Competitive Gerrymandering and the Popular Vote[*]

Felix J. Bierbrauer[†] Mattias Polborn[‡]

October 1, 2021

Abstract

Gerrymandering undermines representative democracy by creating many uncompetitive legislative districts, and generating the very real possibility that a party that wins a clear majority of the popular vote does not win a majority of districts. We present a new approach to the determination of electoral districts. We show that there is a dynamic redistricting game, played between two parties who both seek an advantage in upcoming elections, so that every equilibrium of this game implements the popular vote.

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

[*] We benefitted from conversations with Peter Cramton, Tigran Polborn, Axel Ockenfels, 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

Partisan gerrymandering has received much attention by both economists and political scientists. Gerrymandering undermines representative democracy by creating many uncompetitive legislative districts, and generating the very real possibility that a party that wins a clear majority of the popular vote does not win a majority of districts. For example, Republicans took over the Pennsylvania state legislature in the 2010 Republican wave election and used the opportunity to create a district map that is very favorable to them. Even though Democratic candidates received 55 percent of the popular vote in the 2018 elections across all districts, versus 44.4% for Republican candidates, Republicans still controlled 110 out of 203 seats in the Pennsylvania House of Representatives. Many other examples exist, including ones in which the partisan advantage was on the Democrats' side.[1]

Because of these problems, there is substantial backlash against existing gerrymanders and also the institutions that allow for it to happen. District assignments engineered by legislatures can be challenged in courts, and some state supreme courts have granted injunctive relief against maps considered to be so unfair that they violate democratic principles in the respective state's constitution. However, in the 2004 *Vieth v. Jubilirer* decision, the US Supreme Court has refused to rule against partisan gerrymanders, arguing that "partisan gerrymandering claims were nonjusticiable because there was no discernible and manageable standard for adjudicating political gerrymandering claims."[2]

Any "ideal" measure of gerrymandering faces the problem of drawing a necessarily somewhat arbitrary line between "still legal" and "sufficiently outrageous to be illegal." In this paper, we therefore explore an alternative approach: Can we find rules for redistricting that yield a desirable outcome, simply by relying on the self-interest of the parties participating in this process? This is similar in spirit to the classical problem of how to fairly divide a cake between two children – we let one child cut the cake in two pieces and the other one choose which one she wants to have. This participation of two self-interested and antagonistic agents is more likely to lead to a fair outcome than the alternative of devising general rules and constraints under which only one child chooses both their own and the other child's piece.

[1] At the federal level, McCarty et al. (2009) show that gerrymandering has increased the Republican seat share in the House of Representatives.

[2] See `https://en.wikipedia.org/wiki/Vieth_v._Jubelirer`, and also the US Supreme Court *Gill v. Whiford* and *Benisek v. Lamone* decisions in 2018, upholding *Vieth.*

Theoretical models of gerrymandering frequently invoke the following setup: There are two parties D and R, voters, and a set of districts. Voters differ in their probabilities to vote for either party, and these probabilities are known by the parties. One party is in control of assigning voters to districts[3] and does so to maximize its advantage in future elections. The characterization of the optimal partisan gerrymander often invokes the notions of *packing* (concentrating one's opponent's supporters in few districts so that the opponent has a smaller vote share in the remaining districts) and *cracking* (spreading the opponent's supporters evenly over the remaining districts, together with a majority of the gerrymanderer's own supporters, so that these districts are very likely won by the gerrymanderer's party).

Our analysis is also based on this "canonical" setting, but we are asking a different question: Is it possible to neutralize the distortions due to partisan gerrymandering by having both parties participate in the redistricting process. More specifically, can we design a redistricting game so that in equilibrium , the party that wins the popular vote also wins a majority of seats in the legislature? The answer is "yes." We specify a dynamic game in which parties take turns in assigning voters to districts. We prove two Theorems. Theorem 1 shows that, when the number of rounds is large, any equilibrium of this game implements the popular vote, i.e., the party that wins the popular vote also wins a majority of districts. Theorem 2, moreover, shows that almost every district is competitive, in the sense of not being distorted away from the popular vote in the entire polity.

To prove our results, we derive bounds on equilibrium payoffs. These bounds follow from an analysis of a sequential strategy for cracking and packing which we refer to as a *water-level-and-towers* strategy:[4] Whenever a party is called upon to play, for every district, there is a particular mix of voters inherited from the previous rounds of play. A party can then order districts according to how favorable they are for itself. A water-strategy assigns more opponent supporters to the most favorable half of districts, to the effect that they all end up being equally good – in the sense of having a common and rather high probability of being won. A tower strategy, in contrast, concentrates

[3] In most U.S. states, state legislatures are charged with drawing up new electoral district maps (for both state and federal races) after each decennial census. In a few states, there are special "non-partisan" commissions formed for this purpose.

[4] As will become clear, the terms *water-level* and *towers* are inspired by a graphical representation of this strategy by means of a bar diagram, where the height of a bar represents how favorable a district is.

opponent supporters in the least favorable districts (i.e., those that already have a high concentration of such voters). We show that a skillful combination of water and tower strategies guarantees, for each party, that it wins a majority of districts whenever it has majority support in the electorate at large. Since any equilibrium strategy must do at least as well as this particular strategy, in any equilibrium, each party wins the election whenever it wins the popular vote.

As mentioned above, Theorems 1 and 2 are based on a model in which voters are distinguished only by how likely they are to vote for either party. When this framework is used for an analysis of partisan gerrymandering, the success of a strategy is measured by the extent to which a party can win a majority of districts even though it is not supported by a majority of the electorate. Thus, in the context of this formal framework, an alignment of the overall election outcome with the popular vote is a natural benchmark for successful institution design. Our main results show, moreover, that this benchmark can be reached.

While this possibility result is based on a sequential game with a particular protocol, the protocol does not have be taken literally as a specific proposal for how redistricting should be done in practice. It is of theoretical value in that it provides an upper bound for what is in principle achievable when the rules governing the redistricting process are well designed. Presumably, there are other protocols that also implement the popular vote. 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 there is no possibility for the other party to interfere, there is also no hope to implement the popular vote.

We proceed as follows. Section 2 reviews related literature. Our model and the main results are in Section 3. Section 4 presents several extensions, as well as a discussion of the results. Formal statements are proved in the Appendix.

2 Related Literature

There is a very large literature on gerrymandering, both empirical and theoretical, which we will review below. However, most of the existing theoretical literature is on "optimal" gerrymandering from the point of view of the party that is in control of the gerrymandering process; that is, how to cheat democracy most effectively if given the opportunity to cheat.

There are, to our knowledge, only a couple of papers that deal with the question of how one could implement a better redistricting system. The earliest such paper is William Vickrey's 1961 paper that argues that "the process [of redistricting] should be completely mechanical so that, once set up, there is no room at all for human choice." He proceeds to propose an algorithm that produces geographically-compact districts, but does not study whether elections governed by the generated map have any desirable properties beyond the fact that districts look natural.

The fundamental idea at the core of our paper – letting parties control each other, rather than putting restrictions on partisan gerrymandering – also features prominently in recent work by Ely (2019). However, our models differ drastically in both the implementation of this central idea, and in the objective pursued, and are therefore best seen as complementary.

While the key objective in Ely (2019) is to prevent "obviously abusive" districts (such as PA-7, the famous "Goofy kicking Donald Duck" district),[5] the majority party is still able to select the map that is best for their winning probability, conditional on only drawing convex districts.

In contrast to Ely (and like most of the literature on partisan gerrymandering), our paper neglects the geographic features of maps and instead focuses on the effects of the district map on the identity of the party that wins a majority in the legislature. However, in Section 4.2, we discuss the extent to which our analysis carries over to a richer model with a spatial distribution of voters.

A different normative perspective on gerrymandering can be found in Coate and Knight (2007). The focus is not on the implementability of the popular vote, but of outcomes that are optimal from a utilitarian perspective. Specifically, they define a seat-vote curve (mapping vote shares into seats for each party) and ask whether a benevolent social redistricting planner can implement an optimal seat-vote curve through the appropriate choice of a district map.[6]

[5] Specifically, in Eli's mechanism, the majority party proposes a partition of a state into districts. The minority party then can either accept the partition, or undo any partisan disadvantage caused by "irregular" boundaries (i.e., those resulting in nonconvex districts). Thus, the mechanism ensures that, to the extent that irregular districts result from the process, the minority is never harmed by them.

