Competitive Fair Redistricting*

Felix J. Bierbrauer† Mattias Polborn‡

December 29, 2022

Abstract

We study political redistricting in a plurality rule electoral system, and ask whether there is a way of structuring this process so that a party that wins the popular vote is guaranteed a majority in the legislature. We present a formal analysis of this problem that departs from the literature on partisan gerrymandering and considers instead a system of competitive gerrymandering, i.e. a process of redistricting that involves both parties. We invoke the theory of zero sum games to show that it is possible to specify the rules of this process in such a way that "majorities cannot be stolen."

Keywords: Gerrymandering, legislative elections, redistricting.
JEL classification: D72, C72.

*We benefited from conversations with Peter Cramton, Matthew Knowles, Axel Ockenfels, Tigran Polborn, and Ashutosh Thakur. Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC2126/1 – 390838866 is gratefully acknowledged.

†University of Cologne, E-mail: bierbrauer@wiso.uni-koeln.de

‡Vanderbilt University and University of Cologne, E-mail: mattias.polborn@vanderbilt.edu

1 Introduction

The two most important electoral systems for legislative elections are the single-member plurality system in use in Britain and many of its colonies such as the United States, and the proportional representation system that is used in many European countries.

In proportional representation systems, there are usually many parties that are represented in the legislature in proportion to the votes that they receive in the election. On the plus side, this implies that any majority in the legislature represents a majority of voters.[1] On the other hand, coalition formation can sometimes be very challenging in proportional representation systems. This difficulty, as well as the prevalence of powerful special interest parties under proportional representation, may be a reason for why proportional representation is empirically correlated with higher levels of spending and taxation, or more corruption than plurality systems; see Persson et al. (2000); Persson and Tabellini (2002).

In contrast, a stylized fact known as "Duverger's Law" is that plurality rule systems generally lead to a two-party system, such as the one featuring Democrats and Republicans in the United States. In such a system, one of the parties necessarily wins a legislative majority (assuming the number of seats in a legislature is odd), so there is always a clear election outcome. This makes it easier for voters to know whom to blame when there are problems. However, an unfavorable feature of plurality rule systems is that the winner of the legislative majority does not just depend on which party is preferred by the majority of the overall electorate, but also on how these votes are distributed over the different districts. An example are the 2018 elections of the Pennsylvania House of Representatives. While Democratic candidates received 55 percent of the popular vote, versus 44.4% for Republican candidates, Republicans still won 110 out of 203 seats.

Such cases of a divergence between the popular vote and the majority outcome in the legislature do not arise randomly. They are the effects of skillful redistricting, usually done by the party that benefits from this "gerrymandering." For example, the district map that saved the Republican majority in the 2018 Pennsylvania election was created by them in 2011 when they were in control of

[1]In practical implementations of the proportional representation system, many countries use a minimum vote threshold for parties to be represented. In this case, any legislative majority represents a majority of those voters that voted for one of the parties represented in the legislature.

the redistricting process. In the United States, legislative districts are redrawn after each decennial census in order to ensure that each legislator represents the same number of residents. The task of redistricting falls usually to the current state legislature, a body composed of individuals who have a high degree of self-interest in the outcome of the redistricting process. Partisan gerrymandering thus undermines the legitimacy of election outcomes under plurality rule elections.

In this paper, we therefore ask whether there exists a redistricting system such that, in subsequent elections, the popular vote and the election outcome are aligned. In trying to find a better system, we impose that it cannot involve outsourcing decisions to a "benevolent social planner" who is only concerned with fairness; rather, the redistricting process is to be carried out by the two major parties themselves. This requirement is one of practicality. In a highly partisan world, it appears implausible that the parties could find such a highly competent and, at the same time, completely disinterested individual to perform the redistricting. Even when redistricting is in the hands of a notionally independent commission, its members are likely to have preferences over which party wins a majority under the maps they create.

Instead, we use a competitive – as opposed to partisan – redistricting system, i.e., one that involves both parties. The basic idea is to design an institution in which the parties keep each other in check. As a main result, we show that such a system can protect parties against stolen majorities, and the majority of the electorate against having their preferences in future elections subverted. Broadly, the logic is familiar from the classical problem of how to fairly divide a cake between two children – one child cuts the cake in two pieces and the other one chooses which one she wants to have. Every child has a strategy that ensures getting at least fifty percent of the cake, and this procedure is arguably preferable to an alternative one that attempts to design general rules and constraints under which only one child chooses both the own and the other child's piece.

We consider the standard setting of the theoretical literature on partisan gerrymandering. Voters differ in how likely they are to vote for either party, and need to be assigned to districts.[2] We define a redistricting system as fair if each party can ensure that it wins a majority in the legislature whenever it wins the popular

[2]The central theme of this literature is how an optimal partisan gerrymander involves "packing" (i.e., concentrating opponents in few districts) and "cracking" (distribute one's supporters evenly over the remaining majority of districts).

vote. We specify rules for a competitive gerrymandering game, and then prove that each party has a simple strategy with which it can ensure that it will get a legislative majority whenever it wins the popular vote in a future election. We refer to this strategy as a *pecking order strategy.* It is based on an order of districts according to how likely they can be won, and assigns priority to the weakest district among those that are needed for a legislative majority. Favorable precincts are assigned to that district until the chances of winning it have risen to the level of the next district in the order. From that point on, the priority becomes to lift the chances in these two districts simultaneously until they are both as good as the next district in the order, and so on. The pecking order strategy is not only simple, but is also robust in the sense that it offers protection against stolen majorities even when the other party does not play its pecking order strategy.

Related Literature. There is a large literature on gerrymandering, both empirical and theoretical. However, most of the existing theoretical literature is on "optimal" gerrymandering from the point of view of the party that controls the gerrymandering process; that is, how to cheat democracy most effectively if given the opportunity to do so. Few papers deal with the question of how one could implement a better redistricting system. The earliest such paper is William Vickrey's (1961) paper arguing that "the process [of redistricting] should be completely mechanical so that, once set up, there is no room at all for human choice."[3] Similarly, Ely (2019) proposes a mechanism designed to prevent weirdly-shaped districts. Like our paper, his mechanism relies on the participation of both parties in the redistricting process, and he also appeals to the cake-division problem. There are also important differences: Ely takes convexity as the key desideratum. Our analysis, by contrast, focuses on the alignment of election outcomes with the popular vote, and it abstracts from spatial considerations.

Fundamentally, our paper contributes to a small theoretical literature on how to improve political systems (Gersbach, 2004; Myerson, 2006; Gersbach and Liessem, 2008, e.g.). Like these papers, our objective is to think about possible changes to democratic institutions that improve the equilibrium performance of the system for the average voter.

Our paper is related to the theory of mechanism design and to implementation

[3] He proposes an algorithm that produces geographically-compact districts, but does not study whether elections governed by the generated map have any desirable properties.

theory, which is applied to numerous problems, ranging from auction design over redistributive income taxation to the design of social choice rules. In all these applications, the basic question is whether one can find a game that implements a desirable outcome. At an abstract level, we ask the same question in this paper. Applications of mechanism design and implementation theory differ, however, in what games they look at and in how they define a desirable outcome. In both dimensions, this paper takes an approach that is without precedent in the previous literature: First, our system of competitive redistricting can be interpreted as a dynamic Colonel Blotto game (for applications of static divide-the-dollar or Colonel Blotto games, see, for instance, Myerson (1993), Lizzeri and Persico (2001, 2005), Laslier and Picard (2002), Konrad (2009) and Kovenock and Roberson (2020)). To the best of our knowledge, using a dynamic version of this class of games is novel in the literature on mechanism design and implementation theory.[4] Second, applications of mechanism design in economics often aim at the maximization of economic surplus, social welfare or profits. Our approach, by contrast, takes political legitimacy to be the objective. Formalizing this objective may be difficult in general, but for election rules there is a natural choice: Political legitimacy requires that the party that wins the popular vote gains control over policy.

The proof of our main result uses results from the analysis of zero-sum games. More specifically, we define a fictitious zero-sum game in which one of the parties gets a payoff of 1 when it has enough supporters in half of the districts in the critical state of the world, with the implication that it wins a majority of seats whenever it wins the popular vote. Otherwise the payoff is zero. We then show that the equilibrium payoff for this party is one. By the min-max-theorem due to von Neumann (1928),[5] this implies that the party has a successful strategy – in the sense of winning a majority of districts, conditional on winning the popular vote – *for every strategy of the opposing party.*[6]

[4]Groseclose and Snyder (1996) study coalition formation within a legislature on the assumption that there are two competing vote-buyers. While they also look at a sequential mechanism, their focus is positive rather than normative in that they seek an explanation for the frequent occurrence of supermajorities – as opposed to minimal winning coalitions.

[5]See Osborne and Rubinstein (1994) for a textbook treatment.

[6]Our results also mirror a well-known Theorem by Zermelo (1913) on the game of chess. According to Zermelo's theorem, either *White* has a strategy that guarantees a victory, or *Black* has a strategy that guarantees a victory, or both have a strategy that guarantees a draw. While Zermelo, of course, cannot characterize these strategies for chess, we do not just show that there exist strategies that guarantee winning the election (conditional on winning the popular vote),

As is standard in the theoretical literature pioneered by Owen and Grofman (1988), our formal framework is geography-free, so that parties do not face geographic constraints. Leaving out spatial considerations provides conceptual clarity and keeps the paper directly comparable to the existing theoretical literature.[7]

Furthermore, geographical constraints, such as the requirement that all districts must be contiguous, are arguably best interpreted as second-best constraints in traditional redistricting: A partisan gerrymanderer free of any geographic restrictions would be able to subvert the will of the electorate to an outrageous degree.[8] Thus, geographic constraints prevent the worst excesses under partisan gerrymandering. However, since our proposed redistricting systems guarantees fair outcomes, the second-best justification for geographic constraints is less compelling. Furthermore, while geographic proximity may create one possible *"community of interest,"* there are certainly also other, and oftentimes more compelling, criteria that define other communities of interest. A system that does not require legislative districts to be contiguous has the advantage that it becomes much easier to bundle such communities.[9]

Outline. Section 2 presents a stylized example and discusses our main results in this context. Rigorous game-theoretic analyses of competitive redistricting can be found in Sections 3 and 4. In Section 3, we show that a simple game of competitive redistricting offers protection against stolen majorities – if the order of moves can be tailored to party characteristics. In Section 4, we show that a many-rounds version of the simple game works irrespectively of how the order of moves is specified. Remarks on the relevance of geographic constraints for gerrymandering are in Section 5. The last section contains concluding remarks. Formal proofs are relegated to Appendix A. Appendix B presents an extension of our analysis to a setting with an explicit geography.

---

but we also describe them.

[7]That said, in part B of the Online-Appendix we show how our analysis can be extended to a setting with an explicit geography.

[8]Indeed, a party needs only slightly more than 1/4 of the overall votes to secure a legislative majority if the votes are allocated to districts "optimally" (from the party's point of view).To achieve such an outcome, allocate the party's voters such that they constitute a bare majority in a bare majority of districts, while the remaining districts vote unanimously for the opposition.

[9]For example, generating a *contiguous* majority-Muslim district in Germany, or a majority-Asian district in most U.S. states would be extremely difficult, even though these communities constitute a significant minority in many places.

2 A simple example

Partisan gerrymandering. Consider a polity that consists of a large number of *precincts* (i.e., indivisible geographic units, several of which make up a legislative district). We refer to the two parties as *Republicans* and *Democrats*, though these are purely labels. There are two types of precincts: In *Republican-leaning* precincts (which constitute one-half of all precincts), the Republican vote share is $0.6+0.1\omega$, while in *Democratic-leaning* precincts (the other half), the Republican vote share is $0.3 + 0.1\omega$. Thus, the margins of victory fluctuate in both types of precincts, e.g. because of changes in the popularity of political leaders, and a higher state of the world $\omega \in [0, 1]$ captures times that are more favorable to Republicans.

Partisan redistricting can be very effective in this setting. Suppose, for example, that Republicans control the redistricting process. Observe that a district is guaranteed to be won by the Republican candidate if the share of Republican-leaning precincts is at least 2/3 because then, even in the worst case, the Republican vote share is $(2/3) \times 60\% + (1/3) \times 30\% = 50\%$. Also note that Republicans can make sure that 75 percent of all districts have a share of Republican-leaning precincts equal to 2/3. Thus, they can make sure that they win (at least) 3/4 of all districts in all states of the world. See Figure 1 for an illustration.

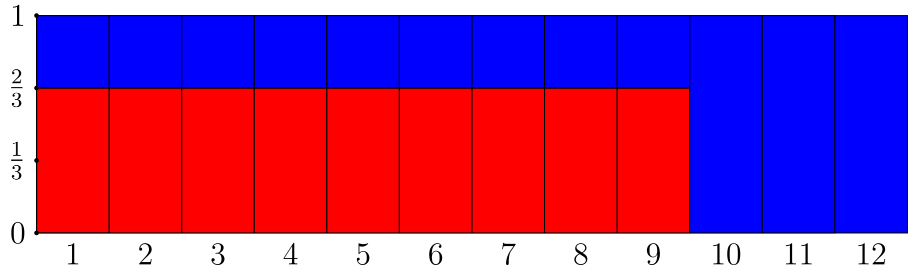


Figure 1: 12 districts. For any district, Republican-leaning precincts in red and Democratic-leaning precincts in blue. The total numbers of blue and red precincts are equal. Partisan gerrymandering enables the Republicans here to create 9 out of 12 districts that have a 2/3 share of Republican-leaning precincts and are won in every state of the world.

