, the Warren Court of the 1960s consistently appeals to the Equal Protection Clause of the 14th Amendment to the U.S. Constitution. To the economist, the Equal Protection Clause reads as an equal-treatment axiom and suggests two alternative interpretations: procedural and welfarist. The procedural interpretation requires all voters to have the same voting power, which has been operationalized, for example, by the Banzhaf power index (proposed by

Penrose, 1946 and further studied by Banzhaf III, 1965, 1966; Dubey and Shapley, 1979). The welfarist interpretation requires the social welfare function to treat voters symmetrically, as in (3) and (4). This paper's calling is to concern itself exclusively with the welfarist perspective.

It is natural for the economist to imagine that voters, being consequentialists, should care about the legislature's final product, the policy, as they are assumed to do in (4). This economic view was endorsed by Chief Justice Warren when 1M1V was debated in *Reynolds v. Sims (1964)*: "Since legislatures are responsible for enacting laws by which all citizens are to be governed, they should be bodies which are collectively responsive to the popular will." In the model, "laws by which all citizens are to be governed" correspond to the policy, which out to be "responsive to the popular will," captured by voter preferences.

Voter concern for representation, as in encoded in (3), figures prominently in political science (Chamberlin and Courant, 1983; Monroe, 1995). A voter may care about representation because she may believe that a representative with ideology like hers would be a better advocate for the local public goods that she likes or would be more inclined to help her with personal issues. Hinckley (1980, Table 7, p. 451) lists a variety of such representation concerns and finds that they resonate with voters. Grant and Rudolph (2004, Table 1, p. 436) find that voters prioritize U.S. Representatives' "work on local issues" over "work on national issues," and do indeed expect their representatives to "help people ... [with] personal problems with the government."

The model's benevolent regulator is a stand-in for the Supreme Court in the 1960s. No agreement on what it means to be benevolent in the context of districting exists. Some goals, such as disenfranchisement of minorities, are forbidden. But nothing is explicitly recommended. The model's identification of benevolence with the maximization of voter welfare is therefore as good any.

The model's partisan districter is a stand-in for the partisan legislature. In practice, typically, the political party that controls the legislature draws the maps. Neither the Founding Fathers nor the Supreme Court believed political parties to be benevolent. In the Federalist Paper No. 10, James Madison judges a "faction"—a political party—rather harshly as a "number of citizens, whether amounting to a majority or a minority of the whole, who are united and actuated by some common impulse of passion, or of interest, adverse to the rights of other citizens, or to the permanent and aggregate interests of the community." Madison goes on to insist that parties do not maximize the welfare of all voters: "Complaints are everywhere .... that the public good is disregarded in the conflicts of rival parties, and that measures are too often decided, not according to the rules of justice and the rights of the minor party, but by the superior force of an interested and overbearing majority." The Court was aware of, and endorsed, Madison's attitude. Writing in *Rucho v. Common Cause* (2019), Justice Kagan notes that the Framers viewed parties, "with deep suspicion, as fomenters of factionalism and symptoms of disease in the body politic." The model's assumption that parties are partisan is therefore justified.

When ruling on 1M1V, the Court was well aware that future districters, scattered over time and space, would be much better informed about the local political landscape than the Court, which would therefore have to formulate a detail-free general principle for districting. Writing for the majority in *Reynolds v Sims* (1964), Justice Warren observed that "[T]he complexions of societies and civilizations change, often with amazing rapidity. A nation once primarily rural in character becomes predominantly urban. Representation schemes once fair and equitable become archaic and outdated. But the basic principle of representative government remains, and must remain, unchanged—the weight of a citizen's vote cannot be made to depend on where he lives." This asymmetry in information is reflected in the model's assumption that the districter knows the affiliation function, whereas the regulator may perceive ambiguity about it, as well as a host of other variables.

Ambiguity in the context of constitutional decision-making has been a matter of concern among scholars of jurisprudence (Barron, 1970; Kavanaugh, 2016). No consensus guidance for passing judgements in ambiguous legal environments exists (Farnsworth, Guzior and Malani, 2010). Viscusi (1999) reports limited evidence that judges are ambiguity averse in the precise sense used by economists, consistent with the assumption that we impose on the regulator who, when faced with ambiguity, engages in robust maximization.

Our model is counterfactual in that it is not spatial. It has no language to speak about contiguity of electoral districts, even though contiguity is a constraint that a districter must respect in practice. This omission is inconsequential, however, for Sherstyuk (1998, Proposition 4) shows that, under mild conditions, the demands of contiguity are not restrictive.

Our model is incapable of requiring districts to be "compact," a desideratum for Congressional districts as noted by the Courts. This omission is not grave either, for measurement and enforcement of compactness have proved to be elusive in practice.

Equating the district representative's ideology to the district mean is no more counterfactual than equating it to the median. In fact, doing so may be a notch more realistic. Empirical evidence on the validity of both the mean voter and the median voter theorems is poor. Schofield and Sened (2005) do find some evidence for the mean voter theorem "for empirical multinomial logit and probit models of a number of elections in the Netherlands and Britain." Clinton (2006) observes that "empirical support for the median voter theory has been found lacking" but finds some support for the mean voter theory. The adoption of the district mean for its representative's ideology liberates us from the counterfactual assumption that candidates run on rigid platforms: in practice, a Democrat campaigning in Alabama espouses a rather different ideology from a Democrat campaigning in New York.

4 The Main Result

Our main result, Proposition 1, describes when 1M1V is optimal, either exactly or approximately. The proposition invokes a regularity condition that we call single-crossing.

Condition 1 (Single-Crossing)**.** The function

$$x \mapsto \frac{\int_x^1 \rho\left(l\right) \mathrm{d}l}{1-x} + \frac{\int_0^x \rho\left(l\right) \mathrm{d}l}{x} - 2\rho\left(x\right)$$

crosses zero once on $(0, 1)$.

Condition 1 embodies regularity because it tends to hold empirically. In Figure 1, the condition holds for all but one (North Dakota) U.S. state for which enough data to construct ρ are available. Condition 1 holds when the probability distribution of ideologies across locations has "light" tails. In order to see this, define $Y \equiv \rho\left(l\right)$ to be the ideology at a location l that is drawn uniformly at random. Then, the condition is equivalent to the requirement that the function

$$y \mapsto \mathbb{E}\left[Y - y \mid Y > y\right] + \mathbb{E}\left[Y - y \mid Y \leq y\right] \tag{5}$$