[6] Like in our paper, there are no geographic constraints in Coate and Knight (2007). A difference between our notion of welfare and theirs is that we assume that what matters for voters is the identity of the majority party. In contrast, Coate and Knight (2007) assume that the seat share in the legislature matters for voters (i.e., moderate voters prefer an (almost) evenly split legislature over one in which

Our objective is to find a game so that *every* equilibrium implements the popular vote, as opposed to the more modest objective of showing that there is some game with some equilibrium that implements the popular vote, see Jackson (2001) for a discussion. Observe that this problem has also a trivial "solution," namely to have proportional representation of parties based on a single national district. However, such a "solution" would eliminate the connection between local constituencies and their representatives, whose legitimacy is based on majority support in that district. So, we take as given that there have to be many districts and that all voters have to be assigned to a new district map from time to time. Working with these predetermined institutional constraints is a similarity to many papers on market design.[7]

Our paper develops a redistricting institution which is similar to a dynamic Colonel Blotto or divide-the-dollar game. In such games, players have an endowment with soldiers or money that needs to be assigned to battlefields or voters in an attempt to win a war or an election. In our setup, parties have endowments with voters of various types and assign them to districts with the objective of winning a majority of them. Applications of static divide-the-dollar or Colonel Blotto games, include, 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.[8]

Related to our paper is also most of the existing theoretical literature on optimal partisan gerrymandering. Here the existing redistricting institution is taken as given and the analysis focusses on how the party in control of redistricting optimally exerts their power to gerrymander. The initial paper analyzing how an optimal gerrymander involves "packing" and "cracking" is Owen and Grofman (1988). For an excellent review of this literature in a very general framework, see Kolotilin and Wolitzky (2020). Other papers in this line of work include Friedman and Holden (2008), who study optimal partisan gerrymandering with noisy signals about voters' party preferences, and Gul

the majority party has a more substantial majority), even though all legislators of each party are homogeneous in their model and so implemented policy is independent of the size of the majority.

[7] See Roth (2002) for an outline of the market design agenda.

[8] 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.

and Pesendorfer (2010), who analyze partisan gerrymandering when each party has some territory that it controls (as in U.S. House gerrymandering).

Apart from the obvious representation problem resulting from the disconnect between the political preference of the majority of the electorate and the election outcome, gerrymandering can also lead to insufficient incentives for good behavior, both in terms of valence provision and in terms of positioning (See, e.g., Callander (2005), Van Weelden (2015) and Krasa and Polborn (2018)).

3 The Model

There are $2N$ districts, indexed by $k \in \{1, 2, \ldots, 2N\}$ and an at-large district. The electorate consists of voters who always vote Republican (R partisans), voters who always vote Democrat (D partisans) and independent voters. The mass of Republican partisans, Democrat partisans and independent voters in the electorate at large is, respectively, given by

$$b_R = 2N\,\beta_R\ , \quad b_D = 2N\,\beta_D\ , \quad b_I = 2N\,\beta_I\ , \quad \text{where} \quad \beta_R + \beta_D + \beta_I = 1\ .$$

We assume, without loss of generality, that $\beta_R \leq \beta_D$. We also assume $\beta_D \leq \frac{1}{2}$, which is required so that it is uncertain which party wins the popular vote.

Let p_R be the probability that an independent votes for the Republicans and p_D the probability that she votes for the Democrats. We denote the difference of these probabilities by $\omega = p_R - p_D$. Thus, $\omega \in \Omega = [-1, 1]$. With an appeal to a law of large numbers for large economies, we interpret p_R and p_D also as the fraction of independents voting, respectively, for Republicans and Democrats. Consequently, ω is the Republican's margin of victory in the pool of independent voters. In the following, ω is taken to be the realization of a random variable with *cdf* denoted by F. We assume F to be continuous.

Popular vote. We denote the set of states in which party D or party R wins the popular vote, respectively, by

$$\Omega_D := \left\{\omega : \omega < \frac{\beta_D - \beta_R}{\beta_I}\right\} \quad \text{and} \quad \Omega_R := \left\{\omega : \omega > \frac{\beta_D - \beta_R}{\beta_I}\right\}\ .$$

The probability that party D wins the popular vote is denoted by

$$\pi_D^* = \text{pr}(\omega \in \Omega^D) = F\left(\frac{\beta_D - \beta_R}{\beta_I}\right)\ .$$

District outcomes. As we describe in more detail below, we consider games so that voters are allocated to districts over various rounds. In such a process, both parties $P \in \{D, R\}$ send voters to any district k. There is then a particular mix of Republican partisans, Democratic partisans and independent voters in that mass of voters. More formally, a strategy for party D is a collection $\sigma_D = (\sigma_{Dk})_{k=1}^{2N}$, and a strategy for party R is a collection $\sigma_R = (\sigma_{Rk})_{k=1}^{2N}$. In this collection,

$$\sigma_{Dk} = (\sigma_{Dk}^D, \sigma_{Dk}^R, \sigma_{Dk}^I) \quad \text{with} \quad \sigma_{Dk}^D + \sigma_{Dk}^R + \sigma_{Dk}^I = 1 \ ,$$

is the assignment of party D for district k, consisting of the shares of D partisans (σ_{Dk}^D), R partisans (σ_{Dk}^R)and independent voters (σ_{Dk}^I). We use analogous notation for party R. Party D wins district k if

$$\sigma_{Dk}^D + \sigma_{Rk}^D \geq \sigma_{Dk}^R + \sigma_{Rk}^R + \omega \left(\sigma_{Dk}^I + \sigma_{Rk}^I\right) \ , \tag{1}$$

or, equivalently, if

$$\frac{\sigma_{Dk}^D + \sigma_{Rk}^D - \left(\sigma_{Dk}^R + \sigma_{Rk}^R\right)}{\sigma_{Dk}^I + \sigma_{Rk}^I} \geq \omega \ . \tag{2}$$

Hence, the probability that party D wins district k is

$$\pi_{Dk} = F\left(\frac{\sigma_{Dk}^D + \sigma_{Rk}^D - \left(\sigma_{Dk}^R + \sigma_{Rk}^R\right)}{\sigma_{Dk}^I + \sigma_{Rk}^I}\right) \ .$$

and the probability that party R wins district k is $\pi_{Rk} = 1 - \pi_{Dk}$. In the following, when we seek to emphasize the dependence of winning probabilities on the parties' strategies, we write $\pi_{Dk}(\sigma_D, \sigma_R)$.

If $\pi_{Dk} = \pi_D^*$, for all districts k, then the popular vote determines outcomes both at the local district level and at the aggregate state or national level. This avoids constellations so that one party wins the popular vote and the other party wins a majority of seats. Theorem 2 below shows that, for N large we can come very close to the ideal of all districts replicating the at-large-district.

Winning the election. A party wins the election if it wins a majority of the seats. Recall that there $2N$ districts and an at-large-district. Thus, there are $2N+1$ seats in total and winning a majority requires to win at least $N+1$ of them. We denote by π_D^V, the probability of such a Victory for party D. We define π_R^V analogously. Our focus is on whether we can ensure an alignment of the party that wins a majority of seats with the party that wins the popular vote.

For a formal treatment of this question the following notation proves helpful. Given a pair of strategies (σ_D, σ_R), we denote the probability of a victory for party R, conditional on party R winning the popular vote, by $\Pi_R^V(\sigma_D, \sigma_R \mid \omega \in \Omega_R)$. A system that guarantees the "correct" outcome has

$$\Pi_R^V(\sigma_D, \sigma_R \mid \omega \in \Omega_R) = 1 \quad \text{and} \quad \Pi_D^V(\sigma_D, \sigma_R \mid \omega \in \Omega_D) = 1 \ ,$$

for every pair of equilibrium strategies (σ_D, σ_R).

Our main result in Theorem 1 shows that we can indeed achieve this outcome through a sequential mechanism in which parties assign voters to districts over many rounds, alternating which party moves first and which one moves second. We now turn to formal description of this sequential protocol.

3.1 The protocol

Each party assigns every voter to one of the districts. As a consequence, any one voter is assigned twice, once by D and once by R. If a voter is assigned to district k by party D and to some other district $k' \neq k$ by party R, he simply casts one vote in each district election. If $k' = k$ (i.e., both parties assign the voter to the same district), then his vote is counted twice in that district.

Sequence of moves. Voters are assigned to districts over L rounds. In each round l, any party P specifies $\sigma_{Pl} = (\sigma_{kPl}^D, \sigma_{kPl}^R, \sigma_{kPl}^I)_{k=1}^{2N}$ so that

$$\sigma_{kPl}^D + \sigma_{kPl}^R + \sigma_{kPl}^I = \frac{1}{L} \ .$$

In words: Party P assigns a mass of $\frac{1}{L}$ voters to any one district k. The percentage shares of D partisans, R partisans and independents in that mass of voters are then, respectively, given by

$$\beta_{kPl}^D \ := \ L\,\sigma_{kPl}^D \ , \quad \beta_{kPl}^R \ := \ L\,\sigma_{kPl}^R \quad \text{and} \quad \beta_{kPl}^I \ := \ L\,\sigma_{kPl}^I \ .$$

For concreteness, we assume that, for l odd, R moves first and D second. For l even, D moves first and R second. Thus, the second-mover advantage, if any, alternates. D has this advantage in odd rounds and R has it in even rounds.

Feasibility. Let the total number of D partisans assigned by party P to district k over L rounds be denoted

$$\sigma_{Pk}^D := \sum_{l=1}^{L} \sigma_{Pkl}^D .$$

Analogously, let

$$\sigma_{Pk}^{R} := \sum_{l=1}^{L} \sigma_{Pkl}^{R} \quad \text{and} \quad \sigma_{Pk}^{I} := \sum_{l=1}^{L} \sigma_{Pkl}^{I} \; .$$

For any party P, $(\sigma_{Pk})_{k=1}^{2N}$ has to be consistent with the distribution of voters in the electorate at large, i.e.,

$$\frac{1}{2N} \sum_{k=1}^{2N} \sigma_{Pk}^{D} = \beta_D \; , \quad \frac{1}{2N} \sum_{k=1}^{2N} \sigma_{Pk}^{R} = \beta_R \; , \quad \text{and} \quad \frac{1}{2N} \sum_{k=1}^{2N} \sigma_{Pk}^{I} = \beta_I \; .$$

Winning probabilities. Winning probabilities for specific districts or for a majority of seats depend on the number of rounds L. We use superscript L to indicate this dependence. For instance, we write π_{Dk}^{L} for the probability that party D wins district k when there are L rounds of play, or π_{R}^{VL} for the probability that party R wins a majority of seats when there are L rounds of play.