Competitive gerrymandering. To see how our proposed system works, consider the same example polity, and denote the players D and R. The task is to define $2N$ equal-sized legislative districts. In addition, there is one at-large district that also sends one representative and ensures an odd number of representatives in the legislature.

At the beginning, each party receives a budget set consisting of half of the precincts of each type. D starts and assign each precinct to a district (such that each district consists of the same number of precincts). After D is done, it's R's turn to assign precincts to districts.

It is straightforward to show that each party has a strategy that can guarantee itself a majority in the legislature whenever it wins the popular vote (i.e., D if $\omega < 0.5$, and R when $\omega > 0.5$).

In any district, R, the second mover, can just "mirror" D's move. For example, if D assigned 60 percent Democratic-leaning and 40 percent Republican-leaning precincts to district k, R can produce a perfectly balanced district by assigning 60 percent Republican-leaning and 40 percent Democratic-leaning precincts. Figure 2 illustrates this balancing strategy.

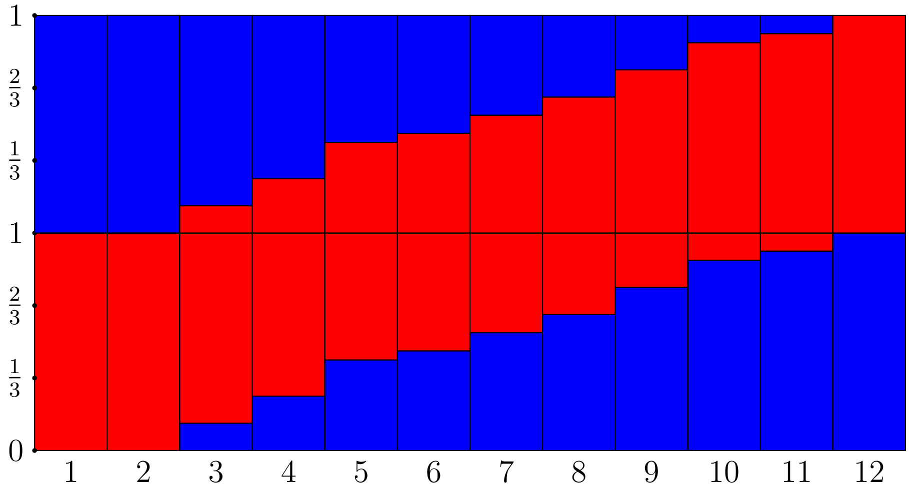


Figure 2: 12 districts. For any district, Republican-leaning precincts in red and Democratic-leaning precincts in blue. The total numbers of blue and red precincts are equal.
The bottom part shows a generic precinct assignment by Democrats, with districts ordered according to their share of Democrat-leaning precincts. The top part shows a feasible Republican balancing strategy that neutralizes, for any district, deviations from the aggregate popular vote.

It is clearly feasible for R to play this balancing strategy for each district, which results in each district going to the winner of the popular vote. Observe, though, that the balancing strategy is not necessarily optimal for R. This depends on how D distributed the precincts, and on R's objective. Thus, a full characterization of best responses or of the subgame-perfect equilibrium would be more cumbersome.

Consider now D, the first mover. Suppose that D assigns only Democratic-leaning precincts to the first N districts, and only Republican-leaning precincts to districts $N+1$ to $2N$. Clearly, this is feasible as it uses up all precincts. Furthermore, no matter what R does in its move, the first N districts will have at least a 50 percent share of Democratic-leaning precincts, so will be won by D whenever $\omega < 0.5$. Since D also wins the at-large district whenever $\omega < 0.5$, D is guaranteed a majority in the legislature whenever $\omega < 0.5$.

An asymmetric setting. In the following analysis, we will show how to generalize this example to the case that the number of Democratic- and Republican-leaning precincts is not the same, and that the average partisan lean of these two types of districts is not the same. This is a relevant generalization because, throughout the United States, Democrats are often very strongly concentrated in urban areas, and Chen and Rodden (2013) suggest that this geographic fact alone provides a significant advantage for Republicans in a traditional redistricting process. We will show that we can maintain a fair system in that setting.

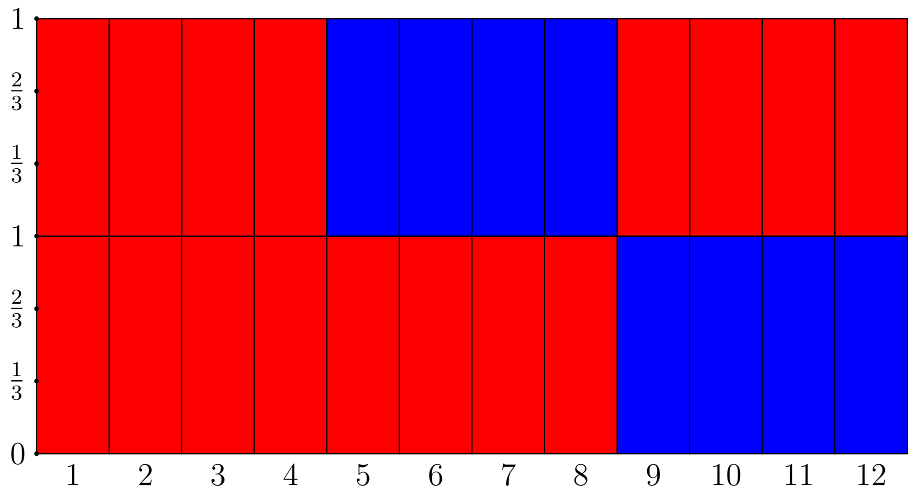


Figure 3: 12 districts. Overall, there are twice as many Republican-leaning precincts (red) as Democratic-leaning precincts (blue).
Bottom part: precinct assignment by R such that 8 districts only contain red precincts, and 4 districts only contain blue precincts.
Top part: Feasible response by D that wins a majority of districts in some states in which the Republicans win the popular vote.

To understand why matters become more involved, consider an example in which 1/3 of precincts are Democratic-leaning, with Republican vote share $0.2 + 0.2\omega$, and 2/3 of precincts are Republican-leaning, with Republican vote share $0.5+0.2\omega$. Compared to before, there are now fewer Democratic-leaning precincts, but in those precincts the Democrat's margin of victory is higher. Again, Democrats win the popular vote if and only if $\omega < 0.5$.[10] As we now argue, the sequence of moves matters in such an asymmetric setting. When D moves first, it is still the case that every party can ensure to win a majority of districts whenever it wins the popular vote. This is not the case when R moves first.

Again, the parties have equivalent budget sets; that is, both players have to assign a set of precincts that has a 2/3 share of Republican-leaning precincts and a 1/3 share of Republican-leaning precincts. Suppose that D moves first. Like in the previous example, D can ensure a win whenever $\omega < 0.5$: By assigning a percentage share of 2/3 of Democratic-leaning precincts to half of the districts, it can guarantee that these districts are won (at least) whenever $\omega < 0.5$, even if R were to add only Republican-leaning precincts to these districts. R can also ensure to win a majority of districts whenever they win the popular vote, i.e. when $\omega > 0.5$. Whatever D does in the first move, half of the districts will have been assigned a share of Democratic-leaning precincts that is below 2/3. If R assigns only Republican-leaning precincts to those districts, it will win those districts whenever $\omega > 0.5$.

In contrast, what would happen if R moves first? As the share of Republican-leaning precincts is greater than one-half, R cannot block all of them together in one-half of the districts. Thus, R cannot play the type of move that is analogous to the one suggested above for D. Blocking Republican-leaning precincts in 2/3 of districts is feasible, but this strategy does not ensure a legislative majority whenever $\omega < 0.5$. To see this, suppose that R creates 2/3 of districts that are composed only of Republican-leaning precincts, and 1/3 of districts that are exclusively Democrat-leaning. Then, D can add only Republican-leaning precincts to the latter, and block their Democrat-leaning precincts in another third of districts, while the remaining third is composed only of Republican-leaning precincts. See Figure 3. Thus, in the two-thirds of districts that consist of an equal share of Democratic and Republican-leaning precincts, the Republican vote share is

$$\frac{1}{2}[0.2 + 0.2\omega] + \frac{1}{2}[0.5 + 0.2\omega] = 0.35 + 0.2\omega,$$

which exceeds 0.5 only if $\omega > 3/4$. Thus, if $\omega \in [0.5, 0.75)$, Democrats win the majority while losing the popular vote. As we show in Section 4, the disadvantage for R can be overcome when voters are assigned over multiple rounds.

[10] Furthermore, observe that we are not imposing any specific probability on the event that $\omega < 0.5$ — this probability can be arbitrary, it does not have to equal 1/2.

Discussion: Which party is the disadvantaged party? The previous discussion may suggest that R is in a weak position. It needs to be given the second mover advantage, otherwise it cannot protect itself against the possibility of a stolen majority. In that case, however, D is put in a disadvantaged position. It has to block its Democratic-leaning precincts in half of the districts, otherwise its majority can be stolen. (Propositions X and Y below contain formal statements of this claim.) But then R can take advantage of this, for instance, by achieving an overall outcome so that

- A quarter of all districts has a share of Democratic-leaning precincts equal to 5/6.
- A quarter of all districts has a share of Democratic-leaning precincts equal to 1/3.
- Half of the districts has a share of Democratic-leaning precincts equal to 1/6.

Consequently, when $\omega > 0.5$, the Republicans win 3/4 of all districts. By contrast, when $\omega < 0.5$, the Democrats win only 1/2 of all districts, and the at-large district is then needed as a tie-breaker. As we show below, this disadvantage for D is also overcome when voters are assigned over multiple rounds.

Discussion: How to make sure that the parties have equivalent budgets? Our analysis rests on the assumption that precincts can be allocated to budget sets for the two parties so that both sets have equal shares of Republican-leaning and Democratic-leaning precincts. This requires a mechanism to determine which precinct is going to which party's budget set. In the context of the model, this is easy. There are only two types of precincts and each party should simply get half of the districts of either type. In practice, it may be more difficult to find an exact doppelganger for each and every district. Still, if the overall number of precincts is large, then a mechanism that assigns precincts at random to the two budget sets would produce two budget sets that are close to equivalent with a very high probability (by the central limit theorem).

An alternative mechanism that makes sure that the parties end up with exactly equivalent budget sets is the following: assign each precinct to both parties, and let each party assign any particular precinct to a district. Voters in that precinct

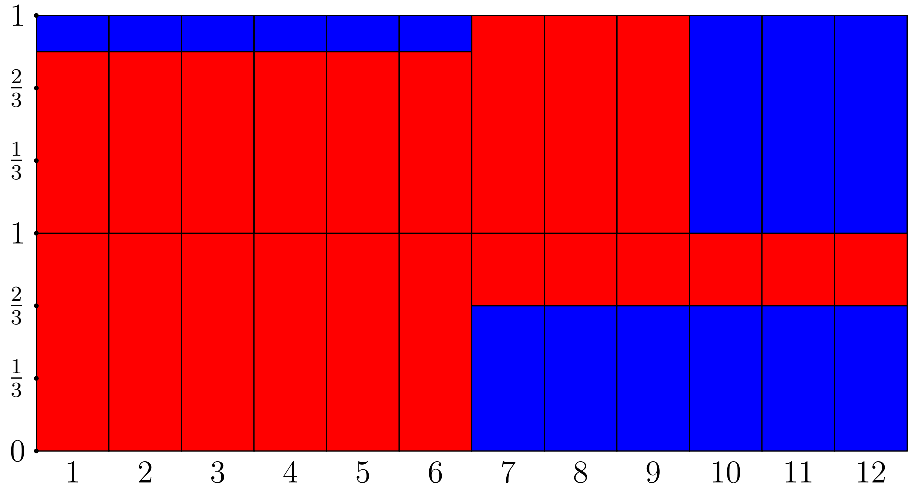


Figure 4: 12 districts. For any district, Republican-leaning precincts in red and Democratic-leaning precincts in blue. Overall, there are twice as many red than blue precincts.
Bottom part: Precinct assignment by D that guarantees a D legislative majority whenever they win the popular vote: 6 districts are assigned a share of 2/3 blue precincts, and are won by D whenever Ds win the popular vote, no matter what R moves.
Top part: Feasible R response such that Rs always win 6 districts, and win another 3 districts whenever R has a majority of the popular vote.

can then vote in both of the districts that they have been assigned to.[11] As a consequence, voters have two representatives. This is some departure from the existing system in the United States, but hardly a radical one. Indeed, most voters already have two legislative state representatives, one in the state house and the other one in the state senate. Likewise, at-large representatives, in addition to district representatives, exist in many cities. Finally, while this system would increase the number of elections each citizen votes in, practically speaking, the required increase in ballot length relative to the status quo would be quite small.