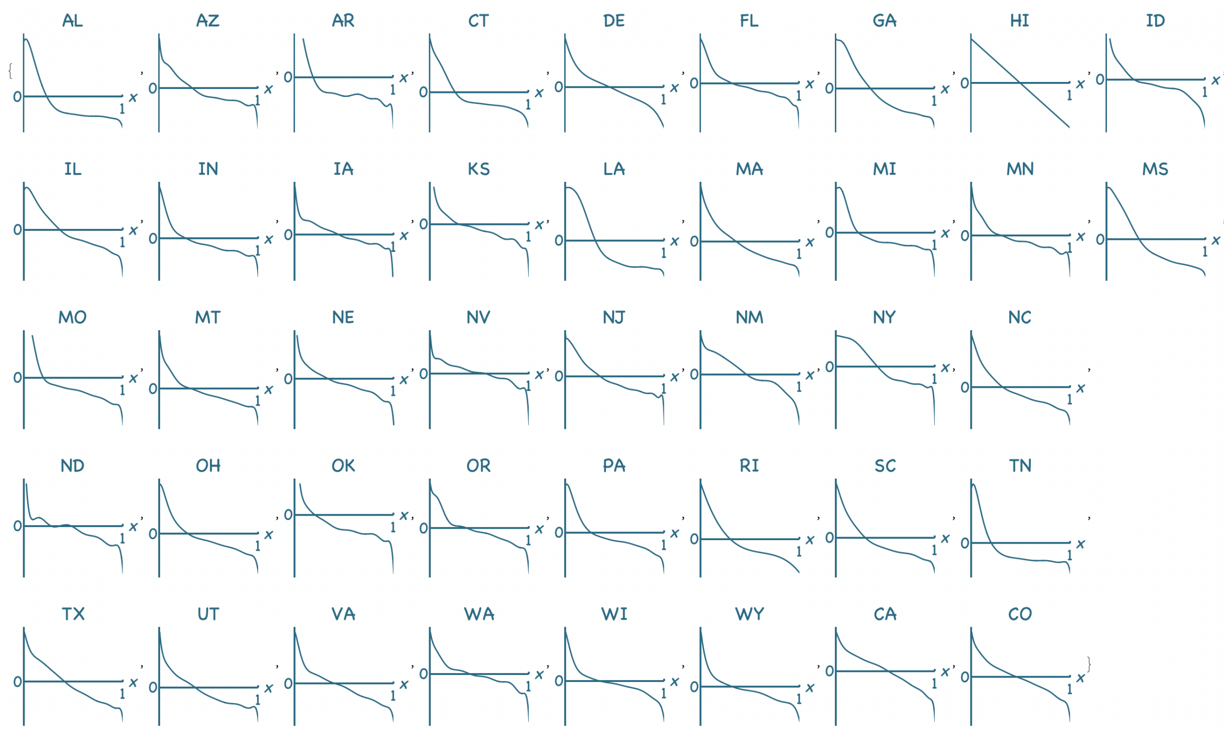

Figure 1: **The function in Condition 1.**

cross zero once on $(0,1)$. In (5), the expectations $\mathbb{E}[Y-y \mid Y>y]$ and $\mathbb{E}[Y-y \mid Y \leq y]$ are so-called mean excess functions. When a mean excess function is decreasing, the corresponding tail of the probability distribution—the left tail for $\mathbb{E}[Y-y \mid Y>y]$ and the right one for $\mathbb{E}[Y-y \mid Y \leq y]$—is "light." When both mean excess functions are decreasing, (5) is decreasing, and Condition 1 holds.

Proposition 1 (The Main Result)**.** *Suppose that the districter is partisan. Then,*

1. *(The Economic View) Under the economic view, for every affiliation function ρ, every partisan bias b, and every resolution of districter's indifference, the regulator uniquely optimally prescribes one man one vote:*

$$S=\left\{\left(\frac{1}{K},\ldots,\frac{1}{K}\right)\right\}.$$

2. *(The Political View) Under the political view, for every affiliation function ρ that satisfies Condition 1, the regulator who perceives ambiguity about the partisan bias b and the resolution of districter's indifference optimally prescribes the approximate one man one vote:*

$$S=\left\{\mathbf{s} \mid \max_{k,k'} \frac{s_k}{s_{k'}} \leq \frac{K+1}{K-1}\right\}.$$

3. *(The Worst-Case View) For every affiliation function ρ, every partisan bias b, and every resolution of districter's indifference, the regulator who perceives ambiguity about whether the economic or the political view is correct uniquely optimally prescribes one man one vote.*

In Proposition 1, the assumption that the regulator knows ρ is not meant to be descriptive. Instead, in the spirit of the dominant strategy, the assumption is meant to clarify that the proposition's conclusions about the optimality of 1M1V are independent of ρ and, therefore, hold whether the districter knows ρ (as stated in the proposition), perceives ambiguity about ρ, or holds an arbitrary probabilistic belief about ρ. The same interpretation applies to the assumptions in parts 1 and 3 that the regulator knows b and knows what the districter will do when indifferent among multiple district maps.

The intuition for the conclusion in part 1 of Proposition 1 is that, under the economic view, voter welfare is maximized by equating the policy to the population mean: $p = R$. This is the compromise policy, not too far from every voter on average. The partisan districter drives the policy away from this compromise and towards the extreme that reflects his partisan bias, b. Take this bias to be $b = min$. Then, the districter selects the smallest number of locations that are enough to form a cluster of $M \equiv (K+1)/2$ districts. He selects these locations with as low ideologies as is possible. He then spreads them across the M districts so that all these districts have the same mean ideology. This mean ideology is also the policy. The smaller the districts are allowed to be, the fewer locations enough to form a cluster of M districts, the lower ideology these locations will have as a result, and the smaller the induced policy will be, to the detriment of the voters. By maximally restricting the size of the smallest districts, 1M1V maximally inhibits the partisan districter's ability to extremize the policy.

In part 2, under the political view, each voter only cares about the ideology of the representative in his district. Thanks to the districter's efforts to extremize the policy as described in the paragraph above, the regulator expects elected representatives to come in only two ideologies: the mean ideology of every district in the M-district cluster and the mean ideology in every district outside the M-district cluster. The districts outside the M-district cluster are assumed to have the same ideology because the regulator who perceives ambiguity about the resolution of districter's indifference fears the districter will choose the worst for voter welfare, and equating mean ideologies across districts is the worst for the representation of the voters who live in these districts. Influenced in addition by his perceived ambiguity about the districter's partisan bias, the regulator reckons that voter welfare is maximized if half of them get a representative with one ideology, and the other half get a representative with the other ideology. This equal split is accomplished by requiring that M smallest districts have the same population as the remaining $K - M$ districts, which is what the approximate 1M1V in part 2 is designed to ensure.

The conclusion in part 3 is reached by reducing the problem of the regulator who perceives ambiguity about whether the economic or the political view is correct to the problem of the regulator who knows that the economic view is correct. Because the welfare loss under the economic view can be shown to always exceed the welfare loss under the political view, the ambiguity averse regulator behaves under the economic view, which is analyzed in part 1.

In preparation for a formal proof, let us rewrite voter welfare as an upper welfare bound less a loss term. Because the loss term appears in subsequent claims, we describe the welfare decomposition here, in the main text. The upper welfare bound is the same under both views in (3) and (4) and equals

$$W^* (\rho) \equiv - \int_{\mathcal{L}} \rho (l) (1 - \rho (l)) \, \mathrm{d}l \, . \tag{6}$$

Under the political view in (3), the welfare loss is

$$L^{repr} (\rho, \mathbf{g}) \equiv \int_{\mathcal{L}} \left(\rho (l) - r_{g(l)} \right)^2 \mathrm{d}l \, , \tag{7}$$

where $r_{g(l)}$ denotes the representative's ideology in the district that contains location l. Under the economic view in (4), the welfare loss is

$$L^{plcy} (\rho, \mathbf{g}) \equiv \int_{\mathcal{L}} (\rho (l) - p)^2 \, \mathrm{d}l \, . \tag{8}$$

Because W^* in (6) is independent of $\mathbf{g}$, the welfare functions (3) and (4), which can be written as $W^{repr} \equiv W^* - L^{repr}$ and $W^{plcy} \equiv W^* - L^{plcy}$, depend on $\mathbf{g}$ only through the corresponding loss terms (7) and (8). As a result, loss minimization is equivalent to welfare maximization. We are ready for the proof.

Proof of Proposition 1. The proof proceeds in steps. First it describes the districter's behavior and, then, addresses each part of the proposition.

The Districter's Behavior

Fix the set S of admissible district size distributions chosen by the regulator.

Assume that the partisan districter minimizes the policy: $b = min$. (The case of $b = max$ is treated analogously.) In order to do so, he constructs a cluster of M districts (where, recall, $M = (K+1)/2$) such that all these districts have the same mean ideology, and such that this ideology is the lowest possible. To this end, the cluster is comprised of the locations in an interval $[0, a]$ (a subset of $\mathcal{L}$) for some a, and the locations are mixed across the districts of the cluster so that the mean ideology in each of these districts is

$$\sigma(a) \equiv \frac{1}{a} \int_0^a \rho(l)\, \mathrm{d}l\,. \tag{9}$$

This mean ideology is minimized by making the districts in the cluster as small as the regulator would allow, which means that $a = \min_{s \in S} \sum_{k=1}^{M} s_k$. Because each of the $K - M$ districts outside the M-district cluster has a higher mean ideology than $\sigma(a)$, the policy is dictated by the districts in the M-district cluster: $p = \sigma(a)$. This is the smallest policy that can be induced given S.

The districter does not care about the composition of the remaining $K - M$ districts. Nor do the voters under the economic view, to whom only the policy $\sigma(a)$ matters. Under the political view, however, the voters do care. In this case, the voter-pessimal composition of the remaining $K - M$ districts (which is the resolution of the districter's indifference that will be relevant in further analysis) equates the mean ideology in each of these districts to

$$\delta(a) \equiv \frac{1}{1-a} \int_a^1 \rho(l)\, \mathrm{d}l\,, \tag{10}$$

as can be shown.