3.2 The main results

Our main result, Theorem 1, shows that, with a sufficiently large number of rounds, every equilibrium is such that the "correct" party wins, namely the one that wins the popular vote.

Theorem 1 *For all $\varepsilon > 0$, there is $\hat{L}$, so that, for $L \geq \hat{L}$, in every equilibrium (σ_D, σ_R),*

$$\Pi_{R}^{VL}\left(\sigma_D, \sigma_R \mid \omega \in \Omega_R\right) \geq 1 - \varepsilon \quad \textit{and} \quad \Pi_{D}^{VL}\left(\sigma_D, \sigma_R \mid \omega \in \Omega_D\right) \geq 1 - \varepsilon \; .$$

Theorem 2 complements this finding: It shows that it is possible to achieve this outcome with only small distortions at the district level. With many districts, i.e. for $N \to \infty$, there is an equilibrium, so that almost every district is a replica of the at-large-district.

Theorem 2 *For all $\varepsilon > 0$ and all $\delta > 0$, there is $\hat{L}$ so that for $L \geq \hat{L}$, there exists a pair of strategies (σ_D, σ_R), so that*

$$\Pi_{R}^{VL}\left(\sigma_D, \sigma_R \mid \omega \in \Omega_R\right) \geq 1 - \varepsilon \quad \textit{and} \quad \Pi_{D}^{VL}\left(\sigma_D, \sigma_R \mid \omega \in \Omega_D\right) \geq 1 - \varepsilon \; ,$$

and

$$\#\left\{k : \; \mid \frac{\sigma_{Dk}^{D} + \sigma_{Rk}^{D} - (\sigma_{Dk}^{R} + \sigma_{Rk}^{R})}{\sigma_{Dk}^{I} + \sigma_{Rk}^{I}} - \frac{\beta_D - \beta_R}{\beta_I} \mid \geq \delta\right\} \frac{1}{2N} \quad \leq \quad \frac{2}{N} \; .$$

Formal proofs of Theorems 1 and 2 are in the Appendix. Note that Theorem 1 is a statement about *all* equilibrium strategies, whereas Theorem 2 is a statement about *one* pair of strategies that approximates a particular equilibrium for L large.

The key insights that carry these proofs are Propositions 1 and 2 below. Proposition 1 shows that party D can ensure to win whenever $\omega \in \Omega_D$ by spreading its partisans supporters evenly over at least fifty percent of the districts. Proposition 2 deals with the challenge for party R to counter this strategy in such a way that it wins a majority whenever $\omega \in \Omega^R$. We show that party R can achieve this outcome by a strategy of spreading D partisans over the complementary districts.

3.3 Why are the Theorems true?

In the following, we first focus on the case $\beta_D > 0$ and $\beta_R = 0$, and subsequently extend the analysis to constellations with $0 < \beta_R \leq \beta_D$. We show that either party can unilaterally ensure to win a majority of districts with a probability at least as large as its probability of winning the popular vote.

While there may be some districts that are uncompetitive (e.g., won by party D, independent of the state of the world), for N and L large, as we will show, there is an equilibrium in which the share of such districts is small.

Recall that, if party D wins N local districts and the popular vote (so that it also wins the at-large district), then it wins a strict majority of seats. The following Proposition 1 shows that party D can indeed ensure to win N districts in all states of the world ω in which it wins the popular vote. Put differently, party D can ensure to win the overall election in all states of the world where it "should" win.

Proposition 1 *Suppose that $\beta_R = 0$ and $\beta_D > 0$. For every L, there is σ_D so that, for every σ_R, $\Pi_D^{VL}(\sigma_D, \sigma_R \mid \omega \in \Omega_D) = 1$.*

Proposition1 follows from a simple argument. Suppose that party D assigns, over the course of the whole game, a mass of $2\beta_D$ partisan D voters to half of the districts, say, to any district with an index k larger or equal to $N + 1$. Then, whatever, the strategy of party R, the percentage share of partisan D voters in those district is bounded from below by β_D. Equivalently, all such districts are won whenever the state ω is such that $\beta_D \geq \omega\ \beta_I$. Hence, $\omega \in \Omega^D$ implies that party D wins at least fifty percent of all districts. The at-large district then ensures a majority of seats for party D. Also note that this conclusion does not depend on the assumption that the number of rounds L is large.

For party R it is more challenging to ensure a victory whenever it wins the popular vote. Remember from the optimal partisan gerrymandering literature that, when party R was unilaterally in control of gerrymandering, it could simply pack all partisan D voters in a small subset of districts, and, as a consequence, have a large number of districts with only independent voters. It would then win all of the latter whenever independent voters lean towards party R, i.e., whenever $\omega > 0$. Thus, party R would win in states $\omega \in (0, \frac{\beta_D - \beta_R}{\beta_I})$, i.e. in states in which it does not win the popular vote.

However, trying to play such a packing strategy in our game would be a severe mistake for party R, as party D would be able to respond in a way that is detrimental of R. Specifically, if party R engaged in a packing of partisan D voters in a small number t of districts, party D would not allocate any partisan D voters from its budget to these districts, but would rather use them uniformly in $N + 1 - t$ other districts. By construction, these districts are more Democratic than the state at-large, and winning them will win the election for D. Thus, attempting to pack D-partisans in a few districts would backfire for party R.

Proposition 2 shows that, for L large, party R can overcome these difficulties. By using a different strategy, it can ensure to win a majority of districts whenever it wins the popular vote.

Proposition 2 *Suppose that $\beta_R = 0$ and $\beta_D > 0$. For every $\varepsilon > 0$, there is $\hat{L}$ so that $L \geq \hat{L}$ implies the existence of a strategy σ_R so that, for all σ_D,*

$$\Pi_R^{VL}(\sigma_D, \sigma_R \mid \omega \in \Omega_R) \geq 1 - \varepsilon \ .$$

We now explain the logic of the proof. Our first observation is that we can, without loss of generality, fix the *ranking* of districts such that lower ranked districts have a weakly lower share of D partisans. All that matters in any subgame is the current content of each district, and the compositions of the two parties' remaining budget sets. Thus, if, at any move, the ranking of districts were to change, there is an equivalent move that produces the same ranking of districts as before the move.[9] Consequently, we can at the beginning of the game, impose an arbitrary ranking of districts and focus on

[9] For example, suppose that, before the move, district 1 has 1 D partisan and district 2 has 2, but then the player who moves adds 4 D partisans to district 1 and only 1 D partisan to district 2, so that the ranking of districts changes (district A now has 5, and B 3 D partisans). However, an alternative move,

the game being played in such a way that districts with lower numbers have a (weakly) lower share of Democratic partisans.

For party R to secure a majority whenever $\omega \in \Omega_R$, it needs to ensure that there are at least N districts so that, after L rounds of play, the percentage share of D partisans is not higher than β^D. Thus, the objective of party R is to minimize, and party D's objective is to maximize, the share of D partisans in district N.

Since party D seeks to maximize the share of D partisans in that district, it will not waste partisan D voters in lower ranked ones. Thus, party D concentrates partisan D voters in the $N+1$ top-ranked districts. More specifically, whenever it is called upon to play in some round l, and plans to assign a certain mass of D partisans, the following pecking order is optimal: Assign D voters to the district with rank N until its mass of D partisans is equal to the one in the district with rank $N+1$. From that point on, keep these two districts at a joint level and add further D partisans until this joint level equals the one in the district with rank $N+2$. From then on, the districts with ranks N, $N+1$ and $N+2$ are raised to the level of district $N+3$ and so on, until no further D partisans are left, see Figures 1 and 2 for an illustration.

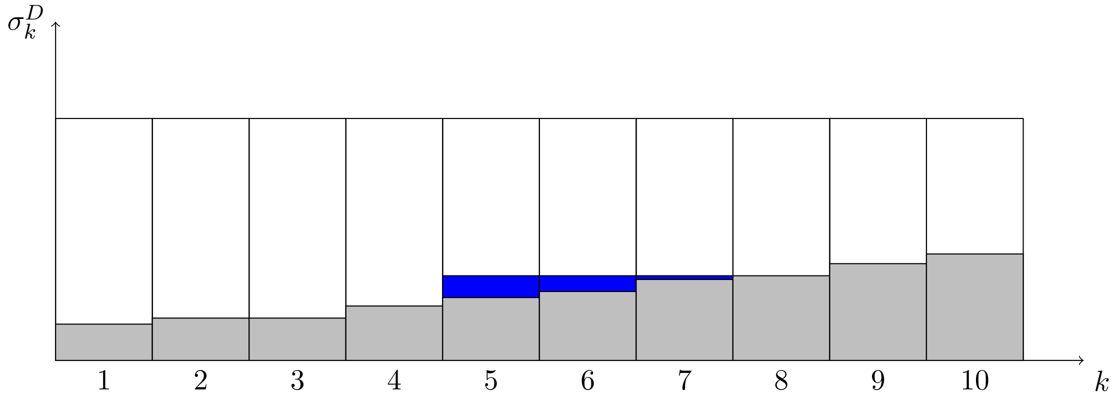


Figure 1: 10 Districts, $0 = \beta_R < \beta_D$. In round l, party D inherits, for every district, a stock of D partisans, illustrated in gray. It then adds further D partisans in round l, illustrated in blue. This figure is drawn under the assumption that party D assigns only few partisan D voters in round l, so that, when assigning them optimally, its budget allows to raise the level of partisan D voters only in districts 5,6, and 7.

