

# Computational Redistricting and the Voting Rights Act

Amariah Becker, Moon Duchin, Dara Gold, and Sam Hirsch

April 2021

## Abstract

In recent years, computers have been used to generate *ensembles* of districting plans: collections of large numbers of electoral maps that are used to assess a proposed map in the context of valid alternatives. Ensemble-based outlier analysis has played a central role in recent redistricting disputes, especially regarding partisan gerrymandering. Until now, methods for generating these ensembles have enforced districting rules that are relatively simple to assess, such as population equality, but have not contended with more complex ones, such as the prohibitions against racial gerrymandering and minority vote dilution that flow from the Constitution and the Voting Rights Act (VRA). We take up the task of building ensembles of plans that respect those legal constraints. Rather than relying on demographic data alone, our method uses precinct-level returns from a large collection of recent primary and general elections. With this electoral history, we build *effectiveness scores* that identify districts where members of minority groups have had realistic opportunities to nominate and elect their preferred candidates. In a case study of Texas congressional districts, we find that detailed election data is indispensable to assessing a map's effectiveness for minority voters. Purely demographic targets, such as demanding some specific number of majority-minority districts, not only raise constitutional concerns but also are inadequate proxies for empirical effectiveness. Beyond the primary task of building VRA-conscious ensembles for comparison, we also repurpose the same algorithmic search methods to find plans that dramatically increase minority electoral opportunities. In Texas, for example, the current enacted 36-district congressional plan has perhaps 11 to 13 districts that are effective for Latino voters, Black voters, or both. We find that better mapmaking could raise that number to at least 16 without sacrificing traditional principles such as contiguity and compactness. This would nearly eliminate the historic underrepresentation of both groups throughout the state.

## 1 Introduction

Today, only 107 Representatives in Congress—fewer than a quarter of all House members—belong to a racial or language minority group.<sup>1</sup> If those groups were represented in proportion to their share of the Nation’s adult citizen population, that number would increase to 144 Representatives.<sup>2</sup> And this sub-proportional representation is not confined to Congress, but is replicated today in 47 of the 50 state legislatures.<sup>3</sup> There are two strands of conventional wisdom on the causes of this shortfall in minority representation. Either districters simply are not trying hard enough, or entrenched patterns of racial polarization in housing and voting make proportionality impossible to attain.

<sup>1</sup>Bialik (2019). This figure refers to the 116th Congress (2019–2021).

<sup>2</sup>This number is based on 2019 one-year ACS data, U.S. Bureau of the Census (2019a), figured as the share of citizen voting-age population comprising those who are either Hispanic/Latino or from a non-White racial group.

<sup>3</sup>See U.S. Bureau of the Census (2019b); National Conference of State Legislatures (2020). Putting those sources together, the three exceptions are Arizona (34.4% minority citizen voting-age population vs. 38% minority legislators), Hawaii (73.2% vs. 76%), and Ohio (16.7% vs. 18%).

This Article explores a third option: perhaps better tools can bring better results. Our algorithmically generated *ensembles*—collections of thousands or millions of alternative maps—show that better-designed redistricting plans could close much (though not all) of that gap and ensure that the House of Representatives and state legislatures “look more like America” than at any time in our history.

The tools to study this issue comprehensively did not exist as recently as a decade ago, when the 50 states last redistricted. Since then, algorithmic innovation and steadily improving computational power have revolutionized our ability to understand the variety of redistricting plans that could plausibly be enacted. It is now possible to generate a multitude of diverse, valid plans on a laptop overnight—and to describe how they are distributed in the universe of all possibilities. That in turn allows any plan, including one proposed for adoption, to be compared meaningfully to the available alternatives.

Not surprisingly, work in this direction has come to dominate some types of redistricting litigation in the last few years, especially lawsuits claiming that a districting plan is excessively partisan. But until now, ensemble methods have not seriously grappled with issues of race in redistricting. And these tend to be the most heavily litigated issues in the field, due to the demands imposed by the Voting Rights Act (VRA) and the Constitution’s Equal Protection Clause. The legal rules addressing race in redistricting are much more complex than, say, the “one person, one vote” doctrine in federal constitutional law, or the contiguity requirements in state constitutional law. Modeling the racial rules is far from straightforward.

This Article takes up that task. **First**, we develop methods that incorporate the legal rules involving the consideration of race in redistricting into the algorithms that generate redistricting ensembles. The main applications of these VRA-conscious ensembles would be to study the normal range of attributes of lawful plans, for instance to assess claims of partisan gerrymandering. **Second**, we show that the methods used to accomplish that task can also be used to draw maps that increase opportunities for minority groups to elect candidates of their choice. As it turns out, there is the potential to provide much more opportunity, at least in some states, than was previously recognized. In short, the algorithmic creation of redistricting ensembles holds the promise of not only sharpening our understanding of redistricting choices and tradeoffs, but also better fostering the aims of the Voting Rights Act, “a statute meant to hasten the waning of racism in American politics” (*Johnson v. De Grandy*, 1994, 1020).

To that end, one of our strongest findings deserves particular emphasis. In the past, the dominant method of looking for effective minority electoral opportunity has been to use district demographics as a proxy, such as by seeking majority-Black districts to secure effective electoral opportunities for Black voters. But in our case studies, demographic share alone is a poor proxy for effectiveness; relying too heavily on demographics could inadvertently disempower minority citizens by packing them into too few districts.

Our methods will be most helpful for proactive legislatures and commissions that wish to draw legally defensible maps that will prove effective for racial and language minority groups while upholding other criteria simultaneously. The tools described here will generate examples of maps with valuable properties and will help elucidate the cost in minority electoral opportunity, if any, that results from strict application of lower-ranked criteria. Although these tools also may be helpful to plaintiffs who wish to challenge existing maps under the VRA, that use case is not our main focus.

We will use three main elements: a Markov chain procedure that proposes successive modifications to districting plans, an ecological-inference procedure that identifies minority-preferred candidates based on precinct-level historical election data matched to demographics, and a benchmark plan from which we can establish a presumptively acceptable number of effective districts.

Below, for our proof of concept, we will use a spanning-tree recombination procedure for the first element, a hierarchical Bayesian model for the second, and an enacted plan that has survived VRA scrutiny for the third<sup>4</sup>—but we emphasize that the main contribution of the current Article is the overarching protocol, which is designed to be **modular**, letting users substitute in other alternatives to play these three roles. Combining these elements, our protocol defines *effective* districts for minority groups at any given threshold of confidence.

**Article Outline.** We begin in Section 2 with a review of the burgeoning science of redistricting ensembles. Section 3 summarizes the legal rules governing the consideration of race and racial data in redistricting. Section 4 sets forth our VRA-conscious ensemble protocol, relying on recent election data to generate effectiveness scores that rate each district’s likelihood of nominating and electing minority-preferred candidates. Section 5 applies this protocol to congressional redistricting in Texas, where both Latino and Black residents are numerous enough to require VRA attention. Section 6 applies techniques from statistics and machine learning to the Texas results to show the importance of using detailed electoral data. And Section 7 concludes with a clear proof of concept showing that the long-standing underrepresentation of minority voters in Texas, far from being an immutable fact, can be addressed through proactive mapmaking.

Finally, we have made the corresponding software tools available for public use in our GitHub ([MGGG Redistricting Lab, 2020a](#)) and through a user-friendly portal at [districtr.org/VRA](#).

## 2 Ensemble methods: algorithms for creating districting plans

As Justice Kagan explained in her dissent in *Rucho v. Common Cause* (2019, 2517–23), a computer equipped with an algorithm that generates a huge number of redistricting plans could potentially create a baseline to help answer questions like:

- *What is an extreme, or unfair, number of Republican (or Democratic) districts, given the partisan composition and political geography of the state’s voters?* or,
- *What would be a typical number of competitive districts, given those same parameters?* or,
- *Given the new census data, can a plan comply with the “one person, one vote” principle without pairing two incumbents’ homes in the same district?*

And as we will soon demonstrate, an ensemble approach also can help us address questions like:

- *What is a fair map for Latino and Black voters?*

### 2.1 Illustrative example: Iowa

To see the power of redistricting ensembles, let’s consider the case of Iowa. According to the 2010 census, Iowa’s 99 counties contained 216,007 census blocks and 3,046,355 residents—enough for four congressional districts. Iowa’s constitution simplifies the redistricting problem by mandating that “no county shall be divided in forming a congressional district,” so drawing our four districts requires assigning only the 99 counties ([Iowa Const. art. III, § 37](#)). We might hope to approach the task of finding fair plans by first building *all* possible plans, and comparing a particular plan to the full set.

---

<sup>4</sup> As described below, we use an implementation called **GerryChain** for plan generation, we use **eiPack** for ecological inference, and we use the current enacted Texas congressional map as our VRA benchmark.

But even this modest problem of dividing 99 counties into four connected parts (four contiguous districts) is currently out of reach: no one has yet been able to find a precise answer for this problem by computer, even with a clever enumeration algorithm and a month of computing time.<sup>5</sup>

This problem is only compounded in most states, which build their districts from census blocks (on average, there are more than 2,000 blocks per county). The full enumeration is subject to what is called *combinatorial explosion*, and the associated counting problem has forbidding complexity. This means not only that we lack the computing power to enumerate all plans today, but that computers likely will never be able to do so.

A second issue is that most plans in a complete enumeration would be irrelevant to the practical problem of redistricting because they would be blatantly unlawful. This is illustrated in Figure 1. The plan on the left, in which the biggest district has more than 750 times the population of the smallest one, would patently violate the Federal Constitution’s “one person, one vote” doctrine.<sup>6</sup> This means that districting plans with large population inequalities are of no practical interest, so a useful ensemble should exclude them.

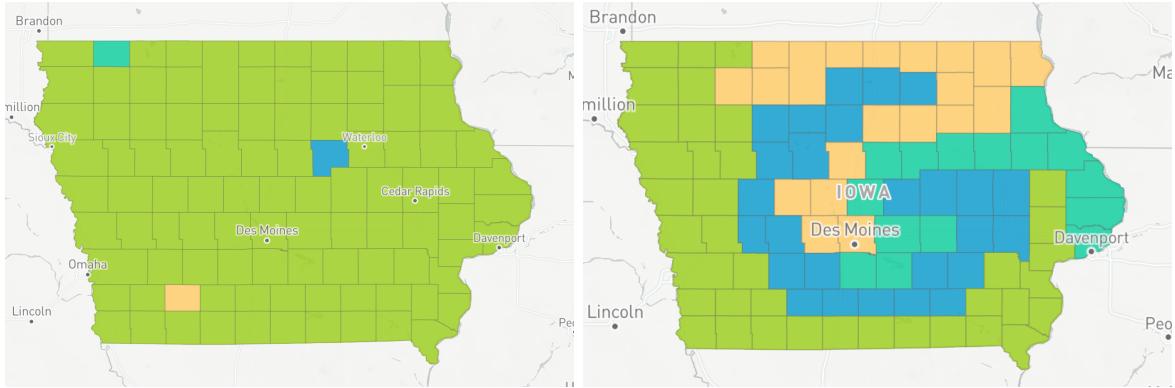


Figure 1: These two partitions of Iowa into four connected pieces are not plausible for adoption as districting plans. The first has nearly all the state’s population in a single large (green) district. The second more closely balances each district’s population, but would likely violate Iowa law’s compactness requirement.

The map on the right has much better population balance, but it also falls outside the plausible zone for plans. Its blue G-shaped district (“G” for gerrymandering) flaunts the mapmaker’s disrespect for the traditional districting principle of compactness, which Iowa law explicitly safeguards ([Iowa Code § 42.4.4](#)).

Good ensemble methods allow us to draw a *representative sample* of compact, contiguous, population-balanced plans from the full space of possibilities—that is, a sample distributed in a known way that is suited to the law. By appealing to this sample, we can hope to address questions of partisan fairness, competitiveness, racial fairness, and all the other concerns and values we bring to bear on redistricting. To illustrate this methodology, we generated a sample of 100,000 valid Iowa congressional maps by the *recombination* method explained below in §4.2, without taking partisan

<sup>5</sup>Indeed, even the simpler problem of partitioning a  $9 \times 9$  grid into nine districts of nine units each has 706,152,947,468,301 solutions. See [mggg.org/table.html](http://mggg.org/table.html).

<sup>6</sup>A district-to-district population difference greater than 10% of the ideal district size is presumptively unconstitutional under the Fourteenth Amendment; for congressional districts, the standard is far stricter, under Article I of the Constitution ([Brown v. Thomson](#), 1983, 842–48; [Karcher v. Daggett](#), 1983, 730–44). The malapportioned plan in Figure 1 has top-to-bottom deviation nearly as large as the whole state, or close to 400% of ideal district size.

data into account.<sup>7</sup> This lets us compare the enacted plan against these alternatives in terms of votes cast for President in the November 2016 election, say. In our ensemble of compact, contiguous, population-balanced plans, nearly 75% have one safe Republican seat and three competitive seats (using a 55% majority as the line between competitive and safe). The current enacted plan has one heavily Trump-favoring district and three competitive ones, putting it in the largest category. This does not tell us by any stretch that the current plan is ideal or fair, but it does tell us that this plan is not an outlier by this way of measuring partisanship. This illustrates an elementary use of ensembles to benchmark partisan lean and competitiveness.

Similarly, ensembles can help us study how plans made without regard to race might tend to distribute a state’s minority populations across districts, merely as a function of human geography. This racial baseline has been studied in a range of reports and papers, including [MGGG Redistricting Lab, 2018d,a, 2019b,a; DeFord and Duchin, 2019; Duchin and Spencer, 2021](#). But exploring the distribution of racial-group members in an ensemble is a different task from building an ensemble that takes VRA compliance into account. We will turn to that task shortly.

## 2.2 Building ensembles

Ensemble methods backed by powerful computers have proliferated in the last decade. Large ensembles of alternative plans proved critically important in federal-court cases invalidating extreme partisan gerrymanders in Ohio and Michigan (before the Supreme Court in *Rucho* held these claims nonjusticiable in federal courts) and more recently in similar state-court cases in Pennsylvania and North Carolina (*Rucho v. Common Cause*, 2019, 2493–508; *League of Women Voters of Mich. v. Benson*, 2019, 893–908; *Ohio A. Philip Randolph Institute v. Householder*, 2019, 1025–62, 1082–85; *League of Women Voters v. Commonwealth*, 2018, 770–81; *Common Cause v. Lewis*, 2019, 17–43, 80–96).

Past ensemble methods used in litigation have focused on generating plans while controlling population balance, contiguity, compactness, and sometimes county and municipality integrity. Generating large ensembles while accounting in some way for these legitimate districting criteria helped judges decide whether one political party’s disproportionate successes were due to the state’s geographic features and the distribution of its voters—or to partisan manipulation of district lines. But in building their ensembles, the experts who testified in these cases did not seriously grapple with the legal requirements involving the consideration of race in redistricting.

In the Wisconsin case, for example, Democratic plaintiffs brought partisan-gerrymandering claims against a state Assembly plan that had resulted in Republicans winning 60 or more of the 99 seats, even in elections where Democratic candidates collectively received more votes than their Republican counterparts. In work prepared for the litigation and described in a subsequent article ([Chen, 2017](#)), political scientist Jowei Chen built an ensemble of alternative Assembly plans to help evaluate the enacted plan and to demonstrate that the heavy advantage that Republicans enjoyed under that plan did not result inevitably from the political geography of the state’s voters. Chen generated an ensemble of plans that altered boundaries for 92 of the 99 districts, while “freezing” seven heavily minority districts in and around Milwaukee, one of which had been ordered into effect to remedy a VRA violation.

Likewise, in the North Carolina cases, the experts’ ensembles relied on proxies for districts’ effectiveness for minority voters. For example, consider the work of one plaintiffs’ expert, mathematician Jonathan Mattingly, as described in a subsequent article by his research group ([Herschlag et al., 2020](#)). Mattingly’s work in North Carolina used demographic targets of 44.48% and 36.20%

---

<sup>7</sup>ReCom always produces contiguous, balanced districts, and favors compact districts for reasons explained below in Section 4.2.

Black population for two congressional districts—the precise levels found in the enacted plan that the plaintiffs were challenging. He then built an ensemble by iterating a random step biased to favor plans that hit those demographic targets.<sup>8</sup> In addition to the effects of this tilted search, he discarded plans that fell short of those targets from the final ensemble presented in court, so that the prescribed population levels served as a minimum for all included plans.

In the context of these mid-decade partisan-gerrymandering cases, the experts’ decisions to de-emphasize VRA complexities were understandable. The litigation, after all, focused on party, not race, and lawful VRA-compliant districts were already in place. But at the beginning of a new decade, with fresh census results available, that option will be foreclosed, as the minority districts from the previous map will have become either over- or under-populated due to population shifts and will thus violate “one person, one vote.” So the minority districts (like all other districts) will have to be redrawn to accommodate the new census data. When generating alternative plans to create a baseline for comparison, redistricters will need to account for the delicate legal requirements imposed by the VRA and the Constitution.

For techniques that have been implemented to build VRA requirements into redistricting ensembles, the literature review is brief. In a new *Yale Law Journal* article called *The Race-Blind Future of Voting Rights* (Chen and Stephanopoulos, 2021), Jowei Chen and legal scholar Nick Stephanopoulos take the problem of identifying suitable VRA districts head-on, defining a minority opportunity district by using a combination of partisan data (returns from the 2012 presidential general election) and demographic data (voting-age population from the 2010 census). In particular, they define a minority opportunity district to be one in which (1) the candidate of choice (typically Obama) carried the district in the general election and (2) most of the candidate’s support is estimated to have come from minority voters. This is somewhat closer in spirit to the method proposed here, though this Article draws dramatically different conclusions from theirs.<sup>9</sup>

Our method for measuring district effectiveness, described in §4 below, will draw on a much larger collection of recent elections, pairing a primary with each general. The outcomes from these elections are the essential components of our effectiveness scores. And in §6 we will show that the scores we develop cannot be well approximated by considering only a district’s partisan lean and demographics.

### 2.3 Using ensembles

As we develop techniques for building VRA-conscious ensembles, there are two important general caveats about how and how not to use these ensembles.

**Comparison, not selection.** Our protocol is not designed to simulate the nuanced judgment of a seasoned voting-rights attorney. Rather, as we generate a chain of thousands of maps, we need a fast and reliable rough cut for VRA compliance. Our protocol uses a random iterative process in which districting plans are proposed, weighed, and potentially accepted into our ensemble of plans. We will be designing an in-or-out criterion that can be assessed in a fraction of a second. It is too much to expect perfection in excluding all unlawful maps and including all lawful ones, partly because the law itself is hardly a bright-line field. For example, even what seems like a rule with a clear threshold, such as the constitutional prohibition against state-legislative plans with population deviations greater than 10%, has exceptions in case law (*Cox v. Larios*, 2004;

<sup>8</sup>Mattingly’s method used a search procedure weighted to favor plans with better scores, based on a combination of population balance, compactness, county integrity, and nearness to his demographic targets for Black population.

<sup>9</sup>For their method’s details, see the full description in Chen and Stephanopoulos (2021). For a critique of their definition of minority opportunity districts and its application, see Duchin and Spencer (2021).

*Unger v. Manchin*, 2002). Nonetheless, an ensemble that includes most of the lawful maps that are proposed in the chain and rejects most of the unlawful ones will suffice for our goals of comparison and benchmarking. Ensembles should not be regarded as supplies of plans ready for immediate adoption; they are not likely to be good plans without extensive human vetting and adaptation.

**Normal range, not ideal.** We advocate using redistricting ensembles to learn a *normal range* for metrics and measures under the constraints of a set of stated redistricting rules and priorities. Ensembles allow us to justify statements such as *Plan X is an outlier in its partisan lean, taking all relevant rules into account*. While talking about normal ranges and outliers, we should avoid the temptation to valorize the top of the bell curve (or its center of mass, or any other value) as an ideal. By analogy, we can talk about people who are unusually tall or short without believing that any height is most desirable or ideal. If the 50th percentile height for American women is 5'4" and the 99th percentile height is 5'10", we can conclude that a woman who is six feet tall is unusual, and we can look for reasons (family history, diet, and so on) to explain her height. But it would be quite strange to decide that a woman who is 5'4" is a "better" height than one who is 5'5".

Justice Kagan's *Rucho* dissent skirted the edge of this temptation. She mostly reasoned from ensembles just as we will recommend here, envisioning a bell curve (in that case, of partisan advantage) and describing plans far from the bulk of the curve as presumptively impermissible: "The further out on the tail, the more extreme the partisan distortion and the more significant the vote dilution" (*Rucho v. Common Cause*, 2019, 2518). But in the course of describing the outlier logic, she implied that plans "at or near the median" are the best of all. An outcome "smack dab in the center" (in Justice Kagan's words) may not be in any sense the most fair, however. For instance, turning to the November 2012 Obama-Romney election as a touchpoint, Obama received nearly 53% of the major-party vote in Iowa. Even if just over half the congressional plans in our ensemble have three Obama-favoring districts out of four (making that the median outcome), we might still reasonably consider a map with two Obama-favoring and two Romney-favoring districts to have at least as strong a claim on fairness, given the nearly even vote split.

Likewise, there would be no reason to prefer a map that preserves intact a *median* number of whole counties or municipalities. Indeed, some States' redistricting laws expressly demand keeping the greatest practicable number of counties or municipalities intact.

The same warning, to be wary of the magnetic attraction to the middle of a bell curve, surely applies as well to racial fairness. If a state's Latino, Black, Asian-American, and Native American residents have historically been (and currently remain) underrepresented, we should gravitate toward solutions that fix the shortfall rather than perpetuate it. Fortunately, federal law pushes redistricters in the right direction.

### 3 The law of race and redistricting

The rules regarding the consideration of race in redistricting flow primarily from two sources of federal law: the Fourteenth Amendment's Equal Protection Clause and Section 2 of the Voting Rights Act, which Congress, exercising its power to enforce the Fifteenth Amendment, enacted in 1965 and significantly revised in 1982.