Let $\mathbf{g}^{a,min}$ denote the district map whose construction is described above, the map that induces the policy-minimizing legislature $(\sigma(a), \ldots, \sigma(a), \delta(a), \ldots, \delta(a))$, where $\sigma(a)$ is repeated M times, $\delta(a)$ is repeated $K - M$ times, and $p = \sigma(a)$ is the policy.

In the analogous case of $b = max$, let $\mathbf{g}^{a,max}$ denote the analogously constructed district map, the map that induces the policy-maximizing legislature $(\sigma(1-a), \dots \sigma(1-a), \delta(1-a), \dots, \delta(1-a))$, where $\delta(1-a)$ is repeated M times, $\sigma(1-a)$ is repeated $K - M$ times, and $p = \delta(1-a)$ is the policy.

Note that, by varying S, the regulator can induce any value of $a = \min_{s \in S} \sum_{k=1}^{M} s_k$ in $\left(0, \frac{1}{2} + \frac{1}{2K}\right]$. For instance, the interval's upper boundary is uniquely induced by $S = \left\{\left(\frac{1}{K}, \dots, \frac{1}{K}\right)\right\}$, which corresponds to 1M1V. This range of the values of a that the regulator can induce determines the range of the policies that he can induce the partisan districter to implement.

Part 1

Under the economic view, the welfare loss in (8) as a function of a is

$$L^{plcy}\left(\rho, \mathbf{g}^{a,min}\right) \equiv \int_0^1 (\rho(l) - \sigma(a))^2 \, \mathrm{d}l \,, \tag{11}$$

whose derivative with respect to a is $2\sigma'(a)(\sigma(a) - R)$. This derivative is negative because σ is strictly increasing and because $\sigma(a) < R = \sigma(1)$ for $a < 1$. Therefore, in order to minimize L^{plcy}, the regulator induces the maximal value of a in $\left(0, \frac{1}{2} + \frac{1}{2K}\right]$, which is accomplished by choosing $S = \left\{\left(\frac{1}{K}, \dots, \frac{1}{K}\right)\right\}$, 1M1V.

An analogous argument establishes that 1M1V is uniquely optimal when $b = max$. Indeed, the welfare loss $L^{plcy}\left(\rho, \mathbf{g}^{a,min}\right) \equiv \int_0^1 (\rho(l) - \delta(1-a))^2 \, \mathrm{d}l$ is decreasing in a and, therefore, is minimized at the upper boundary of $\left(0, \frac{1}{2} + \frac{1}{2K}\right]$, which corresponds to 1M1V. Figure 2 illustrates both functions $L^{plcy}\left(\rho, \mathbf{g}^{a,min}\right)$ and $L^{plcy}\left(\rho, \mathbf{g}^{a,min}\right)$.

Part 1 of the proposition follows.

Part 2

Under the political view, when $b = min$, the welfare loss in (7) as a function of a is

$$L^{repr}\left(\rho, \mathbf{g}^{a,min}\right) \equiv \int_0^a (\rho(l) - \sigma(a))^2 \, \mathrm{d}l + \int_a^1 (\rho(l) - \delta(a))^2 \, \mathrm{d}l \,, \tag{12}$$

whose derivative with respect to a is $(\rho(a) - \sigma(a))^2 - (\delta(a) - \rho(a))^2$. The derivative has the same sign as $2\rho(a) - \sigma(a) - \delta(a)$ does. The sign is first negative and then positive, by Condition 1. As a result, $L^{repr}\left(\rho, \mathbf{g}^{a,min}\right)$ is single-dipped.

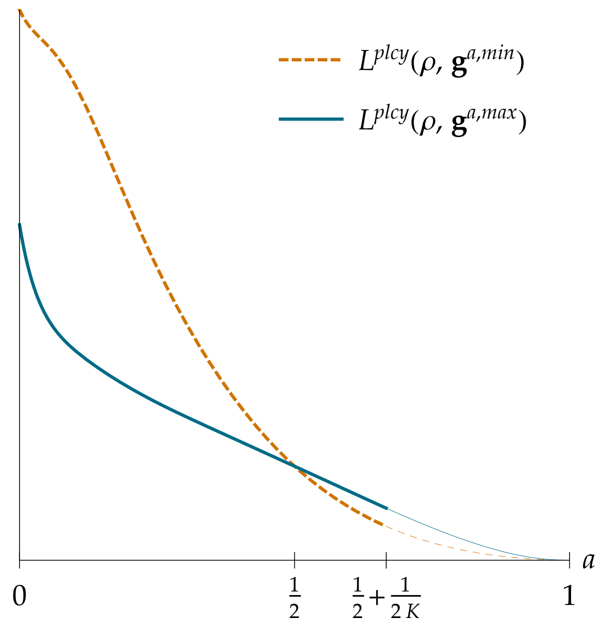

Figure 2: **One man one vote minimizes the welfare loss under the economic view for any $b \in \{min, max\}$.** The losses when the districter minimizes the policy (the dashed curve) and when he maximizes the policy (the solid curve) are each decreasing in a and, therefore, are each minimized at the largest possible value of a, which is $\frac{1}{2} + \frac{1}{2K}$.

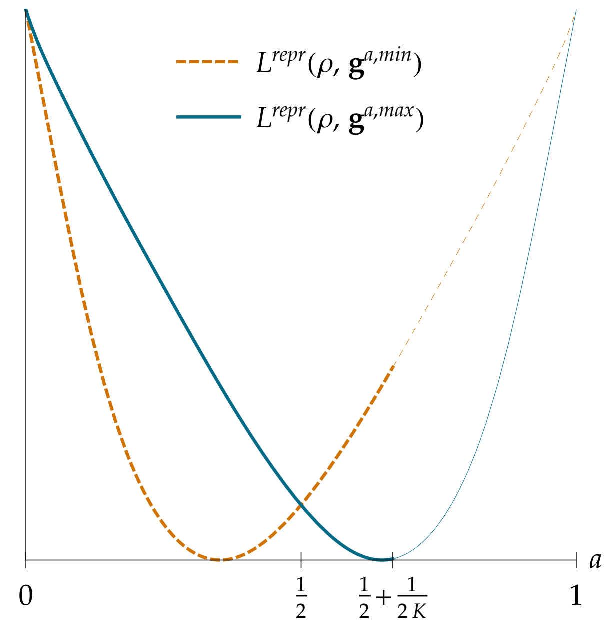

Figure 3: **The approximate one man one vote minimizes the welfare loss under the political view for the worst-case $b \in \{min, max\}$.** The upper envelope of the losses when the districter minimizes the policy (the dashed curve) and when he maximizes the policy (the solid curve) is minimized at $a = \frac{1}{2}$.

When $b = max$, the counterpart of the loss in (12) is

$$L^{repr}\left(\rho, \mathbf{g}^{a,max}\right) = \int_0^{1-a} \left(\rho\left(l\right) - \sigma\left(1-a\right)\right)^2 \mathrm{d}l + \int_{1-a}^{1} \left(\rho\left(l\right) - \delta\left(1-a\right)\right)^2 \mathrm{d}l\,,$$

which is a horizontal reflection of $L^{repr}\left(\rho, \mathbf{g}^{a,min}\right)$ about the vertical axis corresponding to $a = \frac{1}{2}$ and, therefore, is also single-dipped.

The regulator who perceives ambiguity about b and about the resolution of districter's indifference chooses an a in $\left(0, \frac{1}{2} + \frac{1}{2K}\right]$ to minimize the maximal loss

$$\max\left\{L^{repr}\left(\rho, \mathbf{g}^{a,min}\right), L^{repr}\left(\rho, \mathbf{g}^{a,max}\right)\right\}.$$

This maximal loss (an upper envelope of two single-dipped functions, one a reflection of the other around $a = \frac{1}{2}$) is uniquely minimized at $a = \frac{1}{2}$, as Figure 3 illustrates. The value $a = \frac{1}{2}$ is induced by the approximate 1M1V in part 2 of the proposition. Indeed, if the districter chooses M districts of size $\frac{1}{K+1}$ and the remaining $K - M$ districts are of size $\frac{1}{K-1}$, then $M \times \frac{1}{K+1} = \frac{1}{2}$ (which is the target value for a), $M \times \frac{1}{K+1} + (K-M) \times \frac{1}{K-1} = 1$ (the total population of voters adds up), and $\frac{1}{K-1} / \frac{1}{K+1} = \frac{K+1}{K-1}$ (the ratio of the largest district to the smallest one is as specified in part 2). Part 2 of the proposition follows.