What is an optimal response for party R? Its problem is to dispose of a total mass of $2N\beta^D$ partisan D voters in such a way that they contribute as little as possible to

adding 2 D partisans to district-1 and 3 D partisans to district 2, leads to an equivalent distribution of partisans over districts, and uses the same number of D partisans (5) and therefore leads to the same budget set, while preserving the initial ranking.

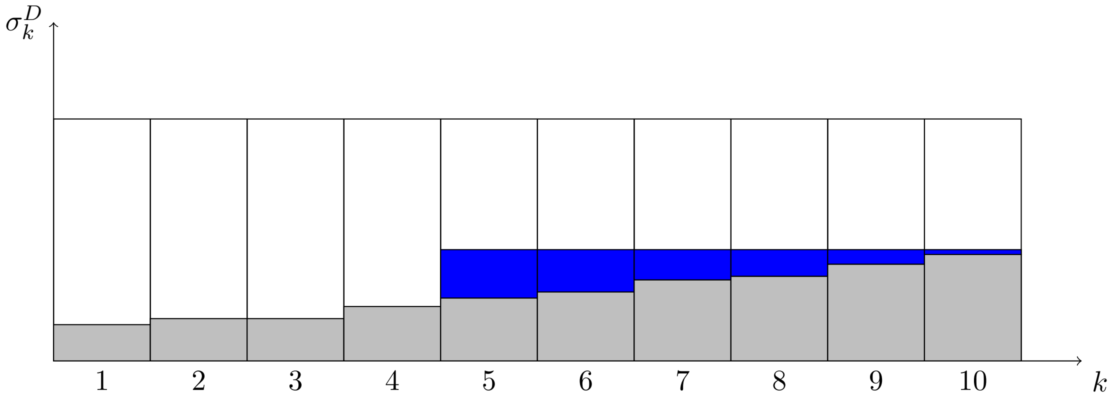


Figure 2: 10 Districts, $0 = \beta_R < \beta_D$. In round l, party D inherits, for every district, a stock of D partisans, illustrated in gray. It then adds further D partisans in round l, illustrated in blue. This figure is drawn under the assumption that party D assigns many partisan D voters in round l, so that, when assigning them optimally, its budget allows to raise the level of partisan D voters in all districts with a rank weakly larger than 5.

the mass of partisan D voters in district N. What is clearly harmless is to add partisan D 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 plans to assign some mass of partisan D voters in some round, it will first fill the bottom $N-1$ districts up to the point where a common level of D partisans is reached in the bottom N districts. This ensures a minimal level of partisan D voters in all districts; see Figure 3 for an illustration under the assumption that the mass of partisan D voters assigned in round l does not suffice to bring the bottom 4 districts to the level of district 5. Figure 4 is based on the alternative assumption that the mass exceeds what would be needed for that purpose.

Figure 4 illustrates the following logic: When further D partisans 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, it is optimal to assign D partisans, starting with the top-ranked district. If the capacity constraint of $\frac{1}{L}$ for that district in that round is reached, party R starts to fill the district with the second highest rank, and so on. Thus, party R concentrates on the top-ranked districts when assigning D partisans.

Party R discards the extra D partisans in very few districts in order to make it as difficult as possible for party D to "use" these partisan voters in an attempt to raise the D content of the median district. To see intuitively why the distribution over non-median districts matters at all, suppose instead that party R distributes the D partisans

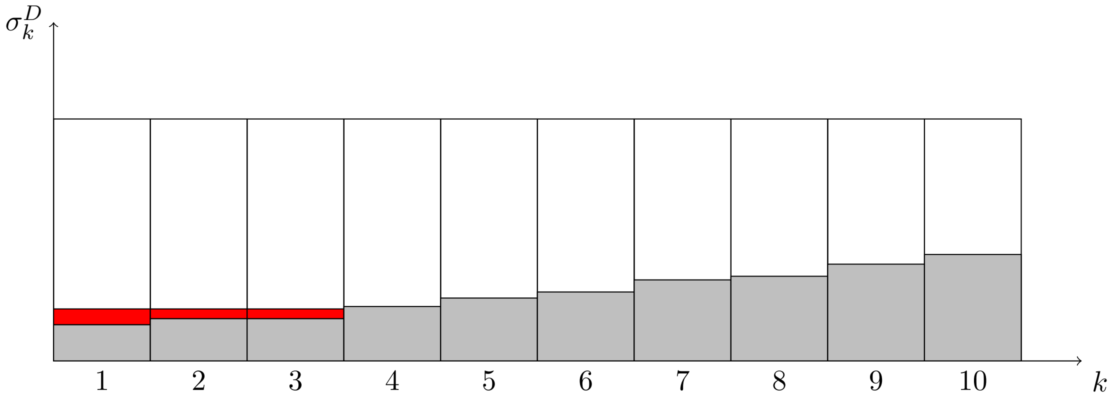


Figure 3: 10 Districts, $0 = \beta_R < \beta_D$. In round l, party R inherits, for every district, a stock of D partisans, illustrated in gray. It then adds further D partisans in round l, illustrated in red. This figure is drawn under the assumption that party R assigns few partisan D voters in round l, so that, when assigning them optimally, the level in the districts with a rank below 5 cannot be raised to the level in the district with rank 5.

uniformly over districts $N+2$ to $2N$. That makes it easier in the next round for party D to raise the D content of district $N+1$: Remember that, when district $N+1$ reaches the level of district $N+2$, party D needs to allocate D partisans to both of these districts in order to avoid a district rank reversal. By allocating D partisans to the highest-ranked districts, R can insure that this rank-reversal constraint for party D kicks in as early as possible.

Using a more colorful language, we also refer to R's strategy for the bottom half as a *water-level-strategy*: The level in the basin consisting of the bottom $N-1$ districts is raised to the level prevailing in district N. We will refer to R's strategy for the upper half as *building-towers-strategy.* A tower is a district in the upper half with a level of D partisans that sticks out. If additional D partisans need to be assigned, party R assigns them with priority to the district that sticks out most, i.e. it will make the highest tower even higher. It will then move to the second highest tower, and so on.

For a complete equilibrium characterization, we would also need to describe how many D partisans are assigned by whom and when, i.e. we would need to characterize, for any party P and any round l the equilibrium value of β_{Pl}^D, defined as the percentage share of D partisans in the total mass of $\frac{2N}{L}$ voters assigned by party P in round l. We do not provide such a complete equilibrium characterization, but show that party R can choose the sequence $\{\beta_{Pl}^D\}_{l=1}^L$ so that the share of D partisans in district N remains below β_D. To this end, assume that party R chooses $\beta_{R1}^D = 0$, and for any $l \geq 2$,

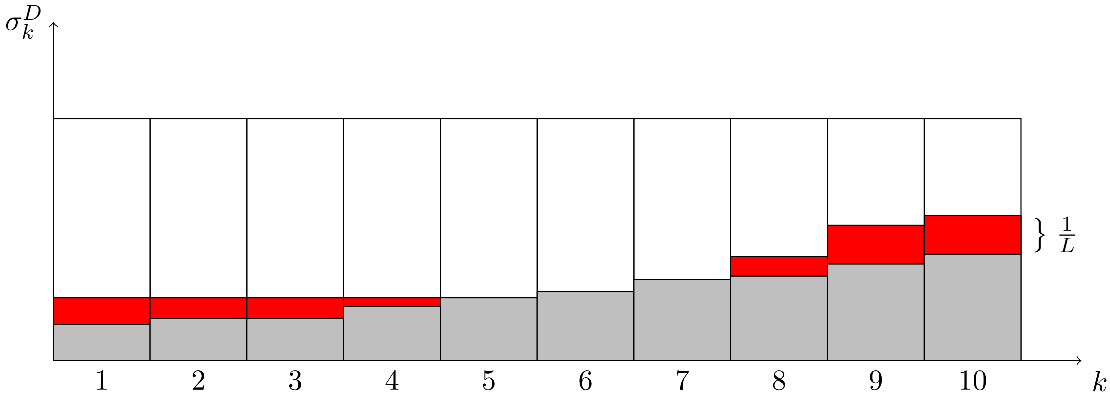


Figure 4: 10 Districts, $0 = \beta_R < \beta_D$. In round l, party R inherits, for every district, a stock of D partisans, illustrated in gray. It then adds further D partisans in round l, illustrated in red. This figure is drawn under the assumption that party R assigns many D partisans in round l, so that, when assigning them optimally, the level in the districts with a rank below 5 is raised to the level in the district with rank 5. Additional D partisans are then assigned to the top-ranked districts.

$\beta_{Rl}^D = \beta_{Dl-1}^D$. Thus, party R waits until party D starts to assign D partisans and then assigns in, any round, as many D partisans as party D assigned in the round before.