3 Formal Analysis

This section contains propositions that complement, in a more general setting, the informal discussion in the preceding section. Specifically, we do not impose the assumption that the parties' vote shares depend linearly on the state of the world, but just impose that vote shares are monotonous is ω. Furthermore, we assume that the parties differ in how concentrated their support is, referring to the one

[11]If both parties happen to assign a precinct to the same district, the votes of these voters would simply count twice in that district's election.

with more concentrated support as *Democrats* (labeled D), and the other party as *Republicans* (labeled R).

We first consider a protocol with one round in which D moves first and R second. Theorem 1 then establishes that the more popular party (whichever it is) has a strategy that protects its majority from being stolen. Proposition 1, moreover, shows that there is one and only one such strategy for D.

Subsequently, we turn to an alternative protocol with many rounds. We show that Theorem 1 and Proposition 1 extend to this setting. In addition, there is now also a version of Proposition 1 that applies to R. Thus, the many rounds protocol is fair in that both parties have essentially only one strategy that protects them against stolen majorities. Moreover, when both parties play those strategies, almost all districts have the same shares of Democrat- and Republican-leaning districts as the electorate at large. Thus, races at the district level are as competitive as the race for the popular vote.

3.1 Setup

There are $2N$ local districts, indexed by $k \in \{1, 2, \ldots, 2N\}$, and one at-large district. There are two types of "voters," $t \in \{t_1, t_2\}$, that we interpret either as individuals, or as the smallest unit that can be assigned to a district, such as a precinct. The mass of type t_j voters is given by

$$b_j = 2N\, \beta_j\ , \quad \text{where} \quad \beta_1 + \beta_2 = 1 \quad \text{and} \quad \beta_1 \leq \frac{1}{2}\ .$$

The state of the world $\omega \in \Omega \subset \mathbb{R}$ is the realization of a real-valued random variable and affects $v(t, \omega)$, the probability that a type t unit votes for R in state ω. The function v is strictly increasing in both arguments; i.e., in any given state ω, type 2 is more likely to vote R than type 1, and higher ω increases the share of R voters among both types. We adopt a law of large numbers convention and also interpret $v(t, \omega)$ as the share of type t voters voting for R in state ω.

The popular vote. Let $\hat{\omega} \in \Omega$ denote the state that yields a tie in the popular vote, i.e.,[12]

$$\beta_1\, v(t_1, \hat{\omega}) + \beta_2\, v(t_2, \hat{\omega}) = \frac{1}{2}\ . \tag{1}$$

[12]To assume the existence of $\hat{\omega}$ is without loss of generality because it may have zero probability.

R wins the popular vote if $\omega > \hat{\omega}$, while D wins the popular vote if $\omega < \hat{\omega}$. Conditional on state $\hat{\omega}$, type 1 voters are more likely to vote for D and type 2 voters are more likely to vote for R,

$$v(t_1, \hat{\omega}) \quad < \quad \frac{1}{2} \quad < \quad v(t_2, \hat{\omega}) \ .$$

We also assume that type 1 voters are weakly more partisan than type 2 voters in the sense that, in the critical state $\hat{\omega}$, type 1 votes D with a probability that is at least as high as the probability that type 2 votes R,

$$1 - v(t_1, \hat{\omega}) \quad \geq \quad v(t_2, \hat{\omega}) \ .$$

Interpretation. One special case of this setup has $v(t_1, \hat{\omega}) = 0$ and $v(t_2, \hat{\omega}) = 1$ and $\beta_1 = \beta_2$. In this case a "voter" is really an individual whose vote, conditional on the state, the parties can perfectly predict. In state $\hat{\omega}$, type 1 (2) votes for D (R). For states $\omega > \hat{\omega}$, some type 1 voters – formally, a fraction that is increasing in ω – vote R. Likewise, for $\omega < \hat{\omega}$, some type 2 voters vote D.

By contrast, when $v(t_1, \hat{\omega}) \in \left(0, \frac{1}{2}\right)$ or $v(t_2, \hat{\omega}) \in \left(\frac{1}{2}, 1\right)$ a "voter" can be interpreted as a precinct or a census block that needs to be treated as an indivisible unit for the purposes of redistricting. Any such unit of type t_j contains a fraction $v(t_j, \hat{\omega})$ of individuals who vote R, and a fraction $1 - v(t_j, \hat{\omega})$ of individuals who vote D. As ω increases above $\hat{\omega}$, R's vote share increases in both types of blocks, and vice versa.

When $1 - v(t_1, \hat{\omega}) \quad > \quad v(t_2, \hat{\omega})$, i.e., type t_1 is strictly more partisan than type 2, then $\beta_2 > \frac{1}{2}$. Hence, while there are equal numbers of D and R voters at the aggregate level in state $\hat{\omega}$, D voters are more concentrated: Fewer units mostly vote for D, $\beta_1 < \beta_2$, but in those units, D's vote share is higher than R's vote share in the R-leaning units.

District outcomes. After precinct assignments are done, every district k contains some mix of type t_1 and type t_2 voters. More formally, a voter assignment by party $P \in \{D, R\}$ is a collection $\sigma_P = (\sigma_{Pk})_{k=1}^{2N}$, where

$$\sigma_{Pk} = (\sigma_{Pk}^1, \sigma_{Pk}^2) \quad \text{with} \quad \sigma_{Pk}^1 + \sigma_{Pk}^2 = 1 \ ,$$

is the voter assignment to district k by party P. To be consistent with the overall distribution of voters, across districts we must have

$$\frac{1}{2N} \sum_{k=1}^{2N} \sigma_{kP}^1 = \beta_1 \quad \text{and} \quad \frac{1}{2N} \sum_{k=1}^{2N} \sigma_{kP}^2 = \beta_2 \ .$$

R wins district k in state ω if

$$(\sigma^1_{Dk} + \sigma^1_{Rk})\ v(t_1, \omega) + (\sigma^2_{Dk} + \sigma^2_{Rk})\ v(t_2, \omega) \quad > \quad \frac{1}{2}\ . \tag{2}$$

D wins if the reverse inequality holds.

Winning a majority of seats. Recall that there are $2N$ districts and an at-large-district. Thus, the party that wins at least $N+1$ seats wins a majority in the legislature. Given a pair of voter assignments (σ_D, σ_R), we denote the probability that R wins a majority of seats, conditional on it winning the popular vote, by $\Pi_R(\sigma_D, \sigma_R \mid \omega > \hat{\omega})$. We define $\Pi_D(\sigma_D, \sigma_R \mid \omega < \hat{\omega})$ analogously.

3.2 Achieving fair outcomes

D, the party with the more concentrated support, moves first and chooses $\sigma_D = (\sigma_{Dk})_{k=1}^{2N}$. R observes this choice and chooses $\sigma_R = (\sigma_{Rk})_{k=1}^{2N}$. Theorem 1 shows that, each party has a strategy that guarantees winning a legislative majority whenever it wins the popular vote, no matter what the opponent does.

Theorem 1

1. *Party D has a strategy that guarantees that it wins a legislative majority whenever $\omega < \hat{\omega}$: There is σ_D so that $\Pi_D(\sigma_D, \sigma_R \mid \omega < \hat{\omega}) = 1$, for all σ_R.*

2. *Party R has a strategy that guarantees that it wins a legislative majority whenever $\omega > \hat{\omega}$: For every σ_D, there is σ_R so that $\Pi_R(\sigma_D, \sigma_R \mid \omega > \hat{\omega}) = 1$.*

The Theorem shows that the given protocol for competitive redistricting is fair in the sense that both parties can protect their majorities.[13]

Theorem 1 does not contain the characterization of a game-theoretic equilibrium. A proper game-theoretic analysis requires a specification of party objectives and so far we have remained agnostic about what the parties actually want. This said, the strategies in Theorem 1 are equilibrium strategies when all that the parties care about is whether they win a legislative majority; more formally a game

[13]As discussed in the introduction, in reality, parties do not necessarily play to minimize the probability of a stolen majority. They may also have other objectives in redistricting, for example incumbent protection or the representation of party wings or ethnic groups. Whether the maximal protection of its majorities actually is in a party's interest is an issue that the party has to deal with internally.

in which either party's payoff is 1 when it has a legislative majority and zero otherwise.

The proof of part 1 of Theorem 1 is along the lines of the example in the previous section: D can block all its supportive precincts into one half of all districts. After D's move, these districts all have double the percentage of Democratic-leaning precincts as the state at-large. Even if R puts only Republican-leaning precincts in those districts, this only dilutes the percentage of Democratic-leaning precincts down to the state-wide average. Thus, if $\omega < \omega^*$, D does not just win the at-large district, but also these N districts. Proposition 1 below shows, moreover, that this strategy for D is unique; i.e. there is no other strategy with the property $\Pi_D(\sigma_D, \sigma_R \mid \omega < \hat{\omega}) = 1$, for all σ_R.

The proof of part 2 of Theorem 1 relies on the fact that R can react to whatever D did. Essentially, R can pick off those N districts in which D put the fewest Democratic-leaning precincts; specifically, after D's move, all of these precincts contain a share of Democratic-leaning precincts that is no more than $2\beta_1$. But any district with such a share can be diluted down to β_1 (by not allocating further Democratic-leaning precincts to it).[14]

Proposition 1 *Suppose that $\beta_1 < 1/2$. Then, up to a relabeling of districts, there is one and only one strategy σ_D so that $\Pi_D(\sigma_D, \sigma_R \mid \omega < \hat{\omega}) = 1$, for all σ_R: choose $\sigma^1_{Dk} = 0$ for all $k \in \{1, \ldots, N\}$ and $\sigma^1_{Dk} = 2\beta_1$ for all $k \in \{N+1, \ldots, 2N\}$.*

When D blocks all its strongholds in half of the districts, none of them is wasted on a district that D doesn't win in state ω^*. There are two, not mutually exclusive ways in which D could deviate from this strategy, but Proposition 1 shows that both are bad for D. First, they could have a non-even distribution of their strongholds in half of the districts. Second, they could allocate their strongholds over more than half of the districts.

In the first case, the least-Democratic district in the targeted half has a lower content of Democratic-leaning precincts than $2\beta_1$ after the Democrats' move. R then can add some more Democratic strongholds to some of the most Democratic districts (essentially giving up on them), and can then spread the remaining Democratic strongholds uniformly on its own half. This strategy guarantees

[14]Observe that it is feasible for R to get rid of its Democratic-leaning precincts by allocating them entirely to the top-half of districts (i.e., those that were endowed with the largest percentage of Democratic-leaning precincts by D).

that the Democratic content of all districts in the Republican half, as well as in the least-Democratic district in the Democrats' targeted half, has fewer than β_1 Democratic-leaning precincts.

Second, if D uses some of its strongholds in the other half of districts, the budget constraint forces D to allocate fewer than $2\beta_1$ Democratic strongholds to the least-Democratic district in its half. R has sufficient flexibility in terms of allocating its Democratic strongholds to make that district Republican-leaning in state ω^*, while also maintaining an advantage in their half of districts. Again, the key to this is that R can give up on some districts and fill them to the brim with Democratic strongholds, thereby easing its problems in all other districts.

With one round, there is no analogue to Proposition 1 for R. There are numerous ways in which R can ensure to win a majority of districts whenever $\omega > \omega^*$. For one, R can balance every district; that is, they distribute the Democratic-leaning precincts uniformly over the other half of districts, while allocating only Republican-leaning districts in those districts targeted by D. This creates $2N$ districts that all look exactly like a replica of the electorate at large.

On the other hand, R can also double down and allocate all Democratic-leaning precincts to the same districts as D. This amounts to creating N districts that are essentially secure for D, and N that are essentially secure for R, with the at-large district being decisive for which party wins a legislative majority.

Since there are multiple ways in which R can maximize its probability of controlling the legislature, it is free to choose an option that maximizes any secondary objective, such as maximizing its expected number of seats (subject to achieving the maximal winning probability). This generates an asymmetry between parties with respect to these secondary objectives; however, Section 4 shows that this asymmetry is eliminated when voters are assigned to districts over multiple rounds.

3.3 Neutral rules

Another reason why it is useful to look at multiple rounds is that the rules of the game above condition the order of moves on party characteristics. The party whose strongholds are more concentrated is designated as the first mover. What if the rules have to be written in a neutral way? Suppose that the rules can refer to two parties, but must treat them symmetrically, that is, who moves when can be decided by randomization, or by which party is the current majority party (i.e., some exogenous criterion), but cannot condition on which party's support is more concentrated. As illustrated in Section 2, this leads to problems in a one-round system. When R moves first, its protection against stolen majorities is gone. As we will now show, with many rounds, we can have both neutrality and a protection against stolen majorities.