#### 3.1 The Voting Rights Act prohibits minority vote dilution

Section 2 of the VRA prohibits a redistricting plan that abridges any citizen's right to vote "on account of race or color [or membership in a language-minority group]" (VRA, §§ 10301(a),

10301(f)(2)). Minority plaintiffs can establish a violation of amended Section 2 by showing, “based on the totality of circumstances,” that members of their racial or language-minority group “have less opportunity than other members of the electorate” to “nominat[e]” and “elect representatives of their choice” ([VRA](#), § 10301(b)).

In assessing whether a redistricting plan provides equal electoral opportunity under amended Section 2, Congress expressly permitted state redistricters and federal judges alike to consider recent election outcomes, namely “[t]he extent to which members of a protected class have been elected to office” ([VRA](#), § 10301(b)). Nothing in Section 2, however, “establishes a right to have members of a protected class elected in numbers equal to their proportion in the population.” While electoral success for minority candidates is important, even more important under Section 2 is that the candidate be the “chosen representative” of a particular racial or language-minority group, regardless of the candidate’s race or ethnicity (*Thornburg v. Gingles*, 1986, 68 (plurality opinion)). And Section 2’s lodestar is “equality of opportunity, not a guarantee of electoral success for minority-preferred candidates of whatever race” (*Johnson v. De Grandy*, 1994, 1014 n.11). As the Supreme Court has explained, “minority citizens are not immune from the obligation to pull, haul, and trade to find common political ground, the virtue of which is not to be slighted in applying a statute meant to hasten the waning of racism in American politics” (*Johnson v. De Grandy*, 1994, 1020).

In redistricting cases “the ultimate question [under Section 2] is whether a districting decision dilutes the votes of minority voters” (*Abbott v. Perez*, 2018, 2332). District lines can dilute the voting strength of politically cohesive minority-group members either by “cracking,” or dispersing, them among multiple districts where they are routinely outvoted by a bloc-voting majority or by “packing,” or concentrating, them into too few districts, wasting votes that could have mattered in neighboring districts (*Johnson v. De Grandy*, 1994, 1007). Section 2 prohibits both cracking and packing whenever district lines combine with social and historical conditions to impair the minority group’s ability to elect its preferred candidates “on an equal basis with other voters” (*Voinovich v. Quilter*, 1993, 153).

In jurisdictions where all sizable demographic groups (majority and minority alike) consistently favor the same candidates, a redistricting plan cannot dilute minority citizens’ voting strength, so Section 2 plays no role (*Thornburg v. Gingles*, 1986, 51). But in most states, where voting is in varying degrees racially polarized, Section 2 can require replacing one or more districts that elect candidates preferred by the majority (usually, a White majority) with districts that would elect candidates preferred by one or more minority groups (*Johnson v. De Grandy*, 1994, 1008). To prevail, Section 2 plaintiffs must prove that, under the challenged plan, a bloc-voting majority usually will defeat “candidates supported by a politically cohesive, geographically insular minority group” (*Thornburg v. Gingles*, 1986, 49). But even with such proof, plaintiffs’ challenge to a state districting plan ordinarily will fail if the plan provides effective opportunities to nominate and elect minority-preferred candidates in a number of districts *roughly proportional* to the minority group’s share of the state’s citizen voting-age population, or CVAP (*LULAC v. Perry*, 2006, 436–38; *Johnson v. De Grandy*, 1994, 1000).

One particularly useful—and simple—method for assessing minority electoral opportunities under a districting plan is to add up the votes cast for each candidate in recent *statewide* primary and general elections by district, to learn which districts gave more votes to the minority-preferred candidate than to any other candidate (*LULAC v. Perry*, 2006, 428 (majority opinion), 493–94, 499–501 (Roberts, C.J., dissenting in part); *Session v. Perry*, 2004, 499–501). This approach is particularly straightforward if each precinct is kept intact within a single district: simply adding up the votes for each candidate in all of a district’s precincts shows, for each election, which candidate carried the district. The most difficult part of these analyses, especially in primaries, is

identifying the candidate who was minority-preferred in each election, which is typically performed by a statistical-inference procedure comparing demographic patterns to voting patterns (King, 1997; King, Rosen and Tanner, 1999; Elmendorf, Quinn and Abrajano, 2016). But we will take care to place actual electoral history at the center of our assessment of district effectiveness, keeping the role of statistical inference to a minimum.

### 3.2 The Equal Protection Clause prohibits excessive attention to race

Regardless of what techniques are used to assess minority electoral opportunities, compliance with Section 2 necessarily requires detailed consideration of race and racial data. But a State's consideration of race is constrained by the Fourteenth Amendment mandate that “[n]o State shall . . . deny to any person within its jurisdiction the equal protection of the laws” (U.S. Const. amend. XIV; see *Bethune-Hill v. Virginia State Bd. of Elections*, 2017, 802). Starting in the 1990s in its *Shaw* line of cases, the Supreme Court has identified at least two ways that the excessive use of race can give rise to a presumptively unconstitutional *racial gerrymander* under the Equal Protection Clause (*Miller v. Johnson*, 1995, 904–05, 910–17; *Shaw v. Reno*, 1993).

First, a bizarrely noncompact district is subject to strict scrutiny under that Clause if the district's boundary is “so irrational on its face that it can be understood only as an effort to segregate voters into separate voting districts because of their race” (*Shaw v. Reno*, 1993, 658). This type of racial predominance most often arises where a district's perimeter is defined not by the boundaries of intact precincts, for which electoral data exists, but by the boundaries of (much smaller) census blocks that have been conspicuously sorted into or out of districts according to their racial composition (Hebert et al., 2010, 66–68 & n.21; *Alabama Legislative Black Caucus v. Alabama*, 2015, 274).

Second, although only a minority of Justices have stated that the intentional creation of a majority-minority district should always be presumptively unconstitutional, a majority of the Court has held that districts violated the Equal Protection Clause because they were drawn to “maintain a particular numerical minority percentage” or to meet arbitrary or “mechanical racial targets.” The Court has thus rejected a bald mandate that certain districts must have at least a 50% or a 55% Black voting-age population regardless of whether that percentage was actually shown to be necessary for the district to nominate and elect minority-preferred candidates (*Cooper v. Harris*, 2017, 1469; *Bethune-Hill v. Virginia State Bd. of Elections*, 2017, 799, 801–02; *Alabama Legislative Black Caucus v. Alabama*, 2015, 267, 275; *Bush v. Vera*, 1996, 969–72).

### 3.3 Implications for redistricting ensembles

These legal points have major implications for an ensemble-creation protocol keyed to compliance with the VRA and the Constitution. As an initial matter, recalling the earlier point about ensembles being far more useful for comparison than for selection, the focus here is on drawing a collection of maps that would be relatively safe from challenges under VRA Section 2, rather than on crafting a map for plaintiffs to propose when suing the State.

As a gatekeeping function before ultimately assessing the “totality of circumstances,” courts generally require Section 2 plaintiffs to present an illustrative map showing that the minority group in question could constitute a literal arithmetic majority of the voting-age population (VAP) in a proposed district.<sup>10</sup> The Supreme Court has noted, however, that a district that falls short of the

---

<sup>10</sup> See *Bartlett v. Strickland*, 2009, 6, 9–11, 20, 24–25, 26 (plurality opinion). *Bartlett* also may be satisfied with a majority of the proposed district's *citizen* voting-age population (CVAP). And *Bartlett*'s 50% rule may not apply if the defendant drew the challenged districts with discriminatory intent, as might well be the case when a State

50% threshold yet can still nominate and elect minority-preferred candidates “can … [and] should” count as a minority-effective district when assessing a State’s compliance with Section 2 (*Bartlett v. Strickland*, 2009, 24 (plurality opinion); see also *Cooper v. Harris*, 2017, 1470). So actual electoral opportunity for minority groups—a track record of effectiveness in elections—is what matters when defending a map against a VRA challenge.

Taken together, the legal points elucidated above in Sections 3.1 and 3.2 suggest three crucial design principles for a VRA-conscious ensemble protocol.

- (1) **Ensure effectiveness in both primaries and generals.** Aiming to weed out of an ensemble plans that violate Section 2, while retaining plans that comply, a protocol must assess whether particular districts will or will not be effective for minority-preferred candidates seeking both nomination (in primaries) and election (in generals). This assessment requires attention to both demographic data and actual election results, including precinct-level returns from primary and general elections.
- (2) **Avoid *a priori* demographic targets.** Threshold decisions about the composition of districts should not be based on purely demographic targets—for example, requiring a certain number of districts that are at least, say, 55% Latino or 50% Black. That approach not only could lead to false positives or false negatives for district effectiveness, but could leave the methodology vulnerable to constitutional attack for excessive race-consciousness.
- (3) **Maintain reasonable compactness.** To further reduce constitutional exposure, the ensemble-generating technique should admit few or no plans with bizarre district shapes.

We note that both the first and the third principles recommend the use of precincts, rather than the much smaller census blocks, when assembling districts. Precinct-based plans promote compactness and facilitate more accurate assessment of electoral history, which is fundamental to evaluating district effectiveness. And though they may not achieve perfect population equality, that fact usually should not present significant constitutional concerns.<sup>11</sup>

## 4 Design of a VRA-conscious ensemble protocol

In this section, we will describe the design of a protocol for generating redistricting plans that comply with not only the criteria of population equality, contiguity, and reasonable compactness, but also the race-related rules mandated by the VRA and the Equal Protection Clause. The protocol begins with data preparation and culminates in the use of a constrained recombination algorithm for generating plans that meet VRA-related requirements. We propose this as a sound and detailed *VRA-conscious algorithm*, but not as the *authoritative VRA algorithm*. There may well be other ways to incorporate the legal requirements around race, and to do it well. But the methods laid out in this section come closer to the big-picture goal—building a representative sample of lawful maps—than any previous work we know. We believe that this elaborated example of one concrete, reasonable way to take account of race and the law helps illuminate some key decisions.

---

dismantles an existing minority-effective district.

<sup>11</sup>Using whole precincts will rarely raise “one person, one vote” concerns for state-legislative maps. However, the Constitution imposes stricter population-equality standards for congressional maps (*Karcher v. Daggett*, 1983, 740-41). Although the most common current practice is to draw congressional plans so that the largest and smallest districts differ by only one person, the Supreme Court has upheld plans with significantly larger deviations (*Tenant v. Jefferson County Comm’n*, 2012, 762, 764–65; *Abrams v. Johnson*, 1997, 99–100). In any event, a map built from whole precincts can usually be readily modified into a map with a minimal deviation by swapping a limited number of census blocks between adjacent districts.

We recall from above that the protocol is modular with respect to three ingredients: a procedure for iteratively modifying districting plans (here, spanning-tree recombination), a procedure for identifying minority-preferred candidates (here, a Bayesian hierarchical model of ecological inference), and a benchmark that prescribes a threshold number of effective districts for each minority group (here, an enacted plan that has evaded or withstood VRA scrutiny). Our choices can be swapped out for others as new methods or special circumstances warrant, leaving the overall structure intact.

## 4.1 Preparing data

### 4.1.1 Electoral and demographic data

We will require a cleaned precinct *shapefile* for the state, with election returns and demographic data joined to those precincts.<sup>12</sup> This can be difficult to obtain because precincts change from year to year and a longitudinal precinct shapefile is needed for the span of years covered by the election dataset. Furthermore, we may need to clean the precinct shapes to get suitable topology: to be usable as building blocks for plans, precincts must tile the state, with every resident located in one and only one precinct.<sup>13</sup>

The shapefile allows us to match reported vote totals to geographic units and to record which pairs of precincts are adjacent, which will be needed to ensure that districts are contiguous. For each precinct, we have joined data on total population from the 2010 decennial census, adult citizen population by race and ethnicity from the American Community Survey (ACS) five-year rolling estimates ending in each election year, and counts of votes received by each candidate for statewide election in a large set of primary and general elections.

Although our modeling concern is with districted elections for Congress and state legislatures, our analysis is based primarily on statewide (exogenous) contests. This is because the choices facing voters in districted elections vary across the state: in any given election year, some districts are uncontested, some have strong incumbents or other idiosyncrasies. When district boundaries are moved to create alternative plans, the newly proposed districts will be composed of voters who faced completely different candidate choices. It is not clear how votes for one candidate would translate to votes for a different candidate. By contrast, statewide elections allow us to make apples-to-apples comparisons across different parts of the state, since the same set of candidates competed everywhere. Ideally, we would include all statewide contests for the last ten years, but this is not always possible because of data availability and precinct instability. As we will discuss further below, this protocol is not intended for use with fewer than five general elections, grouped with the primaries (and, where applicable, primary runoffs) that preceded them.

Because our main concern here is whether minority-preferred candidates are ultimately elected to office, we *link* the primary (and primary runoff) for a given office in a given year to the general election for that same office that same year, and define success by whether the candidate who was minority-preferred in the primary succeeded at all stages of the electoral process.

We use a simplified set of racial groups: every person who identified as Hispanic/Latino on the census or ACS is classified as *Latino*. We use the term *Black* for non-Hispanic respondents who selected Black as their single racial category, and we use *White* similarly. All other respondents (those non-Hispanic persons selecting two or more races, Asian-American, Native American, and so on) are grouped together and designated as *Other*. In a state with only one sizable minority group,

---

<sup>12</sup>Shapefiles store data about the position and attributes of a geographic unit, such as a precinct.

<sup>13</sup>Cleaned and vetted shapefiles that are suitable for longitudinal data are easier to create in some states than others. For instance, the Louisiana shapefile used in this study required hundreds of person-hours of data preparation from members of the MGGG Redistricting Lab. It would be extremely difficult to obtain an analogous data product in Mississippi, for example.

all other minority groups may be merged into the Other category for purposes of this VRA protocol. Citizen voting-age population is denoted by CVAP, and we use HCVAP, BCVAP, WCVAP, and OCVAP to denote Hispanic/Latino, Black, White, and Other CVAP. We focus on Latino and Black voters as minority groups because our main case study involves congressional redistricting in Texas. In other states, like California, Hawaii, or Alaska, or in certain local districting projects, we might specify different racial groups for analysis.

Importantly, we make no prior assumptions about whether the voting behavior of Latino, Black, White, or Other groups will align. This is a case-by-case empirical question addressed with statistical inference.

#### 4.1.2 Candidates of choice

As explained above, the linchpin of a vote-dilution claim under the VRA is the right to replace districts where minority-preferred candidates usually lose with districts where they have a realistic opportunity to win, so long as they “pull, haul, and trade to find common political ground” (*Johnson v. De Grandy*, 1994, 1020). To assess whether a district falls into the former category or the latter requires determining which candidates are preferred by members of each sizable minority group.

Because vote totals are not reported by racial group, we cannot directly determine which candidates are minority-preferred. Instead, this effort falls under the umbrella of *ecological inference* (EI). Voting preferences are never monolithic, but techniques for measuring racial polarization have been refined for decades, and they can help us estimate the degree of bloc voting. The techniques in the ecological-inference family, like all statistical-inference methods in the presence of missing data, give imperfect and uncertain answers (Elmendorf, Quinn and Abrajano, 2016). It is fundamentally important to estimate the error that is produced by techniques and keep track of how it compounds or cancels out in our high-level conclusions. As much as possible, we will opt to make gradated and not bright-line determinations from the outputs of EI.

Our VRA-conscious ensemble protocol requires identifying the candidate who was preferred by each sizable minority group in each election, together with confidence measures that these preferred candidates are correctly identified. To perform the check for minority control of a district, as well as to identify district-wide candidates of choice for newly proposed districts, we make use of not only statewide but also precinct-level vote estimates by race for each candidate (with variance estimates). Users can employ various methods to generate these estimates (e.g., using King’s EI, Ecological Regression, exit polls, or voter files). Notably, this allows our protocol to immediately incorporate any future advances in inference techniques.

In the implementation described here, we generate estimates using a version of King’s EI, specifically the `ei.MD.bayes` function from eiPack (Lau, Moore and Kellermann, 2020) which is based on the Bayesian hierarchical Multinomial Dirichlet model for  $R \times C$  tables proposed in King, Rosen and Tanner, 1999.<sup>14</sup> For each election we run EI at the statewide level, using precinct-level input tables. The inputs for each precinct are the row and column *sums* for the  $R \times C$  table of vote counts. The row sums correspond to the precinct’s estimated number of adult citizens in each racial group (HCVAP, BCVAP, WCVAP, and OCVAP). The column sums are the precinct’s vote totals for each candidate as well as a *None* count, which is the sum of the four CVAP figures minus the sum of the recorded vote totals for all candidates, estimating the number of nonvoters. EI then infers values for the internal cells of these tables, i.e., estimated vote counts by racial group and candidate. Inclusion of the *None* column allows the underlying model to estimate differential turnout by race; without this, EI would rely on the unrealistic assumption that adult citizens from all demographic groups were equally likely to have cast a ballot.

---

<sup>14</sup>Here,  $R \times C$  stands for the number of rows (or racial groups)  $R$  and columns (or candidates)  $C$ .

Each EI run generates a large random sample of estimated precinct vote counts; we can sum these across the entire state to get statewide estimates. For each racial group, the candidate with the highest average estimated vote total for a given election is identified as the group’s “candidate of choice.” For a measure of confidence that Candidate X was the candidate of choice for a racial group in a given election, we first take repeated draws from the EI distribution and record the frequency with which X receives the most votes from that group. We then transform this to a confidence score.<sup>15</sup>

## 4.2 Building new plans by recombination

The science of representative sampling has advanced greatly in the past few years as ensemble methods for redistricting have matured. Using a technique known as *Markov chain Monte Carlo* (MCMC), it is now possible to efficiently create an ensemble of thousands or millions, even billions, of plausible maps. We can even sample while keeping control of the weighting that makes some kinds of plans appear more often than others. For example, we can be sure that a preference for more compact plans is designed to depend *only* on a prescribed score of compactness and on no hidden factors.<sup>16</sup>

The engine of our district-generation process is a Markov chain known as recombination, abbreviated ReCom, whose central idea of using spanning trees to split districts is fast becoming the standard in the field (DeFord, Duchin and Solomon, 2021; Carter et al., 2019; McCartan and Imai, 2020). We will apply it to plans built from whole precincts, the smallest geographic units for which we have accurate, detailed electoral data. Earlier MCMC methods for redistricting reassigned a single geographic unit (such as a precinct) from District A into adjacent District B at each step, creating a new plan that agreed with its predecessor on the assignment of every unit except one. (If Texas, for example, had 9000 precincts, 8999 would stay in their districts at each step.) By contrast, ReCom typically proposes a much larger change: at each step, two entire (adjacent) districts are merged and then re-split in a new way that is completely independent of the division in the previous plan. This means that a single ReCom step can reassign hundreds of precincts at a time. (Each of Texas’s 36 congressional districts, for instance, has roughly 9000/36, or 250, precincts, so each recombination step performs a random division of roughly 500 precincts into two new districts.) By iterating this transformation hundreds of times per minute, the map soon loses any resemblance to its starting configuration.

A ReCom step merges a random pair of adjacent districts and splits the region in a new way. Under the hood, each ReCom step uses a *spanning tree*, which is a kind of “skeleton” of the double-district created by the random merger, and then searches for a place to cut that tree to leave behind two population-balanced, connected pieces. So, by construction, all plans proposed by

---

<sup>15</sup>Let  $p$  be the frequency in a batch of trials with which X is observed to be the preferred candidate. We logistically transform this to a confidence score using  $C(p) = 1/(1 + \exp(18 - 26p))$  to weight the election in the compound score of district effectiveness (see Table 1 below). The parameters 18 and 26 were chosen so that an election in which the draws have Candidate X ahead only 50% of the time should receive almost no weight (because it is a toss-up); but if Candidate X comes out ahead in, say, 85% of trials, the confidence should be nearly 100%. It is certainly possible to use other parameters, to skip this step and just use  $C(p) = p$  as a measure of confidence, or even to forgo confidence altogether. Without some factor of this kind, however, the resulting score will have more noise due to cases where the candidate of choice is uncertain. If we do not strongly down-weight the uncertain elections, we risk a situation in which just rerunning the EI with identical settings could produce a significantly different answer. We discuss this and other robustness checks in footnote 29.

<sup>16</sup>To be precise, the recombination algorithm used here approximately targets a known distribution called the *spanning-tree distribution*, where the probability of selecting a particular plan is proportional to a certain measure of compactness. A modified algorithm called *reversible recombination* exactly targets that steady state. See DeFord, Duchin and Solomon 2021; Duchin and Tenner 2018; Cannon, Duchin, Randall and Rule 2020.

recombination are contiguous and maintain the desired population balance. What is less obvious is that ReCom’s use of spanning trees also places an automatic priority on districts that have more internal adjacencies: so *compactness*, or a preference for plump, regular forms over thin necks or stringy appendages, is also a structural feature of the algorithm (see Figure 2) and does not have to be set as a manual choice by the programmer (DeFord, Duchin and Solomon, 2021). In fact, when the district boundaries of a plan generated by ReCom look ragged to the eye, it is often because the building-block units themselves (such as precincts) have jagged edges.<sup>17</sup>



Figure 2: If all contiguous, population-balanced plans were made equally likely, the compact plans (left) would be enormously outnumbered by bizarrely noncompact ones (right). The ReCom algorithm prefers the compact one, with a relative weight dictated *only* by its compactness score.

Over thousands or millions of iterations, this simple method can undertake far-reaching exploration of the universe of possible plans subject to population balance, contiguity, and reasonable compactness. We will call a set of plans collected in a recombination chain an *ensemble* of plans.

Additional features and constraints can be incorporated into ReCom either with hard thresholds (i.e., validity checks) or by using probabilistic acceptance. To illustrate this, consider the traditional districting principle that counties should be kept intact when practicable. We could enforce a maximum allowable number of county splits by adding an instruction to automatically reject as invalid any proposed plan that exceeds some level of county-splitting, creating a *constrained* ensemble. A different option would be to impose a bias to the probability of acceptance, essentially flipping a weighted coin each time a proposal is generated that makes it rare but not impossible to accept plans with a large number of county splits. This would create a *biased* (or *tilted*) ensemble favoring fewer county splits.