Given the partial characterization of equilibrium behavior above, this implies that, after any move of party R, the bottom $2N-2$ districts have the same level of D partisans, while there are some further D partisans in the top ranked district, and, possibly, also in the district with the second highest rank. To see this, suppose for concreteness, that party D chooses $\beta_{D1}^D > 0$. Then, it will spread a mass of $\beta_{D1}^D \frac{2N}{L}$ D partisans evenly over $N+1$ districts. In round 2, party R will use the mass of voters previously assigned to $N-1$ of those districts to have an equal water-level in the bottom half. The remaining mass of D partisans is then assigned to at most two top districts. See Figure 5 for an illustration. This pattern is now repeated over various rounds, with the implication that, after any move of party R there is a joint level of D partisans in the bottom $2N-2$ districts.

It is now easy to see that the percentage share of D partisans in the pivotal district N cannot be strictly above β_D. This would imply a percentage share above β_D in all districts and this is incompatible with the fact that the share if D partisans in the electorate at large equals β_D. Also note that there is a common level of D partisans in all districts, with the possible exception of the two top ranked ones. Thus, party R's has a strategy that ensures winning a majority whenever $\omega \in \Omega_R$, and moreover, implies that there are at most two districts that are "safe" for party D. For $N \to \infty$

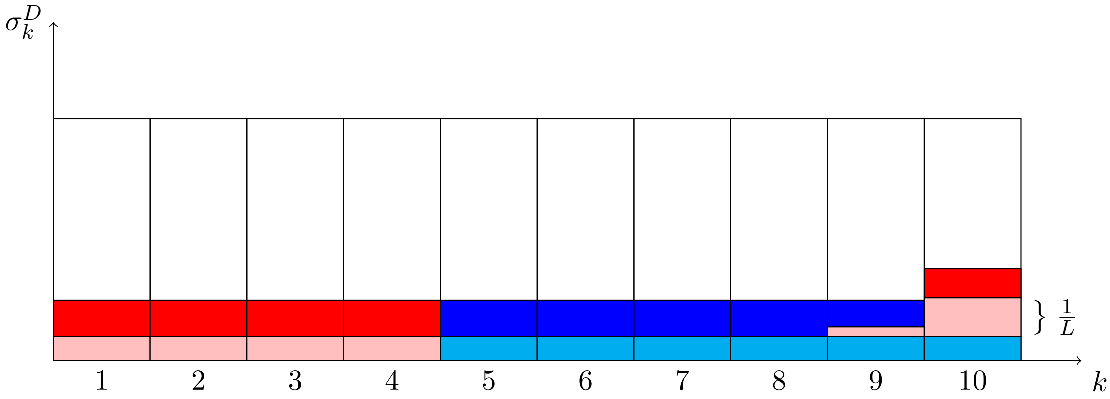


Figure 5: 10 Districts, $0 = \beta_R < \beta_D$. Party R assigns as many D partisans as party D did in the previous round. In light blue is the first round in which party D assigns a positive mass of D partisans. Party R's response is in light red. In blue is the second round in which party D assigns a positive mass of D partisans, and party R's response is in red. As a consequence, there is a common level in the bottom eight districts, both after R's first and second response.

the fraction of districts where the outcome deviates from the popular vote is negligible.

By Proposition 2, and for L large, party R can ensure to win a majority of seats whenever $\omega \in \Omega_R$ and, by Proposition 1, party D can ensure to win whenever $\omega \in \Omega_D$. Thus, for N and L large, a constellation where party D concentrates its supporters in fifty percent of the districts and party R concentrates them in the other half approximates the equilibrium. Also, the share of districts in which outcomes are not approximately equal to the popular vote then becomes negligible.

Relaxing the assumption that $\beta^R = 0$. Proposition 3 below is a generalization of Proposition 2 that allows for the possibility that there are both R partisans and D partisans, but maintains the assumption that there are (weakly) more of the latter.

Proposition 3 *Suppose that $\beta_R < \beta_D$.*

a) For every $\varepsilon > 0$, there is $\hat{L}$ so that $L \geq \hat{L}$ implies the existence of a strategy σ_R so that, for all σ_D,

$$\Pi_R^{VL}(\sigma_D, \sigma_R \mid \omega \in \Omega_R) \geq 1 - \varepsilon \,.$$

b) For every $\varepsilon > 0$, there is $\hat{L}$ so that $L \geq \hat{L}$ implies the existence of a strategy σ_D so that, for all σ_R,

$$\Pi_D^{VL}(\sigma_D, \sigma_R \mid \omega \in \Omega_D) \geq 1 - \varepsilon \,.$$

The key for the proof of Proposition 3 is the insight that each party has an incentive to use its own partisan supporters so that they are spread evenly over fifty percent of the districts. The other party then has an incentive to respond to this attempt using a water-level and building-towers-strategy. When this logic is squared with the assumption that either party assigns in a round l as many rival partisans as the rival party used in the previous round – so that $\beta^D_{Rl} = \beta^D_{Dl-1}$ and $\beta^R_{Dl-1} = \beta^R_{Rl-2}$ – then there are at most 2 safe districts for party D and at most two safe districts for party R. Theorem 2 follows from this last observation.

4 Discussion and Extensions

In Section 4.1, we relate our model to the standard "cracking and packing" terminology in partisans gerrymandering models; discuss whether there is monotone convergence to the limit results in Theorems 1 and 2; and discuss how our results would extent to a generalized setting with more than three different voter types. In Section 4.2, we discuss what happens when we impose geographic constraints, and in Section 4.3, we discuss how our system can be modified if we want to insure a minimum opposition presence in the legislature.

4.1 Theoretical Discussion

Packing and cracking. The terms "packing" and "cracking" are frequently used to describe the optimal plan for a party that can determine districts unilaterally; see, e.g., Owen and Grofman (1988) or Kolotilin and Wolitzky (2020). The party in control assigns its own supporters to $N + 1$ districts in order to win those with a maximal probability (this is referred to as "cracking"), and "packs" the supporters of their opponents into the smaller complementary set of districts.

Propositions 1 - 3 show how this logic extends to a game of competitive districting. Our analysis shows that each party cracks the own partisan supporters. For the partisan supporters of the competing party it uses the "water-level" and "building towers" strategy. The "water level" part deals with the complementary set of districts, left untouched by the rival. If even more partisan supporters of the opponent need to be assigned, there is some extra packing in the two districts that the rival is most likely to win.

Monotone convergence? According to the limit result in Theorem 1, with a large number of rounds, competitive gerrymandering implements the popular vote. This raises the question whether the convergence to this outcome is monotone: Does any increase of the number of rounds L yield a better approximation of the popular vote? The answer is "no". The following example illustrates this: Suppose there are only D partisans and independent voters, also let β_D be "small."

Then, when $L = 1$, party R is called upon to move first in the one and only round. It will then pack the D partisans in a minimal number of, say, m "hopeless" districts. When party D responds, it can take the m packed districts for granted (i.e., will not allocate any partisans from its own budget there), and will crack its budget of D-partisans over $N - m + 1$ districts. Consequently, the content of D-partisans in the pivotal district is $2N\beta_D/2(N-m+1) > \beta_D$, and thus the outcome favors party D that wins the election in some states in which it loses the popular vote.

For $L = 2$, party D (which moves first) can guarantee itself a victory whenever it wins the popular vote, by blocking its partisans in N districts; R, the last mover, will have kept all D-partisans in its budget in their first move, and can then (in the last move) allocate all of them to the other half of the districts. This implements the popular vote and so moving from $L = 1$ to $L = 2$ brings progress.[10]

For $L = 3$, the progress is undone, however. Now party D is the last mover and will not assign any D partisans earlier than that. Party R will therefore in its last move pack D partisans in only few districts. The outcome ultimately is the same as when $L = 1$.

Similar arguments show why the outcome for $L = 4$ equals the one for $L = 2$ and why the outcome for $L = 5$ is equal to the one for $L = 3$, etc. Thus, it is not generally true that adding further rounds makes things better. What is true, by contrast, is that having a sufficiently large number of rounds leads to improvements.

A richer set of voter types Our analysis allows for three voter types: Voters who vote for party D with probability 1, voters who vote for party R with probability 1, and, finally, independent voters who vote for party D with probability $p_D = \frac{1}{2} - \frac{\omega}{2}$, where

[10] Can D do better? The answer is negative. If D attempts to allocate its supporters over $N + 1$ districts instead, the level of D-partisans in those districts is lower than before. Party R is still able to dispose of all D-partisans in its budget set in m districts, which leaves one of the districts only fill by D as the median district. Thus, this outcome is strictly worse for Party D.

ω is the realization of an aggregate shock. We now sketch the lines along which our analysis can be extended to a richer set of voter types.