4 Many rounds

In this section, we analyze the game when rules specify that there are L rounds, where L is large, and in each round, parties alternate, and each party distributes a fraction $1/(2LN)$ of their precincts to each district. We show that each party can again ensure that it wins in all states of the world in which it has a majority of the popular vote. Moreover, the strategies that achieve this objective have the property that, if both parties play them, then almost all districts will be replicas of the voter preference distribution of the polity at-large, and thus, competitive.

We show first that Theorem 1 extends to the multi-round setting; both parties continue to be able to protect themselves against majorities being stolen from them. The difference is that many rounds create a level playing field in the following sense: R no longer has a second-mover advantage.

More precisely, we show that R has to play a particular strategy – which we refer to as a *pecking order strategy* – to protect itself against stolen majorities. We show, moreover, that when D also plays a pecking order strategy of their own, then (i) it wins a majority of districts whenever it wins the popular vote, and hence, is also protected against stolen majorities and (ii) almost all districts are turned into replicas of the at-large district.

Notation. Voters are allocated to districts over L rounds such that, in each round l, each party P specifies $\sigma_{Pl} = (\sigma^1_{kPl}, \sigma^2_{kPl})_{k=1}^{2N}$ so that

$$\sigma^1_{kPl} + \sigma^2_{kPl} = \frac{1}{L} \,.$$

Thus, per round, each party P assigns a mass of $\frac{1}{L}$ voters to any one district. For concreteness, we alternate the second-mover role between D and R so that, for l odd, R moves first and D second; for l even, D moves first and R second. However, interchanging this move order would not affect our results.

Denote the total mass of type t_1 partisans assigned by party P to district k over the L rounds by $\sigma^1_{kP} := \sum_{l=1}^{L} \sigma^1_{kPl}$. Analogously, let $\sigma^2_{kP} := \sum_{l=1}^{L} \sigma^2_{kPl}$. To be consistent with the overall distribution of voters, $(\sigma_{kP})_{k=1}^{2N}$ must satisfy

$$\frac{1}{2N}\sum_{k=1}^{2N}\sigma_{kP}^{1}=\beta_1 \quad \text{and} \quad \frac{1}{2N}\sum_{k=1}^{2N}\sigma_{kP}^{2}=\beta_2 \ .$$

Theorem 2 *Let $N \geq 3$. For every $\varepsilon > 0$, there is $\hat{L}$, so that, for $L \geq \hat{L}$: There is a strategy σ_R so that*

$$\Pi_R\left(\sigma_D, \sigma_R \mid \omega > \hat{\omega}\right) = 1 \ , \quad \textit{for every} \quad \sigma_D \ ,$$

and there is a strategy σ_D so that

$$\Pi_D\left(\sigma_D, \sigma_R \mid \omega < \hat{\omega}\right) = 1 \ , \quad \textit{for every} \quad \sigma_R \ .$$

As in the one-round game, the party with more concentrated support, D, can protect its majorities simply by assigning, over the course of the whole procedure, a mass of $2\beta_1$ type 1 voters to half of the districts. Then, for any strategy of R, the percentage share of type 1 voters in those district is at least β_1, which is the share necessary to win a district whenever $\omega < \hat{\omega}$. Furthermore, as D also wins the at-large district whenever $\omega < \hat{\omega}$, this guarantees a legislative majority for D.

The more difficult part of the Theorem is to show that R can also protect its majorities form being stolen. With one round, R's task was facilitated by the fact that R could allocate all of its voters after having observed D's voter assignment; however, this second-mover advantage also generated an asymmetry enabling the creation of a Republican supermajority in expectation. In the multi-round game, things are more difficult for R, but, as we explain now, there still is a strategy that it can use to protect its majorities from being stolen.

4.1 How to protect Republican majorities

A zero-sum game. Our analysis builds on a fictitious zero-sum game. The sequence of moves is as outlined as above, but payoff are as follows: R gets a payoff of $\pi_R = 1$ when there are at least N districts with a type t_2 voter share of at least β_2. Otherwise, R's payoff is $\pi_R = 0$. D's payoff is given by $\pi_D = 1 - \pi_R$.

In the following, we order districts according to their share of type 1 voters, so that District 1 has the (weakly) lowest, and District $2N$ has the (weakly) highest share of type 1 voters. Lemma 2 in the Appendix shows that this is without loss of generality throughout all rounds because strategies that lead to a reordering of districts (say, adding voters in a way such that District 5 after the move has strictly more type 1 voters than District 6) are weakly dominated by order-preserving strategies. Thus, we can restrict parties to only consider assignments that preserve the district ranking.

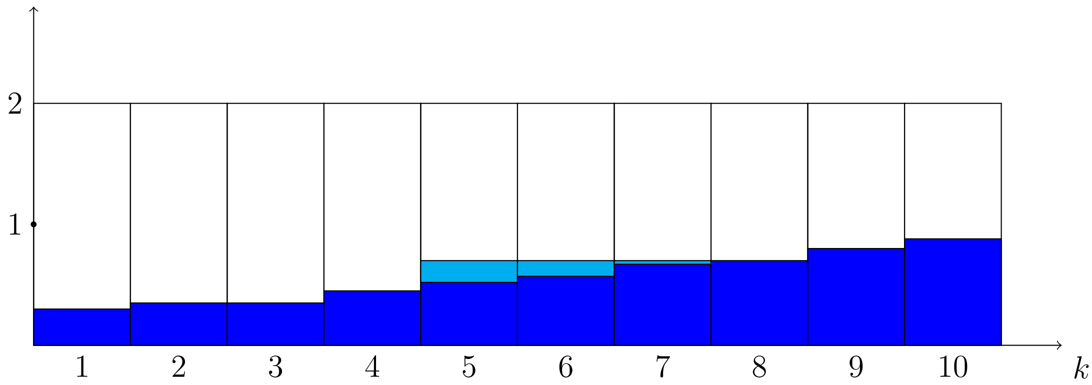
Lemma 1 *Let $N \geq 3$. In the zero-sum game, there is $\hat{L}$ so that $L > \hat{L}$ implies $\pi_R = 1$ in equilibrium.*

Before we illustrate the main argument in the proof, we explain the significance of Lemma 1 for the proof of Theorem 2. R's equilibrium strategy in the zero-sum game allows it to hold the share of type t_1-voters in half of the districts (weakly) below β_1, for any strategy of D. Consequently, if R plays the same strategy in the original redistricting game, it wins all of these districts whenever $\omega > \hat{\omega}$. Could D prevent this outcome by deviating from its equilibrium strategy in the zero-sum game? The answer is negative because any equilibrium strategy for R in the zero-sum game solves a maximin-problem, i.e., it maximizes R's payoff under the assumption that D's strategy is chosen to minimize the maximum attained by R; see e.g. Osborne and Rubinstein (1994). Thus, if D does not behave this way, R's payoff cannot decrease. We therefore obtain the following Corollary to Lemma 1. This completes the proof of Theorem 2.

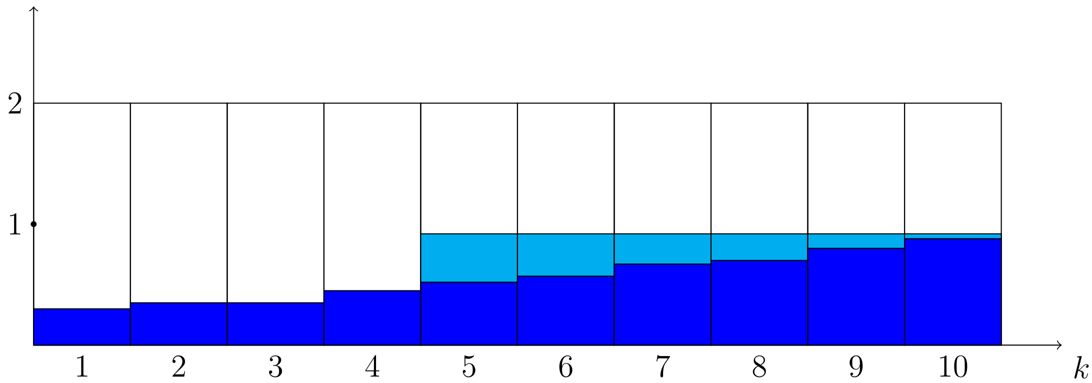
Corollary 1 *Let $N \geq 3$. In the zero-sum game, there is $\hat{L}$ so that $L \geq \hat{L}$ implies the existence of a strategy σ_R so that $\Pi_R(\sigma_D, \sigma_R \mid \omega > \hat{\omega}) = 1$, for all σ_D.*

On the proof of Lemma 1: Pecking order strategies. Because we can focus on the zero-sum game being played in such a way that districts with lower numbers have a (weakly) lower share of type t_1-voters, R needs to ensure that, after L rounds of play, the percentage share of type t_1-voters in district N does not exceed β_1.

To maximize this share, D should not waste type t_1-voters in lower-ranked districts, but rather concentrate type t_1-voters in the $N+1$ top-ranked districts. Specifically, whenever D moves in round l, and plans to assign a certain mass of t_1-voters, the following pecking order is optimal: Assign t_1-voters to district N until its mass of t_1-voters is equal to the one in district $N+1$. From that point on, keep these two districts at a joint level and add further t_1-voters until this joint level equals the one in district $N+2$, and so on, until no further t_1-voters are left, see Figure 5 for an illustration in a setting with ten districts ($N = 5$).



(a) Few t_1 precincts to distribute

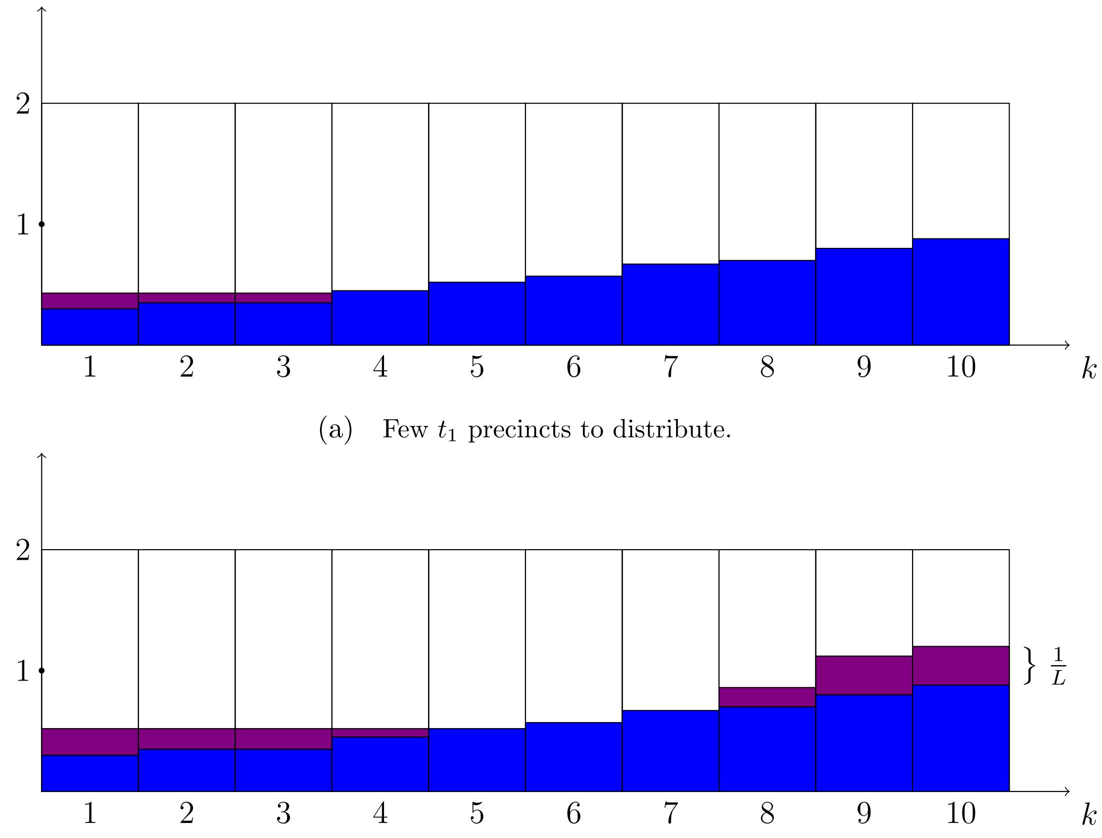


(b) Many t_1 precincts to distribute

Figure 5: 10 Districts. Begin-of-round stock of t_1-precincts in blue. Optimal additions by D in light blue.

What is an optimal response for R? Its problem is to dispose of a total mass of $2N\beta_1$ t_1-voters in such a way that they contribute as little as possible to the mass of t_1-voters in district N. What is clearly harmless is to add t_1-voters to districts with ranks up to $N-1$, provided they are not yet at an equal level with the district that has rank N. Thus, when party R assigns some mass of t_1-voters, it first fills the bottom $N-1$ districts up to the point where a common level of t_1-voters is reached in the bottom N districts; see Figure 6a for an illustration under the assumption that the mass of t_1-voters assigned in round l does not suffice to bring the bottom 4 districts to the level of district 5.