When a proposed plan is rejected, a new plan is proposed by merging and re-splitting a freshly chosen pair of adjacent districts. This continues until some proposed plan passes the necessary tests to be accepted, at which point it is added to our ensemble. The next step proceeds from this newly accepted map, and so on until the Markov chain reaches its stopping condition (such as by collecting a prescribed number of plans). Our ensembles contain every valid plan rather than *sub-sampling*, or thinning out by accepting only every 1000th or 10,000th plan as previous authors have done (Herschlag et al., 2020; Fifield et al., 2020). The long-range statistical properties are the same whether we use continuous sampling or sub-sampling, and we employ standard convergence heuristics from the scientific computing literature to provide evidence that our chains are run long enough for the statistics we collect to approach stationarity.<sup>18</sup> For more information

<sup>17</sup>The reasons spanning-tree partition methods produce compact districts are explored in Duchin and Tenner (2018) and DeFord, Duchin and Solomon (2021).

<sup>18</sup>Markov chains that take large steps, like ReCom, require many fewer steps to achieve approximate independence than methods that iterate very small changes.

about spanning-tree recombination and for comparisons to other methods, see DeFord, Duchin and Solomon, 2021; Becker and Solomon, 2020; DeFord and Duchin, 2020; Cannon, Duchin, Randall and Rule, 2020; McCartan and Imai, 2020; Carter et al., 2019.

Below, we will refer to district-level as well as statewide EI estimates as we build scores of district effectiveness. The district-level procedure requires some thought because of the computational cost of any calculation that occurs while the algorithm runs, rather than being performed in advance. It is not feasible to rerun EI to determine district-level candidate preferences with each newly proposed plan in a ReCom chain. We need a highly efficient calculation to retrieve both a point estimate and an estimated confidence level when a new district is formed. To handle this, we make use of the hierarchical structure of EI. The EI algorithm generates large random samples for each precinct from the distribution of possibilities produced by the underlying Bayesian model. This means that we can store outputs for each precinct in the state. Ideally, we would save the full *detailed histogram* describing the frequency with which various vote counts were estimated for each candidate and racial group in that precinct. Because this is too much information to store, we instead record the point estimate for each group's support of each candidate in addition to a simplified *coarse histogram* of vote counts, compressed down to just nine values, which turns out to be enough to recover the shape of the detailed histogram with remarkable fidelity, as shown in Supplement A. During the run of the ReCom Markov chain, we can re-draw samples from these coarse distributions and aggregate to the district level for each newly generated plan to determine the confidence that we have correctly identified candidates of choice.

### 4.3 Building raw scores of district effectiveness

We next lay out three ways to use prior election results in assigning a minority-effectiveness score to a proposed district: an unweighted score, a score that weights elections based on statewide voting patterns, and a score that weights elections based on voting patterns restricted to the proposed district itself. We will denote these scores by  $s^{\text{unw}}$ ,  $s^{\text{state}}$ , and  $s^{\text{dist}}$ , respectively. Although election-weighting schemes differ across the three effectiveness scores, each score captures the same underlying idea: the effectiveness of a district for a minority group is keyed to the district's history of voting for minority-preferred candidates running for statewide offices. Importantly, because our districts are built from whole precincts and we have prior election results matched to those precincts, no statistical inference is required to determine which candidate prevailed in each district. We simply total up the votes cast in the district for each candidate and note which candidate got the most support.

First, we need to settle on the meaning of a successful outcome for the voters of a minority group in a particular election and district. If the candidate of choice from the primary does not advance to the runoff or general, then the outcome of the general is less informative with respect to the group's preferences. Therefore, we group elections by pairing primary and general (or grouping primary–runoff–general if applicable) as Table 3 illustrates for our Texas case study. A successful election is one in which the minority-preferred candidate in the primary prevailed in both elections in the grouping (or all three, if there was a primary runoff).<sup>19</sup>

---

<sup>19</sup>To be precise, suppose the primary candidate of choice is Candidate X and the runoff candidate of choice is Candidate Y (who might or might not be the same person as Candidate X). Then there are three cases we count as primary success. Case one: X won the primary (in the district) and there was no runoff. Case two: X received over 50% of the vote in the primary (in the district), whether or not there was a runoff. Case three: X ranked first or second in the primary (in the district) and Y won the runoff (in the district). An election set that meets one of these primary-success conditions and in which the minority-preferred nominee wins the general election in the district is counted as a successful election in the scores below.

Our weighting scheme is keyed to the *probative* value of each statewide election in determining minority effectiveness—its value as evidence. The unweighted score treats each election equally; no election is considered more probative than any other in determining a district’s effectiveness. By contrast, the statewide weighted score  $s^{\text{state}}$  and the district weighted score  $s^{\text{dist}}$  treat some statewide elections as more probative than others and weight them accordingly. These *election weighting factors* each fall on a scale from zero to one. Their product is the final weight for an election. In keeping with case law, we up-weight elections if they have certain features:

- **Recent.** More recent elections provide stronger evidence of future electoral opportunity.
- **Clear candidate of choice.** As described above in Section 4.1.2, our ecological-inference outputs come with estimates of the probability that the minority-preferred candidate in the primary election has been correctly identified. Translating this to a *confidence* that EI has identified the correct candidate gives greater weight to elections in which the minority group has a clearly preferred candidate.
- **Group member preferred.** An outcome gives stronger evidence of electoral opportunity when the minority-preferred candidate is a member of the particular minority group.

Score/Factor	Recent	Clear candidate of choice	Group member preferred
Unweighted ( $s^{\text{unw}}$ )	1	1	1
Weighted/ Statewide ( $s^{\text{state}}$ )	$\begin{cases} 1 & \text{Most recent} \\ .841 & 1 \text{ year prev.} \\ .707 & 2 \text{ years} \\ .595 & 3 \text{ years} \\ .500 & 4 \text{ years} \\ .421 & 5 \text{ years, etc.} \end{cases}$	Confidence from statewide EI	$\begin{cases} 1 & X \text{ belongs to group,} \\ .5 & \text{otherwise} \end{cases}$
Weighted/ District ( $s^{\text{dist}}$ )		Confidence from district-level EI	

Table 1: The weighting factors for the unweighted, statewide, and district-based effectiveness scores ( $s^{\text{unw}}$ ,  $s^{\text{state}}$ , and  $s^{\text{dist}}$ , respectively). All of these are computed with respect to the primary election in an election set, because the runoff and general may not contain the most-preferred candidate for the minority group. Here, Candidate X is the minority group’s candidate of choice. These factors will be combined into an election-weighting term  $w$  for all elections in the dataset.

The weighting factors are summarized in Table 1. We discount elections for each year of age by a multiplicative factor of  $2^{1/4} \approx .841$ , so that if any one election is four years older than another,

it weighs half as much. The confidence that we have correctly identified the minority-preferred candidate is the same confidence score  $C(p)$  described above (see footnote 15), using draws at the state level for  $s^{\text{state}}$  and drawing from the district-level coarse histogram for  $s^{\text{dist}}$ . When gauging Latino effectiveness, we place twice as much weight on elections in which the Latino-preferred candidate is Latino; and the analogous statement holds for other minority groups. Of course, these detailed weights are choices made by the modeler. We will introduce a calibration step for our effectiveness scores in the next section that makes our outputs more robust to these parameters, and we tested this by re-running the protocol several times with slightly different choices (see footnote 29).

These weighting factors are important for the legal interpretation we intend. More recent elections are up-weighted because the predictive value of election results tends to erode over time, as older voters pass away, younger citizens reach voting age, immigrants are naturalized, people move into or out of the district, and voters change their political preferences and behaviors. Confidence in correctly identifying candidates of choice is clearly pertinent, because a wrongly identified candidate of choice undermines all subsequent conclusions we will draw. Elections where the minority-preferred candidate belongs to the minority group in question are up-weighted because they are more probative: in the words of the late Judge Richard Arnold, the VRA’s guarantee of equal opportunity is not met when “[c]andidates favored by [a minority group] can win, but only if the candidates are white” (*Smith v. Clinton*, 1988, 1318).

We now have all the ingredients for the *raw effectiveness score* for a given district and racial group, multiplying the three factors above to get a weight  $w = w(E, D)$  for each election and district. For instance, if we have 20 elections, then each  $w$  will be .05 for the  $s^{\text{unw}}$  score, no matter the election. For the statewide score  $s^{\text{state}}$ , the elections will not all count equally, so that, for example, a recent election with an in-group candidate will weigh four times as heavily as a four-year-old election with only White candidates.

Each effectiveness score is computed similarly:

$$\text{score of district } D = s(D) = \sum_{E \in \mathcal{E}} w \cdot \delta = \text{weighted share of elections won by candidate of choice},$$

where  $\delta$  is 1 if the minority-preferred candidate carried the district and 0 otherwise. This expression applies to all three kinds of effectiveness scores  $s = s^{\text{unw}}, s^{\text{state}}, s^{\text{dist}}$ . For example, suppose there are two election groupings separated by four years, both have equal confidence weights and feature group members, and the candidate of choice is successful in one of those two election sets. Then the statewide and district raw scores of effectiveness would be 1/3 if the success was in the earlier election and 2/3 if the success was in the later election, while the unweighted score would be 1/2.

The strength of using an approach that centers on electoral effectiveness rather than demographics is that we do not make evidence-free assumptions about how large a Latino population is needed to nominate and elect Latino-preferred candidates, or similarly for other minority groups. Rather, we directly and empirically answer that question, by totaling up votes, district by district. Our direct, empirical approach is better keyed to actual minority electoral opportunities, and so also comports better with federal law. The VRA’s plain text does not equate a minority-effective district with a majority-minority district; rather, it demands an assessment of whether minority citizens have an equal opportunity to “nominat[e]” and “elect representatives of their choice.” And our empirical approach also respects the Equal Protection Clause’s prohibition against relying on racial-percentage targets when drawing districts.

## 4.4 Calibrating effectiveness scores

The raw effectiveness scores described above combine election results in three different, reasonable ways. Each score ranges from zero (never electing minority-preferred candidates) to one (always electing them). We next convert these to calibrated scores that we will use when deciding whether to accept plans into the ensemble.

At this stage, we take a *group-control factor* into account, combining it with the raw effectiveness score because it is relevant to predicting future performance and to ensuring an emphasis on electoral success for larger numbers of minority voters. It is clear from redistricting case law that majority-minority districts are not required for VRA compliance, and indeed that setting out to draw districts with a demographic target is sometimes prohibited. At the same time, a district that has only 5% Black CVAP would not be reasonably viewed as an effective opportunity district for Black voters, on par with a district with more significant Black population. We have chosen to address this issue with a factor based on the minority group's share of district CVAP.<sup>20</sup> Group control of the district is relevant for two reasons. First, Section 2 of the VRA focuses on a minority group's ability to play a controlling or "decisive ... role in the electoral process" and not merely one of "influence" (*LULAC v. Perry*, 2006, 446 (plurality opinion) (citation and quotation marks omitted)). Second, because Section 2 protects the voting rights of a minority group's individual members, the effectiveness of a district should in part depend on the number of those members represented by their candidate of choice.

The goal of the calibration step is to bolster the *probabilistic* interpretation of the scores, so that, for example, a district with  $s = .5$  can be described as having a 50/50 chance to perform for the minority group under consideration. To lend justification to this probabilistic interpretation, we apply a standard logistic regression to normalize the raw scores based on observed success data from actual enacted districts (specifically, all congressional, state Senate, and state House elections in the last decade).<sup>21</sup>

By design, the calibration step helps ensure that although the elections that are used in constructing the raw effectiveness scores are statewide contests, they still reflect election outcomes in *local* (districted) elections. We think of the logistic transformation as producing a score that best captures the observed performance of congressional, state Senate, and state House districts in the last decade. Each input (raw) score falls between zero and one; after applying the logit function we obtain an output (calibrated) effectiveness score that still falls between zero and one, but is now easier to interpret. We will reuse the same notation  $s^{\text{unv}}, s^{\text{state}}, s^{\text{dist}}$  for the outputs, taking care to refer to the scores as raw or calibrated when there is a possibility of confusion.

## 4.5 Counting effective districts

To assess whether a proposed plan complies with the VRA, we will need to *count* effective districts, and not just report scores. We elect to define a *Latino-effective* (or *Black-effective*) district as one

<sup>20</sup>Namely, our group-control factor for a district is  $c = \min(2k, 1)$  where  $k$  is the group's share of CVAP. Alternatively, the modeler could set an election-specific group-control factor in several reasonable ways: as the minority group's estimated share of votes for the candidate of choice; the group's estimated share of the district's Democratic primary electorate; or the estimated group votes for the minority-preferred candidate divided by the total votes for all candidates, for example.

<sup>21</sup>We tune logit curves  $f(x) = 1/(1 + \exp(-(ax + b)))$  so that  $f(0) \geq 0$ ,  $f(1) \leq 1$ , and  $f(c \cdot s_i) \approx \delta_i$  where  $s_i$  are the raw effectiveness scores of enacted districts,  $c$  is group control, and  $\delta_i \in \{0, 1\}$  are the ground-truth outcomes (with 1 for success) for the corresponding candidates of choice. The aim is to input a raw effectiveness score  $s$  and a group-control factor  $c$  and update  $s$  to a probability of effectiveness  $f(cs)$ . For details and examples, see Supplement B.

whose calibrated effectiveness score estimates at least a certain threshold chance of both nominating and electing a Latino-preferred (or Black-preferred) candidate.

This threshold is a parameter to be set by the modeler, and it may involve considerable discretion. One consideration may be the mapmaker’s level of risk aversion, since setting a lower threshold may result in a higher number of qualifying districts that can be simultaneously drawn, but some or all of those districts will be less certain to nominate and elect minority-preferred candidates. A second consideration may be how particular districts in the current enacted map have been characterized by judges and victorious litigants in prior redistricting litigation, or how they have actually performed in prior elections. A third consideration may be the number of statewide elections in the dataset: we may choose a higher effectiveness threshold if we have a smaller set of available elections, to account for the possibility that the signal from any single election is misleading.

In our Texas case study below, we have adopted the threshold condition  $s > .6$ —that is, to be deemed an *effective district*, we require a greater than 60% estimated chance of nominating and electing a minority-preferred candidate. We chose this figure in view of the above considerations, and because we found that districts with  $s > .6$  in any one of our three scores were quite likely to have  $s > .5$  in the other two versions, increasing our confidence that the districts selected in this way are likely to perform more often than not.<sup>22</sup>

## 4.6 Assembling the ingredients to build a VRA-conscious ensemble

Running on a standard laptop, ReCom generates new plans at a pace of hundreds of plans per minute in the Python implementation in ([MGGG Redistricting Lab, 2018b](#)), and runs about 40 times faster in the Julia implementation in ([MGGG Redistricting Lab, 2020b](#)), depending on the size of the districting problem and the tightness of the constraints.<sup>23</sup> The VRA-conscious protocol implemented here in Python ([MGGG Redistricting Lab, 2020a](#)) reassesses district effectiveness scores at each step, which slows the process somewhat, so that our runs take about 35 steps per minute for the unweighted and statewide scores and about 15 steps per minute for the district-level score on a state the size of Texas. For a smaller state like Louisiana, the speed more than doubles.

The last question to specify our protocol is how to set the numbers of effective districts that a proposed map must contain for each minority group, to be presumptively valid under the VRA and the Constitution, and thus to be included in our ensemble. Our first guide in answering this question is the state’s most recent districting plan, which may have been in effect for up to a decade and either has gone unchallenged in court or has withstood legal challenges, including VRA claims.<sup>24</sup> The second guide, discussed above, is *rough proportionality*, within the meaning of the Supreme Court’s important VRA decisions in *Gingles* and *De Grandy*: plans are frequently judged by whether the share of effective districts is similar to each group’s share of statewide CVAP.

Considering these guides, we will reject proposed plans that have fewer minority-effective districts than the benchmark plan; in other words, we will treat this threshold level of effectiveness

---

<sup>22</sup>Case law does not dictate how certain we must be of district effectiveness. When analyzing Texas districts, we found that rejection sampling for effectiveness ran as efficiently at the  $s > .7$  threshold as it did at  $s > .6$ , suggesting that a modeler could exercise considerable discretion in setting the effectiveness threshold.

<sup>23</sup>To be more precise, we conducted non-VRA trial runs on Texas, Virginia, and Pennsylvania congressional plans built out of precincts using identical machines (Intel(R) Xeon(R) CPU E5-2660 v2 @ 2.20GHz [Ivy Bridge, late 2013]), allowing districts to deviate from ideal population by only 1%. Over runs of various lengths and with various seeds, the Python implementation generated 3 to 8 valid plans per second, while the Julia implementation generated 120 to 320 valid plans per second.

<sup>24</sup>Numbers derived from this benchmark may need to be adjusted if the state’s political geography or demographics or the number of districts in a state’s plan has changed (for example, due to reapportionment of congressional seats). Our protocol can be run using a different map as a benchmark if there is reason to believe the current plan violates the VRA or the Constitution.

as a *validity check* in the district-generation algorithm. For instance, if we are considering a single minority group and the benchmark plan has three districts that are effective for that group, then each plan included in the ensemble must have at least three effective districts as well. On the other hand, we would reject a proposed plan if it had so many effective districts for one minority group that it would relegate another sizable demographic group to substantially sub-proportional representation.

Surveying the protocol described in this section, the key to our approach is its close reliance on detailed, precinct-level election results from both primary and general elections. We do not assume that some *a priori* demographic threshold will cleave districts that provide minority voters with realistic electoral opportunities from districts that will not. The approach is deeply empirical, focusing on whether a specific district, regardless of its precise demographic percentages, has a recent history of consistently supporting minority-preferred candidates in both primary and general elections. To quote Justice Kagan, our protocol is “evidence-based, data-based, statistics-based. Knowledge-based, one might say” (*Rucho v. Common Cause*, 2019, 2519 (Kagan, J., dissenting)).

## 5 Case study: Congressional districting in Texas

We applied the VRA-conscious protocol described in Section 4 of this Article to build 36-district Texas congressional plans.

### 5.1 Data

We downloaded the 2018 Texas precinct shapefile and statewide election returns from the Texas Legislative Council’s website ([Texas Legislative Council, 2020](#)). Table 2 shows summaries of the demographic data obtained from the 2010 decennial census and the American Community Survey (ACS) rolling average for the five-year span ending in 2018. (We used CVAP from ACS five-year spans ending 2016, 2014, and 2012 when assessing elections from those years.) While election data could be directly joined to the shapefile, we used the `maup` package to disaggregate ACS data from block groups (the smallest unit for which CVAP is available) down to census blocks and then aggregated the block-level data up to precincts ([MGGG Redistricting Lab, 2018c](#)). Total population and voting-age population (VAP) were collected from the 2010 decennial census; and because these data are available at the block level, they required no proration and could be directly aggregated up to the precinct level.

Racial group	Share of total population	Share of VAP	Share of CVAP
Latino	37.62%	33.61%	29.36%
Black	11.48%	11.36%	13.08%
White	45.33%	49.64%	52.28%
Other	5.57%	5.39%	5.28%
<b>Total count</b>	<b>25,145,561</b>	<b>18,279,737</b>	<b>17,858,066</b>

Table 2: Latino, Black, White, and Other shares of Texas residents by total population, voting-age population (VAP), and citizen voting-age population (CVAP). Total population and VAP data are taken from the 2010 decennial census, while CVAP data comes from the ACS five-year rolling average ending in 2018.

We then analyzed 21 statewide Texas elections conducted from 2012 to 2018, which are recorded

in Table 3. These were all the statewide elections conducted since the last round of redistricting almost a decade ago—for federal and state offices, both executive and legislative, omitting only state judicial elections.

	2012	2014	2016	2018
President	P/G		P/G	
U.S. Senator	P/R/G	P/R/G		P/G
Governor		P/G		P/R/G
Lieutenant Governor		G		P/G
Attorney General		G		G
Comptroller		G		P/G
Land Commissioner		G		P/G
Ag. Commissioner		P/R/G		G
RR Commissioner	G	P/G	P/R/G	P/G

Table 3: The 14 election sets in our Texas data (5 of which included a primary runoff), and the 7 general elections that we omitted because the Democratic nominee lacked any primary opposition. P means Democratic primary; R means Democratic primary runoff; and G means general election.

Ultimately, we eliminated from consideration 7 of those 21 elections (struck through in the table) because there was no contest in the Democratic primary, which in Texas is a critically important stage of the electoral process for determining which candidates are minority-preferred. We were left with 14 contests: nine primary/general sets and five primary/runoff/general sets, where the runoff was conducted because no candidate garnered an outright majority of the vote in the Democratic primary.

We also compiled district-level data for the 36 U.S. House, 31 Texas Senate, and 150 Texas House of Representatives seats, including the race and party of the winning candidates in all elections from 2012 to 2018, as well as demographic data for the districts, for use in the score calibration described in §4.4 and carried out in §5.3 (Klarner, 2019; History, Art & Archives, U.S. House of Representatives, Office of the Historian, 2020a,b).

## 5.2 Racial polarization and candidates of choice

The statewide results for general elections in Texas show a stark pattern of racial polarization. Across 14 separate contests in four election cycles, all three minority groups consistently voted Democratic, and White voters consistently voted Republican, as shown in Figure 3. In Texas, it is commonplace for more than three-quarters of White voters to vote Republican and more than three-quarters of minority voters to vote Democratic in the same election. Furthermore, this basic pattern appears to hold, to a greater or lesser degree, in every region of the state.

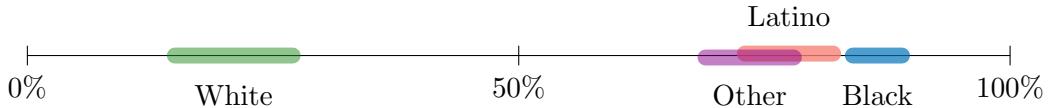


Figure 3: The highest and lowest EI point estimates for each racial group’s support of the 14 Democratic nominees in statewide general elections: White (15–27%), Other (69–78%), Latino (73–82%), and Black (84–89%).