Consider the following generalization of our setup that follows recent work by Kolotilin and Wolitzky (2020): Voters differ in their type $s \in [0,1]$ and $v(s,\omega)$ is the probability that a type s person votes for party R when the aggregate shock takes the value ω. We also interpret $v(s,\omega)$ as the fraction of type s voters who vote for party R in state ω. The function v is taken to be increasing in both arguments; i.e. higher types are more likely to vote R, and higher states increase the population share of R voters. The remarks that follow suggest that squaring this richer model of voter behavior with our game of competitive districting will ultimately yield conclusions similar to Theorems 1 and 2. Where does this confidence come from? First note that, in the richer setup, a strategy for party P in round l specifies for every district k a function $\sigma_{Pkl} : s \mapsto \sigma_{Pkl}(s)$, with the interpretation that $\sigma_{Pkl}(s)$ is the mass of type s voters that are assigned by party P to district k in round l. Let $\sigma^l_{Dk} = (\sigma_{Dkj})^l_{j=1}$ and $\sigma^l_{Pk}(\sigma_{Dkj})^l_{j=1}$ be the history of play over periods 1 to l for district k. If the game were to end after round l, party R would win district k whenever the state ω is such that

$$\int_0^1 v(s,\omega)\,\frac{L}{2l}\,\left(\Sigma^l_{Dk}(s)+\Sigma^l_{Rk}(s)\right)ds \quad \geq \quad \frac{1}{2}\,, \tag{3}$$

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

Second, note that the assumption that v is increasing in both arguments implies that there is a cutoff value $\hat{\omega}_{kl}(\sigma^l_{Dk},\sigma^l_{Pk})$ so that inequality (3) holds if and only if $\omega \geq \hat{\omega}_{kl}(\sigma^l_{Dk},\sigma^l_{Pk})$. As a consequence, after any round l, we can order districts according to the value of $\hat{\omega}_{kl}$, or, equivalently, according to their probability to turn Republican. With this observation, it is now easy to show how the previous analysis extends: To win a majority, parties need to target the median district. Party D will therefore crack the voter types who increase its winning probabilities over the $N+1$ districts that will most likely turn democrat and use the "water-level and building towers"-strategy for the voter types who increase the winning probability of party R. The strategy of party R is the mirror image. We conjecture that, as a consequence, after L rounds of play, the cutoff values are equalized in almost all district with the implication that the popular vote determines the outcome in almost all districts.[11]

4.2 Geography

Our main results in Theorems 1 and 2 provide an equilibrium characterization in terms of the shares of partisan D and R voters and independents in a district. This leaves a degree of freedom. Typically, there will be different assignments of voters to districts that all give rise to the equilibrium composition of districts. This degree of freedom can be used to deal with further concerns in the determination of district boundaries. For instance, in a recent paper, Ely (2019) has used convexity to formalize a requirement of non-crazy district boundaries. In the following, we discuss how such concerns can be related to our analysis.

Suppose we enrich the characterization of voters in our setup by also endowing them with an address (a location in a two-dimensional plane). Figure 6 provides an exemplary spatial distribution of R voters (red dots), D voters (blue dots) and independent voters (gray dots), in a ratio of 1:2:3. For this particular distribution in space, it is easy to find an assignment of voters to districts so that all districts are convex sets and so that the popular vote is implemented.

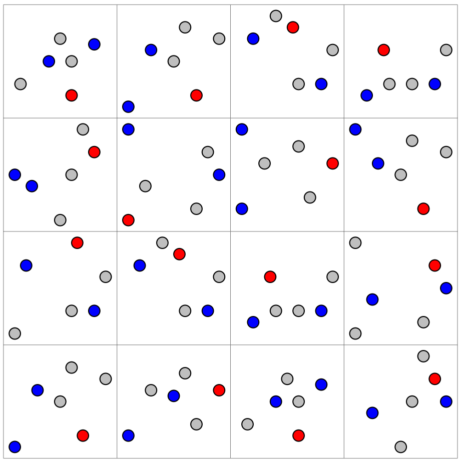
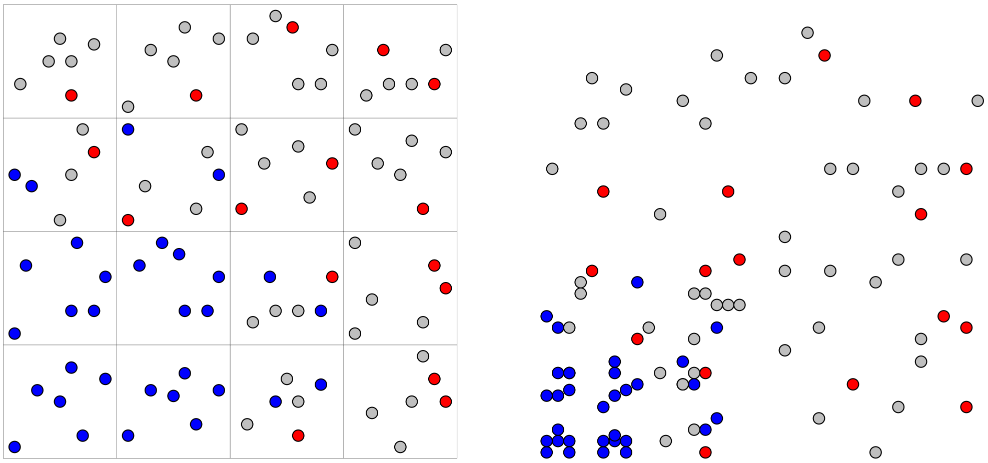


Figure 6: Spatial distribution of voters. Red dots represent R voters, blue dots D voters, and gray dots I voters. Here, voters can be assigned to districts in such a way that (i) districts are convex (ii) the election outcomes in districts are aligned with the popular vote.

When the spatial distribution is more segregated as in Figures 7a (where the population density is uniform across the polity, but addresses and preferences are strongly

[11] These remarks are tentative though. We felt that spelling out the formal arguments would make the paper's length excessive, without really adding a new insight.

correlated) or 7b (where we can think of the concentration in the lower left corner as a city with mostly Democratic voters), then the objectives of generating a map with aesthetically pleasing district shapes, and implementing the popular vote are necessarily in conflict (see also Chen and Rodden (2013) for an argument that this is realistically the case in the United States). So far, the question how to deal with this tradeoff in a satisfactory has not been taken up by the literature on gerrymandering. It is an interesting avenue for future research.



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

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

Figure 7: Redistricting these polities such that the election outcomes in districts are aligned with the popular vote *and* geographical district shapes are aesthetically pleasing is impossible.

4.3 Opposition and minority representation

With the sequential redistricting game that we propose, by Theorem 2, most or all districts may ultimately be replicas of the electorate at-large. Thus, in most states of the world, the majority-preferred party wins a very large percentage of seats, with few or none going to the minority party.

Even though the minority party has very limited influence on which policies are enacted even if it is represented in the legislature, this representation my 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 oversight and information about legislative proposals to the public.

Finally, 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 who do not have to learn from scratch how a legislature works.

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 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. 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.[12]

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

Another conceivable objective is that there is a certain subset of districts whose majority has to be composed of a certain demographic type such as African Americans or Hispanics ("majority-minority districts"). Generally, there is a tension between imposing this constraint and an implementation of the popular vote: If demographic minorities are extremely likely to vote for Democrats, then generating a subset of districts in which minority voters are a majority necessarily implies that the remaining districts have a below-average share of Democratic partisans. Thus the objectives of ensuring a fair election outcome in terms of a correspondence between the outcome of the popular vote, and creating a large set of "majority-minority" districts, may be logically

[12] 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.

incompatible.

This said, with the system of competitive gerrymandering described in the previous section, if a party wants to generate districts that overrepresent certain demographic groups, it can plausibly do so that without negatively impacting its winning probability. For example, suppose that the Democratic party has its core supporters among Blacks and certain urban Whites, while they are weaker among other groups of voters, e.g., rural Whites. How the Democrats mix these voter types into legislative districts is their choice – in particular, it seems well feasible to create some districts in which the Democratic partisans allocated to these districts are predominantly Black, so that they would have a strong influence on the outcome of the Democratic primary.

References

Callander, Steven, "Electoral Competition in Heterogeneous Districts," *Journal of Political Economy*, 2005, *113* (5), 1116–1145.

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.

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

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

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

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

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

Jackson, Matthew, "A crash course in implementation theory," *Social Choice and Welfare*, 2001, *18* (4), 655—708.

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

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.

Krasa, Stefan and Mattias K. Polborn, "Political competition in legislative elections," *American Political Science Review*, 2018, *112* (4), 809–825.

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.

McCarty, Nolan, Keith T Poole, and Howard Rosenthal, "Does gerrymandering cause polarization?," *American Journal of Political Science*, 2009, *53* (3), 666–680.

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

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.

Roth, Alvin E., "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics," *Econometrica*, 2002, *70* (4), 1341–1378.

Van Weelden, Richard, "The welfare implications of electoral polarization," *Social Choice and Welfare*, 2015, *45* (4), 653–686.

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

A Proofs (Online Appendix)

A.1 Proof of Proposition 1

Party D wins the popular vote whenever $\omega < \frac{\beta_D}{\beta_I}$. Hence, to ensure winning a majority of seats whenever $\omega < \frac{\beta_D}{\beta_I}$, party D needs to ensure that there are N districts so that, after L rounds of play,

$$\frac{\sum_{l=1}^{L} \sigma_{kDl}^{D} + \sum_{l=1}^{L} \sigma_{kDl}^{R}}{2 - \sum_{l=1}^{L} \sigma_{kDl}^{D} - \sum_{l=1}^{L} \sigma_{kDl}^{R}} \quad > \quad \omega \, ,$$

whenever $\frac{\beta_D}{\beta_I} > \omega$. Since the distribution F of ω is continuous, the probability of this event (with a strict inequality) is the same as when the inequality holds weakly.