Part 3

Consider the case of $b = min$. Then, for any a in $\left(0, \frac{1}{2} + \frac{1}{2K}\right]$, we have

$$\begin{aligned} L^{plcy}\left(\rho, \mathbf{g}^{a,min}\right) - L^{repr}\left(\rho, \mathbf{g}^{a,min}\right) &= a\sigma\left(a\right)^2 + \left(1-a\right)\delta\left(a\right)^2 - \left(2\sigma\left(a\right)R - \sigma\left(a\right)^2\right) \\ &> a\sigma\left(a\right)^2 + \left(1-a\right)\delta\left(a\right)^2 - R^2 \\ &= a\sigma\left(a\right)^2 + \left(1-a\right)\delta\left(a\right)^2 - \left(a\sigma\left(a\right) + \left(1-a\right)\delta\left(a\right)\right)^2 \\ &> 0, \end{aligned}$$

where the first equality combines the expressions for L^{plcy} and L^{repr} in (11) and (12) with the definitions of σ and δ in (9) and (10) and rearranges; the first inequality uses the fact that $2\sigma\left(a\right)R - \sigma\left(a\right)^2$ is strictly increasing in $\sigma\left(a\right)$ because $\sigma\left(a\right) < R$ (by $a < 1$), and replaces $\sigma\left(a\right)$ with R; the second equality uses the definitions of σ, δ, and R; and the second inequality is Jensen's.

From the display above, conclude that $L^{plcy}\left(\rho, \mathbf{g}^{a,min}\right) > L^{repr}\left(\rho, \mathbf{g}^{a,min}\right)$, where, recall, $\mathbf{g}^{a,\min}$ is the policy-minimizing map that maximizes the welfare loss from representation. If $\hat{\mathbf{g}}^{a,min}$ is an arbitrary policy-minimizing map, then $L^{repr}\left(\rho, \hat{\mathbf{g}}^{a,min}\right) \leq L^{repr}\left(\rho, \mathbf{g}^{a,min}\right)$ and $L^{plcy}\left(\rho, \hat{\mathbf{g}}^{a,min}\right) = L^{plcy}\left(\rho, \mathbf{g}^{a,min}\right)$, which combined with the display above implies that $L^{plcy}\left(\rho, \hat{\mathbf{g}}^{a,min}\right) > L^{repr}\left(\rho, \hat{\mathbf{g}}^{a,min}\right)$.

Letting $\hat{\mathbf{g}}^{a,max}$ denote an arbitrary policy-maximizing map, one can similarly conclude that $L^{plcy}\left(\rho, \hat{\mathbf{g}}^{a,max}\right) > L^{repr}\left(\rho, \hat{\mathbf{g}}^{a,max}\right)$.

As a result, the regulator who observes ρ, b, and the resolution of districter's indifference but perceives ambiguity about the correct view chooses an S to induce the value of a that minimizes

$$\max\left\{L^{plcy}\left(\rho, \hat{\mathbf{g}}^{a,b}\right), L^{repr}\left(\rho, \hat{\mathbf{g}}^{a,b}\right)\right\} = L^{plcy}\left(\rho, \hat{\mathbf{g}}^{a,b}\right),$$

which by part 1 is uniquely minimized at $S = \left\{\left(\frac{1}{K}, \ldots, \frac{1}{K}\right)\right\}$, 1M1V. The conclusion in part 2 follows. ■

5 Extension I: What Can Go Wrong?

Proposition 1 assumes that the districter is partisan and recommends 1M1V, exact or approximate. While partisan districters dominate U.S. politics, some states (e.g., Alaska and California) have made an effort to appoint purportedly nonpartisan districting commissions. The generous interpretation is that these commissions are benevolent. If they indeed are, then what, if any, is the welfare cost of imposing 1M1V on them? Clearly, there can be no benefit in constraining the districter whose objective is aligned with the regulator's. But can there be harm?

Proposition 2 shows that 1M1V imposed on a benevolent districter is harmless under the economic view but may harm voters under the political view. In the proposition, $L^{repr}_{1M1V}(\rho) \equiv \inf_{\mathbf{g}:s_1=\ldots=s_K} L^{repr}(\rho, \mathbf{g})$ denotes the welfare loss under the political view when the benevolent districter is constrained by 1M1V (i.e., $s_1 = \ldots = s_K$). The corresponding loss when the districter is unconstrained is denoted by $L^{repr}_{OPT}(\rho) \equiv \inf_{\mathbf{g}} L^{repr}(\rho, \mathbf{g})$.

Proposition 2 (Benevolent Districter)**.** *Suppose that the districter is benevolent. Then, the reduction in voter welfare due to the imposition of one man one vote:*

1. *Is zero for any affiliation function ρ under the economic view.*

2. *Under the political view,*

   (a) *is positive and is of the order* $1/K$ *for some* ρ*, as is implied by*

$$\sup_{\rho} L^{repr}_{1M1V}(\rho) = \frac{1}{4K} \qquad \textit{and} \qquad \sup_{\rho} L^{repr}_{OPT}(\rho) \leq \frac{1}{4K^2};$$

   (b) *is positive and is of the order* $1/K^2$ *for every nonlinear and smooth* ρ*, as is implied by:*

$$L^{repr}_{1M1V}(\rho) \approx \frac{1}{12K^2}\int_0^1 \rho'(l)^2 \, \mathrm{d}l \qquad \textit{and} \qquad L^{repr}_{OPT}(\rho) \approx \frac{1}{12K^2}\left(\int_0^1 \rho'(l)^{\frac{2}{3}} \, \mathrm{d}l\right)^3.$$

Part 1 of Proposition 2 says that 1M1V is not restrictive under the economic view. The conclusion holds because, no matter the district size distribution, all districts can be made ideologically "representative" of the entire population as a consequence of the measure-theoretic result known as the Dvoretzky–Wald–Wolfowitz purification theorem. Intuitively, in order to construct one "representative" district of a desired size, one can sample uniformly at random a continuum of locations from the set $\mathcal{L}$ of all locations. The informal "law of large numbers for the continuum" implies that the mean ideology in this representative district is the mean ideology in the population, which is R. The remaining districts can be made representative by iterating on the same random sampling procedure. Because all districts so constructed have the same mean ideology R, so does the median district, leading to the policy R.

In order to animate the assertions in parts 2a and 2b of Proposition 2, we define regret. **Regret** is how much higher welfare would have been if the regulator had not imposed 1M1V on the benevolent districter under the political view:

$$\textit{regret}(\rho) = L^{repr}_{1M1V}(\rho) - L^{repr}_{OPT}(\rho).$$

Part 2a can be read to say that one can find a ρ for which the regret from imposing 1M1V is of the order $1/K$ when K is "large." The regret is of the order $1/K$ because $L^{repr}_{1M1V}(\rho) = 1/(4K)$ is possible for some ρ, whereas $L^{repr}_{OPT}(\rho) \leq 1/\left(4K^2\right)$ holds for every ρ. Part 2b can be read to say that, for every smooth nonlinear ρ, the regret from imposing 1M1V is of the order $1/K^2$ when K is "large." Indeed,

$$regret\,(\rho) \approx \frac{1}{12K^2}\left[\int_0^1 \rho'(l)^2\,\mathrm{d}l - \left(\int_0^1 \rho'(l)^{\frac{2}{3}}\,\mathrm{d}l\right)^3\right],$$

which is positive when ρ is nonlinear by Jensen's inequality and is of the order $1/K^2$ by inspection. The maximal regret is larger in part 2a than in part 2b because part 2b requires ρ to be smooth, whereas part 2a imposes no such restriction.