In Figure 6b, instead, the mass exceeds that quantity. When additional t_1-voters need to be assigned after a common level in the bottom N districts has been achieved, party R continues with districts in the upper half. Here, concentrating on the top-ranked districts is optimal. R starts with the top-ranked District $2N$ and fills it up as much as possible. If the per-round capacity constraint of $\frac{1}{L}$ for



(a) Few t_1 precincts to distribute.

(b) Many t_1 precincts to distribute.

Figure 6: 10 Districts. Begin-of-round stock of t_1-precincts in blue; R's additions in purple.

that district is reached, R starts to fill District $2N-1$, and so on. Intuitively, R discards the extra t_1-voters in very few districts in order to make it as difficult as possible for D to "use" these t_1-voters in an attempt to raise the t_1-share in the pivotal district N.

Why does the distribution of voters in non-pivotal districts $k \neq N$ matter at all? Suppose instead that R distributes the t_1-voters uniformly over districts $N+2$ to $2N$. That makes it easier for D to raise the t_1-content of district $N+1$ in the next round: Remember that, when district $N+1$ reaches the level of district $N+2$, D allocates t_1-voters to both of these districts in order to avoid a district rank reversal. By allocating t_1-voters to the highest-ranked districts, R ensures that this no-rank-reversal constraint for D kicks in as early as possible, thereby preventing D from concentrating more of its most loyal voters in the pivotal district.

For a complete characterization of equilibrium strategies we would also need to describe how many t_1-voters are assigned by whom and when, i.e., we would need to characterize, for each party P and any round l, the equilibrium value of β^1_{Pl}, defined as the percentage share of t_1-voters in the total mass of $\frac{2N}{L}$ voters assigned by party P in round l. We do not provide such a complete characterization, but show that R can choose the sequence $\{\beta^D_{Pl}\}^L_{l=1}$ so that the share of t_1-voters in district N remains below β_1. To this end, assume that R chooses $\beta^D_{R1} = 0$, and for any $l \geq 2$, $\beta^D_{Rl} = \beta^D_{Dl-1}$. Thus, R waits until D starts to assign t_1-voters and then assigns in, any round, as many t_1-voters as D assigned in the round before.

Thus, after any of R's moves, the bottom $2N-2$ districts have the same level of t_1-voters, while there are some further t_1-voters in the top ranked district, and, possibly, also in the district with the second highest rank. To see this, suppose for concreteness, that D chooses $\beta^1_{D1} > 0$. Then, it will spread a mass of $\beta^1_{D1}\ \frac{2N}{L}$ t_1-voters evenly over $N+1$ districts. In round 2, R will use the mass of voters previously assigned to $N-1$ of those districts to equalize the level in the bottom half. The remaining mass of t_1-voters is then assigned to at most two top districts. This pattern is now repeated over various rounds, with the implication that, after any move of R there is a joint level of t_1-voters in the bottom $2N-2$ districts.

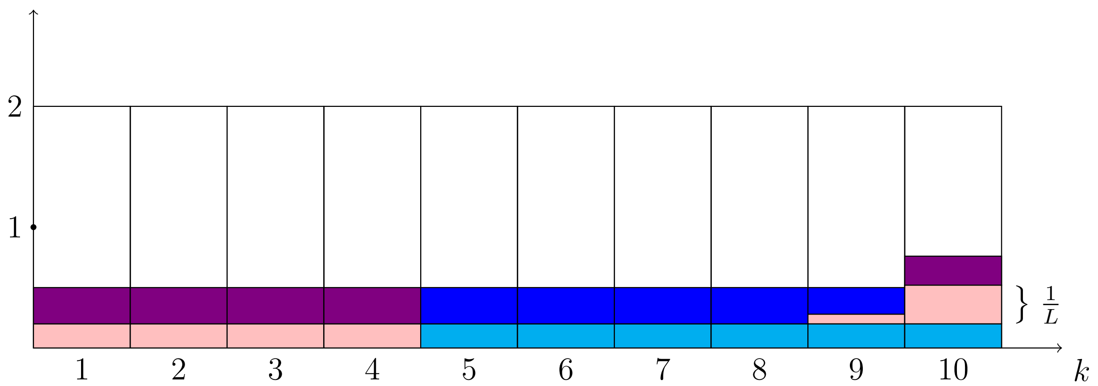


Figure 7: 10 Districts. R assigns as many t_1-precincts as D did in the previous round. Light blue: First-round assignments of t_1-precincts by D. Light red: R's first-round response. Blue: Second-round assignments of t_1-voters. Purple: R's second-round response. As a consequence, there is a common level in the bottom eight districts, both after R's first and second response.

It is now evident that the share of t_1-voters in the pivotal district N cannot strictly exceed β_1. This would imply a percentage share above β_1 in all districts, and is incompatible with the fact that the overall share of t_1-voters is β_1.

Also note that there is a common level of t_1-voters in all districts, with the possible exception of the two top ranked ones, see Figure 7. Thus, R's equilibrium strategy guarantees winning a majority whenever $\omega \in \Omega_R$.

4.2 Further implications of the zero sum game

The above analysis has shown that the fictitious zero sum game has an equilibrium where both parties play pecking order strategies. Moreover, the equilibrium is such that the share of t_1 voters in the pivotal district N is bounded from above by β_1. (Since the game is zero-sum, every equilibrium needs to have that property.)

Moreover, in the equilibrium characterized, R does not assign any t_1-voters to the pivotal district N. While there may be other equilibrium strategies for R – which might differ by the order in which type t_1 voters are assigned to non-pivotal districts – every equilibrium strategy for R has the property that no type t_1 voters are assigned to district N. Otherwise D could exploit this and raise the share of t_1 voters above β_1, so that Democrats would win $N + 1$ districts, and hence a majority in the legislature, in some states $\omega > \omega^*$.

Thus, to protect its majority, R has to follow the pecking order strategy, or, one that has equivalent implications for the pivotal district. Mutatis mutandis, the same is true for D. If it deviated from its pecking order strategy in such a way that it assigned less than $2\beta_1$ type t_1-voters to the pivotal district N, then R could win in some states $\omega < \omega^*$.

We interpret these observations as analogues to Proposition 1 which showed, for the game with one round, D has only one choice if it seeks to protect its majority from being stolen; in contrast, R had more degrees of freedom in the one-round game and could use them to further expand its majority whenever $\omega > \omega^*$ while winning almost the same number of seats as D when $\omega < \omega^*$. In contrast, the multi-round game eliminates this asymmetry; both parties (essentially) have to play pecking order strategies in order to protect their majorities, and so the consequences of $\omega > \omega^*$ and $\omega < \omega^*$ for the seat shares of the parties are symmetric.

Supermajorities. An implication of both parties playing pecking order strategies is that R distributes its t_1-voters evenly over the bottom half of districts, while D distributes its t_1-voters evenly over the top half of districts. Consequently, all districts are turned into replicas of the overall electorate. Thus, whichever party is more popular in an election will win almost all districts.

For reasons outside of our model, an outcome in which the winning party wins in almost all districts may be problematic. Even though the minority party has only limited influence on which policies are enacted even if it is represented in the legislature, this representation may have beneficial effects. For one, the minority can at least participate in the discussion of legislative proposals and provide additional information in this context, and, to the extent that they can persuade the majority party, they can have (possibly Pareto-improving) influence on policy. A strong opposition within the legislature may also be useful for providing information about legislative proposals to the public.

Furthermore, if legislative experience matters for performance, then the voters' opportunity to replace the current majority (if either voters' political preferences shift, or if the current majority party "misbehaves" and needs to be replaced for incentive reasons) is better if the opposition party contains at least some experienced legislators.

So, how could we adjust our system if we wanted to guarantee a substantial opposition representation in the legislature? One simple possibility is to turn each single-member district into a multi-member district. For example, suppose that each district is represented by 3 legislators. Within each district, there is proportional representation (or some transferable vote system), so that the party that gets more votes in the district receives 2 representatives, and the other party the remaining seat if its vote share is above a threshold.[15]

In this case, the redistricting game between the parties remains exactly the same, and the losing party is essentially guaranteed a representation of one-third in the legislature. In contrast to a system with one representative per district, this system would also guarantee that each voter is represented, in the legislature, by (at least) one member of his favorite party from his district.

5 Discussion

Our model shows that we can specify a dynamic game in which both parties assign voters to districts such that each party has a strategy that guarantees winning a majority in the legislature whenever it wins the popular vote. We now discuss some extensions.

[15]The percentage of votes that is required to win one seat in a district of three representatives depends on the specific rules that map the votes obtained by the parties in the district to a seat allocation. For example, with both the Hare-Niemeyer procedure and the Webster/Sainte-Lague procedure (the methods used in German federal elections from 1987 to 2005, and after 2005, respectively), obtaining more than 1/6 of the vote entitles the weaker party in a district with three representatives to one seat. The methods would differ in the vote share that is required to guarantee the stronger party two seats if there are three or more parties.

Communities of interest and geography. As is standard in the formal literature on gerrymandering, we do not impose geographic restrictions on the players.

In a partisan redistricting system, there are two justifications for requiring contiguity of districts. First, and probably most importantly, the contiguity requirement can be interpreted in the current system as a second-best constraint to the gerrymanderer's power that is supposed to limit his ability to implement a biased map. Our mechanism directly gets rid of that power to distort election outcomes, so that indirect constraints for that purpose are unnecessary.

Second, other things equal, both voters and representatives/parties have an interest that voters are located closely to each other as this facilitates the provision of constituency service by the representatives. Observe, however, that contiguity of districts, by itself, is neither necessary nor sufficient for this objective being satisfied – the average distance between voters can be large in contiguous districts, and small in non-contiguous ones. See Figure 8 for an example.[16]

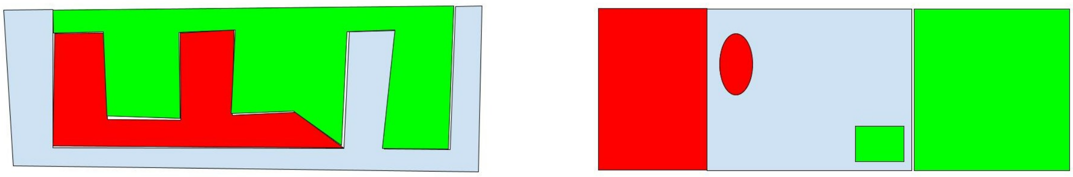


Figure 8: Contiguous districts (left) may have a larger distance between their voters than non-contiguous districts (right)

Furthermore, while it is desirable to keep "communities of interest" together in districts, any type of geographic constraint on drawing districts helps with this objective only as far as the "community of interest" is defined in a geographic way. While neighbors often share some interests, say with respect to local safety and infrastructure, it is not clear that this is necessarily the most important way in which communities of interest can be defined. For example, groups that are defined by their ethnic origin, their religion or their socio-economic position might also be relevant communities that parties might want to combine. Without geographic

[16]The fact that contiguity does not guarantee that voters live close to each other is vividly illustrated by many districts under the current redistricting system. For example, district TX-35 during the 2012-2020 time period stretched all the way from San Antonio to Austin even though both of these metropolitan areas have more than enough population to fill several complete congressional districts.

constraints, it is easier to create districts for non-geographic communities.

For example, suppose that the Democratic party wants to create a district in which a heavily Democratic-voting, but not very concentrated minority group (such as Asian Americans) has the opportunity to select a candidate. With geographic contiguity constraints, this is very difficult to achieve,[17] while it would be easy if precincts can be combined without contiguity restrictions.

This said, suppose that parties, in addition to caring about their chance of winning in future elections, also prefer districts where voters live geographically close to each other. In this case, one could easily consider the outcome of the redistricting game only as a default endowment. If both parties agree, then reassignments of precincts (for example, to generate more compact districts) can certainly be permitted.

Incorporating geography. While there are good arguments for dispensing with geographic constraints, in part B of the Online-Appendix, we sketch how our analysis can be extended to a setting with an explicit geography. In particular, we show that to any strategy in the basic setup, there is a class of "equivalent" strategies in a model where voters are not only distinguished by their voting behavior, but also by their locations on a map.

Moreover, we discuss the trade-offs between a *representation objective* which requires that the more popular party wins the election and a *closeness objective* stipulating that the geographic distance between two voters in the same district must not become too large. We formalize this trade-off by means of a loss function and obtain the following insights: First, there is no conflict between the *representation objective* and the *closeness objective* when one can find the same mix between Republican-leaning and Democrat-leaning voters everywhere on the map. That is, there is an extension of the pecking order strategies from the geography-free setup that does not require "crazy" districts.

Second, when party supporters are clustered in different areas, the "pecking order outcome" from the basic model remains optimal in the degenerate case where the loss function puts all the weight on the *representation objective*.[18]

[17] In the entire U.S., there are currently only 2 Asian majority districts, and 2 more in which there is an Asian plurality. This is less than 1 percent of Congressional districts, while Asians make up about 6 percent of the U.S. population.

[18] Whether competitive gerrymandering yields desirable outcomes with positive weight on the closeness objective is an important follow-up question, but beyond the scope of this paper.