It therefore is not surprising that the great majority of Texas’s non-White officeholders are

Democrats. From 2012 through 2018, there were only two exceptions for Representatives in Congress (out of 15 Latino or Black members) and eight exceptions for Texas state Senators or Representatives (out of 83 Latino or Black state legislators).

No Democratic candidate has won a statewide general election in Texas since 1994. So none of the Latino- or Black-preferred candidates in our 14 recent contests prevailed statewide. But the vote patterns show that each of them carried a significant number of *districts* in general elections under the current Texas congressional plan and under every plan in our ensembles.

Just as the Latino-preferred and Black-preferred candidates in all 14 statewide elections were Democrats (see Figure 3), the same has held true in congressional elections. The success of Latino- and Black-preferred congressional candidates in Texas therefore has hinged on their ability to win Democratic primaries (and, where applicable, primary runoffs) and then win general elections. A large majority of White voters in Texas primary elections participate in the Republican primary, while most people of color who participate in Texas primaries vote in the Democratic primary. So, for VRA purposes, we can currently forgo analysis of voting patterns in Republican primaries or Republican primary runoffs in Texas.

In Democratic primaries and primary runoffs, we found a high degree of cohesion across demographic groups. Because all 14 contests were for single-member offices (like Governor), we focused on the one candidate in each Democratic primary who was preferred by each of the four demographic groups. In 9 of the 14 Democratic primaries and in 4 of the 5 Democratic primary runoffs, the three minority groups (Latino, Black, Other) preferred the same candidate, as shown in Supplemental Table 7.

Given this cohesion in Democratic primaries and runoffs and especially in general elections, it might well be possible to treat Latino and Black voters, or Latino/Black/Other, as a single coalition group for VRA purposes (*Campos v. City of Baytown*, 1988, 1244–45). Our main analysis will treat Latino and Black voters as separate minority groups, but the same method could be adapted (and indeed simplified) for coalitional analysis.

As a final and important point relating to our EI setup, we note that we do not need to run EI on small geographies to detect regional difference.

For example, in the 2018 gubernatorial runoff, former Dallas County Sheriff Lupe Valdez and Houston's Andrew White are identified as the statewide candidates of choice for Latino voters and Black voters, respectively. But in the Dallas-Fort Worth Metroplex, Valdez carried both minority groups. As Figure 4 shows, that effect is visible in our EI outputs from a statewide run, because the hierarchical model works by computing distributions of support on each precinct. This lets us identify Valdez as the Black-preferred candidate in the Dallas-Fort Worth Metroplex while White is seen to have carried the Black vote in the Houston area.

### 5.3 Effectiveness scores and inclusion criteria

In Texas, we have the benefit of seeing results from 33 separate contests (14 primaries, 5 primary runoffs, and 14 generals), so that 14 potential successes make up the raw effectiveness score.<sup>25</sup>

According to recent CVAP data (shown in Table 2 above), rough proportionality would require 10.6 districts and 4.7 districts that are effective for Latino voters and Black voters, respectively, given Texas's current congressional apportionment of 36 seats. We will round these to 11 and 5

<sup>25</sup>To perform the logit calibration step described in §4.4, we used all congressional and state-legislative winners from 2012 to 2018. This includes 145 congressional contests (36 districts), 600 state House contests (150 districts), and 77 state Senate contests (31 districts), for a total of 822 data points. This includes one special election for Congress.

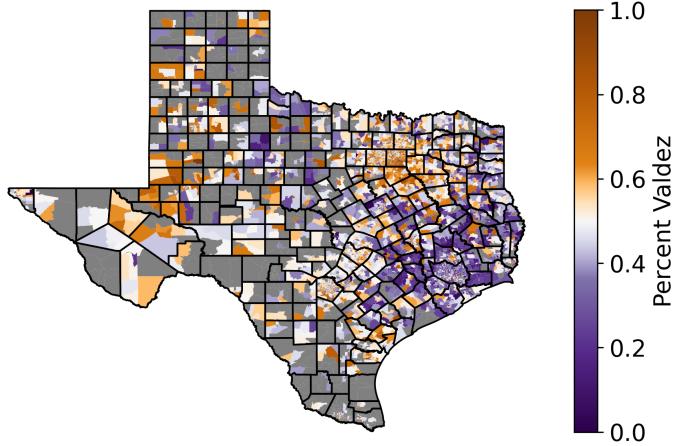


Figure 4: The distribution of EI-estimated Black support for former Dallas County Sheriff Lupe Valdez in the 2018 gubernatorial runoff. The Dallas-Fort Worth area, in northeastern Texas, is mostly orange in this map, while the Houston area, in southeastern Texas, is mostly purple. (The map's gray areas contain few, if any, Black voters.) This map shows that even statewide EI can find significant regional variation in a group's voter preferences.

districts, respectively. If Latino, Black, and Other voters were treated as a coalition, that coalition's proportional share would exceed 17 districts.

Using any of our three calibrated scores, Texas currently has 11 effective districts for minority groups at the 60% threshold: seven Latino-effective districts, three Black-effective districts, and one district that is effective for both groups (see Table 4). If our protocol focused solely on the most recent elections (e.g., 2018), however, two additional districts—District 7, currently represented by Lizzie Fletcher, a White Democrat, and District 32, currently represented by Colin Allred, a Black Democrat—might meet the effectiveness thresholds for Latino voters or Black voters under some or all of our three calibrated scores. But in the early years of the decade (e.g., 2012 and 2014) both districts were still reliably voting for Republicans in statewide and congressional elections.

Since the current map has withstood judicial scrutiny under both the VRA and the Equal Protection Clause (*Abbott v. Perez*, 2018, 2324–34), we require plans in our VRA-conscious ensemble to meet or exceed that map's level of effectiveness: so we require at least eight Latino-effective districts, at least four Black-effective districts, and a total of at least 11 districts that are effective for at least one of the groups. So, for example, a plan whose (Latino, Black, Both, Neither) effective-district count was (4, 0, 4, 28) would not qualify for the ensemble because it falls short of 11 minority-effective districts. In effect, this approach allows plans whose effective-district counts are (7, 3, 1, 25) or (8, 4, 0, 24), as well as plans that dominate one of those outcomes from the minority perspective by shifting districts from Neither to any of the other categories.<sup>26</sup>

<sup>26</sup>Although a map with fewer than 18 Neither districts could potentially give rise to a Section 2 claim by White plaintiffs and thus merit exclusion from an ensemble, our chain runs did not generate any such plan.

## Current Enacted Plan

CD	Location	HCVAP %	Latino Effective			BCVAP %	Black Effective			WCVAP %	Representative	Race
			$s^{\text{unw}}$	$s^{\text{state}}$	$s^{\text{dist}}$		$s^{\text{unw}}$	$s^{\text{state}}$	$s^{\text{dist}}$			
9	Houston	24.7	44	38	43	46.7	96	96	94	16.1	Al Green	Black
15	South Texas	73.7	95	97	97	2.5	8	9	7	22.1	Vicente Gonzalez	Latino
16	EI Paso	76.0	99	99	97	4.2	11	12	10	17.5	Veronica Escobar	Latino
18	Houston	26.9	51	44	51	44.9	95	95	95	22.8	Sheila Jackson Lee	Black
20	San Antonio	65.0	97	97	97	5.6	12	12	12	25.8	Joaquin Castro	Latino
28	South Texas	69.2	86	93	96	5.5	10	12	8	23.2	Henry Cuellar	Latino
29	Houston	64.0	98	97	97	16.2	49	48	46	16.7	Sylvia R. Garcia	Latino
30	DFW	22.7	44	38	39	52.1	99+	99+	99	21.7	Eddie Bernice Johnson	Black
33	DFW	46.5	98	98	95	24.1	78	75	64	25.6	Marc A. Veasey	Black
34	South Texas	78.5	98	99	93	1.6	8	9	6	19.1	Filemon B. Vela	Latino
35	Austin/San Antonio	52.2	97	97	97	10.3	22	20	24	34.4	Lloyd Doggett	White

Table 4: The population shares and calibrated effectiveness scores for the 11 districts in the current Texas congressional map that are labeled effective for Latino and/or Black voters. Scores over 60% have darker shading, and scores in the 50–60% range have lighter shading. Mark Veasey’s District 33 is the only one that registers as effective for both Latino and Black voters, though Sheila Jackson Lee’s District 18 and Sylvia Garcia’s District 29 are close. All 11 Representatives are Democrats.

### 5.4 Basic results

In this section we first present evidence to support the claim that our chains of districting plans have produced VRA-conscious ensembles whose statistics have stabilized after 100,000 steps. We then look at how the statistics from these ensembles compare to an ensemble built with no consideration of race and to an ensemble generated with demographic thresholds as a potential stand-in for VRA compliance. Put differently, we compare ensembles generated by our VRA-conscious protocol, which uses both racial and electoral data, with an ensemble built with racial but not electoral data and an ensemble built with neither racial nor electoral data.

We built five ReCom ensembles, by running each of the following kinds of chain until 100,000 maps are accepted.

(non-VRA) **No VRA consideration.** Only population equality is an explicit validity check, since contiguity is required and compactness is weighted into ReCom ensembles by construction, so the algorithm does not have to be manipulated to produce reasonably compact districts.

(unw) **Constrained by  $s^{\text{unw}}$  effectiveness.** Ensemble inclusion additionally requires at least eight districts over 60% Latino-effective, at least four districts over 60% Black-effective, and at least 11 total districts effective for one or both groups, using unweighted effectiveness scores.

(state) **Constrained by  $s^{\text{state}}$  effectiveness.** (Same as above, but using statewide weighted scores.)

(dist) **Constrained by  $s^{\text{dist}}$  effectiveness.** (Same as above, but using district weighted scores.)

(CVAP) **Constrained by CVAP shares.** A plan must have at least eight districts over 45% HCVAP and at least four districts over 25% BCVAP to pass the validity check.<sup>27</sup>

#### 5.4.1 Convergence heuristics and robustness checks

Neither ReCom nor any other MCMC method will work properly if it is not allowed to run long enough, or if designed in a way that thwarts convergence. In this Article we have used ensembles built by including every plan that passes the validity checks and continuing until 500,000 maps are collected. We used two kinds of evidence to arrive at the conclusion that 500,000 plans is probably sufficient: first, we have confirmed that chains of that length have aggregate statistical properties that are approximately independent of their starting points, or “seeds,” even when the seeds are quite different. This test is sometimes called the *multistart heuristic*. Second, for selected instances we have confirmed that an ensemble ten times as large has similar aggregate statistics. Passing these tests is not a rigorous proof of approximately representative sampling, but these are standard convergence heuristics used across applied statistics. If any ensemble method fails these tests, we can be sure that either the setup violates the conditions for a unique steady state, or we have not run the chain long enough to approach it.

For the multistart heuristic to have high value, we should choose plans that are initially very different and check to see that the ensembles converge to find the same summary statistics nevertheless. The first seed plan used for the multistart test for this Texas case study is the enacted congressional plan that is currently in effect, which came out of the court proceedings challenging the early-decade plan of the Republican legislature. To find two other seeds with exaggerated differences from the enacted plan, we turned to the Atlas of Redistricting project conducted by the politics team at FiveThirtyEight ([FiveThirtyEight, 2018](#)). Seed 2 is their Texas plan drawn to favor Democrats, which is visibly quite different from the enacted plan and of course has very different partisan properties as well. Seed 3 is based on the plan FiveThirtyEight drew with an eye to compactness scores and county integrity.<sup>28</sup>

For the ensemble using the statewide effectiveness score, Figure 5 shows that a simple partisan statistic—the Clinton share of the major-party presidential vote from November 2016 across the 36 districts—gives roughly the same answers after 100,000 steps, whether the chain commences with the enacted plan or with either of the two other seed plans. Similar charts for  $s^{\text{unw}}$  and  $s^{\text{dist}}$  are found in Supplemental Figure 17. These are boxplots (or “box-and-whiskers plots”) where for each plan the districts have been sorted from 1 (the district with the lowest Clinton share) to 36 (highest Clinton share). The boxes show the values at the 25th to 75th percentiles, with the median marked, and the whiskers are set at the 1st and 99th percentiles. Colored circles show the initial values for the enacted congressional plan (red) and the two additional seed plans (blue and green). The aggregate data collected from the three differently initialized runs is broadly consonant: across the districts, the three ensembles have medians, quartiles, and overall ranges within one or two percentage points of each other, even when the seeds began over 15 points apart.

We can also compare spatialized statistics such as the one shown in Figure 7, a record of the number of times that each precinct appeared in a district with  $s^{\text{state}} > .6$ . Just 1000 steps from the starting point, the heatmaps are visibly different, showing that the chain has not run long enough

---

<sup>27</sup>To build a demographic-target ensemble, we searched for maps with at least eight majority-Latino districts and at least four majority-Black districts by CVAP. Initial attempts did not produce any such maps. We then lowered the thresholds to 45% for Latino CVAP and 25% for Black CVAP. While those thresholds are somewhat arbitrary, they roughly track Table 4, as well as the results of Section 6 shown in Figure 9.

<sup>28</sup>The FiveThirtyEight compact plan did not initially meet our VRA effectiveness requirements, so we used a heuristic-optimization run as in Supplement H to get it past the thresholds. Both FiveThirtyEight plans had to be transferred onto our precinct units with the `maup` package ([MGGG Redistricting Lab, 2018c](#)).

for this statistic to converge. Much nearer visual correspondence is achieved after 10,000 steps, and the heatmaps are nearly indistinguishable after 100,000 steps.

Beyond the multistart trials, we also checked the same statistics (Clinton vote distribution and cut-edges score) after 1 million steps. We found minimal difference in partisan or district-shape metrics when comparing the initial 100,000 steps, a sub-sampled 100,000-plan ensemble containing every tenth map from the set of 1 million, or the full million-plan ensemble. This raises our confidence both that the size of the sample is adequate to this level of statistical detail and that a run length in the hundreds of thousands is sufficient for convergence. Finally, we conducted slightly altered runs to confirm whether the general findings are robust to reasonable perturbations in the methodology laid out in §4.3, §4.4, and §4.5.<sup>29</sup>

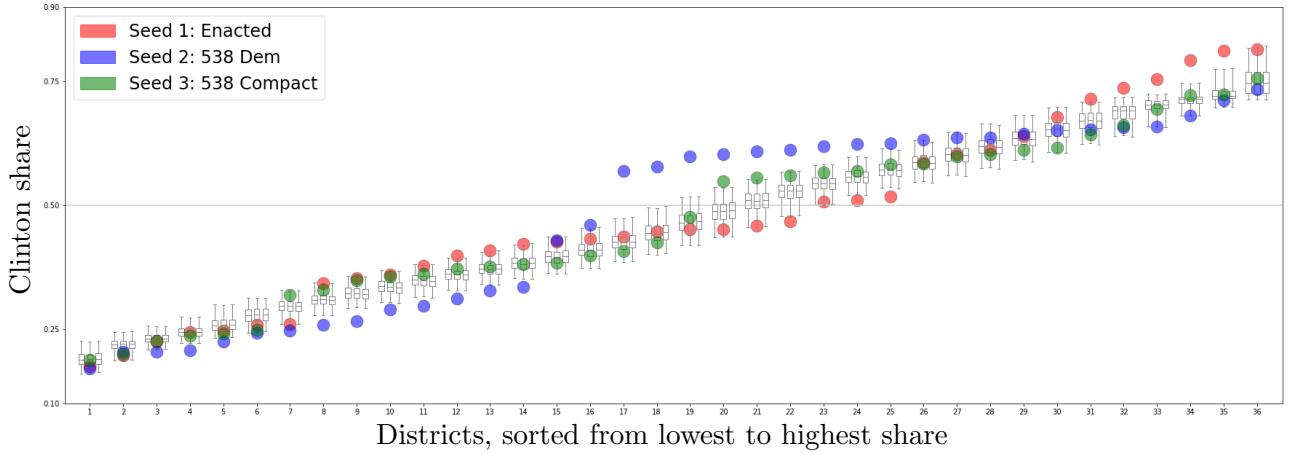


Figure 5: In this multistart heuristic convergence test, the VRA-conscious chain for the statewide weighted effectiveness score  $s^{\text{state}}$  is run for 500,000 steps from three very different starting points. The colored dots show the Clinton share of the major-party vote from the 2016 presidential general election, district by district, in the three seed plans described in the text (with the districts sorted from lowest Clinton share to highest). The boxes and whiskers show Clinton share by district for each of the three ensembles—they have converged to within one or two percentage points in each district, even though the seed plans sometimes differ by 15 points or more.

#### 5.4.2 Comparing ensembles

In this section we compare the five ensembles defined in §5.4 to each other, considering whether those created using our VRA-conscious protocol differ significantly from those created without

<sup>29</sup>We conducted the following tests: using estimated share of candidate support rather than CVAP share of the district as the group-control factor  $c$ ; replacing the confidence term for correctly identifying candidates of choice  $C(p)$  with the simpler term  $p$ ; and dropping both the group-control factor and the calibration entirely. For the alternative group-control measure, the changes to scores on Texas congressional plans were minor for both the enacted plan and generated plans. Changes also were typically small with the simplified confidence factor, but the scores became more unstable because outcomes with high EI-based uncertainty had more weight relative to clear outcomes, producing an illusion of greater electoral success on some reruns of EI. The logit calibration was valuable largely to correct for the reduction of scores by group control; we find that if we drop both of them, districts with significant shares of both Latino and Black voters are rated higher for both groups than recent electoral history warrants. Finally, we confirmed that the rate of ensemble generation is similar whether the effectiveness threshold is set at 60%, 70%, or even 75%. Taken together, these robustness runs increase our confidence that each of these parameters that requires user choice is indeed doing work in constructing a stable score that comports with electoral history, but that some of the details could be altered without breaking the protocol.

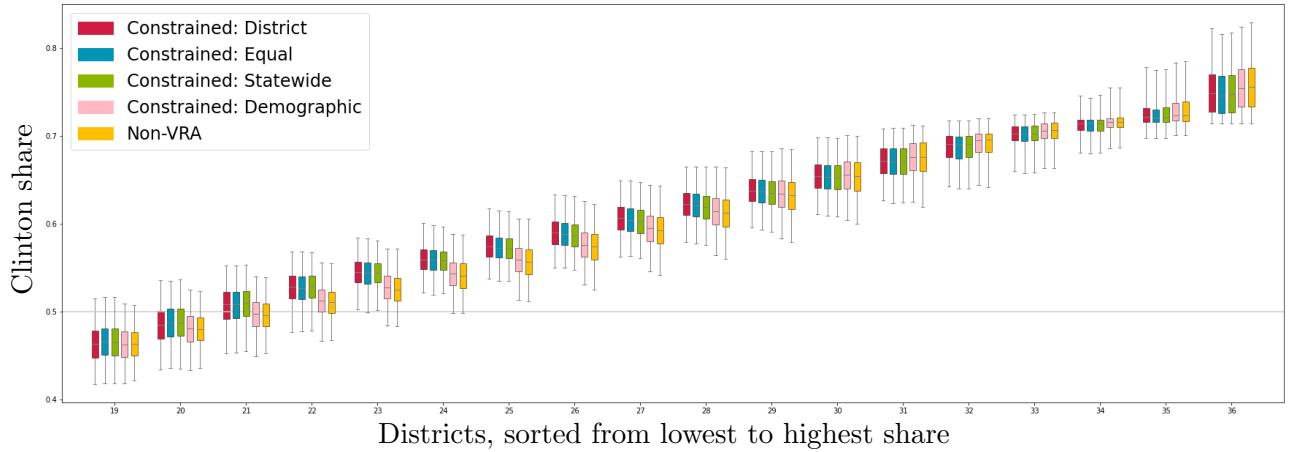


Figure 6: Comparing the three kinds of VRA-conscious ensembles, constrained by the  $s^{\text{dist}}$ ,  $s^{\text{unw}}$ ,  $s^{\text{state}}$  scores, respectively, to the alternatives described in the text. Here, the Clinton share is plotted across 500,000 steps and displayed for the 18 most Democratic districts. There is a small but discernible difference that separates the partisan statistics of the VRA-conscious ensembles from those of the control ensembles, which are interestingly similar.

electoral data or without both electoral and racial data. The answer is a definitive YES. We have already seen that the three effectiveness scores are similar to each other for the enacted plan’s minority-effective districts (Table 4). Using summary statistics, we can confirm that the constrained ensembles using the three scores are similar to each other as well. But the three VRA-conscious ensembles do not resemble either the non-VRA ensemble (which uses neither electoral nor racial data) or the CVAP-shares ensemble (which uses racial, but not electoral, data as a purported stand-in for VRA compliance).

The upshot of rejecting plans with not enough effective districts is seen in Figure 8 with respect to the  $s^{\text{state}}$  score: no plan in the ensemble has fewer than eight Latino-effective or fewer than four Black-effective districts. This number of effective districts rarely happens by chance without a VRA-conscious method. Interestingly, enforcing the demographic threshold condition (bottom row) makes it somewhat more common to get at least four Black-effective districts but does not make an appreciable difference in the likelihood of creating an eighth Latino-effective district. (Supplemental Section F contains analogous plots for the  $s^{\text{dist}}$  and  $s^{\text{unw}}$  scores.)

Table 5 is another view of the comparison. A significant share of the plans in all the VRA-conscious ensembles pass the demographic test set forth above, but relatively few plans in the non-VRA and the CVAP-shares ensembles pass our effectiveness tests.<sup>30</sup> This suggests that Texas ensembles built without rich electoral data—or by imposing a racial threshold—are unlikely to reflect VRA compliance and might well contain far too many maps that violate federal law. And this problem likely cannot be cured simply by changing the threshold levels for the CVAP-shares ensemble: if the CVAP thresholds are raised, it will become harder to find plans with enough qualifying districts, and many effective districts will be missed.

Comparing the three score-based ensembles against each other shows some differences but also substantial alignment in the determinations of validity. We should not be surprised that scores that

<sup>30</sup>That only about half the maps in the three VRA-conscious ensembles satisfy the demographic criteria implies that it is not uncommon in Texas for Latino-effective districts to have less than 45% HCVAP or for Black-effective districts to have less than 25% BCVAP. That fact in turn suggests that, at least in some parts of the state, there is significant coalitional voting between different minority groups.