A corollary to parts 2a and 2b of Proposition 2 is that, should the political view be correct, voters are bound to be worse off with the partisan districter than they would have been with the benevolent one, in spite of the regulator's best efforts. This is so because voter welfare drops once 1M1V is imposed on the benevolent districter (parts 2a and 2b), and the partisan districter can only do weakly worse at maximizing welfare than the benevolent one.

6 Extension II: Beyond the Quadratic Case

Replace the quadratic disutilities $(1-r_k)^2$ and $(0-r_k)^2$ in the welfare from representation function (3) by $u\,(1-r_k)$ and $u\,(-r_k)$ for any function $u : [-1,1] \to \mathbb{R}_+$ that is **nicely convex** in the sense of being strictly convex, as well as bounded, continuously differentiable, and minimized at zero. Perform the same replacement for the social welfare function in (4). The probabilistic voting model (Duggan, 2017, Theorems 7 and 10) predicts that the representative elected from a district with a mean ideology μ will have the ideology that maximizes the utilitarian welfare (by minimizing welfare from misrepresentation),

$$r\,(\mu) \equiv \arg\min_{x\in[0,1]} \left\{\mu u\,(1-x) + (1-\mu)\,u\,(-x)\right\}. \tag{13}$$

We assume the relationship in (13). The special case of $r\,(\mu) \equiv \mu$ corresponds to the quadratic u. For realism, one can choose that u which induces the function $r\,(\mu)$ that matches best the data on how a candidate's ideology responds to voter ideology. One can verify that, for every u that is nicely convex, the candidate's ideology function r is strictly increasing.

It turns out that, as long as u is nicely convex, the conclusions in Proposition 1 continue to hold provided the single-crossing Condition 1 is replaced with Condition 2.

Condition 2 (Generalized Single-Crossing)**.** The function

$$x \mapsto (1-\rho(x))\left[u(-r(\sigma(x))) - u(-r(\delta(x)))\right] + \rho(x)\left[u(1-r(\sigma(x))) - u(1-r(\delta(x)))\right]$$

crosses zero once on $(0,1)$.

Condition 2 reduces to Condition 1 when u is quadratic and plays the same role: it ensures that voter welfare from representation is single-peaked (equivalently, welfare loss from misrepresentation is single-dipped, as in Figure 3) in the stringency of the constraint the regulator imposes on the districter. The condition is presented here for completeness, not because it is easy to interpret, which it is not. The condition suffices to ascertain, though, that the main result, Proposition 1, survives departures from the quadratic specification of voter preferences.

Proposition 3 (The Main Result Revisited)**.** *Proposition 1 continues to hold for all nicely convex disutility functions if Condition 1 in part 2 is replaced with Condition 2.*

7 Concluding Remarks

The paper has shown that 1M1V and the maximum population deviation criterion (1) used by the courts of law admit a formal justification grounded in concerns for voter welfare. The justification relies on the assumption that the districter is partisan, which is the empirically relevant case. If the districter were benevolent, then 1M1V would damage voter welfare from representation by as much as the reduction from K electoral districts down to $\sqrt{K}$ would. (The welfare from the policy would remain intact.) The charitable interpretation of the historical opposition to 1M1V in the United States is that voters (who rejected 1M1V in ten referenda between 1946 and 1962, as noted by Ansolabehere and Issaacharoff, 2003) cared about representation and believed their districter to be benevolent.

The general thrust of the paper's main result, the optimality of 1M1V, is that a partisan districter ought to be constrained maximally, or almost maximally. If so, can the districter be further constrained by means other than setting admissible district sizes? One possibility is to garble the districter's information about the distribution of voter ideologies. This garbling can be achieved by committing to collect coarser information about voter behavior or to collect no information at all. For instance, in the United Kingdom, there is no analogue of the precinct-level voting data that is available in the United States. Garbling would benefit the regulator under the economic view by making it harder for the districter to assemble districts with sufficiently disparate ideologies in order to affect the policy substantially. By contrast, under the political view, garbling would harm the regulator by impairing voter representation. The increase in the granularity of locations (infinitesimal in our model) would have effects similar to garbling. Both garbling and granularity can be modeled as "flattening" of the affiliation function, ρ. Because the conclusion of Proposition 1 holds for all ρ, the optimality of 1M1V, exact or approximate, can be analyzed separately from garbling, granularity, and other policies that deform ρ.

A Appendix: Omitted Proofs

A.1 Proof of Proposition 2

Part 1

We define two measures, μ_{voters} and $\mu_{Republicans}$, on $\mathcal{L}$. Measure μ_{voters} is the Lebesgue measure and describes the population of voters on a subset of $\mathcal{L}$. The measure $\mu_{Republicans}$ is induced by ρ and describes the measure of ideology-1 voters on a subset of $\mathcal{L}$. With K districts, let $f_k(l)$ denote the fraction of the voters at a location l that are placed into a district k, with $\sum_{k\in\mathcal{K}} f_k(l) = 1$. In particular, set $f_k(l) = \frac{1}{K}$ for all l and k, so that each location is split evenly across all districts. Then, every district k has the same measure of voters and the same mean ideology:

$$\int_{\mathcal{L}} f_k(l)\,\mathrm{d}\mu_{voters}(l) = \frac{1}{K} \qquad \text{and} \qquad K\int_{\mathcal{L}} f_k(l)\,\mathrm{d}\mu_{Republicans}(l) = R.$$

The functions $(f_k)_{k\in\mathcal{K}}$ do not describe a district map because they split individual locations, whereas a district map is not allowed to split. However, the Dvoretzky–Wald–Wolfowitz purification theorem reported by Khan, Rath and Sun (2006, Theorem DWW, p. 93) implies existence of characteristic functions $(f_k^*)_{k\in\mathcal{K}}$ with $f_k^*:\mathcal{L}\to\{0,1\}$ that, for each k, satisfy

$$\int_{\mathcal{L}} f_k^*(l)\,\mathrm{d}\mu_{voters}(l) = \int_{\mathcal{L}} f_k(l)\,\mathrm{d}\mu_{voters}(l) \quad \text{and} \quad \int_{\mathcal{L}} f_k^*(l)\,\mathrm{d}\mu_{Republicans}(l) = \int_{\mathcal{L}} f_k(l)\,\mathrm{d}\mu_{Republicans}(l)\,.$$

The collection $(f_k^*)_{k\in\mathcal{K}}$ induces a district map $\mathbf{g} = (g_k)_{k\in\mathcal{K}}$ by letting $l \in g_k$ if and only if $f_k^*(l) = 1$. Combining the two displays above implies that the induced district map $\mathbf{g}$ satisfies 1M1V and induces the same mean ideology R in every district. The implied policy is also R, which is the policy that maximizes welfare under the economic view with no 1M1V imposed. Hence, the imposition of 1M1V does not lower welfare under the economic view.

An Auxiliary Observation for Parts 2a and 2b

Let $X(\mathbf{s}) \subset \mathbb{R}_+^K$ denote the set of **implementable legislatures**, that is, the legislatures each of which can be induced by some district map with the district size distribution $\mathbf{s} \equiv (s_1, \ldots, s_K)$. The districting problem that characterizes $X(\mathbf{s})$ can be reinterpreted as the Bayesian persuasion problem in which the sender wants to characterize the set of all posterior mean beliefs that he can induce the receiver to hold by devising a signal structure with K signal realizations such that the probability that a signal k in $\mathcal{K}$ is realized is s_k. Gentzkow and Kamenica (2016) have solved this Bayesian persuasion problem. Adapting their results, we assert that a legislature $\mathbf{r}$ is in $X(\mathbf{s})$ if and only if $\mathbf{r}$ is $\mathbf{s}$-majorized by the **extreme legislature** $\mathbf{r}^e(\mathbf{s})$ defined as

$$\mathbf{r}^e(\mathbf{s}) = \left(\frac{1}{s_1}\int_0^{s_1} \rho(l)\,\mathrm{d}l, \frac{1}{s_2}\int_{s_1}^{s_1+s_2} \rho(l)\,\mathrm{d}l, \ldots, \frac{1}{s_K}\int_{s_1+\ldots+s_{K-1}}^{1} \rho(l)\,\mathrm{d}l\right). \tag{14}$$

The legislature $\mathbf{r}^e(\mathbf{s})$ **s-majorizes** a legislature $\mathbf{r}$ if $\sum_{k\in\mathcal{K}} s_k r_k = \sum_{k\in\mathcal{K}} s_k r_k^e$ and, for each k' in $\mathcal{K}$, $\sum_{k=1}^{k'} s_k r_k \geq \sum_{k=1}^{k'} s_k r_k^e$.