These observations lend themselves to tentative remarks about *optimal political systems*. If the heterogeneity in political preferences across districts gets larger and larger, then the resolution of the trade-off between the *representation objective* and the *closeness objective* becomes more and more painful: The total loss gets large even if the two components of the loss function are appropriately balanced. At some point, sticking to a single-member plurality system may just become too painful and a proportional representation system may then be the more attractive alternative.

6 Concluding remarks

This paper provides one of the first normative analyses of redistricting in a plurality rule electoral system. The analysis is based on the setting which is "canonical" in the large positive literature on partisan gerrymandering. We show how to neutralize the distortions due to partisan gerrymandering by having both parties participate in the redistricting process. In our system, each party has a strategy that guarantees winning a legislative majority whenever it wins the popular vote.

While this possibility result is based on a particular sequential game, the protocol does not have be taken literally as a specific practical redistricting system. It is of theoretical value in that it provides an upper bound for what is achievable when the rules governing the redistricting process are well designed. Presumably, there are other protocols that also implement the popular vote, or at least something closer than the current system. Any such protocol must, however, have the property that the parties can keep each other in check. As the literature on partisan gerrymandering has shown, when one party unilaterally controls the redistricting process, there is no hope to implement the popular vote.

References

Chen, Jowei and Jonathan Rodden, "Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures," *Quarterly Journal of Political Science*, June 2013, *8* (3), 239–269.

Ely, J, "A Cake-Cutting Solution to Gerrymandering," 2019. Northwestern University.

Gersbach, Hans, "Competition of politicians for incentive contracts and elections," *Public choice*, 2004, *121* (1), 157–177.

_ **and Verena Liessem**, "Incentive contracts and elections for politicians with multi-task problems," *Journal of Economic Behavior & Organization*, 2008, *68* (2), 401–411.

Groseclose, Tim and James M. Snyder, "Buying Supermajorities," *The American Political Science Review*, 1996, *90* (2), 303–315.

Konrad, Kai A, *Strategy and dynamics in contests*, Oxford University Press, 2009.

Kovenock, Dan and Brian Roberson, "Generalizations of the General Lotto and Colonel Blotto games," *Economic Theory*, 2020, *7* (1), 5 – 22.

Laslier, Jean-François and Nathalie Picard, "Distributive Politics and Electoral Competition," *Journal of Economic Theory*, 2002, *103* (1), 106 – 130.

Lizzeri, Alessandro and Nicola Persico, "The provision of public goods under alternative electoral incentives," *American Economic Review*, 2001, *91* (1), 225–239.

_ **and** _ , "A drawback of electoral competition," *Journal of the European Economic Association*, 2005, *3* (6), 1318–1348.

Myerson, Roger, "Incentives to Cultivate Favored Minorities Under Alternative Electoral Systems," *American Political Science Review*, 1993, *87* (4), 856–869.

_ , "Federalism and incentives for success in democracy," *Quarterly Journal of Political Science*, 2006, *1* (1), 3–23.

Osborne, Martin and Ariel Rubinstein, *A course in Game Theory*, MIT Press, Cambridge, MA., 1994.

Owen, Guillermo and Bernard Grofman, "Optimal partisan gerrymandering," *Political Geography Quarterly*, 1988, *7* (1), 5 – 22.

Persson, Torsten and Guido Tabellini, "Political economics and public finance," *Handbook of public economics*, 2002, *3*, 1549–1659.

_ **, Gerard Roland, and Guido Tabellini**, "Comparative politics and public finance," *Journal of Political Economy*, 2000, *108* (6), 1121–1161.

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

von Neumann, John, "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen*, 1928, *100* (1), 295—-320.

Zermelo, Ernst, "Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels," in "Proceedings of the fifth international congress of mathematicians," Vol. 2 Cambridge University Press 1913, pp. 501–504.

A Appendix

A.1 Proof of Theorem 1

1. Observe first that, if there are N districts in which the percentage of type 1 voters is at least β_1, then D wins these districts whenever $\omega < \hat{\omega}$, and thus, together with the at-large district, a legislative majority.

   Suppose D chooses $\sigma^1_{kD} = 2\beta_1$ for all districts $k \in \{N+1, N+2, \ldots, 2N\}$ and $\sigma^1_{kD} = 0$ for all districts $k \in \{1, 2, \ldots, N\}$. Such a strategy is feasible because $\frac{1}{2N} \sum_{k=1}^{2N} \sigma^1_{kP} = \beta_1$. Furthermore, since $\sigma^1_{kR} \geq 0$, the percentage of type 1 precincts is at least $\frac{2\beta_1+0}{2} = \beta_1$ for all districts $k \in \{N+1, N+2, \ldots, 2N\}$. Thus, this strategy guarantees $\Pi_D(\sigma_D, \sigma_R \mid \omega < \hat{\omega}) = 1$.

2. Without loss of generality, let the σ^1_{kD} be ordered in a weakly increasing way such that $k < k'$ implies $\sigma^1_{kD} \leq \sigma^1_{k'D}$. Thus, districts 1 to N have a (weakly) lower percentage of type 1 voters than districts $N+1$ to $2N$. We want to show that R can choose $\sigma^1_{kR} \geq 0$ for districts 1 to N such that $\frac{\sigma^1_{kD}+\sigma^1_{kR}}{2} \leq \beta_1$.

   Assume, to the contrary, that there is a feasible strategy for D such that this is not possible for R because the percentage of Democrat-leaning precincts in at least some of these districts is so high after D's move that $\frac{\sigma^1_{kD}+0}{2} > \beta_1$, or, equivalently, it must be true that, at least in district N, we have $\sigma^1_{ND} > 2\beta_1$.

   Because the districts are ordered in an increasing way, for all districts $k > N$, we have $\sigma^1_{kD} \geq \sigma^1_{ND} > 2\beta_1$. Thus, summing over all districts, we have

   $$\frac{1}{2N} \sum_{k=1}^{2N} \sigma^1_{kD} > \frac{N+1}{2N} 2\beta_1 = \frac{N+1}{N} \beta_1 > \beta_1.$$

   Thus, D's strategy is not feasible, which gives the desired contradiction. ■

A.2 Proof of Proposition 1

Without loss of generality, assume that the districts are numbered in weakly decreasing order of their type-1 voter content after the Democrat's move, and that, to the contrary of the statement, $\sigma^1_{D,N} < 2\beta_1$, either because Democrats allocated some type-1 precincts to the other half of districts ($\sigma^1_{D,N+1} > 0$), and/or because they allocated more than $2\beta_1$ to some district $k < N$. We prove that Republicans can respond so that they reduce the total share of Democratic-leaning type

1 precincts to less than β_1 in $N+1$ districts. As a consequence, Republicans win $N+1$ districts whenever $\omega < \hat{\omega}$ for $\hat{\omega} < \omega^*$.

Consider the following strategy by the Republicans: Assign as many type 1 precincts as possible to district 1, i.e. $\sigma^1_{R,1} = \min\{1, 2N\beta_1\}$. If type 1 precincts are left, i.e. if $1 < 2N\beta_1$ assign as many precincts as possible to district 2, $\sigma^1_{R,1} = \min\{1, 2N\beta_1 - 1\}$, proceed analogously until district $N-1$ is reached. If $2N\beta_1 > N-1$, then further type 1 districts are left. Those are allocated to the district with rank $2N$, possibly until the fraction of type 1 of districts is equal to the one in district $2N-1$. If a common level is reached, and further precincts are left, they are spread evenly over districts $2N$ and $2N-1$ until the fraction of type 1 precincts is equal to the one in district $2N-2$. If type 1 precincts are left, they are spread evenly over districts $2N$, $2N-1$, and $2N-2$ until the fraction of type 1 precincts is equal to the one in district $2N-3$, and so on.

Now suppose that after the Republicans move, the mass of type 1 precincts in district N is at least $2\beta_1$. Then the same is true for all districts with an index larger than N. Moreover, for all districts smaller than N the mass of type 1 precincts is strictly larger than $2\beta_1$, since the Republicans assigned a mass of $1 > 2\beta_1$.

Consequently, the total mass of type precincts that have been assigned is bounded from below by

$$N - 1 + (N+1)2\beta_1 > 4N\beta_1 \ ,$$

which is a contradiction, since, by the parties' budget constraints,

$$\sum_{k=1}^{2N} \sigma^1_{kD} + \sum_{k=1}^{2N} \sigma^1_{kR} = 4N\beta_1 \ .$$

■

A.3 Proof of Theorem 2

We first show that party D has a strategy that guarantees a majority in the legislature whenever $\omega < \omega^*$. Consider the following strategy for party D: In all rounds l, choose $\sigma^1_{kDl} = 0$, for $k \leq N$ and $\sigma^1_{kDl} = \frac{2\beta_1}{L}$, for all $k > N$. We seek to show that, with this strategy, for all σ_R and for all districts with an index $k > N$,

$$(\sigma^1_{kD} + \sigma^1_{kR})\ v(t_1, \omega) + (\sigma^2_{kD} + \sigma^2_{kR})\ v(t_2, \omega) \quad < \quad \frac{1}{2} \ , \tag{3}$$

whenever $\omega < \hat{\omega}$. Since the left-hand side of equation (3) decreases in ω, it suffices to show that

$$(\sigma^1_{kD} + \sigma^1_{kR})\ v(t_1, \hat{\omega}) + (\sigma^2_{kD} + \sigma^2_{kR})\ v(t_2, \hat{\omega}) \quad \leq \quad \frac{1}{2}\ , \tag{4}$$

or, equivalently, that

$$\sigma^1_{kD} + \sigma^1_{kR} \quad \geq \quad \frac{v(t_2, \hat{\omega}) - \frac{1}{2}}{v(t_2, \hat{\omega}) - v(t_1, \hat{\omega})} \quad = \quad \beta_1\ , \tag{5}$$

where the inequality in the left part of (5) follows from (4) upon using that $\sigma^2_{kD} = 1 - \sigma^1_{kD}$ and $\sigma^2_{kR} = 1 - \sigma^1_{kR}$. The equality in the right part of (5) then follows from (1).

After L rounds, the total mass of voters assigned by the two parties to any one district k equals 2. Under party D's strategy the share of type 1 voters is in any district with an index $k > N$ is bounded from below by β_1. To see this note that

$$\sigma^1_{kD} + \sigma^1_{kR} \quad \geq \quad \frac{L\ \frac{2\beta_1}{L}}{2} \quad = \quad \beta_1\ .$$

In the remainder of the proof we show that party R has a strategy that guarantees a majority in the legislature whenever $\omega > \omega^*$.

A.3.1 On the ranking of districts

Ordering districts. If the game were to end after round l, party R would win district k in state ω when

$$\sum_{j=1}^{2} v(t_j, \omega)\ \frac{L}{2l}\ \left(s^l_{Dk}(t_j) + s^l_{Rk}(t_j)\right) \quad > \quad \frac{1}{2}\ , \tag{6}$$

where $s^l_{Dk}(t_j) := \sum_{l'=1}^{l} \sigma_{Dkl'}(t_j)$ and $s^l_{Rk}(t_j) := \sum_{l'=1}^{l} \sigma_{Rkl'}(t_j)$ are the stocks of type t_j voters who have been assigned by parties D and R, respectively, over the first l rounds of play. To interpret this inequality, note that $\frac{L}{2l}\ (s^l_{Dk}(t_j) + s^l_{Rk}(t_j))$ is the share of type t_j voters among those voters who have been assigned to district k in the first l periods. Thus, if ω is such that the above inequality holds, then party R has majority support in district k after round l.

Let $s^l_k(t_j) := s^l_{Dk}(t_j) + s^l_{Rk}(t_j)$. We define a rank order of districts according to their republican vote share after l rounds of play. Thus, the rank of district k is higher than the rank of district k' if, for some ω,

$$\sum_{j=1}^{2} v(t_j, \omega)\ \frac{L}{2l}\ s^l_k(t_j) \quad \geq \quad \sum_{j=1}^{2} v(t_j, \omega)\ \frac{L}{2l}\ s^l_{k'}(t_j)\ . \tag{7}$$

Using that, in any district the shares of type 1 and type 2 voters add up to 1, inequality (7) can equivalently be written as

$$\begin{aligned} & v(t_1,\omega) + \tfrac{L}{2l}\, s^l_k(t_2)\Big(v(t_2,\omega) - v(t_1,\omega)\Big) \\ & \geq v(t_1,\omega) + \tfrac{L}{2l}\, s^l_{k'}(t_2)\Big(v(t_2,\omega) - v(t_1,\omega)\Big) \ . \end{aligned} \tag{8}$$

or, more simply, as

$$s^l_k(t_2) \geq s^l_{k'}(t_2) \ .$$

Thus, ordering districts according to their republican vote share is equivalent to ordering them according to the share of type 2 voters. Also, the Republican vote share in any district k is, for every state ω, a monotonic function of the mass of type 2 voters.