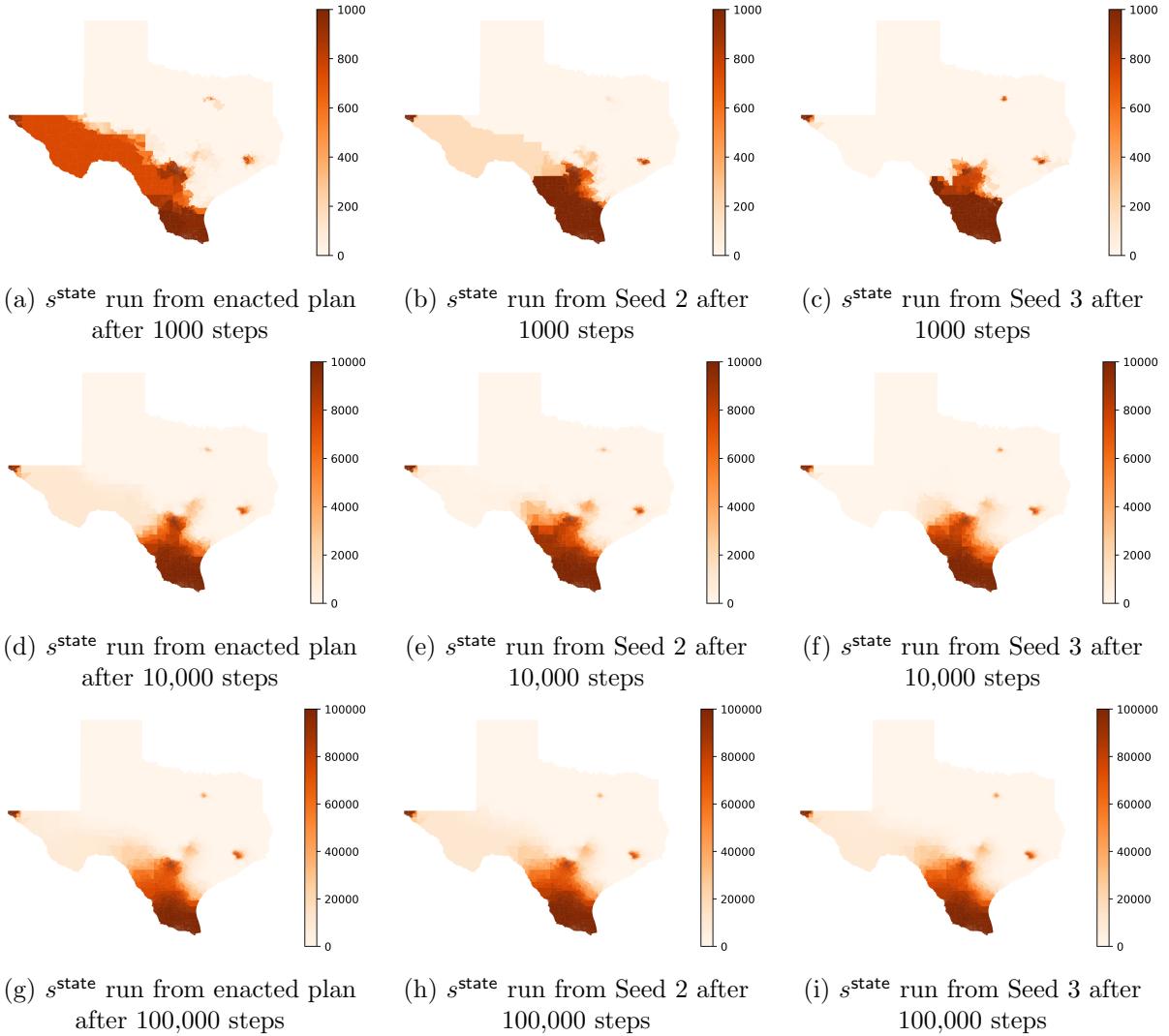


Figure 7: The color of each precinct shows how many times it had appeared in a Latino-effective district after 1000, 10,000 and 100,000 steps. These VRA-conscious ensembles are drawn with respect to the  $s^{\text{state}}$  score from the same three seed maps described in the text. There are initially significant differences across the three seeds (top row), but the plots converge over the course of the run (bottom row).

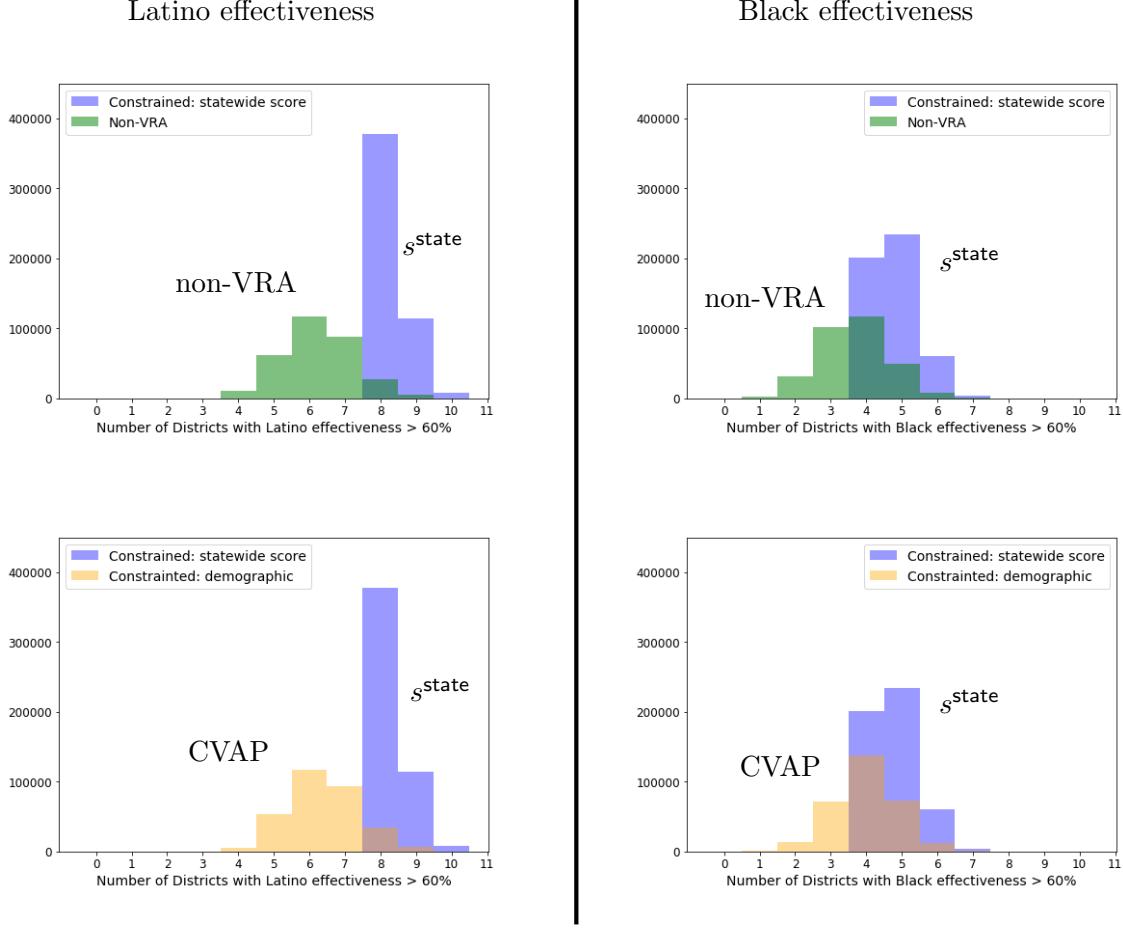


Figure 8: The distribution of Latino- and Black-effective districts in a VRA-conscious ensemble (purple), compared to the non-VRA alternative (top, in green) and the CVAP-shares, demographics-based alternative (bottom, in orange). All are shown with respect to the  $s^{\text{state}}$  score. Note the very modest improvement in effectiveness for the CVAP-shares ensemble compared to the non-VRA ensemble.

typically track each other within a few percentage points can fall on the other side of a bright-line threshold: if  $s^{\text{unw}}$  is just over .6, it can certainly happen that  $s^{\text{dist}}$  is just below that level. But most districts for which one score is over .6 have the other scores over .5, making them more likely than not to be effective for the group in question. This standard is met by more than three-quarters of the  $s^{\text{state}}$  and  $s^{\text{dist}}$  ensembles. (Again, this is part of the justification to set the effectiveness threshold for ensemble inclusion at a level buffered safely above 50%).

Considering all the evidence so far, one might ask whether any of the three calibrated effectiveness scores is to be preferred to the other two. Our determination is that all three scores can be useful. The unweighted score has the weakest claim of the three, because on its face it omits factors that are legally and factually relevant. As for the other two scores, we think it can be valuable to consider both. The district-weighted score has more regional discernment and a more sophisticated incorporation of EI outputs; the statewide-weighted score has a simpler explanation and still takes uncertainty into account. While results for different scores are not identical, the modeling methodology is robust across three reasonable ways of weighting elections to measure

	<b>Unconstrained</b> (non-VRA)	<b>Constrained</b>		
	$(s^{\text{unw}})$	$(s^{\text{state}})$	$(s^{\text{dist}})$	<b>Constrained</b> (CVAP)
<b>Satisfies effectiveness criteria</b>	15%	(100%)	88%	81%
	20%	98%	(100%)	94%
	16%	72%	78%	(100%)
<b>Satisfies demographic criteria</b>	30%	39%	46%	51%
				(100%)

Table 5: The share of maps in the five ensembles (columns) satisfying various criteria (rows). For the effectiveness criteria, maps must have at least eight Latino-effective districts (effectiveness over 50% for the indicated score), at least four Black-effective districts, and at least 11 distinct districts that are effective (for one or both groups) overall. Note that each VRA-conscious variant is built to satisfy effectiveness in a chosen score at the 60% level, making it likely to pass at least 11 district effectiveness tests for the other scores at the 50% level, since the scores are similar but not identical. The demographic test in the bottom row requires a map to have at least eight districts over 45% HCVAP and at least four districts over 25% BCVAP.

district effectiveness.

## 6 Learning patterns in district effectiveness

We have just seen that Texas congressional ensembles using demographic data but no electoral data do not resemble ensembles generated by our VRA-conscious, heavily data-driven protocol. But what about a method that uses both demographics and electoral data but in a limited way, needing only a smaller and simpler dataset? Often, scores that seem to be complicated by taking many things into account can be closely replicated using simpler inputs. In our setting, we would like to see whether our seemingly sophisticated handling of dozens of election contests could be well approximated by pared-down district metrics. To examine this question, we now model the non-linear relationship between effectiveness scores and lower-dimensional combinations of demographic and partisan features.

In statistics and machine learning, numerous techniques have been developed to recognize patterns in data. *Classifier* models use training data to “learn” discrete labels (like yes/no effectiveness), while *regression* models “learn” continuous-valued assignments (like effectiveness scores), on the basis of features in the data. For our examples, we are choosing to classify potential Texas congressional districts on the basis of two kinds of features:

- *Demographics*, using Latino and Black CVAP shares; and
- *Partisan lean*, obtained by averaging the Democratic shares of the 2016 and 2012 major-party presidential vote, with the more recent general election weighted twice as heavily as the older one.

We begin with a (non-VRA) ensemble of 500,000 plans, then extract the districts from each to make a large dataset, containing 997,163 districts after de-duplication. For each district, we compute its statewide weighted effectiveness score  $s^{\text{state}}$ . We randomly separate these districts into training data (80%) and data points held back for testing and validation (20%).

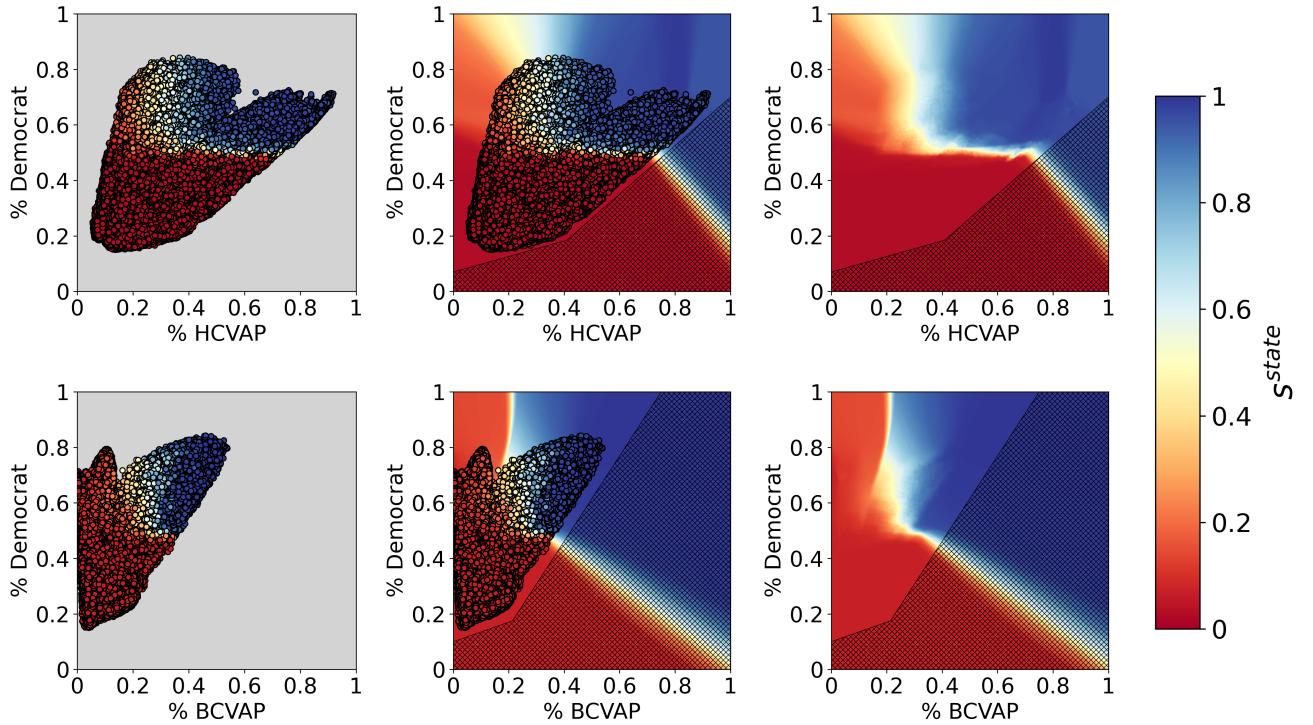


Figure 9: The top row refers to effectiveness for Latino voters and to Latino CVAP; the bottom row to corresponding statistics for Black voters. Two-dimensional scatterplots (left column) show a collection of districts drawn from a non-VRA ensemble, arranged by Latino or Black CVAP share on the  $x$  axis and partisan lean on the  $y$  axis, then colored by their  $s^{\text{state}}$  score for Latino- or Black-effectiveness, respectively. The  $k$ -nearest-neighbors (KNN) method is “trained” on that data to infer approximate scores for all possible positions in the square (shown with the training data in the center figures and without it at right). The hatched areas in the center and right-hand plots contain no labeled data points, so the KNN estimates are less meaningful in those areas.

We attempted several kinds of models. A  $k$ -nearest neighbors (KNN) model assigns a value to each point based on the  $k$  points in the training data that are closest to its location. This can be thought of as a predicted effectiveness score for districts that may be proposed in the future. The choice of  $k$  is made by a validation step that attempts many different values and chooses the one that provides the highest accuracy.<sup>31</sup> For the regression, the learned value assigned to a point is the average value of its  $k$  nearest neighbors, while the yes/no classification is made by selecting the majority label among those neighbors.

The outcomes of two-dimensional KNN regression are shown in Figure 9. They show a complicated district-level relationship between effectiveness (color), Latino or Black CVAP shares ( $x$  axis), and partisan lean ( $y$  axis). If the effectiveness of districts could be captured with CVAP shares alone, we would see a vertical line dividing the effective (blue) from the ineffective (red) zones. If overall partisanship were a good predictor on its own, we might see a horizontal dividing line; this is not the case, but we note that partisanship alone is more predictive for Latino effectiveness. If

<sup>31</sup>To be precise, we use  $m$ -fold cross-validation with  $m = 10$ , then choose the  $k$  for KNN with the best average  $r^2$  and mean squared error (MSE) over those ten-fold trials. Using those values of  $k$ , the final accuracy estimates use the full set of training data and are then corroborated against the withheld testing data.

effectiveness could be expressed in a simple linear relationship between partisan lean and CVAP, we would see a straight line of some slope separating the blue and red regions. Instead, we see a more complicated frontier with a large zone of ambiguity, especially in Latino effectiveness.<sup>32</sup>

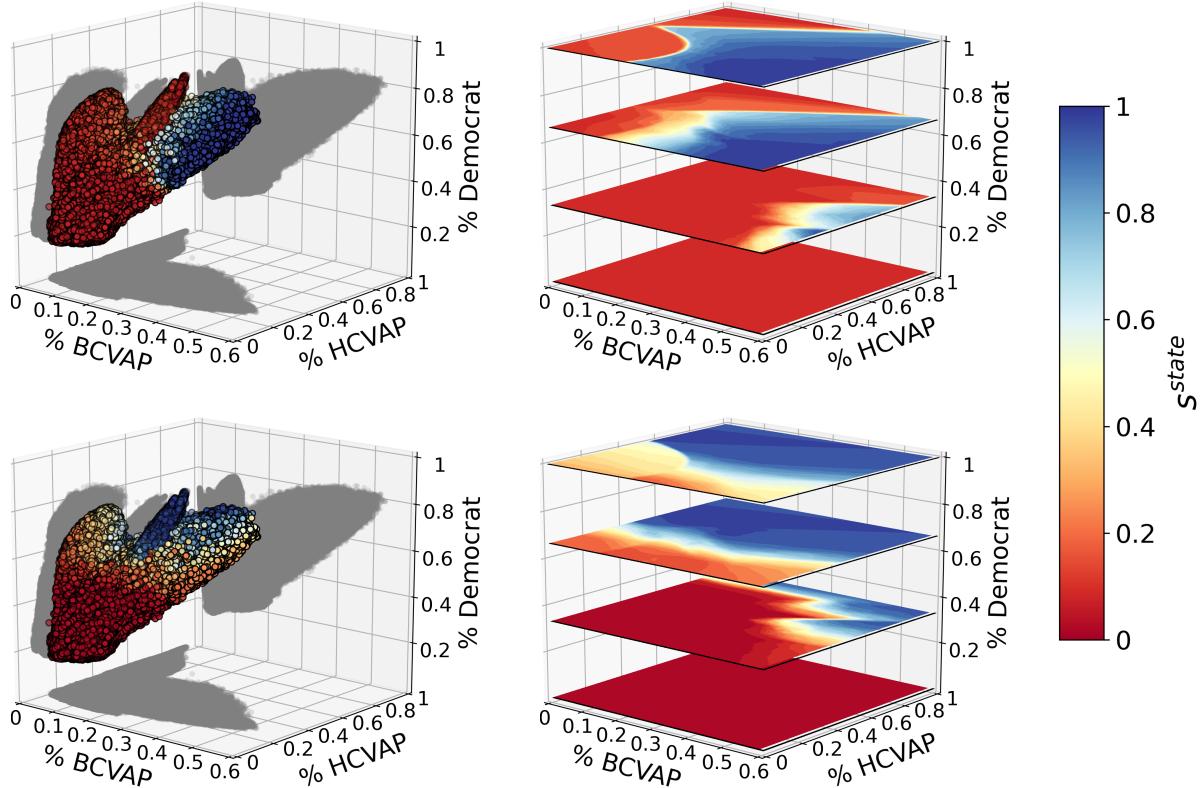


Figure 10: KNN regression for a three-dimensional scatterplot of district effectiveness.

Because Texas has two sizable minority groups, and Latino and Black voters often have overlapping electoral preferences, we might hope to do better by taking both groups' CVAP shares into account simultaneously. To this end, Figure 10 shows the same kind of regressions in three dimensions: Latino CVAP, Black CVAP, and the same measure of partisan lean. These plots still reveal complex, nonlinear frontiers and significant zones of ambiguity.

Further pattern-recognition results using various models for regression and classification are found in Supplement G. Together, these methods indicate that scores built from our involved electoral methodology do not easily reduce to combinations of CVAP demographics and general-election partisan lean. This leads us to conclude that electoral complexity, perhaps especially the dynamics of actual primary elections, is playing an ineliminable role in our determination of district effectiveness.

---

<sup>32</sup>Grofman, Handley, and Lublin (2001) studied what amounts to effectiveness classification in a similar feature space nearly 20 years ago, positing an “elbow” or V-shaped frontier of effectiveness. For a comparison of our classification results with their framework, see Supplement G.

## 7 Closing the representation gap

Finally, we return to where this Article began: the underrepresentation of communities of color at both the federal and state level. The algorithmic techniques described in this Article can be readily reconfigured to point the way to maps that are likely to promote significant gains in minority representation.

### 7.1 Searching for higher effectiveness

Recall first that our VRA-conscious ensembles are made by imposing yes/no validity constraints rather than a probabilistic tilt or bias: the proposal of new plans is made without regard to race, and the validity criteria are given by a threshold test, with no preference for plans that exceed the threshold by a wider margin. It is therefore unsurprising that this procedure does not on its own favor the creation of plans that greatly surpass the status quo in minority electoral opportunities. But—so long as districts are population-balanced, contiguous, reasonably compact, and constructed largely or entirely from intact precincts, as is the case across all our ensembles—maps generating rough proportionality for all sizable minority groups might well be the ones that actually minimize legal exposure under both the VRA and the Equal Protection Clause.