Therefore, $\Pi_D^{VL}(\sigma_D, \sigma_R \mid \omega \in \Omega_D) = 1$ holds when there are N districts so that, after L rounds of play,

$$\frac{\sum_{l=1}^{L} \sigma_{kDl}^{D} + \sum_{l=1}^{L} \sigma_{kDl}^{R}}{2 - \sum_{l=1}^{L} \sigma_{kDl}^{D} - \sum_{l=1}^{L} \sigma_{kDl}^{R}} \quad \geq \quad \frac{\beta_D}{\beta_I} \, .$$

Consider the following strategy for party D: In all rounds l, choose $\sigma_{kDl}^{D} = 0$, for $k \leq N$ and $\sigma_{kDl}^{D} = \frac{2\beta^D}{L}$, for all $k > N$. Consequently,

$$\begin{aligned}
\frac{\sum_{l=1}^{L} \sigma_{kDl}^{D} + \sum_{l=1}^{L} \sigma_{kDl}^{R}}{2 - \sum_{l=1}^{L} \sigma_{kDl}^{D} - \sum_{l=1}^{L} \sigma_{kDl}^{R}} \quad &= \quad \frac{2\beta_D + \sum_{l=1}^{L} \sigma_{kDl}^{R}}{1 - 2\beta_D + 1 - \sum_{l=1}^{L} \sigma_{kDl}^{R}} \\
&\geq \quad \frac{2\beta_D}{2(1-\beta_D)} \\
&= \quad \frac{\beta_D}{\beta_I}
\end{aligned}$$

Thus, whenever D wins the popular vote (i.e., $\beta_D < \omega \ \beta_I$), then D also wins all districts with an index $k \in \{N+1, \ldots, 2N\}$ with probability 1.

A.2 Proof of Proposition 2

Party R wins the popular vote whenever $\omega > \frac{\beta_D}{\beta_I} = \frac{\beta_D}{1-\beta_D}$. To win a majority of seats in all such states, after L rounds of play, there need to be at least N districts with

$$\frac{\sum_{l=1}^{L} \sigma_{kDl}^{D} + \sum_{l=1}^{L} \sigma_{kDl}^{R}}{2 - \sum_{l=1}^{L} \sigma_{kDl}^{D} - \sum_{l=1}^{L} \sigma_{kDl}^{R}} \quad < \quad \frac{\beta_D}{1-\beta_D} \, .$$

Equivalently, there need to be N districts with a percentage share of partisan D voters below β_D. In the following, we show that, for L large, party R indeed has a strategy available that ensures this outcome. The proof will be indirect: We will show that there is no strategy for party D that so that the percentage share of partisan D voters is larger than β_D in at least $N+1$ districts.

District ranks. Parties assign voters to districts over various rounds. We denote by $s^D_{k,l}$ the percentage share of partisan D voters in district k after l rounds of play. We denote the corresponding mass of partisan D voters by $\mu^D_{k,l}$. We will often rank districts according to the share of D voters. Let $r_l(k) \in \{1, \dots, 2N\}$ be the rank of district k after l rounds of play. We assign ranks so that $r_l(k) > r_l(k')$ implies $s^D_{k,l} > s^D_{k',l}$. Hence, the district with largest share of D voters has rank $2N$, the district with the second-largest share has rank $2N - 1$ and so on. The mapping r_l is taken to be injective implying that every rank in $\{1, \dots, 2N\}$ is assigned. Thus, if two districts have the same share of partisan D voters one is (arbitrarily) assigned a higher rank than the other. What matters for the analysis that follows is that, if some district k has rank r, this implies that there are $2N - r$ further districts with a share of at least $s^D_{k,l}$.

Party Objectives. As explained above, we seek to show that there is no strategy for party D that so that the percentage share of partisan D voters is larger than β^D in at least $N + 1$ districts. We therefore assume that it is party D's objective to maximize the percentage share of partisan D voters in the district with rank N after L rounds of play. Specifically, we will show that party R has a strategy under which this percentage share will be (weakly) below β_D, on the assumption that D's objective is to maximize this share. As an implication, party R's strategy implies a percentage share (weakly) below β_D, for any strategy of party D.[13]

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

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

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

[13] This is an implication of the game being zero sum. Any equilibrium strategy of party R 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, the payoff realized by R can only increase.

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 D partisans to the $N-1$ lowest ranked districts until the mass of D voters equals the mass in the district with rank N. From then on, don't add further D partisans to one of the bottom N districts.

iv) Add D voters to the top ranked district.

v) If there are still D voters left in the budget after a mass of $\frac{1}{L}$ D voters has been assigned to the top ranked district, add D 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 D 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 $\{\beta_{Dl}^{D}\}_{l=1}^{L}$.

Note that since party R never affects the ranking of districts, the ranking of districts in any round is entirely due to party D. We now argue that it entails no loss of generality to assume that party D's moves do neither affect the ranking of districts.

To be specific, consider party D's move in a round $l+1$, where l is odd, implying that D moves first in round $l+1$.[14] Suppose that two districts k and k' are such that $\mu_{k,l}^{D} > \mu_{k',l}^{D}$. Also suppose that, after party D's move in round $l+1$, the ranking is reversed, so that $\mu_{k,l+1}^{D} < \mu_{k',l+1}^{D}$. We now argue that an equivalent outcome can be induced without a rank reversal, and with the same implications for the budget of partisan D voters.

- Note first that the the rank reversing move requires a mass of D partisans equal to

$$\sigma_{k,l+1}^{D} + \sigma_{k',l+1}^{D} = \left(\mu_{k,l+1}^{D} - \mu_{k,l}^{D}\right) + \left(\mu_{k',l+1}^{D} - \mu_{k',l}^{D}\right) \tag{4}$$

- Now consider an alternative strategy in round $l+1$, $\bar{\sigma}_{l+1} = (\bar{\sigma}_{k,l+1})_{k=1}^{2N}$ that is the same for all districts, except for the two districts k and k' with the rank reversal.

[14] The same logic applies to party D's moves in odd rounds. Writing this down formally would require some obvious adjustments of notation, taking account of the fact that party D moves second in those rounds.

Under this alternative strategy, the mass of partisan D voters in district k is raised to the high level equal to $\mu^D_{k',l+1}$ and the mass of partisan D voters in district k' is raised to the low level of $\mu^D_{k,l+1}$. Thus, this alternative strategy yields an equivalent outcome as the original strategy: Districts k and k' flip their ranks, but in any case there are $\mu^D_{k',l+1}$ D partisans in the higher ranked district and $\mu^D_{k,l+1}$ D partisans in the lower ranked district.

- The mass of D partisans required under the alternative strategy is

$$\bar{\sigma}^D_{k,l+1} + \bar{\sigma}^D_{k',l+1} = \left(\mu^D_{k',l+1} - \mu^D_{k,l}\right) + \left(\mu^D_{k,l+1} - \mu^D_{k',l}\right), \tag{5}$$

  and therefore equal to the mass required by the rank reversing strategy.

We can therefore assume without loss of generality that, from the initial round onward, party D assigns partisan D voters only to $N+1$ districts. The ranking of these districts can be assumed to remain unchanged throughout the whole game. From now on, we assume for notational ease, that the index k coincides with the ranking of district k, i.e. we let $r_l(k) = k$, for all k and l.

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 more than $N+1$ districts. Thus, we may assume that, for any round l,

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

or, equivalently,

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

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

i) Add partisan D voters to the district with rank N until the mass of D 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 partisan D voters to the two districts with ranks N and $N+1$ until the mass of D 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 partisan D voters, $\beta^D_{R1} = 0$. In any round $l \geq 2$, and as long os this is feasible, party R assigns as many partisan D voters as party D did in the previous round

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

This is clearly feasible in early rounds. If, however, party D keeps some partisan D voters for the last round so that $\beta^D_{DL} > 0$, then party R will have to assign an additional mass of $\beta^D_{DL} \frac{2N}{L}$ late in the game. Otherwise party R would violate its budget constraint. This amount is bounded from above by $\frac{2N}{L}$ and vanishes for L large. Thus, for $L \to \infty$, and given that F is a continuous *cdf*, this will not affect the parties' winning probabilities in any one district.

In the following, we will focus on the limit case $L \to \infty$. For expositional ease, and without loss of generality, we assume that $\beta^D_{DL} = 0$, and that

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

for all $l < L$.

Party R's strategy has the following implication: Whenever party R moves, it brings the mass of D 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 D 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 D voters left. But those fill at most two further districts. According to party R's strategy, as many as possible are assigned to the district with the top rank $2N$. If additional D voters are left, they go to the district with rank $2N-1$.

Now consider rounds 3 and 4:

- In round 3, party D's best response stipulates to assign an equal mass of D voters to the districts with ranks $N, N+1, \ldots, 2N-2$. Those are $N-1$ districts. Possibly, it also assigns D voters to the two 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 added voters to the two top ranked districts, then party R has additional D 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 D voters are left, they go to the district with rank $2N-1$.

Completing the argument. The strategies of parties R and D described above imply that after the last move in round L, there is an equal mass of partisan D 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 D partisans in the district with rank N is strictly larger than β_D. Equivalently, the mass of D voters in that district exceeds 2 β_D. Then, the mass of D voters exceeds 2 β_D in all districts. Hence, the total mass of assigned D voters is strictly larger than $4N\ \beta_D$. But this is infeasible as the two parties' total endowments with partisan D voters only sum to $4N\ \beta_D$. Thus, the assumption that party D can generate $N+1$ districts with a percentage share of partisan D voters strictly larger than β_D leads to a contradiction, and must be false.