Order preserving assignments. Assume without loss of generality that after l rounds of play district 1 has a weakly lower Republican vote share than district 2, that district 2 has a weakly lower Republican vote share than district 3 and so on. District $2N$ is then among those with a maximal republican vote share. Now consider round $l+1$. Suppose that party R moves first in round $l+1$. It then assigns a mass of $\frac{1}{L}$ voters to any district k. Thus, for any district k,

$$\sum_{j=1}^{2} \sigma_{kRl+1}(t_j) \quad = \quad \frac{1}{L} \ .$$

This move of R induces a new order of districts according to

$$s^l_k(t_2) + \sigma_{kRl+1}(t_2) \ .$$

Let $r_\sigma(k) \in \{1, \ldots, 2N\}$ be the new rank of the district with initial rank k.

Lemma 2 *Given a move $\sigma_{Rl+1} = (\sigma_{kRl+1})_{k=1}^{2N}$ of party R in round $l+1$ with a resulting ranking $k \mapsto r_\sigma(k)$ according to the republican vote share, there is an alternative move $\sigma'_{Rl+1} = (\sigma'_{kRl+1})_{k=1}^{2N}$ of party R with the following properties:*

i) The alternative move uses the same voter types: For every j,

$$\sum_{k=1}^{2N} \sigma_{kRl+1}(t_j) \quad = \quad \sum_{k=1}^{2N} \sigma'_{kRl+1}(t_j) \ .$$

ii) The alternative move preserves the old ranking; formally, it induces a new ranking $k \mapsto r_{\sigma'}(k)$ so that $r_{\sigma'}(k) = k$, for every k.

iii) The republican vote share in the district with rank $N+1$ under the alternative move σ'_{Rl+1} is at least as high as in the district with rank $N+1$ under the initial move.

Proof of Lemma 2. Suppose there is some district with initial rank k' that has rank k in the ranking induced by $\sigma_{Rl+1} = (\sigma_{kRl+1})_{k=1}^{2N}$; i.e. $k' = r^{-1}(k)$. The mass of type t_2 voters after R's move under $\sigma_{Rl+1} = (\sigma_{kRl+1})_{k=1}^{2N}$ is given by

$$s^l_{k'}(t_2) + \sigma_{k'Rl+1}(t_2) \ .$$

We now choose $\sigma'_{kRl+1}(t_2)$ so that

$$\sigma'_{kRl+1}(t_2) \quad = \quad \max\left\{0, s^l_{k'}(t_2) + \sigma_{k'Rl+1}(t_2) - s^l_k(t_2)\right\} \ .$$

Proceeding in the same way for all k implies that

$$s^l_k(t_2) + \sigma'_{kRl+1}(t_2) \quad \geq \quad s^l_{r^{-1}(k)}(t_2) + \sigma_{r^{-1}(k)Rl+1}(t_2) \ .$$

The mass of type t_2 voters used by $\sigma'_{Rl+1} = (\sigma'_{kRl+1})_{k=1}^{2N}$ across all districts is such that

$$\sum_{k=1}^{2N} \sigma'_{kRl+1}(t_2) \quad = \quad \sum_{k=1}^{2N} \max\left\{0, s^l_{r^{-1}(k)}(t_2) + \sigma_{r^{-1}(k)Rl+1}(t_2) - s^l_k(t_2)\right\}$$

An upper bound is obtained under the assumption that

$$s^l_{r^{-1}(k)}(t_2) + \sigma_{r^{-1}(k)Rl+1}(t_2) - s^l_k(t_2) > 0 \ ,$$

for all k, i.e. so that type t_2 voters have to be assigned to all districts. Therefore,

$$\begin{aligned} \textstyle\sum_{k=1}^{2N} \sigma'_{kRl+1}(t) &\leq \textstyle\sum_{k=1}^{2N} s^l_{r^{-1}(k)}(t) + \sigma_{r^{-1}(k)Rl+1}(t) - s^l_k(t) \\ &= \textstyle\sum_{k=1}^{2N} s^l_{r^{-1}(k)}(t) - \sum_{k=1}^{2N} s^l_k(t) + \sum_{k=1}^{2N} \sigma_{r^{-1}(k)Rl+1}(t) \\ &= \textstyle\sum_{k=1}^{2N} \sigma_{r^{-1}(k)Rl+1}(t) \ . \end{aligned}$$

Thus, σ'_{Rl+1} does not use more type t_2 voters than σ_{Rl+1}, and it yields, in any district, at least as type 2 voters in total. If there is a strict inequality, i.e. if σ'_{Rl+1} use strictly less type t_2 voters than σ_{Rl+1}, then those voters can be assigned to the districts in such a way that the initial ranking is preserved. □

A.3.2 Proof of Lemma 1

A strategy for party R. In any round l, given a – for now exogenous – budget of $\beta^1_{Rl}\,\frac{2N}{L}$ type t_1-voters to be assigned, proceed sequentially in the following way – until the budget of type t_1-voters for that round is exhausted:

i) Add type t_1-voters to the lowest ranked district until the mass of t_1-voters equals the mass in the district with the second lowest rank. From then on, keep the mass in these two districts equal.

ii) Add type t_1-voters to the two lowest ranked districts until the mass of t_1-voters equals the mass in the district with the third lowest rank. From then on, keep the mass in these two districts equal.

iii) Proceed analogously for all districts with a rank smaller or equal $N-2$. From then on, keep the mass in all these districts equal. Add t_1-voters to the $N-1$ lowest ranked districts until the mass of t_1-voters equals the mass in the district with rank N. From then on, don't add further t_1-voters to one of the bottom N districts.

iv) Add t_1-voters to the top ranked district.

v) If there are still t_1-voters left in the budget after a mass of $\frac{1}{L}$ t_1-voters has been assigned to the top ranked district, add t_1-voters to the district with the second highest rank, etc, then move to the district with the third highest rank, etc.

vi) Stop when no further t_1-voters are left.

Note that, as an implication, R's play in any round leaves the ranking of districts unchanged.

A best response for party D. Consider a – for now exogenous – sequence of budgets for party D's play $\{\beta^1_{Dl}\}_{l=1}^{L}$. Note that since party R never affects the ranking of districts, the ranking of districts in any round is entirely due to party D. As argued above it entails no loss of generality to assume that party D's moves do neither affect the ranking of districts. This also implies that it is never optimal to have a budget of partisan D voters in some round that makes it necessary to assign D voters to strictly more than $N+1$ districts. Thus, we may assume that, for any round l,

$$\beta^1_{Dl}\,\frac{2N}{L} \quad \leq \quad \frac{N+1}{L}\,,$$

or, equivalently,

$$\beta^D_{Dl} \quad \leq \quad \frac{1}{2}+\frac{1}{2N}.$$

Given some budget for moves in round l, the optimal strategy for party D is now as follows:

i) Add type t_1-voters to the district with rank N until the mass of t_1-voters equals the mass in the district with the rank $N+1$. From then on, keep the mass in these two districts equal.

ii) Add type t_1-voters to the two districts with ranks N and $N+1$ until the mass of t_1-voters equals the mass in the district with rank $N+2$. From then on, keep the mass in these three districts equal.

iii) Proceed analogously for all districts with a rank larger or equal $N+2$, until the budget of D voters is exhausted.

Party R's sequence of budgets. We now specify a particular sequence of budgets for party R: As the first mover in the initial round, it does not assign any type t_1-voters, $\beta^1_{R1}=0$. In any round $l\geq 2$, and as long os this is feasible, party R assigns as many t_1-voters as party D did in the previous round

$$\beta^1_{Rl+1} \quad = \quad \beta^1_{Dl}\,.$$

This is clearly feasible in early rounds. If, however, party D keeps some type t_1-voters for the last round so that $\beta^1_{DL}>0$, then party R will have to assign an additional mass of $\beta^1_{DL}\leq\frac{2N}{L}$ type t_1-voters somewhen in the game. Otherwise party R would violate its overall budget constraint. Note that this quantity vanishes for $L\rightarrow infty$.

Thus, there is a subset of rounds L' so that

$$\sum_{l'\in L'}\beta^1_{Dl'}<\sum_{l'\in L'}\beta^1_{Rl'+1} \quad \leq \quad \sum_{l'\in L'}\beta^1_{Dl'}+\frac{2N}{L}\,.$$

and for l not in L' we let

$$\beta^1_{Rl+1} \quad = \quad \beta^1_{Dl}\,.$$

Party R's strategy has the following implication: Whenever party R moves, it brings the mass of t_1-voters in the bottom $N-1$ districts to the level that party D has generated for the district with rank N in the previous round. Moreover, party R adds t_1-voters at most to the two top-ranked districts, and does not assign any D voters to districts with the ranks $N, N+1, \ldots, 2N-2$.

To see this, first consider rounds 1 and 2:

- In round 1, party D assigns an equal mass of D voters to $N+1$ districts.
- In round 2, party R fills the bottom $N-1$ districts. It then has additional t_1-voters left. According to party R's strategy, as many t_1-voters as possible are assigned to the district with the top rank $2N$. If additional t_1-voters are left, they go to the district with rank $2N-1$ and then, possibly, to the district with rank $2N-2$.

Now consider rounds 3 and 4:

- In round 3, party D's best response stipulates to assign an equal mass of t_1-voters to the districts with ranks $N, N+1, \ldots, 2N-2$. Those are $N-1$ districts. Possibly, it also assigns t_1-voters to the three top ranked districts.
- In round 4, party R fills the bottom $N-1$ districts. It can do so by adding to the districts in the bottom $N-1$ exactly the amount of D voters that party D has added to the districts with ranks $N, N+1, \ldots, 2N-2$ in round 3.
- If party D has previously added t_1-voters to the two top ranked districts, then party R has additional t_1-voters left after the bottom $2N-2$ districts have been leveled. Again, by party R's strategy, of these voters as many as possible are assigned to the district with the top rank $2N$. If additional t_1-voters are left, they go to the district with rank $2N-1$.

Completing the argument. Suppose first that, for all l,

$$\beta^D_{Rl+1} \quad = \quad \beta^D_{Dl} \, .$$

The strategies of parties R and D described above then imply that after the last move in round L, there is an equal mass of type t_1-voters for all districts with a rank smaller or equal to $2N-2$. The mass of these voters is (weakly) larger in the two top ranked districts. Now suppose that the percentage share of t_1-voters in the district with rank N is strictly larger than β_1. Equivalently, the total mass of t_1-voters in that district exceeds 2 β_1. Then, the mass of t_1-voters exceeds 2 β_1 in all districts. Hence, the total mass of assigned t_1-voters is strictly larger than $4N\ \beta_1$. But this is infeasible as the two parties' total endowments with partisan t_1-voters only sum to $4N\ \beta_1$. Thus, the assumption that party D can generate $N+1$ districts with a percentage share of type t_1-voters strictly larger than β_D leads to a contradiction, and must be false.

Now suppose, there needs to be a subset of rounds L' so that

$$\sum_{l' \in L'} \beta^1_{Dl'} < \sum_{l' \in L'} \beta^1_{Rl'+1} \quad \leq \quad \sum_{l' \in L'} \beta^1_{Dl'} + \frac{2N}{L} \ .$$

For L sufficiently large, we can chose the number of such rounds equal to $2N$, i.e. $\#L' = 2N$. Party R can then satisfy its overall budget constraint by assigning, for every round $l' \in L'$, an additional mass of t_1-voters that is bounded from above by $\frac{1}{L}$.

Then, party R's moves in rounds $l' \in L'$ may require to add type t_1-voters to the three highest ranked districts, with the mass going to the district with rank $2N-2$ being bounded from above by $\frac{1}{L}$. The strategies of parties R and D described above then imply that after the last move in round L, there is an equal mass of type t_1-voters for all districts with a rank smaller or equal to $2N-3$. The mass of these voters is (weakly) larger in the three top ranked districts. Again, the assumption that the percentage share of t_1-voters in the district with rank N is strictly larger than β_1 leads to a contradiction, and must be false. □

B Geography: The trade-off between the representation objective and the closeness objective

The previous analysis was based on the workhorse model for theoretical analyses of gerrymandering due to Owen and Grofman (1988). In that model voters or precincts are distinguished only by how likely they vote for the Democrats or the Republican party. Thus, one can think of a gerrymanderer as having an urn that contains all Republican-leaning precincts and an urn that contains all precincts that are leaning to the Democrats. District are like slots that need to be filled with precincts. There are no constraints on feasible voter assignments, any precinct can go to any district. This model has proven useful to obtain a basic understanding of how majorities can be stolen by gerrymandering. Presumably, this explains why it has become popular in theoretical analyses. This being said, the model provokes the question whether it is "as simple as possible" or "too simple." Practically, a gerrymanderer has to draw district boundaries on a map. She is not having urns and not every precinct can go to every district.

In this section, we discuss the extent to which our analysis can be extended to a richer model with an explicit account of geography. We consider a setting in which voters or precincts differ in two dimensions. As before, one dimension is the voting behavior. Some precincts are Republican-leaning and some lean towards the Democrats. In addition, precincts now have a location on a map. This setting gives rise to a tradeoff: On the hand, there is a *representation objective*: The party that wins the popular vote should win a majority in the legislature. On the other hand, there is a *closeness objective*: it is desirable that voters who belong to the same district do not live far apart. We formalize this trade-off with a loss function. There is a loss when the "wrong party" wins the election and there is a loss when the geographical distance between the voters in one district becomes too large.