By shifting to an algorithm that has a tilted acceptance function favoring increased minority electoral opportunities, we found it to be straightforward to create maps that fully meet (or even exceed) rough proportionality simultaneously for multiple minority groups. For example, in Texas we were able to create maps that are effective enough to typically meet rough proportionality simultaneously for both Latino and Black voters, while not sacrificing districts to double-counting—i.e., while achieving near-proportionality for people of color overall as well as for each group individually. A *heuristic optimization* algorithm can preferentially accept maps with higher minority effectiveness. We carried this out with the general “short bursts” strategy outlined in [Cannon, Goldbloom-Helzner, Gupta, Matthews and Suwal, 2020](#); for details, see Supplemental Section H.

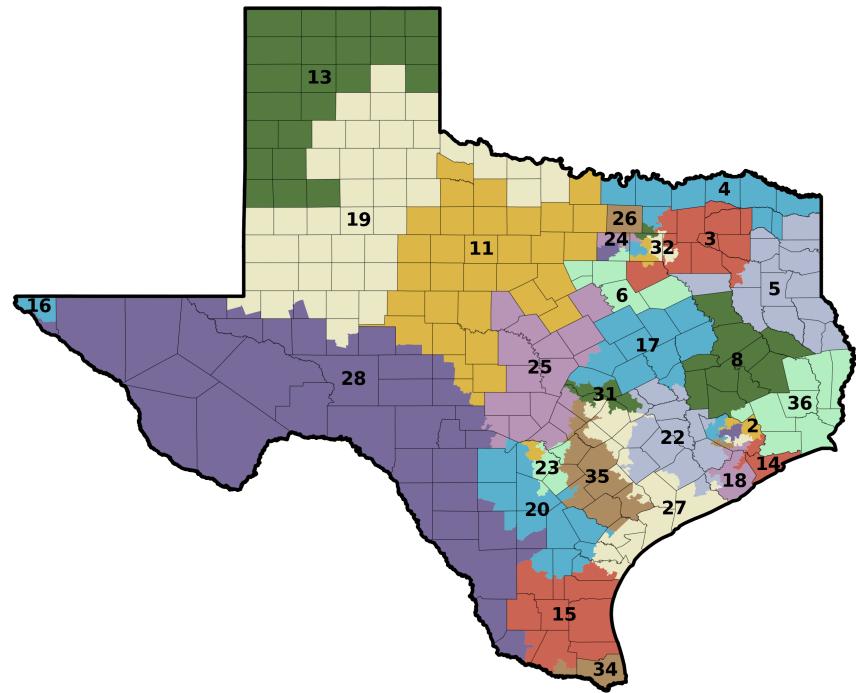
To be clear: maps proposed for adoption should be developed through human deliberation based on significant community input and a broader range of criteria and values than our algorithm incorporates. No map plucked from an ensemble is likely to satisfy all human desiderata off the shelf. But just to demonstrate that a map with eight Latino-effective districts and four Black-effective districts can be replaced by one with (at least) ten and five such districts, respectively, we examine one demonstration plan found in a local search.

### 7.2 A demonstration plan

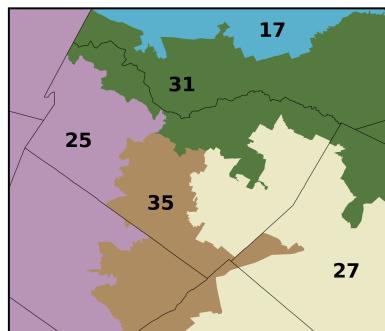
Our demonstration plan is depicted in Figure 11, and its effectiveness statistics by district are shown in Table 6.

We emphasize that this map is not intended to be an ideal map. But it does show that a carefully drawn plan could be dramatically fairer for historically underrepresented minority groups in Texas. We call it a “demonstration map” because it demonstrates that the shortfall of minority representation in the status quo map can be cured. The failure to do so can be attributed not to geography or law, but only to line-drawing.

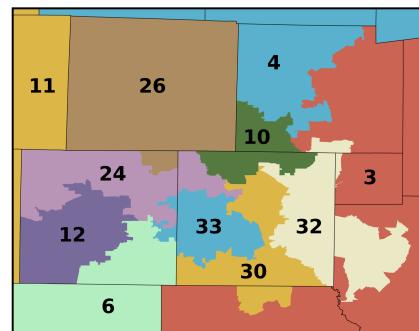
In Table 6, we have *uncoupled* the primary and general elections, to give a more detailed view of the electoral history of these districts. In other words, this table shows the primary/runoff success independent of the general-election outcome, while our effectiveness-scoring system requires wins in both the primary (or primary and runoff) and the general, to be counted as a success. The table shows that, using any of the three scores, the demonstration plan contains at least 11, and



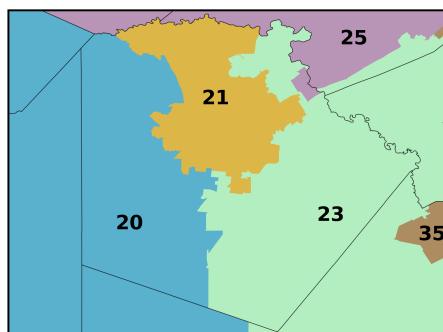
(a) Entire Plan



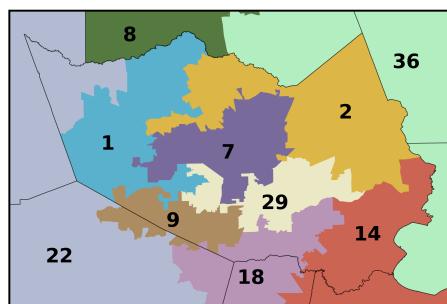
(b) Austin Area



(c) Dallas-Fort Worth Area



(d) San Antonio Area



(e) Houston Area

Figure 11: An interesting demonstration plan found by heuristic optimization.

## Demonstration Plan

CD	Location	HCVAP	Latino Effective			BCVAP	Black Effective			WCVAP	14 Primaries		14 Gen (Dem.)
			$s_{unw}$	$s_{state}$	$s_{dist}$		$s_{unw}$	$s_{state}$	$s_{dist}$		Latino	Black	
7	Houston	36.5	77	65	77	25.5	70	58	31	31.4	9-13	9-10	14
9	Houston	23.3	40	30	33	28.6	78	66	75	31.5	10-12	10-12	14
15	South Texas	78.8	97	98	96	1.7	8	9	6	17.5	12-14	10-11	14
16	EI Paso	76.1	99	99	97	4.2	11	12	10	17.4	13-14	11-14	14
18	Houston	32.0	66	59	63	30.7	76	77	69	30.4	10-13	10-12	14
20	San Antonio	60.6	77	82	76	5.5	10	11	9	30.9	12-14	12-13	9
21	San Antonio	47.5	35	74	79	5.6	8	8	8	42.9	12-14	10-14	7
23	San Antonio	51.1	77	82	79	10.7	14	15	14	34.7	12-14	10-12	9
27	Austin/Gulf Coast	39.8	84	85	85	8.8	17	16	18	47.7	12-13	10-14	13
28	South/West Texas	81.4	91	95	96	1.0	7	8	6	16.6	11-14	9-11	14
29	Houston	33.4	70	57	75	25.5	70	58	52	35.5	9-11	9-12	14
30	DFW	15.5	20	15	13	31.8	85	84	69	48.5	9-10	10-11	14
32	DFW	24.1	24	26	28	24.4	52	67	62	44.9	10-13	12-14	10
33	DFW	37.0	85	80	66	32.9	96	97	88	25.1	10-11	13	14
34	South Texas	86.7	97	98	97	0.4	6	7	5	12.3	11-14	9-11	14
35	Austin	30.7	62	62	67	4.8	10	10	9	60.6	11-13	9-10	14

### District 27 (with statewide candidates of choice)

		Primary Election		Primary Runoff Election		General Election	
		Latino-Pref.	Winner	Latino-Pref.	Winner	Latino-Pref.	Winner
President	2012	Obama	Obama ✓			Obama	Obama ✓
U.S. Senator	2012	Sadler	Sadler ✓	Sadler	Sadler ✓	Sadler	Sadler ✓
U.S. Senator	2014	Alameel	Alameel ✓	Alameel	Alameel ✓	Alameel	Cornyn ×
Governor	2014	Davis	Davis ✓			Davis	Davis ✓
Ag. Commissioner	2014	Friedman	Friedman ✓	Hogan	Hogan ✓	Hogan	Hogan ✓
RR Commissioner	2014	Brown	Brown ✓			Brown	Brown ✓
President	2016	Clinton	Clinton ✓			Clinton	Clinton ✓
RR Commissioner	2016	Yarbrough	Yarbrough ✓	Yarbrough	Yarbrough ✓	Yarbrough	Yarbrough ✓
U.S. Senator	2018	O'Rourke	O'Rourke ✓			O'Rourke	O'Rourke ✓
Governor	2018	Valdez	Valdez ✓	Valdez	Valdez ✓	Valdez	Valdez ✓
Lieutenant Governor	2018	Cooper	Collier ×			Collier	Collier ✓
Comptroller	2018	Mahoney	Chevalier ×			Chevalier	Chevalier ✓
Land Commissioner	2018	Suazo	Suazo ✓			Suazo	Suazo ✓
RR Commissioner	2018	McAllen	McAllen ✓			McAllen	McAllen ✓

Table 6: The demonstration plan has up to 16 minority-effective districts, as shown in the top table, while the enacted plan has no more than 11 to 13 (compare Table 4 and accompanying text). Scores over 60% have darker shading, and scores in the 50–60% range have lighter shading. The frequency of primary and general election wins by minority-preferred candidates is shown in the last two columns. Because different candidates of choice can be identified by the statewide and district-specific method, the number of successes is given as a range. The bottom table shows that candidates preferred by Latino voters statewide prevailed in District 27 in 12 of the 14 primaries, 5 of the 5 runoffs, and 13 of the 14 general elections. (With the candidates of choice inferred from the district-specific method, there are 13 primary successes.)

perhaps as many as 13, effective districts for Latino voters and at least 5, and perhaps as many as 7, effective districts for Black voters. Because one district in the Dallas area (District 33) and at least one in the Houston area (District 18) appear to be effective for both Black and Latino voters, the total number of minority-effective districts in the demonstration plan is 14, 15, or 16, depending on whether you rely on the unweighted, statewide, or district scores, respectively. Only 1 of the 16 districts is majority-White by CVAP.

Several of these 16 highlighted districts have demographics and effectiveness scores similar to those of the minority-effective districts in the current enacted plan (compare Table 4). However, in the current enacted plan, every district except Congressman Veasey's District 33 follows the rule that districts marked effective for Latino voters have HCVAP over 50% and those marked effective for Black voters have BCVAP over 40%. By contrast, the demonstration plan presented here features several effective districts with lower Latino and Black population percentages. For example, the Austin-based District 27 is a Latino-effective district with an HCVAP a shade under 40%, and the Houston-based District 9 is a Black-effective district with a BCVAP of only 28.6%. We emphasize that each of those demonstration districts earned its effectiveness score by voting for the Latino- or Black-preferred candidates, respectively, in nearly every statewide election conducted in the last decade.

This map refutes the notion that demographics is destiny when it comes to Texas congressional districts. It contains districts that are majority-minority but not minority-effective (District 2), majority-White but Latino-effective (District 35), plurality-White but Black-effective (Districts 9, 30, and 32) or Latino-effective (Districts 27 and 29), and plurality-Latino but Black-effective (the two coalition districts, 18 and 33). There are also districts that are reliably Democratic but are not effective for either Latino voters or Black voters (Districts 12 and 31).

Table 6 takes a single district and brings us back to the most basic facts about it: whether the minority-preferred candidates actually won the most votes. We use as an example the plurality-White but Latino-effective District 27, which starts in East Austin and stretches south toward the Gulf Coast. For 11 of the 14 offices, the candidate preferred by Latino voters statewide prevailed at every step in District 27: primary, runoff (when there was one), and general. In the 2014 general election, however, the Latino-preferred Democratic nominee David Alameel failed to carry District 27 against Republican incumbent U.S. Senator John Cornyn; and in the 2018 Democratic primaries for Lieutenant Governor and Comptroller, the candidates preferred by Latino voters statewide (Michael Cooper and Tim Mahoney, respectively) failed to carry the district. This district generated Latino-effectiveness scores of about 84 or 85%, far above our threshold for effectiveness (60%) but below the scores for the map's four most heavily Latino districts, which consistently exceeded 90%.

### 7.3 Aggregate effectiveness

The use of a search technique tailored to raise the number of minority-effective districts might lead us to wonder about the effect on the rest of the map. With respect to demographics alone, redistricting is a fixed-sum activity: there are only so many Latino citizens of voting age in the state, so building more districts with high HCVAP means there is less remaining HCVAP to distribute across the other districts. We might worry that we can only secure a larger number of effective districts by draining opportunities for coalitional influence from the rest of the state. But this is not the case. Because of the highly nonlinear relationship between demographics and effectiveness (see §6), it is possible to create some plans with a greater overall effectiveness than others.

To see this, let us consider the sum of the effectiveness scores for all 36 Texas congressional districts. Because each district has a score between 0 and 1, the sum will fall between 0 and 36. To

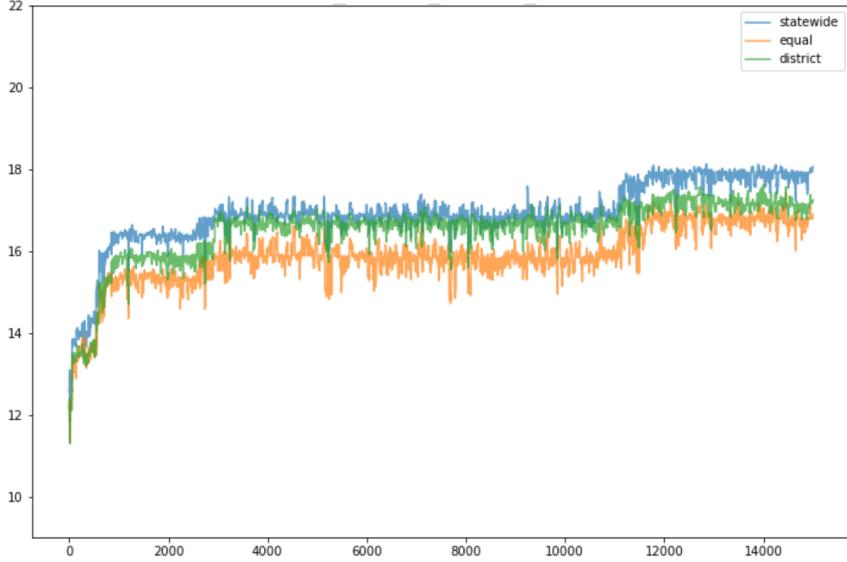


Figure 12: This trace plot shows a kind of aggregate effectiveness for Latino and Black voters, formed by summing Latino and/or Black effectiveness scores over all 36 districts. This aggregate effectiveness trends up markedly over the course of a heuristic-optimization run that preferentially accepts plans with more districts effective for at least one minority group under the  $s^{\text{state}}$  score. This drives up the  $s^{\text{state}}$  score (in blue) most, with the other two scores following behind. (See Supplement H for details on related optimization runs.)

the extent that a group’s effectiveness scores behave like probabilities of electoral success, the sum over the 36 districts can be regarded as the *expected value* for the group in a given election. This expected-value score takes into account the probability but not certainty of electoral success in the effective districts, and also includes contributions from other districts in which an effectiveness score could fall well below .5 yet still reflect real political influence and a chance to win.

The enacted plan has an expected-value score a bit under 12, driven by 11 highly effective districts. After a few thousand steps of a heuristic-optimization run (shown in Figure 12), the expected-value score is well over 15, usually over 16, and it is possible to drive the expectation up near 18 in the score being optimized. Our demonstration plan has an expectation of nearly 17, which tracks with the 16 districts highlighted in Table 6.

We find that, with respect to electoral opportunity, districting is not a fixed-sum game. We can find plans that combine Latino and Black voters with other population (including Asian-American and White voters who tend to support the same candidates) in ways that lead to effective combinations. We can create safe minority districts, likely-to-elect minority districts, and some minority influence districts in a way that is especially beneficial in aggregate. This is a departure from the narrower focus on effectiveness that is directly relevant for VRA compliance, but may still point the way to a more coalitional expansion of minority opportunities beyond the demands of the law.

## 8 Conclusion

The principal goal of this project is the design and study of a protocol for building ensembles of alternative districting plans, taking closely into account the law of race and redistricting. We do

this by using longitudinal electoral data, one of a choice of effectiveness scores, and a constrained district-generation algorithm.

No inclusion criterion assessed by a computer could perfectly track the conclusions of a court (not least because of variation in the judiciary itself), but ours is constructed to give us strong justification for describing it as a *representative sample* of the universe of VRA-compliant plans. We have pursued this objective in a way that also avoids overreliance on purely demographic targets that might run afoul of the Equal Protection Clause.

The structure of our protocol is described in §4, and a detailed case study for Texas congressional districts is detailed in §5. In §6 we confirm that the role played by the extensive electoral data is not easily replaced by simpler proxies. And in §7 we explore the use of similar techniques to minimize underrepresentation for minority groups—showing in particular that pushing to find plans that go the farthest to cure long-standing underrepresentation is a markedly different task from creating collections of alternatives that pass VRA muster. Studying the conditions of political and human geography that make it possible to attain near-proportionality is an interesting direction for future work.

With a detailed case study in the large, complex state of Texas, we confirm that our implementation lets us carry out the work on a time scale suitable for all stages of redistricting, from considering plans for possible adoption all the way to challenging them in litigation. We have made careful use of error estimates, performed tests of quality for ensemble generation, and confirmed robustness of the method across reasonable variations in the steps. By making our code and data public ([MGGG Redistricting Lab, 2020a](#)), we aim to make it possible for other researchers and practitioners to use this method on the ground.

This tool now makes it possible to assess proposed districting plans in racially diverse states against a baseline that takes the Voting Rights Act and the Equal Protection Clause into account. The computational tools for redistricting are continually becoming both more powerful and more refined, facilitating the creation of new maps that better meet our ideals of fairness and helping to understand maps in the context of realistic alternatives. By using novel tools in combination with renewed commitment to safeguarding minority representation, we can come closer than ever to the goal articulated by John Adams almost 250 years ago, in the midst of the American Revolution: to make our representative assemblies “in miniature an exact portrait of the people at large” ([Adams, 1776](#), 108).

## Acknowledgments

The authors thank Jessie Amunson, Ash Bharatkumar, Charlotte Hirsch, Justin Levitt, JN Matthews, James Murphy, Heather Rosenfeld, Parker Rule, Gabe Schoenbach, Doug Spencer, Adam Unikowsky, and Thomas Zeitzoff for extremely helpful and stimulating conversations and in some cases for invaluable technical support of this project. Many members and student interns affiliated with the MGGG Redistricting Lab and the Voting Rights Data Institute have been involved in long-term data preparation, and we are warmly grateful for their collaboration.

Racial identification of candidates was made possible by a dataset purchased from Carl Klerner ([klernerpolitics.org](#)).

*This Article is dedicated to the memory of Rice University sociology professor Chandler Davidson (1936–2021), who fought successfully for a half century to protect Latino and Black voting rights and to expand minority electoral opportunities in Texas and throughout the United States.*

## References

- Abbott v. Perez.* 2018. 138 S. Ct. 2305.
- Abrams v. Johnson.* 1997. 521 U.S. 74.
- Adams, John. 1776. Thoughts on Government. In *The Founders' Constitution*, ed. Philip B. Kurland and Ralph Lerner. University of Chicago Press (1987).
- Alabama Legislative Black Caucus v. Alabama.* 2015. 575 U.S. 254.
- Bartlett v. Strickland.* 2009. 556 U.S. 1.
- Becker, Amariah and Justin Solomon. 2020. Redistricting Algorithms. In *Political Geometry*, ed. M. Duchin and O. Walch. Birkhäuser. Under contract.
- Bethune-Hill v. Virginia State Bd. of Elections.* 2017. 137 S. Ct. 788.
- Bialik, Kristen. 2019. “Pew Research Center: For the Fifth Time in a Row, the New Congress Is the Most Racially and Ethnically Diverse Ever.” <https://www.pewresearch.org/fact-tank/2019/02/08/for-the-fifth-time-in-a-row-the-new-congress-is-the-most-racially-and-ethnically-diverse-ever/>.
- Brown v. Thomson.* 1983. 462 U.S. 835.
- Bush v. Vera.* 1996. 517 U.S. 952.
- Campos v. City of Baytown.* 1988. 840 F.2d 1240 (5th Cir.).
- Cannon, Sarah, Ari Goldbloom-Helzner, Varun Gupta, JN Matthews and Bhushan Suwal. 2020. “Voting Rights, Markov Chains, and Optimization by Short Bursts.” On file with authors.
- Cannon, Sarah, Moon Duchin, Dana Randall and Parker Rule. 2020. “A reversible recombination chain for redistricting.” On file with authors.
- Carter, Daniel, Gregory Herschlag, Zach Hunter and Jonathan Mattingly. 2019. “A Merge-Split Proposal for Reversible Monte Carlo Markov Chain Sampling of Redistricting Plans.” *arxiv:1911.01503* .
- Chen, Jowei. 2017. “The Impact of Political Geography on Wisconsin Redistricting: An Analysis of Wisconsin’s Act 43 Assembly Districting Plan.” *Election Law Journal* 16(4).
- Chen, Jowei and Nicholas O. Stephanopoulos. 2021. “The Race-Blind Future of Voting Rights.” *Yale Law Journal* 130(4):862–946. Available at [https://www.yalelawjournal.org/pdf/CS\\_fkm4m4bz.pdf](https://www.yalelawjournal.org/pdf/CS_fkm4m4bz.pdf).
- Common Cause v. Lewis.* 2019. No. 18 CVS 014001, 2019 WL 4569584 (N.C. Super. Ct., Wake Cty. Sept. 3) (three-judge court).
- Cooper v. Harris.* 2017. 137 S. Ct. 1455.
- Cox v. Larios.* 2004. 542 U.S. 947.
- DeFord, Daryl and Moon Duchin. 2019. “Redistricting Reform in Virginia: Districting Criteria in Context.” *Virginia Policy Review* XII:120–146.