A.3 Proof of Proposition 3

We explain how the proofs of Propositions 1 and 2 need to be adapted when there is a non-negligible fraction of partisan R voters.

On the pivotal district. When we seek to show that party D has a strategy that ensures a victory whenever it wins the popular vote, we need to show that party D can ensure to win all districts with a rank larger or equal to $N+1$ whenever $\omega \in \Omega^D$; where ranks, after some round l, now refer to the order of districts according to

$$\Delta^D(k,l) := \frac{\mu^D_{k,l} - \mu^R_{k,l}}{2\frac{l}{L} - (\mu^D_{k,l} + \mu^R_{k,l})}\ .$$

In this expression, $\mu^D_{k,l}$ denotes, as before, the total mass of partisan D voters assigned to district k over the first l rounds of play, and $\mu^R_{k,l}$ is the analogously defined mass of partisan R voters. To show that party D has such a strategy, we assume that party D seeks to maximize $\Delta^D(k,l)$ in the district with rank $N+1$, and that party R seeks to minimize this quantity. This strategy of R is the one that makes it most difficult for party D to achieve, in the district with rank $N+1$, a value of $\Delta^D(k,l)$ that exceeds $\frac{\beta_D-\beta_R}{\beta_I}$, and hence ensures winning a majority of seats whenever $\omega \in \Omega_D$. We thereby construct an equilibrium on the assumption that the pivotal district is the one with rank $N+1$.

When we seek to show that party R has a strategy that ensures a victory whenever it wins the popular vote, we provide an indirect proof. We show that there is a strategy for party R which prevents party D from winning all districts with a rank larger or equal to N whenever $\omega \in \Omega_R$. This analysis amounts to constructing an equilibrium on the assumption that the pivotal district is the one with rank N.

A.3.1 Proof of statement b) in Proposition 3

Party R's strategy. We adapt party R's strategy in the following way: The strategy for the assignment of partisan D voters is the same as in the proof of Proposition 2, except that there is an adjustment for the pivotal district which instead of being the district with rank N is now the district with rank $N+1$. In any round l, given a – for now exogenous – budget of $\beta^D_{R,l}\,\frac{2N}{L}$ partisan D voters to be assigned, proceed sequentially in the following way – until the budget of partisan D voters for that round is exhausted:

i) Add D partisans to the lowest ranked district until $\Delta^D(k,l)$ equals the value for the district with the second lowest rank. From then on, keep $\Delta^D(k,l)$ in these two districts equal.

ii) Add D partisans to the two lowest ranked districts until the joint level of $\Delta^D(k,l)$ equals the value in the district with the third lowest rank. From then on, keep $\Delta^D(k,l)$ in these three districts equal.

iii) Proceed analogously for all districts with a rank smaller or equal $N-1$. From then on, keep $\Delta^D(k,l)$ in all these districts equal. Add D partisans to the Nth lowest ranked districts until the value of $\Delta^D(k,l)$ equals the one for the district with rank $N+1$. From then on, don't add further D partisans to one of the bottom $N+1$ districts.

iv) Add D voters to the top ranked district.

v) If there are still D partisans left in the budget after a mass of $\frac{1}{L}$ D voters has been assigned to the top ranked district, add D 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 D partisans are left.

The assignment of partisan R voters is the mirror image of the assignment of D partisans by party D in the proof of Proposition 2. Party R will focus on bringing down $\Delta^D(k,l)$ in the bottom $N+1$ districts. This also implies that party R will always choose

$$\beta^R_{Rl} \leq \frac{1}{2} + \frac{1}{2N}$$

to avoid having to assign partisan R voters to a district with rank $N+2$ or larger. Now, given a budget of $\beta^R_{R,l}$ $\frac{2N}{L}$ partisan R voters to be assigned, party R proceeds sequentially in the following way – until the budget of partisan R voters for that round is exhausted:

i) Add R partisans to the district with rank $N+1$ until $\Delta^D(k,l)$ falls to the value for the district with rank N. From then on, keep $\Delta^D(k,l)$ in these two districts equal.

ii) Add R partisans to the districts with ranks $N+1$ and N until their joint level of $\Delta^D(k,l)$ equals the value in the district with rank $N-1$. From then on, keep $\Delta^D(k,l)$ in these three districts equal.

iii) Proceed analogously for all districts with a rank smaller or equal $N-1$.

Party D's strategy. Party D can now respond to party R' strategy in the following way:

- Assign R partisans according to a water-level and building-towers-strategy as outlined in the proof of Proposition 2: Assign as many R partisans as party R did in the previous round. Bring the top $2N-2$ districts to a joint level of R partisans and possibly have additional R partisans in the two bottom districts.

- Assignment of D partisans: Over the L rounds of play, assign a mass of $2\beta^D$ voters to any district with a rank lager or equal to $N+1$.

Outcome in the pivotal district. Consequently, after L rounds of play, and for L sufficiently large, in any one of the top $2N-2$ districts, the mass of partisan R voters is bounded from above by $2\beta_R$. Moreover, the districts in the bottom half are filled with partisan D voters assigned by party R, and the districts in the upper half have are filled with the partisan D voters assigned by party D, i.e. $2\beta_D$ per district in the upper half. All this implies that, in the district with rank $N+1$ after L rounds of play,

$$\begin{aligned}\Delta^D(N+1,L) &= \tfrac{\mu^D_{N+1,L}-\mu^R_{N+1,L}}{2-(\mu^D_{N+1,L}+\mu^R_{N+1,L})} \\ &= \tfrac{2\beta_D-\mu^R_{N+1,L}}{2-(2\beta_D+\mu^R_{N+1,L})} \\ &\geq \tfrac{2\beta_D-2\beta_R}{2-(2\beta_D+2\beta_R)} \\ &= \tfrac{\beta_D-\beta_R}{\beta_I}\ .\end{aligned}$$

The inequality in the third line follows from the fact that $\frac{2\beta_D-x}{2-(2\beta_D+x)}$ is a decreasing function of x provided that $\beta_D \leq \frac{1}{2}$.

A.3.2 Proof of statement a) in Proposition 3

The reasoning parallels the one from part b), except that we now seek to show that party R can respond to party D's optimal behavior in such a way that, after L rounds of play, it is ensured that, in the district with rank N,

$$\Delta^D(N,L) \leq \tfrac{\beta_D-\beta_R}{\beta_I}\ .$$

To achieve this outcome, party R can respond with a water-level and building-towers-strategy to party D's assignment of partisan D voters. As a consequence, there is a common level of $2N-2$ partisan D voters in the bottom $2N-2$ districts and possibly a higher level in the two top districts. Consequently, the mass of partisan D voters in any one district in the bottom $2N-2$ is bounded from above by $2\beta_D$. Moreover, R can, over the L rounds of play, assign a mass of $2\beta_R$ partisan R voters to any district with a rank smaller or equal to N. This implies that

$$\begin{aligned}\Delta^D(N,L) &= \frac{\mu^D_{N,L}-\mu^R_{N,L}}{2-(\mu^D_{N,L}+\mu^R_{N,L})}\\ &= \frac{\mu^D_{N,L}-2\beta_R}{2-(\mu^D_{N,L}+2\beta_R)}\\ &\leq \frac{2\beta_D-2\beta_R}{2-(2\beta_D+2\beta_R)}\\ &= \frac{\beta_D-\beta_R}{\beta_I}\ ,\end{aligned}$$

where the inequality follows from the fact that $\frac{x-2\beta_R}{2-(x+2\beta_R)}$ is an increasing function of x.

A.4 Proof of Theorems 1 and 2

Propositions 1-3 imply that for all (β_D, β_R) with $0 \leq \beta_R \leq \beta_D \leq \frac{1}{2}$, the following statements hold true:

a) For every $\varepsilon > 0$, there is $\hat{L}$ so that $L \geq \hat{L}$ implies the existence of a strategy σ_R so that, for all σ_D,
$$\Pi^{VL}_R(\sigma_D, \sigma_R \mid \omega \in \Omega_R) \geq 1 - \varepsilon\ .$$

b) For every $\varepsilon > 0$, there is $\hat{L}$ so that $L \geq \hat{L}$ implies the existence of a strategy σ_D so that, for all σ_R,
$$\Pi^{VL}_D(\sigma_D, \sigma_R \mid \omega \in \Omega_D) \geq 1 - \varepsilon\ .$$

Thus, for $L \geq \hat{L}$, if party R plays the strategy in part a) it realizes a payoff of at least $1-\varepsilon$ conditional on $\omega \in \Omega_R$, whatever the strategy chosen by party D. Therefore, in any equilibrium party R's equilibrium payoff in these states is bounded from below $1-\varepsilon$. (It is also bounded from above by 1.) The same is true for party D. This proves Theorem 1.

For L large, the strategies constructed in the proof of Proposition 3 approximate equilibrium strategies for all (β_D, β_R) with $0 \leq \beta_R \leq \beta_D \leq \frac{1}{2}$: With these strategies party R ensures to win with probability arbitrarily close to 1 when $\omega \in \Omega_R$ and party D ensures to win with probability arbitrarily close to 1 whenever $\omega \in \Omega_D$. Theorem 2 then follows from the observation that, with these strategies, there are at most 4 districts, out of a total of $2N$ districts, that are not replicas of the electorate at large.