Given a district size distribution $\mathbf{s}$, voter welfare from representation can be written as $\sum_{k\in\mathcal{K}} s_k r_k^2$ plus a constant that is independent of $\mathbf{r}$, where each r_k^2 is continuous and convex. Proposition 14.A.2 of Marshall, Olkin and Arnold (2011, p. 580) then implies that $\mathbf{r}^e(\mathbf{s})$ maximizes voter welfare (equivalently, minimizes the welfare loss) from representation on $X(\mathbf{s})$. The import of this result for proving parts 2a and 2b is that it does not matter whether one minimizes the welfare loss from representation subject to a fixed admissible district size distribution (i.e., the one corresponding to 1M1V) or subject to a nonsingleton set of admissible district size distributions (i.e., the set of all feasible distributions), one can restrict attention to the district maps that induce extreme legislatures in (14), which are the maps that partition $\mathcal{L}$ into intervals.

Part 2a

In order to show that $\sup_\rho L^{repr}_{OPT}(\rho) \leq 1/\left(4K^2\right)$, consider first an auxiliary district map, called **uniformly diverse**, or DIV for short, and denoted by $\mathbf{g}^{DIV}$:

$$\mathbf{g}^{DIV} \equiv \{[z_0, z_1], (z_1, z_2], \ldots, (z_{K-2}, z_{K-1}], (z_{K-1}, z_K]\}, \tag{15}$$

where $z_i \equiv \rho^{-1}\left(\frac{i}{K}\right)$ for each $i = 0, 1, \ldots, K$. By construction, DIV in (15) makes each district representative a custodian of the same range of adjacent ideologies but, in general, a different number of voters. With some abuse of notation, let

$$r(z_{i-1}, z_i) \equiv \frac{1}{z_i - z_{i-1}} \int_{z_{i-1}}^{z_i} \rho(l)\, \mathrm{d}l$$

denote the ideology of the district representative in a district $(z_{i-1}, z_i]$. Letting $L^{repr}_{DIV}(\rho)$ denote the welfare loss associated with DIV under the political view, we have

$$\begin{aligned} L^{repr}_{DIV}(\rho) &= \sum_{i=1}^{K} \int_{z_{i-1}}^{z_i} \left(\rho(l) - r(z_{i-1}, z_i)\right)^2 \mathrm{d}l \\ &\leq \sum_{i=1}^{K} \frac{z_i - z_{i-1}}{4K^2} \\ &= \frac{1}{4K^2}, \end{aligned} \tag{16}$$

where only the inequality in the second line requires a justification. The bound $L^{repr}_{DIV}(\rho) \leq 1/K^2$, which is weaker than the bound $L^{repr}_{DIV}(\rho) \leq 1/\left(4K^2\right)$ in (16), follows immediately from $\left(\rho(l) - r(z_{i-1}, z_i)\right)^2 \leq 1/K^2$, which, in turn, is implied by the fact that both $\rho(l)$ and $r(z_{i-1}, z_i)$ are confined to the interval $[\rho(z_{i-1}), \rho(z_i)]$ of the length $1/K$. For the actual inequality in (16), we shall establish that L^{repr}_{DIV} is maximized at an affiliation function that is a step function. (A step function is neither continuous nor strictly increasing but can be approximated by a function that is both continuous and strictly increasing, which are the maintained assumptions on ρ in this paper.)

Given an arbitrary affiliation function ρ, define a corresponding step function $\hat{\rho} : \mathcal{L} \to [0, 1]$ to satisfy

1. $(\forall i = 0, 1, \ldots, K)\, \hat{\rho}(z_i) = \rho(z_i) = \frac{i}{K}$, and

2. $(\forall i = 1, \ldots, K)\,(\forall l \in (z_{i-1}, z_i))\,\hat{\rho}(l) = \mathbf{1}_{\{l<y_i\}}\rho(z_{i-1}) + \mathbf{1}_{\{l \geq y_i\}}\rho(z_i)$, where y_i is defined implicitly by $\int_{z_{i-1}}^{z_i} \rho(l)\,\mathrm{d}l = \int_{z_{i-1}}^{z_i} \left(\mathbf{1}_{\{l<y_i\}}\rho(z_{i-1}) + \mathbf{1}_{\{l \geq y_i\}}\rho(z_i)\right)\mathrm{d}l$.

By construction, for all $i = 1, \ldots, K$ and all $y \in (z_{i-1}, z_i)$, we have $\int_{z_{i-1}}^{y} \hat{\rho}(l)\,\mathrm{d}l \leq \int_{z_{i-1}}^{y} \rho(l)\,\mathrm{d}l$ and $\int_{z_{i-1}}^{z_i} \hat{\rho}(l)\,\mathrm{d}l = \int_{z_{i-1}}^{z_i} \rho(l)\,\mathrm{d}l$. Then, Theorem Marshall, Olkin and Arnold (2011, Theorem D.22, p. 22) implies that $\int_{z_{i-1}}^{z_i} \phi(\rho(l))\,\mathrm{d}l \leq \int_{z_{i-1}}^{z_i} \phi(\hat{\rho}(l))\,\mathrm{d}l$ for any continuous convex function ϕ and, in particular, for the function $\phi(x) \equiv (x - r(z_{i-1}, z_i))^2$. Conclude that $L^{repr}(\rho) \leq L^{repr}(\hat{\rho})$.

Now one can ask what the worst step function on each interval $[z_{i-1}, z_i]$ looks like. If $\hat{\rho}$ jumps from $(i-1)/K$ to i/K at $y_i \in (z_{i-1}, z_i)$, the corresponding component of the welfare loss $L^{repr}(\hat{\rho})$ on $[z_i, z_{i-1}]$ can be verified to be

$$\int_{z_{i-1}}^{z_i} (\hat{\rho}(l) - r(z_{i-1}, z_i))^2 = \frac{(z_i - y_i)(y_i - z_{i-1})}{K^2(z_i - z_{i-1})}.$$

This component is uniquely maximized at $y_i = (z_i + z_{i-1})/2$, reaching the value $(z_i - z_{i-1})/\left(4K^2\right)$. The inequality in (16) then follows. Moreover, the inequality in (16) cannot be improved upon because every step function $\hat{\rho}$ can be approximated arbitrarily closely by a continuous and strictly increasing affiliation function.

The just obtained $L^{repr}_{DIV}(\rho) \leq 1/\left(4K^2\right)$ coupled with $L^{repr}_{OPT}(\rho) \leq L^{repr}_{DIV}(\rho)$ implies $L^{repr}_{OPT}(\rho) \leq 1/\left(4K^2\right)$, as stated in part 2a of the proposition.

The argument used to establish that a step function maximizes the welfare loss from representation under DIV establishes that a step function of the form

$$\hat{\rho}(l) = \sum_{k \in \mathcal{K}} \delta_k \mathbf{1}_{\left\{l \in \left[\frac{k}{K} - \frac{1}{2K}, \frac{k}{K} + \frac{1}{2K}\right)\right\}}, \qquad \text{where} \quad \sum_{k \in \mathcal{K}} \delta_k = 1 \quad \text{and} \quad (\forall k \in \mathcal{K})\,\delta_k \geq 0$$

maximizes the welfare loss from representation under 1M1V. The associated welfare loss is

$$L^{repr}_{1M1V}(\hat{\rho}) = \frac{1}{4K} \sum_{k \in \mathcal{K}} \delta_k^2,$$

which is maximized at $\hat{\rho}(l) = \mathbf{1}_{\left\{l \geq \frac{1}{2K}\right\}}$, which by setting $\delta_1 = 1$ and $\delta_k = 0$ for $k \neq 1$.