- DeFord, Daryl and Moon Duchin. 2020. Random Walks and the Universe of Districting Plans. In *Political Geometry*, ed. M. Duchin and O. Walch. Birkhäuser. Under contract.
- DeFord, Daryl, Moon Duchin and Justin Solomon. 2021. “Recombination: A Family of Markov Chains for Redistricting.” *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.eb30390f>.
- Duchin, Moon and Bridget Tenner. 2018. “Discrete Geometry for Electoral Geography.” <https://arxiv.org/>.
- Duchin, Moon and Douglas M. Spencer. 2021. “Models, Race, and the Law: A Response to ‘The Race-Blind Future of Voting Rights.’” *Yale Law Journal Forum* 130:744–97. Available at [https://www.yalelawjournal.org/pdf/DuchinandSpencerEssay\\_q6t1y3pq.pdf](https://www.yalelawjournal.org/pdf/DuchinandSpencerEssay_q6t1y3pq.pdf).
- Elmendorf, Christopher S., Kevin M. Quinn and Marisa J. Abrajano. 2016. “Racially Polarized Voting.” *University of Chicago Law Review* 83. Available at <https://chicagounbound.uchicago.edu/uclrev/vol83/iss2/2>.
- Fifield, Benjamin, Michael Higgins, Kosuke Imai and Alexander Tarr. 2020. “Automated Redistricting Simulation Using Markov Chain Monte Carlo.” *Journal of Computational and Graphical Statistics*. DOI: 10.1080/10618600.2020.1739532.
- FiveThirtyEight. 2018. “The Atlas of Redistricting.” <https://projects.fivethirtyeight.com/redistricting-maps/>.
- Grofman, Bernard, Lisa Handley and David Lublin. 2001. “Drawing Effective Minority Districts: A Conceptual Framework and Some Empirical Evidence.” *North Carolina Law Review* 79(5):1383–430.
- Hebert, J. Gerald, Paul M. Smith, Martina E. Vandenberg and Michael B. DeSanctis. 2010. *The Realist’s Guide to Redistricting: Avoiding the Legal Pitfalls*. 2d ed. ABA.
- Herschlag, Gregory, Han Sung Kang, Justin Luo, Christy Vaughn Graves, Sachet Bangia, Robert Ravier and Jonathan C. Mattingly. 2020. “Quantifying Gerrymandering in North Carolina.” *Statistics and Public Policy* 7:30–38.
- History, Art & Archives, U.S. House of Representatives, Office of the Historian. 2020a. “Black-American Members by Congress, 1870–Present.”. Available at <https://history.house.gov/Exhibitions-and-Publications/BAIC/Historical-Data/Black-American-Representatives-and-Senators-by-Congress/>.
- History, Art & Archives, U.S. House of Representatives, Office of the Historian. 2020b. “Hispanic-American Representatives, Senators, Delegates, and Resident Commissioners by Congress, 1822–Present.”. Available at <https://history.house.gov/Exhibitions-and-Publications/HAIC/Historical-Data/Hispanic-American-Representatives,-Senators,-Delegates,-and-Resident-Commissioners-by-Congress/>.
- Iowa Code § 42.4.4. N.d.
- Iowa Const. art. III, § 37. N.d.
- Johnson v. De Grandy*. 1994. 512 U.S. 997.

*Karcher v. Daggett.* 1983. 462 U.S. 725.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton University Press.

King, Gary, Ori Rosen and Martin A. Tanner. 1999. “RBinomial-Beta Hierarchical Models for Ecological Inference.” *Sociological Methods and Research* 28:61–90.

Klarner, Carl. 2019. “Racial Identification of State Legislators 2001–2019.” Unpublished data set. Purchased from <http://klarnerpolitics.org/>.

Lau, Olivia, Ryan T. Moore and Michael Kellermann. 2020. “eiPack: Ecological Inference and Higher-Dimension Data Management (R package).”. Version 0.2-1. Available at <https://CRAN.R-project.org/package=eiPack>.

*League of Women Voters of Mich. v. Benson.* 2019. 373 F. Supp. 3d 867 (E.D. Mich.) (three-judge court), summarily vacated and remanded *sub nom. Chatfield v. League of Women Voters of Mich.*, 140 S. Ct. 429.

*League of Women Voters v. Commonwealth.* 2018. 178 A.3d 737 (Pa.).

*LULAC v. Perry.* 2006. 548 U.S. 399.

McCartan, Cory and Kosuke Imai. 2020. “Sequential Monte Carlo for Sampling Balanced and Compact Redistricting Plans.”.

MGGG Redistricting Lab. 2018a. “Comparison of Districting Plans for the Virginia House of Delegates.” *White Paper* . <https://mggg.org/VA-report.pdf>.

MGGG Redistricting Lab. 2018b. “GerryChain Python package.” *GitHub repository* . <https://github.com/mggg/GerryChain>.

MGGG Redistricting Lab. 2018c. “MAUP Python package.” *GitHub repository* . <https://github.com/mggg/MAUP>.

MGGG Redistricting Lab. 2018d. “Study of Voting Systems for Santa Clara, CA.” *White Paper* . <https://mggg.org/santaclara>.

MGGG Redistricting Lab. 2019a. “Findings on the City of Lowell’s Election Systems.” *White Paper* . <https://mggg.org/lowell>.

MGGG Redistricting Lab. 2019b. “Study of Reform Proposals for Chicago City Council.” *GitHub repository* . <https://mggg.org/chicago>.

MGGG Redistricting Lab. 2020a. “Computational VRA.” *GitHub repository* . [https://github.com/mggg/VRA\\_article](https://github.com/mggg/VRA_article).

MGGG Redistricting Lab. 2020b. “GerryChainJulia Julia package.” *GitHub repository* . <https://github.com/mggg/GerryChainJulia>.

*Miller v. Johnson.* 1995. 515 U.S. 900.

National Conference of State Legislatures. 2020. “State Legislator Demographics.” <https://www.ncsl.org/research/about-state-legislatures/state-legislator-demographics.aspx>.

*Ohio A. Philip Randolph Institute v. Householder.* 2019. 373 F. Supp. 3d 978 (S.D. Ohio) (three-judge court), *summarily vacated and remanded*, 140 S. Ct. 101.

*Rucho v. Common Cause.* 2019. 139 S. Ct. 2484.

*Session v. Perry.* 2004. 298 F. Supp. 2d 451 (E.D. Tex.) (three-judge court) (*per curiam*), *vacated and remanded on other grounds sub nom. Jackson v. Perry*, 543 U.S. 941.

*Shaw v. Reno.* 1993. 509 U.S. 630.

*Smith v. Clinton.* 1988. 687 F. Supp. 1310 (E.D. Ark.) (three-judge court), *summarily aff'd*, 488 U.S. 988.

*Tennant v. Jefferson County Comm'n.* 2012. 567 U.S. 758 (*per curiam*).

Texas Legislative Council. 2020. “Capitol Data Portal.” Available at <https://data.capitol.texas.gov>.

*Thornburg v. Gingles.* 1986. 478 U.S. 30.

*Unger v. Manchin.* 2002. 536 U.S. 935.

U.S. Bureau of the Census. 2019a. “American Community Survey: Citizen Voting-Age Population, by Selected Characteristics.” <https://data.census.gov/cedsci/table?q=S2901&tid=ACSST1Y2019.S2901>.

U.S. Bureau of the Census. 2019b. “American Community Survey: Citizen Voting-Age Population, by State.” [https://www.census.gov/library/visualizations/2016/comm/citizen\\_voting\\_age\\_population.html](https://www.census.gov/library/visualizations/2016/comm/citizen_voting_age_population.html).

U.S. Const. amend. XIV. N.d.

*Voinovich v. Quilter.* 1993. 507 U.S. 146.

VRA. 2020. 52 U.S.C. § 10301 et seq.

## A Compressing probability distributions

In this appendix, we detail a method to record precinct-level probabilistic information in a condensed form, so that the distributions can be efficiently recovered at every step of a Markov chain. The strategy is to compress a histogram into octiles, storing only eight “bars” instead of dozens or hundreds.

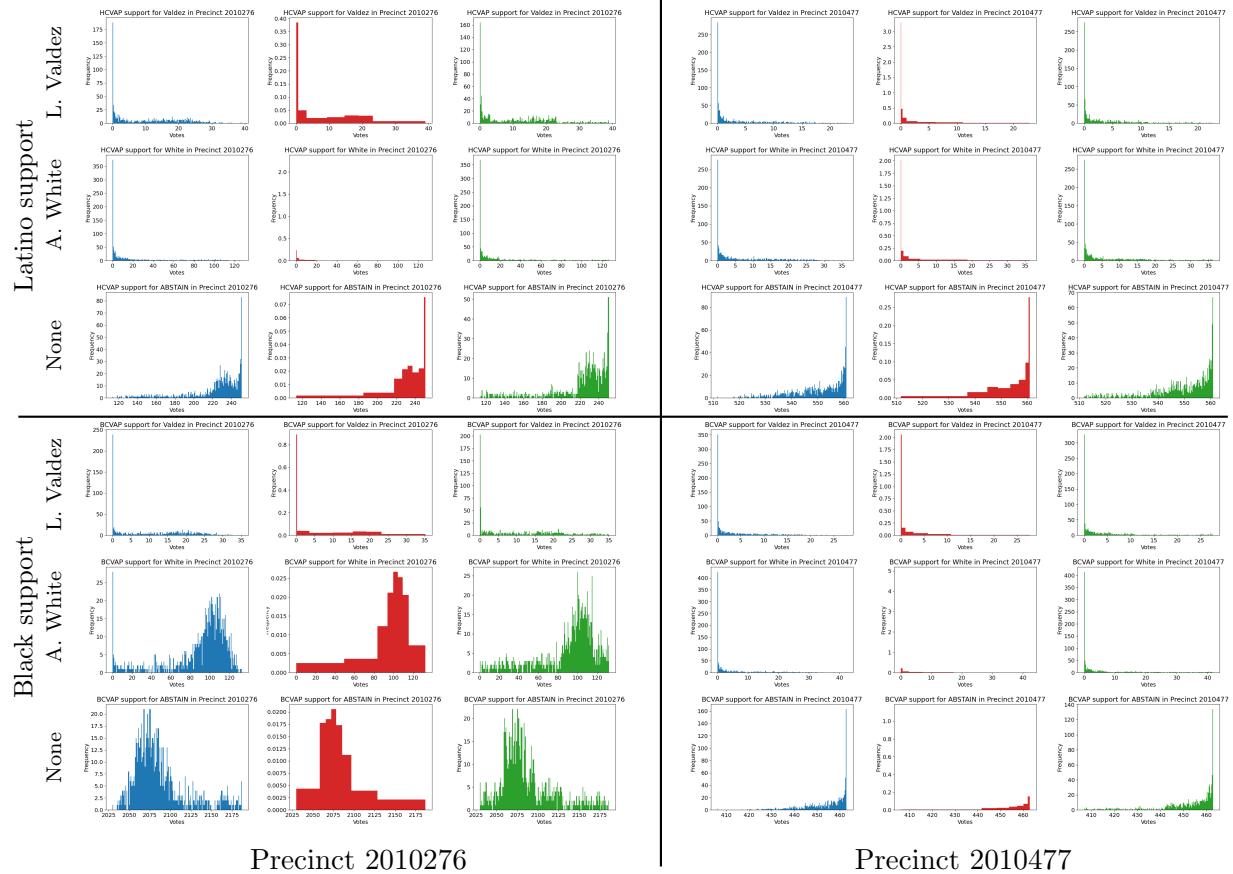


Figure 13: Original (blue), compressed (red), and reconstituted (green) probability distributions. The model incorporates a turnout estimate by including “None” as a candidate. The reconstituted distributions reflect the original histograms remarkably closely, even though only eight histogram bars were stored in each case.

Figure 13 demonstrates the precinct-level EI estimation process for two precincts in Texas. This example comes from the 2018 Democratic gubernatorial primary runoff. The left plot (blue) shows estimated support levels for Valdez, White, and “None” (CVAP minus the vote for the candidates) by Latino and Black voters, shown with a *detailed histogram* in which each number of votes is recorded separately with its observed frequency for 1000 draws from EI. The center plot (red) shows the *coarse histograms* approximating the distributions of these EI draws, by binning them into eighths (octiles). In particular, by saving the values of the end points of each 12.5% interval, we can approximate the vote-count distribution by saving only nine values. In the plot, the vote axis is divided at these endpoints, and each bar has the same mass. The right plot (green) shows the samples re-drawn from the coarse histogram, performed quickly during a ReCom run.

The re-draws closely resemble the original samples, as shown by how closely the reconstituted

histograms (green) match the original detailed histograms (blue). Notably, this is true regardless of the shape of the distributions. By contrast, a common practice is to assume that certain types of random draws are reasonably approximated by normal distributions, which can be saved very efficiently using only two values (a mean and variance) and then easily resampled. But these examples show that vote-count distributions can be highly skewed (and in inconsistent ways), which would not be well approximated by normal distributions. With our resampling methods, however, we can recover a very close estimate to the original distribution from a highly compressed data format, without having to make any assumptions about the shapes of vote-count distributions.

Finally, we note that this kind of EI method does implicitly rely on an assumption of independence between these outcomes. That is, even though we recover the individual vote-count distributions, we do not attempt to recover the *joint distribution* of these counts across candidates. For example, a precinct-level EI draw that has a very high vote share for Valdez is likely to have a low share for White, but these interdependencies are not included in the model, which merely recovers the individual histograms.

## B Logit adjustment

Here, we describe the details for the calibration step (logit adjustment) for Texas. For each score component, we calibrated the raw score with observed performance using logistic regression (logit) models. Specifically, we measured the raw effectiveness scores for each district in each of the three enacted plans (congressional, state Senate, and state House) and began by labeling each district with a 0 or 1 based on observed performance of each of these districts across all elections held using these plans. For example, in Texas, this gives us 822 data points for each score component (145 congressional, 77 state Senate, and 600 state House).

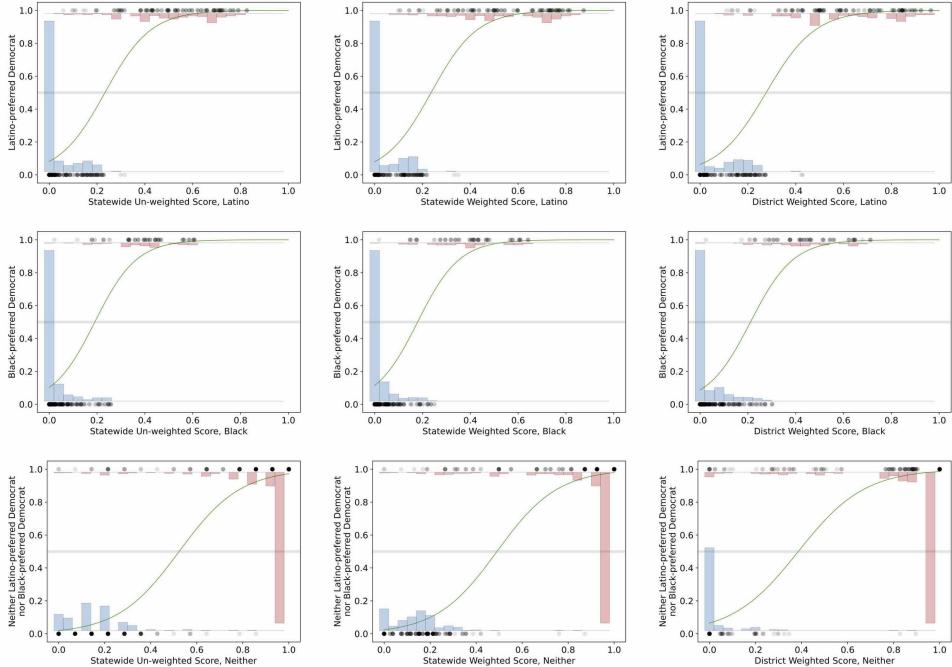


Figure 14: Logit curves calibrating Latino, Black, and Neither effectiveness for the unweighted, weighted/statewide, and weighted/district scores. Raw scores are on the  $x$  axis and calibrated scores are on the  $y$  axis.

We label each of these district–elections with a 1 in the Latino classification if if the candidate with the most votes was either a Latino Democrat or a Democrat in a district that is plurality-Latino by CVAP and a 0 otherwise; and similarly for the Black classification. The Neither label is given by the complement of the union of those success conditions. We then use these classifications to fit a logit model using `LogisticRegression` from the scikit-learn Python machine-learning library, with an L2 penalty and balanced class weights (to account for the large imbalance in class size). The fit logistics are of the form  $f(x) = \frac{1}{1+\exp(-ax-b)}$  for the  $(a, b)$  shown here.

	unw	state	dist
Latino	(10.4, 2.4)	(10.4, 2.5)	(9.8, 2.7)
Black	(11.4, 2.2)	(11.4, 2.0)	(11.2, 2.4)
Neither	(7.5, 3.9)	(7.6, 3.7)	(6.9, 2.7)

## C Candidates of choice in Texas elections

Table 7 shows the list of candidates for the Texas elections in our dataset, with the statewide candidate of choice marked for both Latino and Black voters. In the large majority of contests, these two groups agree on a candidate of choice.

Most statewide candidates of choice in this table are assessed 100% confidence because they came out ahead on each of the 1000 draws from the EI probability distribution computed in `eiPack`. The number of times that each listed candidate was drawn first is recorded in the table.

Year	Election	Type	Candidates	Leading Draws	Race	Party
2012	President	Primary	Obama	1000	1000	Black Democrat
			Richardson	0	0	White Democrat
			Ely	0	0	White Democrat
			Wolfe	0	0	White Democrat
2012	President	General	Romney	0	0	White Republican
			Obama	1000	1000	Black Democrat
2012	U.S. Senator	Primary	Allen	0	110	Black Democrat
			Hubbard	0	0	White Democrat
			Sadler	1000	832	White Democrat
			Yarbrough	0	58	Black Democrat
2012	U.S. Senator	Primary Runoff	Sadler	1000	1000	White Democrat
			Yarbrough	0	0	Black Democrat
2012	U.S. Senator	General	Cruz	0	0	Latino Republican
			Sadler	1000	1000	White Democrat
			Alameel	1000	1000	White Democrat
			Kim	0	0	Asian Democrat
2014	U.S. Senator	Primary	Rogers	0	0	Black Democrat
			Scherr	0	0	White Democrat
			Fjetland	0	0	White Democrat
			Alameel	1000	1000	White Democrat
			Rogers	0	0	Black Democrat
2014	U.S. Senator	Primary Runoff	Cornyn	0	0	White Republican
			Alameel	1000	1000	White Democrat
2014	U.S. Senator	General	Madrigal	0	0	Latino Democrat
			Davis	1000	1000	White Democrat
			Abbott	0	0	White Republican
			Davis	1000	1000	White Democrat
2014	Agriculture Commissioner	Primary	Fitzsimons	0	0	White Democrat
			Hogan	8	1000	White Democrat
			Friedman	992	0	White Democrat
			Friedman	0	0	White Democrat
2014	Agriculture Commissioner	Primary Runoff	Hogan	1000	1000	White Democrat
2014	Agriculture Commissioner	General	Miller	0	0	White Republican
			Hogan	1000	1000	White Democrat
			Henry	0	0	White Democrat
			Brown	1000	1000	Black Democrat
2014	Railroad Commissioner	Primary	Brown	1000	1000	Black Democrat
			Sitton	0	0	White Republican
			Judd	0	0	White Democrat
			De La Fuente	0	0	Latino Democrat
2016	President	Primary	Clinton	1000	1000	White Democrat
			Locke	0	0	White Democrat
			O'Malley	0	0	White Democrat
			Sanders	0	0	White Democrat
			Wilson	0	0	Black Democrat
			Hawes	0	0	Other Democrat
			Clinton	1000	1000	White Democrat
2016	President	General	Trump	0	0	White Republican

Year	Election	Type	Candidates	Leading	Draws	Race	Party
2016	Railroad Commissioner	Primary	Burnam	0	0	White	Democrat
			Yarbrough	900	1000	Black	Democrat
			Garrett	100	0	White	Democrat
2016	Railroad Commissioner	Primary Runoff	Yarbrough	836	1000	Black	Democrat
			Garrett	164	0	White	Democrat
2016	Railroad Commissioner	General	Yarbrough	1000	1000	Black	Democrat
			Christian	0	0	White	Republican
2018	U.S. Senator	Primary	Kimbrough	0	0	Black	Democrat
			Hernandez	0	0	Latino	Democrat
			O'Rourke	1000	1000	White	Democrat
2018	U.S. Senator	General	Cruz	0	0	Latino	Republican
			O'Rourke	1000	1000	White	Democrat
2018	Governor	Primary	Wakely	0	0	White	Democrat
			Clark	0	0	Other	Democrat
			Payne	0	0	White	Democrat
			Yarbrough	0	0	Black	Democrat
			Davis	0	0	Black	Democrat
			Ocegueda	0	0	Latino	Democrat
			Valdez	1000	1000	Latino	Democrat
			White	0	0	White	Democrat
			Mumbach	0	0	White	Democrat
			Valdez	1000	429	Latino	Democrat
2018	Governor	Primary Runoff	White	0	571	White	Democrat
			Abbott	0	0	White	Republican
2018	Governor	General	Valdez	1000	1000	Latino	Democrat
			Cooper	1000	1000	Black	Democrat
2018	Lt. Governor	Primary	Collier	0	0	White	Democrat
			Patrick	0	0	White	Republican
2018	Lt. Governor	General	Collier	1000	1000	White	Democrat
			Mahoney	1000	879	White	Democrat
2018	Comptroller	Primary	Chevalier	0	121	Black	Democrat
			Hegar	0	0	White	Republican
2018	Comptroller	General	Chevalier	1000	1000	Black	Democrat
			Suazo	1000	1000	Latino	Democrat
2018	Land Commissioner	Primary	Morgan	0	0	White	Democrat
			Bush	0	0	Latino	Republican
2018	Land Commissioner	General	Suazo	1000	1000	Latino	Democrat
2018	Railroad Commissioner	Primary	McAllen	1000	0	Latino	Democrat
			Spellmon	0	1000	Black	Democrat
2018	Railroad Commissioner	General	Craddick	0	0	White	Republican
			McAllen	1000	1000	Latino	Democrat

■ Preferred by Latino voters and Black voters   ■ Preferred by Latino voters   ■ Preferred by Black voters

Table 7: Texas elections used for analysis, with the statewide candidate of choice marked for Latino and Black voters. We record the number of times that the candidate got the most votes out of 1000 draws from the Latino voter-preference distribution, followed by the Black voter-preference distribution (for instance, in the 2018 Governor runoff election, Valdez got the most votes from Latino voters in all 1000 draws, while getting the most votes from Black voters in 429 of 1000 draws). Only major-party candidates are included for general elections, and only Democratic primaries and runoff elections are considered. We have also excluded elections with uncontested primaries, as well as judicial contests.