We make two observations. First, suppose one can find the same mix between Republican-leaning and Democrat-leaning precincts everywhere on the map. Then there is no tradeoff. In this case, to any voter assignment obtained in the Owen and Grofman (1988) setup, there is an equivalent one that does not require to compromise with the closeness objective. This implies, in particular, that our analysis of competitive gerrymandering in the previous sections extends to such an ideal geography. Second, if the geography is not ideal, then there is a non-trivial tradeoff. The outcomes obtained from our model of competitive gerrymandering are still optimal when the loss function puts all the weight on the representation objective.

This shows that our analysis is not inconsistent with an enriched setup that takes explicit account of spatial considerations. Our results have an extension into such a setup. At the same time, it shows the limitations of our analysis. Our analysis does not settle the question whether competitive gerrymandering gives rise to desirable outcomes when there are is a non-trivial trade-off between the *representation objective* and the *closeness objective*.

B.1 An enriched setup

Let I be the set of voters or precincts. Any voter i is characterized by a pair (t^i, ℓ^i). As before $t^i \in \{t_1, t_2\}$, where $t^i = t_1$ indicates that voter i leans towards the Democrats and $t^i = t_2$ indicates that i leans towards the Republicans. Voter i now also has a location on a map that we denote by ℓ^i. More formally, ℓ^i is a point in a two-dimensional space. Thus, $\ell^i = (\ell^i_x, \ell^i_y)$, where ℓ^i_x is the x-coordinate and ℓ^i_y is the y-coordinate. We denote by I_1 the set of voters with $t^i = t_1$ and by I_2 the set of voters with $t^i = t_2$.

In this enriched setup, voter assignments can condition on both t^i and ℓ^i. In the previous analysis, they could condition only on t^i. We now define an equivalence relation between voter assignments in the enriched setup and voter assignments in the previous setup. To this end, denote the set of feasible voter assignments for the previous setup by Σ. A generic element σ of Σ specifies, for any district k, a mass of Democrat-leaning voters σ^1_k that are assigned to this district and a mass of Republican leaning voters assigned to this district, σ^2_k. Denote the set of voter assignments in the enriched setup by Σ_I. In the enriched setup, every voter i is represented by a collection of dummy variables $\sigma_i = (\sigma_{ik})_{k=1}^{2N}$, where $\sigma_{ik} = 1$ if i is assigned to district k and $\sigma_{ik} = 0$, otherwise. Thus a generic element of Σ_I is a collection $\sigma_I = (\sigma_i)_{i \in I}$.

For any voter assignment $\sigma \in \Sigma$ we now define an equivalence class of voter assignments $C(\sigma) \subset \Sigma_I$. Specifically, we say that σ_I belongs to $C(\sigma)$ if, for any district k,

$$\sigma^1_k = \sum_{i \in I_1} \sigma_{ik} \quad \text{and} \quad \sigma^2_k = \sum_{i \in I_2} \sigma_{ik} \ .$$

Thus, the different assignments that belong to $C(\sigma)$ have in common that they induce the same vote distribution in any district k. They differ in the spatial distribution of voters across districts.

To study whether insights from the previous setup without geography extend to the current setup we proceed as follows: Recall that, in the setting without geography, both parties need to play pecking order strategies if they seek to protect their majorities. Also recall that this results in a voter assignment so that almost all districts have the same mix of type t_1 and type t_2 voters as the electorate at large. Denote this voter assignment by σ^*. We now study the corresponding equivalence class $C(\sigma^*)$ to see whether the protection against stolen majorities that is offered by σ^* makes it necessary to have "crazy" districts.

As we will now elaborate, the answer to this question depends on the spatial distribution of voters. Suppose, for concreteness, that σ^* implies that 16 districts have the same mix type t_1 and type t_2 voters as the electorate at large. Consider the spatial distribution of voters in Figure 9. Then, as the figure shows, there is an element of $C(\sigma^*)$ which gives rise to 16 "reasonably shaped" districts. The spatial distributions in Figure 10, by contrast, are more challenging. To see this, consider the blue dots far in the north-east of Figure 10a. Any attempt to have any one of these voters in a district with another blue voter implies that the geographical distance between voters in the same district becomes "large." The same is true for the blue dot far in the north-west of Figure 10b.

The spatial distributions in Figure 10 show that there is a trade-off between implementing an element of $C(\sigma^*)$ (the representation objective) and the objective that neighboring voters belong to the same district (the closeness objective.) An optimal resolution of this tradeoff can be formalized by means of a loss function.

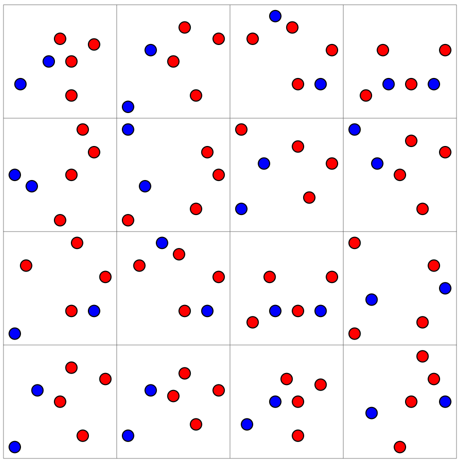
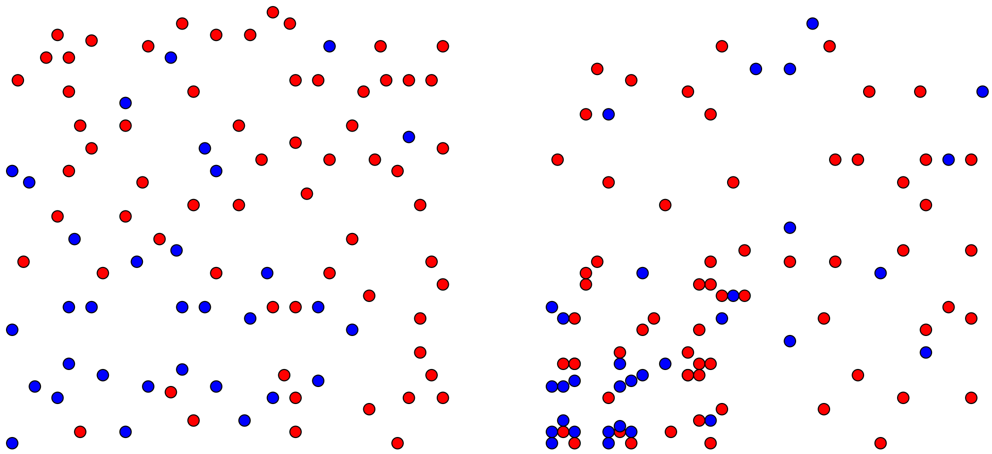


Figure 9: An ideal spatial distribution of voters. Red dots represent R voters and blue dots D voters. Here, voters can be assigned to districts in such a way that the assignment belongs to the equivalence class of the assignment that results when both parties play pecking order strategies in the fictitious zero sum game

B.2 A loss function

We now develop a loss function that makes it possible to trade-off deviations from the representation objective and deviations from the closeness objective. We begin with the losses from missing the representation objective and turn to the losses from missing the closeness objective subsequently.



(a) No spatial concentration of voters, preferences correlated with location

(b) Spatial concentration of voters, and preferences correlated with location

Figure 10: Spatial distributions of voters that are not ideal

Losses from missing the representation objective. Denote the set of states in which party D or party R wins the popular vote, respectively, by

$$\Omega_D := \{\omega : \omega < \omega^*\} \quad \text{and} \quad \Omega_R := \{\omega : \omega > \omega^*\} \ .$$

We denote by the Republican vote share in the popular vote by

$$\Pi_R(\omega) = \beta_1 \ v(t_1, \omega) + \beta_2 \ v(t_2, \omega) \ .$$

The democratic vote share is given by

$$\Pi_D(\omega) = 1 - \Pi_R(\omega) \ .$$

The margin of victory in the popular vote is given by

$$M(\omega) = \max\{\Pi_D(\omega), \Pi_R(\omega)\} - \min\{\Pi_D(\omega), \Pi_R(\omega)\} \ .$$

Any assignment of voters to districts implies an order of the $2N$ districts according to their cutoff values $\hat{\omega}_k$ defined by the property that district k goes to party R when $\omega > \hat{\omega}_k$ and goes to party D when $\omega < \hat{\omega}_k$. Without loss of generality let the district indices reflect this order so that

$$\hat{\omega}_1 \quad \leq \quad \hat{\omega}_2 \quad \leq \quad \ldots \quad \leq \quad \hat{\omega}_{2N} \ .$$

Thus, party D wins the election when ω belongs to

$$V_D = \{\omega : \omega < \hat{\omega}_N \quad \text{or} \quad (\ \omega < \hat{\omega}_{N+1} \ \text{ and } \ \omega \in \Omega_D\ )\} \ .$$

Analogously, party R wins the election when ω belongs to

$$V_R = \{\omega : \omega > \hat{\omega}_{N+1} \quad \text{or} \quad (\ \omega > \hat{\omega}_N \ \text{ and } \ \omega \in \Omega_R\ )\}\ .$$

Note that the sets V_D and V_R depend on the assignment of voters to districts. Henceforth, for any $\sigma_I \in \Sigma_\ell$, we denote by $V_D(\sigma_I)$ the set of states in which party D wins a majority and by $V_R(\sigma_I)$ the set of states in which party R wins a majority. Armed with this notation, we can now define a set of "bad" states, i.e. of states with a discrepancy between the party that wins the election and the party that wins the popular vote. Given an assignment σ_I, the set of bad states is

$$B(\sigma_I) = \Big\{\omega : \omega \in \Big(\Omega_D \cap V_R(\sigma_I)\Big) \cup \Big(\Omega_R \cap V_D(\sigma_I)\Big)\Big\}\ .$$

We are now in the position to formally introduce the first element of our loss function. Losses from missing the representation objective are captured by

$$\mathcal{L}(\mathbf{1}(\omega \in B(\sigma_I))g(M(\omega)))\ ,$$

where $\mathcal{L}$ is an increasing function and $\mathbf{1}$ is the indicator function. We are interacting the indication for a bad state with the function $g : M(\omega) \mapsto g(M(\omega))$. This allows for a range of specific formulations of the representation objective. For instance, suppose that g is an increasing function. Then, when we are close to a tie in the popular vote, it is more acceptable that the "wrong party wins" than otherwise. By contrast, for g decreasing, a victory of the "correct" party is more important when the election is close.

Losses from missing the closeness objective. Let d_{ij} be the geographical distance between two voters i and j who belong to district k. Then define

$$L_k \quad := \quad \max_{ij \in I_k \times I_k} \ d_{ij}\ ,$$

where I_k is the set of voters who have been assigned to district k. Also define

$$L \quad := \quad \max_{k \in \{1,\dots,2N\}} \ L_k\ .$$

Henceforth we write $L(\sigma_I)$ for the losses that result from a voter assignment σ_I.

The loss function. The overall loss from a voter assignment σ_I can now be written as

$$O(\sigma_I) = \lambda\ \mathcal{L}(\mathbf{1}(\omega \in B(\sigma_I))g(M(\omega)) + (1-\lambda)\ L(\sigma_I)$$

where $\lambda \in [0, 1]$ is the weight on the representation objective and $1 - \lambda$ is the weight on the closeness objective. The set of optimal assignments is given by

$$\Sigma_l^*(\lambda) \quad = \quad \text{argmox}_{\sigma_I \in \Sigma_l} O(\sigma_I) \ ,$$

where the notation $\Sigma_l^*(\lambda)$ emphasizes that what is optimal depends on the weight of the representation objective relative to the closeness objective.

B.3 An extension of the main results in Theorems 1 and 2 to a setting with geography

Theorems 1 and 2 imply that, for a setting without geography, there exists a voter assignment $\sigma^* \in \Sigma$ that protects each party against a stolen majority. By these Theorems, σ^* moreover admits a decentralization through a game of competitive gerrymandering on the assumption that the parties' sole objective is to win a majority in the legislature. These strategies can be extended to a setting with an explicit geography using the equivalence class of voter assignments $C(\sigma^*)$ introduced above. The assignments in $C(\sigma^*)$ all have the property that, in every state of the world ω, the party that wins the popular vote wins a majority in the legislature. This implies that these assignments are optimal for a degenerate loss function that puts all the weight on the representation objective. We summarize these observations in the following Proposition.

Proposition 2 $C(\sigma^*) \subset \Sigma_l^*(1)$.

As our discussion above has shown, when the closeness objective receives positive weight, then the assignments in $C(\sigma^*)$ remain optimal in the case of an ideal spatial distribution of voters. Empirically, the ideal case is not the relevant one. This raises the questions (i) what an optimal assignment looks like when there is a non-trivial tradeoff between the closeness and the representation objective and (ii) how close one can get to such an optimum by competitive gerrymandering. We leave these questions to future research.