In order to show $\sup_{\rho} L^{repr}_{1M1V}(\rho) = 1/(4K)$, note that an argument analogous to the one used to establish that a step function maximizes L^{repr}_{DIV} implies that a step function also minimizes the welfare loss from representation under 1M1V. The analogous argument is appropriate because under

1M1V, the welfare-maximizing district map, denoted by $\mathbf{g}^{1M1V}$, must induce the extreme legislature $\mathbf{r}^{e}\left(\left\{\frac{1}{K},\ldots,\frac{1}{K}\right\}\right)$ (by the "Auxiliary Observation for Parts 2a and 2b" above) and, therefore, partitions $\mathcal{L}$ into intervals, just as $\mathbf{g}^{DIV}$ does in (15), except the intervals are known:

$$\mathbf{g}^{1M1V} \equiv \left\{\left[0,\frac{1}{K}\right],\left(\frac{1}{K},\frac{2}{K}\right],\ldots,\left(\frac{K-2}{K},\frac{K-1}{K}\right],\left(\frac{K-1}{K},1\right]\right\}.$$

The step function that maximizes the welfare loss under $\mathbf{g}^{1M1V}$ takes the form

$$\rho\left(l\right)=\sum_{k\in\mathcal{K}}\delta_{k}\mathbf{1}_{\left\{l\in\left[\frac{k}{K}-\frac{1}{2K},\frac{k}{K}+\frac{1}{2K}\right)\right\}},\qquad \text{where}\quad \sum_{k\in\mathcal{K}}\delta_{k}=1\quad\text{and}\quad\left(\forall k\in\mathcal{K}\right)\delta_{k}\geq 0.$$

The associated value of L_{1M1V}^{repr} is $\sum_{k\in\mathcal{K}}\delta_{k}^{2}/\left(4K\right)$ and is maximized by setting $\left(\delta_{1},\delta_{2},\ldots,\delta_{K}\right)=\left(1,0,\ldots,0\right)$, which gives the welfare-pessimal affiliation function $\hat{\rho}\left(l\right)=\mathbf{1}_{\left\{l\geq\frac{1}{2K}\right\}}$. Because this $\hat{\rho}$ can be approximated by a strictly increasing continuous affiliation function, we have $\sup_{\rho}L_{1M1V}^{repr}\left(\rho\right)=1/\left(4K\right)$, as stated in part 2a of the proposition.

Part 2b

The reported approximation of $L_{1M1V}^{repr}\left(\rho\right)$ follows by approximating ρ by a piece-wise linear function:

$$\begin{aligned}
L_{1M1V}^{repr}\left(\rho\right)&=\sum_{k\in\mathcal{K}}\int_{\frac{k-1}{K}}^{\frac{k}{K}}\left(\rho\left(l\right)-r\left(\frac{k-1}{K},\frac{k}{K}\right)\right)^{2}\mathrm{d}l\\
&\approx\sum_{k\in\mathcal{K}}\int_{\frac{k-1}{K}}^{\frac{k}{K}}\left(\rho\left(\frac{k-1}{K}\right)\left(k-Kl\right)+\rho\left(\frac{k}{K}\right)\left(Kl-k+1\right)-\frac{\rho\left(\frac{k-1}{K}\right)+\rho\left(\frac{k}{K}\right)}{2}\right)^{2}\mathrm{d}l\\
&=\frac{1}{12K}\sum_{k\in\mathcal{K}}\left(\rho\left(\frac{k}{K}\right)-\rho\left(\frac{k-1}{K}\right)\right)^{2}\\
&\approx\frac{1}{12K^{2}}\sum_{k\in\mathcal{K}}\rho'\left(\frac{k}{K}\right)^{2}\frac{1}{K}\\
&\approx\frac{1}{12K^{2}}\int_{0}^{1}\rho'\left(l\right)^{2}\mathrm{d}l\,.
\end{aligned}$$

The reported approximation of $L_{OPT}^{repr}\left(\rho\right)$ is a result in the signal processing subfield of engineering and is due to Panter and Dite (1951). In the signal processing problem that corresponds to our districting problem, a continuous signal must be discretized, or quantized, with minimal loss. In the notation of our model, the signal is the random variable $\rho(l)$, which is itself driven by the random variable l distributed uniformly on $\mathcal{L}$. The discrete representation of the signal is restricted to take the values that are the elements of the vector $\mathbf{r} \equiv (r_1, \ldots, r_K)$. The mapping from a signal realization to its discrete representation—the quantizer—is described by the function $g : \mathcal{L} \to \mathcal{K}$ and the associated vector $\mathbf{r}$. The Lloyd–Max quantizer (Lloyd, 1982; Max, 1960) is defined to minimize the mean square error of quantization in the class of quantizers that partition $\mathcal{L}$ into K intervals. The Lloyd–Max quantizer corresponds to the district map that minimizes the welfare loss from representation and to the legislature induced by this map because the search for the optimal map can also be restricted to the maps that partition $\mathcal{L}$ (by the "Auxiliary Observation for Parts 2a and 2b" above). Panter and Dite (1951) work out an approximate loss associated with the Lloyd–Max quantizer. This loss corresponds to the right-hand side of $L^{repr}_{OPT}(\rho) \approx \ldots$ in part 2b of the proposition.

A.2 Proof of Proposition 3

This proof sketch retraces the steps in the proof of Proposition 1 and follows the notation from that proof.

The Districter's Behavior

Under the proposition's assumptions on u, the function r in (13) is strictly increasing. Therefore, the partisan districter extremizes the policy $r(\mu_M)$ by extremizing μ_M, the mean ideology in the median district. This is exactly what the districter did when u was assumed to be quadratic. Moreover, one can show that the districter who lexicographically minimizes voter disutility from representation continues to equate the ideologies across all the $K - M$ districts that do not form the M-district cluster that determines the policy. Thus, the districter's behavior for the general u is the same as his behavior for the quadratic u.

Part 1

Under the economic view, the regulator varies an a in $\left(0, \frac{1}{2} + \frac{1}{2K}\right]$ to maximize the welfare $W^{plcy}\left(\rho, \mathbf{g}^{a,min}\right) \equiv -Ru(1-p) - (1-R)u(-p)$, where $p = r(\mu_M)$. The regulator's ideal policy would be $r(R)$ by the definition of r in (13). Short of that, the regulator is happier the closer the implemented policy is to $r(R)$ because $-Ru(1-p)-(1-R)u(-p)$ is single-peaked in p under the assumptions on u. Because 1M1V minimizes the districter's tendency to extremize the policy, the conclusion in part 1 of Proposition 1 follows.

Part 2

Under the political view, the regulator who faces the partisan districter with $b = min$ maximizes

$$\begin{aligned} W^{repr}\left(\rho, \mathbf{g}^{a,min}\right) \equiv & -a\left[\sigma(a)\, u\left(1-r\left(\sigma(a)\right)\right)+\left(1-\sigma(a)\right) u\left(-r\left(\sigma(a)\right)\right)\right] \\ & -(1-a)\left[\delta(a)\, u\left(1-r\left(\delta(a)\right)\right)+\left(1-\delta(a)\right) u\left(-r\left(\delta(a)\right)\right)\right]. \end{aligned}$$

The regulator who faces the partisan districter with $b = max$ maximizes $W^{repr}\left(\rho, \mathbf{g}^{a,max}\right) \equiv W^{repr}\left(\rho, \mathbf{g}^{1-a,min}\right)$. Condition 2 ensures that $W^{repr}\left(\rho, \mathbf{g}^{a,min}\right)$ and $W^{repr}\left(\rho, \mathbf{g}^{a,max}\right)$ each are single-peaked in a. The regulator who perceives ambiguity about both b and the resolution of districter's indifference chooses an a in $\left(0, \frac{1}{2}+\frac{1}{2K}\right]$ to maximize the minimal welfare

$$\min\left\{W^{repr}\left(\rho, \mathbf{g}^{a,min}\right), W^{repr}\left(\rho, \mathbf{g}^{a,max}\right)\right\}.$$

This minimal welfare—the lower envelope of two single-peaked functions, one a reflection of the other around $a=\frac{1}{2}$—is uniquely maximized at $a=\frac{1}{2}$. The conclusion in in part 2 of Proposition 1 follows.

Part 3

Consider the case of $b = min$. Then, for any a in $\left(0, \frac{1}{2}+\frac{1}{2K}\right]$, we have

$$\begin{aligned} & \frac{W^{repr}\left(\rho, \mathbf{g}^{a,min}\right)-W^{plcy}\left(\rho, \mathbf{g}^{a,min}\right)}{1-a} \\ & =\delta(a)\, u\left(1-r\left(\sigma(a)\right)\right)+\left(1-\delta(a)\right) u\left(-r\left(\sigma(a)\right)\right)-\delta(a)\, u\left(1-r\left(\delta(a)\right)\right)-\left(1-\delta(a)\right) u\left(-r\left(\delta(a)\right)\right) \\ & >\min_x\left\{\delta(a)\, u(1-x)+\left(1-\delta(a)\right) u(-x)\right\}-\delta(a)\, u\left(1-r\left(\delta(a)\right)\right)-\left(1-\delta(a)\right) u\left(-r\left(\delta(a)\right)\right)=0, \end{aligned}$$

where the first equation follows from the definitions of $W^{repr}\left(\rho,\mathbf{g}^{a,min}\right)$ and $W^{plcy}\left(\rho,\mathbf{g}^{a,min}\right)$ and from $R\equiv a\sigma\left(a\right)+\left(1-a\right)\delta\left(a\right)$; the strict inequality follows the definition of r in (13) and from $\sigma\left(a\right)\neq\delta\left(a\right)$; and the last equality follows by the definition of $r\left(\delta\left(a\right)\right)$. Conclude that $W^{repr}\left(\rho,\mathbf{g}^{a,min}\right)>W^{plcy}\left(\rho,\mathbf{g}^{a,min}\right)$. If $\hat{\mathbf{g}}^{a,min}$ is an arbitrary policy-minimizing map (not necessarily the one that minimizes voter welfare from representation), then $W^{repr}\left(\rho,\hat{\mathbf{g}}^{a,min}\right)\geq W^{repr}\left(\rho,\mathbf{g}^{a,min}\right)$ and $W^{plcy}\left(\rho,\hat{\mathbf{g}}^{a,min}\right)=W^{plcy}\left(\rho,\mathbf{g}^{a,min}\right)$, which combined with $W^{repr}\left(\rho,\mathbf{g}^{a,min}\right)>W^{plcy}\left(\rho,\mathbf{g}^{a,min}\right)$ implies that $W^{repr}\left(\rho,\hat{\mathbf{g}}^{a,min}\right)>W^{plcy}\left(\rho,\hat{\mathbf{g}}^{a,min}\right)$. Letting $\hat{\mathbf{g}}^{a,max}$ denote an arbitrary policy-maximizing map, one can similarly conclude that $W^{repr}\left(\rho,\hat{\mathbf{g}}^{a,max}\right)>W^{plcy}\left(\rho,\hat{\mathbf{g}}^{a,max}\right)$. As a result, for all ρ, b, and ways to resolve the districter's indifference, we have

$$\min\left\{W^{plcy}\left(\rho,\hat{\mathbf{g}}^{a,b}\right),W^{repr}\left(\rho,\hat{\mathbf{g}}^{a,b}\right)\right\}=W^{plcy}\left(\rho,\hat{\mathbf{g}}^{a,b}\right),$$

which by part 1 is uniquely maximized at $S=\left\{\left(\frac{1}{K},\ldots,\frac{1}{K}\right)\right\}$, and the conclusion in part 2 of Proposition 1 follows.

References

Ali, S. Nageeb, Nima Haghpanah, Xiao Lin, and Ron Siegel, "How to Sell Hard Information," *The Quarterly Journal of Economics*, 2022, *137* (1), 619–678.

Ansolabehere, Stephen and Samuel Issaacharoff, "Baker v. Carr in Context, 1946-1964," in Michael C. Dorf, ed., *Constitutional Cases in Context*, New York: Foundation Press, 2003.

Balinski, Michel L. and H. Peyton Young, *Fair Representation: Meeting the Ideal of One Man, One Vote*, Yale University Press, 1982.

Banzhaf III, J. F, "Weighted voting doesn't work: A mathematical analysis," *Rutgers Law Review*, 1965, *19.*

_ , "Multi-Member Electoral Districts. Do They Violate the "One Man, One Vote" Principle," *The Yale Law Journal*, 1309-1338 1966, (75), 8.

Barberà, Salvador and Matthew O. Jackson, "On the Weights of Nations: Assigning Voting Weights in a Heterogeneous Union," *Journal of Political Economy*, 2006, *114* (2), 317–339.

Barron, Jerome E., "The ambiguity of judicial review: A response to Professor Bickel," *Duke Law Journal*, 1970, *1970*, 591–604.

Black, Duncan, *The theory of committees and elections*, Cambridge University Press, 1958.

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.

Clinton, Joshua D., "Representation in Congress: Constituents and Roll Calls in the 106th House," *The Journal of Politics*, 2006, *68* (2), 397–409.

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

Dubey, Pradeep and Lloyd S. Shapley, "Mathematical Properties of the Banzhaf Power Index," *Mathematics of Operations Research*, 1979, *4* (2), 99–131.

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

Farnsworth, Ward, Dustin F. Guzior, and Anup Malani, "Ambiguity about ambiguity: An empirical enquiry into legal interpretation," *Journal of Legal Analysis*, 2010, *2* (1), 257–300.

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

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

Gilboa, Itzhak and David Schmeidler, "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics*, 1989, *18*, 141–153.

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

Gomberg, Andrei, Romans Pancs, and Tridib Sharma, "Electoral Maldistricting," 2022.

Grant, J. Tobin and Thomas J. Rudolph, “The Job of Representation in Congress: Public Expectations and Representative Approval,” *Legislative Studies Quarterly*, 2004, *29* (3), 431–445.

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

Hinckley, Barbara, “House Re-Elections and Senate Defeats: The Role of the Challenger,” *British Journal of Political Science*, 1980, *10* (4), 441–460.

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

Holmstrom, Bengt Robert, “On Incentives and Control in Organizations.” PhD dissertation, Stanford University 1977.

_ , “On the Theory of Delegation,” in M. Boyer and R. Kihlstrom, eds., *Bayesian Models in Economic Theory*, 1984.

Jacobsohn, Gary J., “The "Pragmatic Dogma" of the Political Thicket: The Jurisprudential Paradox of "One Man, One Vote",” *Polity*, 1977, *9* (3).

Kavanaugh, B. M., “Two challenges for the judge as umpire: statutory ambiguity and constitutional exceptions,” *Notre Dame L. Rev.*, 2016, *92* (5), 1907–1919.

Khan, M. Ali, Kali P. Rath, and Yeneng Sun, “The Dvoretzky-Wald-Wolfowitz theorem and purification in atomless finite-action gaes,” *International Journal of Game Theory*, 2006, *34*, 91–104.

Koriyama, Yukio, Antonin Macé, Rafael Treibich, and Jean-François Laslier, “Optimal Apportionment,” *Journal of Political Economy*, 2013, *121* (3), 584–608.

Lindbeck, Assar and Jorgen W. Weibull, “Balanced-budget redistribution as the outcome of political competition,” *Public Choice volume*, 1987, *52*, 273–297.

Lloyd, Stuart P., “Least Squares Quantization in PCM,” *IEEE Transactions on Informatin Theory*, 1982, *28* (2), 129–137.

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

Max, J., "Quantizing for minimum distortion," *IEEE Transactions on Acoustics, Speech, and Signal Procesing*, 1960, *6* (1), 7–12.

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.

Panter, P. F. and W. Dite, "Quantization Distortion in Pulse-Count Modulation with Nonuniform Spacing of Levels," *Proceedings of the IRE*, 1951, *39* (1), 44–48.

Penrose, L. S., "The Elementary Statistics of Majority Voting," *Journal of the Royal Statistical Society*, 1946, *109* (1), 53–57.

Santo, Alessia De and Benoît Le Maux, "On the optimal size of legislatures: An illustrated literature review," *European Journal of Political Economy*, 2022, (in press).

Schofield, Norman and Itai Sened, "Modeling the interaction of parties, activists and voters: Why is the political center so empty?," *European Journal of Political Research*, 2005, *44* (3), 355–390.

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

Stigler, George J., "The Sizes of Legislatures," *The Journal of Legal Studies*, 1976, *5* (1), 17–34.

Viscusi, W. Kip, "How Do Judges Think about Risk?," *American Law and Economics Review*, 1999, *1* (1/2), 26–62.