## D Adapting for multiple groups

In a state with only a single salient minority (for example, Black voters in Louisiana, as discussed in Supplement E), a district's effectiveness score is a single number between zero and one, representing the probability that the district is effective for Black voters. For a state like Texas, however, where both Latino and Black voters may raise plausible VRA claims, the effectiveness determination becomes more complex. It may be tempting initially to address these voting groups independently, by simply calculating a Latino effectiveness score and separately calculating a Black effectiveness score. However, scores determined that way could be misleading. Suppose a district was estimated to have a 50% Latino effectiveness score and a 50% Black effectiveness score. That could aptly describe a situation in which a district is *always* effective for either Latino or Black voters but never elects a candidate preferred by both groups, instead alternating from election to election. Or it could equally well describe a scenario where a district elects a consensus candidate preferred by Latino and Black voters alike half the time, and a candidate preferred by neither group the other half of the time. And of course there are scenarios in between those extremes, as depicted in Figure 15. Disambiguating between aligned and mutually exclusive success will help ensure that the plans we judge to be VRA-compliant do not secure effective districts for one group at the expense of the other.

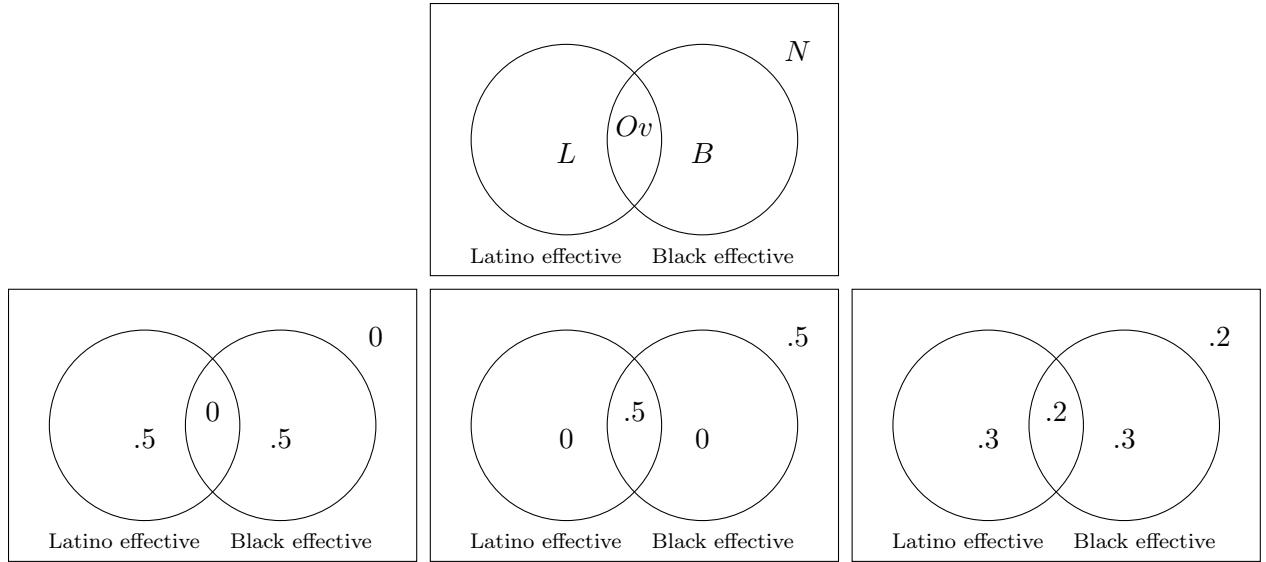


Figure 15: A district may be effective for Latino voters, Black voters, both groups, or neither. The figure shows three different ways that a district could have a 50% chance of performing for Latino voters and a 50% chance of performing for Black voters.

To address this complexity, our scores have four components: first, effectiveness for Latino voters but not Black voters (abbreviated  $L$ ); second, effectiveness for Black voters but not Latino voters ( $B$ ); third, simultaneous effectiveness for both Latino and Black voters ( $Ov$ , for Overlap); and fourth, effectiveness for neither Latino nor Black voters ( $N$ ). Since these four cases are mutually exclusive and exhaustive, the components must sum to one.

Here, raw Latino-effectiveness scores and raw Black-effectiveness scores are calculated exactly as they would be in a single-minority-group state, without regard to interactions between the effectiveness for one group and the other. Neither-effectiveness is handled similarly, designating neither-successful elections as those where the minority-preferred candidate(s) lost.

We then calibrate each of these three raw scores individually, using the adjustment described in §4.4 and Supplement B. Having three calibrated effectiveness scores allows us to solve for the four individual Venn diagram components, which we can denote by four components with  $L + B + Ov + N = 1$ .<sup>33</sup> So the first district in Figure 15 has  $L = .5$ , the second has  $L = 0$  and the third has  $L = .3$ . For a district to be deemed effective for Latino voters, it should have  $L + Ov > T$ , where  $T$  is the effectiveness threshold set in §4.5 (for example,  $T = .6$ ). To be deemed simultaneously effective for Latino and Black voters, it should satisfy both  $L + Ov > T$  and  $B + Ov > T$ .

In a state like Texas, with two sizable minority groups, we would examine effectiveness scores computed for the districts in the benchmark plan to observe its number of districts currently over the prescribed threshold  $T$  of effectiveness for Latino voters, Black voters, both groups, and neither group. We denote these numbers in a four-tuple with the following order: (Latino only, Black only, Overlap, Neither). To be included in our VRA-conscious ensembles, a proposed plan must meet or exceed the number of effective districts for Latino voters *and* for Black voters, separately, while also meeting or exceeding the overall number for both groups. For instance, with 21 districts, suppose the benchmark plan has eight effective districts for Latino and/or Black voters with a (5, 2, 1, 13) split. Then we would require a proposed plan to have at least  $5 + 1 = 6$  Latino-effective districts, at least  $2 + 1 = 3$  Black-effective districts, and at least  $5 + 2 + 1 = 8$  distinct districts overall that are effective for Latino voters, for Black voters, or both.

---

<sup>33</sup>We solve for these via  $L + Ov$  = Latino effectiveness,  $B + Ov$  = Black effectiveness,  $N$  = Neither effectiveness, while  $L + B + Ov + N = 1$ . It is possible to end up with calibrated scores for Latino, Black, and Neither leaving no non-negative solution for the components (e.g.,  $L + Ov = .8$ ,  $B + Ov = .65$ ,  $N = .25$ ). In this case we treat the effectiveness scores for Latino and Black groups as primary and adopt the closest feasible Neither score (in this case  $N = .2$ ). Adjusting the Neither score is preferable to asymmetrically adjusting the minority-group scores or seeking a simultaneous adjustment for both.

## E Case study: Louisiana

To confirm that the protocol here has applicability beyond the Texas congressional districts, we briefly report results from a second trial carried out on congressional and state Senate plans for Louisiana, which have 6 and 39 districts, respectively. Louisiana has 31.4% BCVAP, per 2018 ACS data, which indicates that a proportional share of effective districts would call for 1.9 congressional districts and 12.2 state Senate districts. The current map has one congressional district and 11 state Senate districts that are effective for Black voters.

Data preparation for Louisiana is similar to Texas, except that the precinct shapefiles provided by the State exhibit more frequent changes, calling for careful geodata-matching work to produce a shapefile that can support longitudinal election results from 2015 to 2019. This gives us a somewhat smaller election dataset than in Texas. In addition, Louisiana has a nonpartisan primary (dubbed a “jungle primary” or “Cajun primary”) used in all but presidential contests; if no candidate surpasses 50% of the statewide vote, then the top two vote-getters advance to a general election (effectively, a runoff). Our dataset began with 11 election sets, with primaries paired with generals when appropriate. We excluded one election (State Treasurer in 2015) because the primary featured two Republicans and no Democrats (in a state where Black voters overwhelmingly vote Democratic), leaving us with ten election sets for use in the VRA-conscious protocol. Success for the Black-preferred candidate is assessed in a district when that candidate either (a) receives a majority in the primary, (b) achieves a plurality in the primary and the Democratic candidate receives the most votes in the paired general, or (c) achieves a plurality in the primary but there is no general because of a statewide majority for some (possibly different) candidate.

Other elements of the VRA protocol are similar to Texas, but greatly simplified by needing to consider only one minority group. Effectiveness for Black voters is computed with the same formula ( $\sum w\delta / \sum w$ ) described in §4.3, using the same election weight factors. A logit adjustment step is conducted to calibrate effectiveness for Black voters to the empirical record of representation at all three levels of statewide redistricting. An effectiveness threshold of 65% is now used, rather than 60%, to reflect the greater margin of error that comes with a smaller dataset.

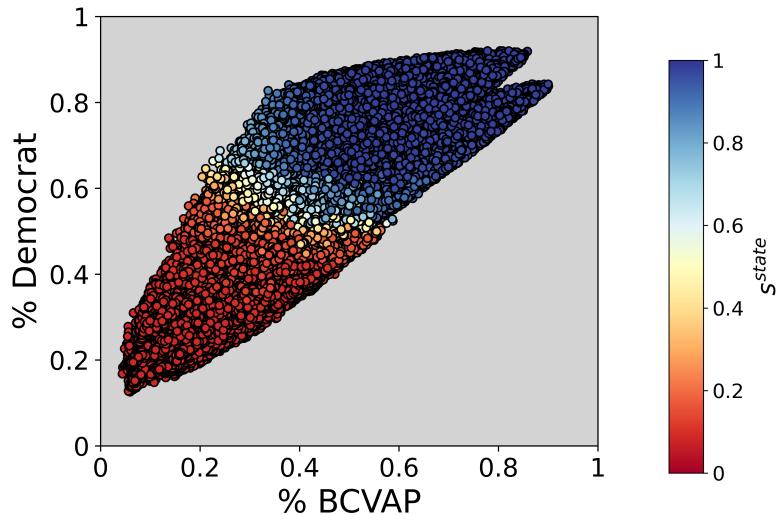


Figure 16: Scatterplot of districts from a Louisiana state Senate ensemble (39 districts), plotted by BCVAP vs. partisan lean as in the Texas examples. Data points are colored by the calibrated  $s_{state}^{state}$  statewide effectiveness score for Black voters.

For the congressional plans, we are easily able to produce a supply of alternative plans with one highly effective district and a second or even third potential “influence” district, while maintaining much better compactness than the enacted plan. In a state Senate run with our VRA-conscious protocol, we find numerous plans with 12 Black-effective districts, even though 11 districts suffice for acceptance.

## F Convergence heuristics

We continue with additional figures in the style of §5.4.1, demonstrating multistart heuristics and ensemble comparisons.

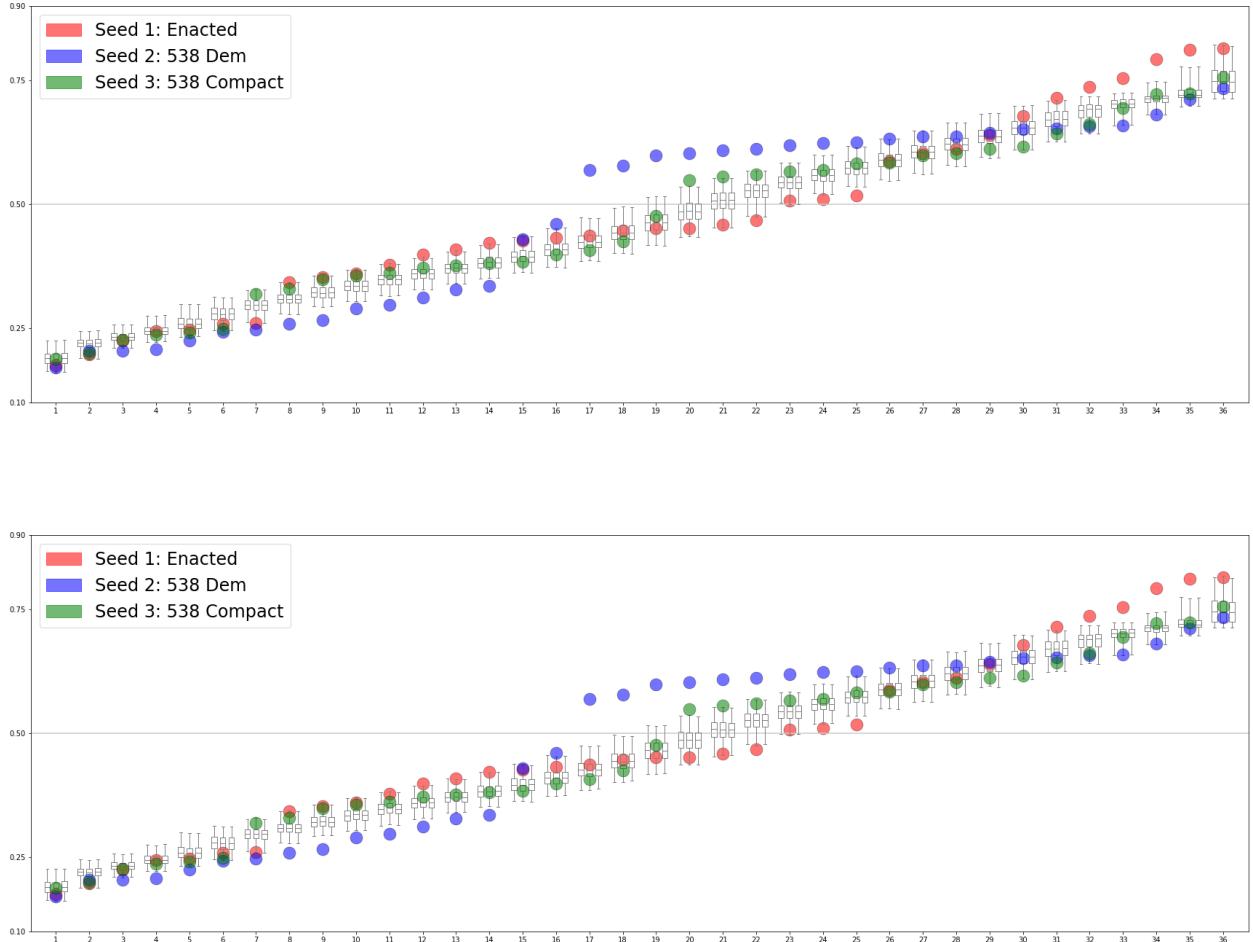


Figure 17: Multistart heuristics for Democratic vote share in the 2016 presidential general election in two sets of ensembles. Top, three constrained ensembles using the  $s^{\text{dist}}$  score, each starting at a different seed map. Bottom, analogous plot for the  $s^{\text{unw}}$  score with the same seeds. Each shows good multistart agreement. Between the top plot and the bottom, the colored markers are exactly identical and the boxes-and-whiskers are nearly indistinguishable.

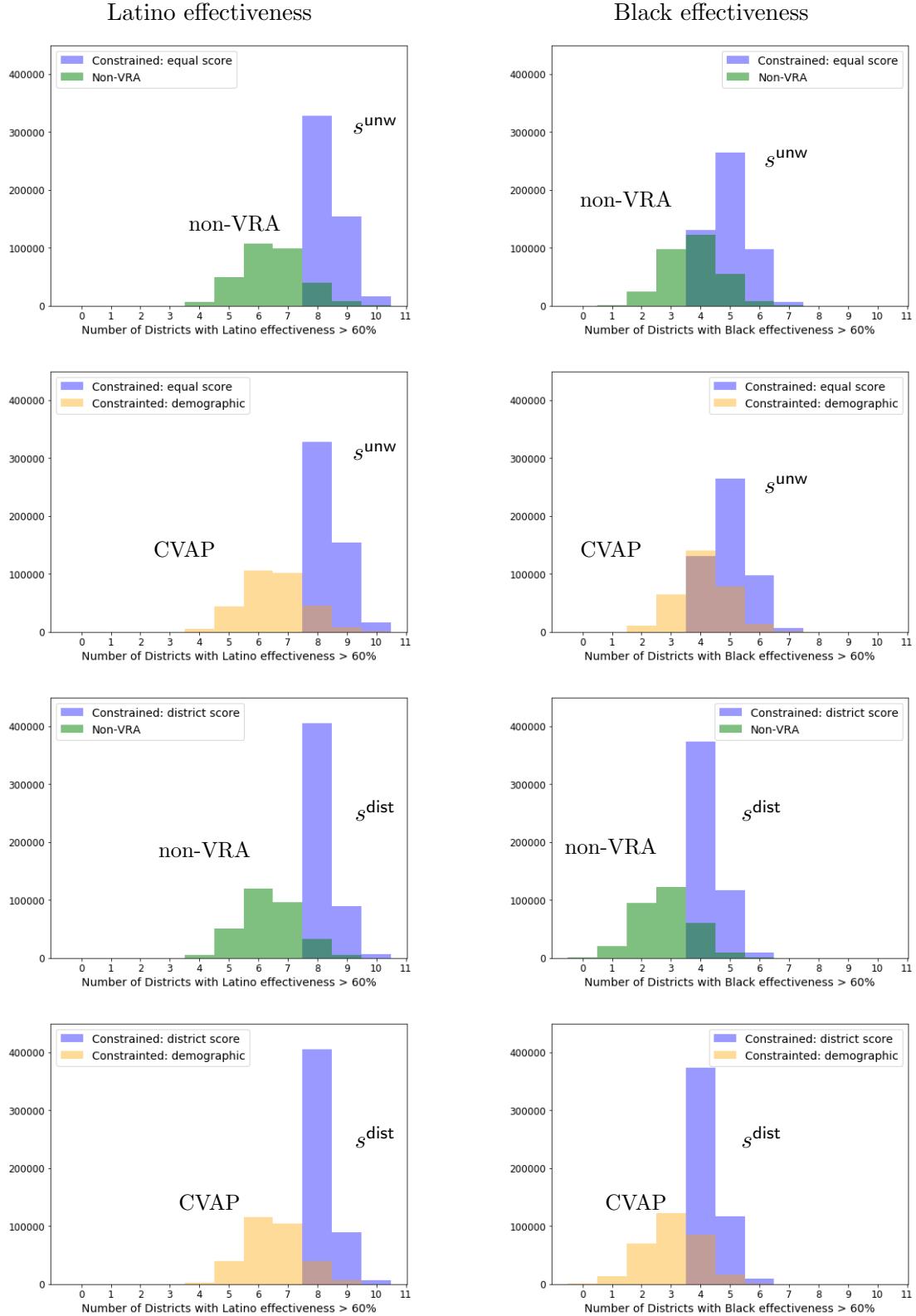


Figure 18: Ensemble comparisons. Together with Figure 8, we see that the non-VRA ensemble and the CVAP-shares ensemble resemble each other; imposing a CVAP threshold has not materially impacted the number of effective districts.

## G Regression and classification results

In this appendix, we present further classification and regression results to extend the discussion in §6. We also compare and contrast our findings to the model described in *Drawing Effective Minority Districts: A Conceptual Framework and Some Empirical Evidence* ([Grofman, Handley and Lublin, 2001](#), 1430 figure 4).

Based on their empirical investigation, Grofman–Handley–Lublin suggested that an elbow-shaped frontier with two linear segments might be needed to cut out the effectiveness zone in a plot of demographics and partisanship like the ones we present here. One line in the frontier would ensure the composition needed to win a Democratic primary and the other would correspond to success in the general election.

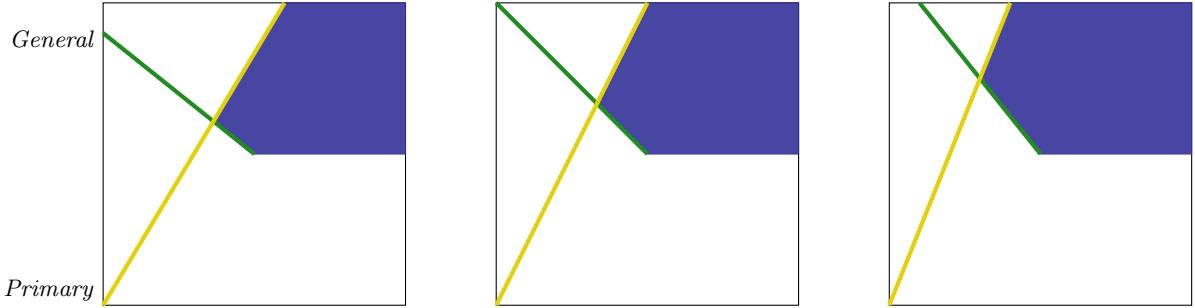


Figure 19: Three possible “electoral success” plots in the Grofman–Handley–Lublin framework, with Black population prevalence on the  $x$  axis and Democratic partisan lean on the  $y$  axis. Compare to Figure 20 below.

The “joint” in their elbow was projected to have a relatively low minority CVAP and a Democratic share over 50%. On this account, the frontier for primary success should run “northeast” from the elbow because of the need for minority demographic control as a district becomes more Democratic. The frontier for general election success would run “southeast” from the elbow because greater minority share would help offset White Democratic defection if a minority-preferred candidate is nominated. Though the correspondence is not perfect, this comports fairly well with the patterns “learned” by KNN classification (Figure 20), especially if we recall that very few effective districts can be found below the square’s midline. Note though that the primary frontier is nearly vertical—this tracks with especially low polarization observed in Texas Democratic primaries.

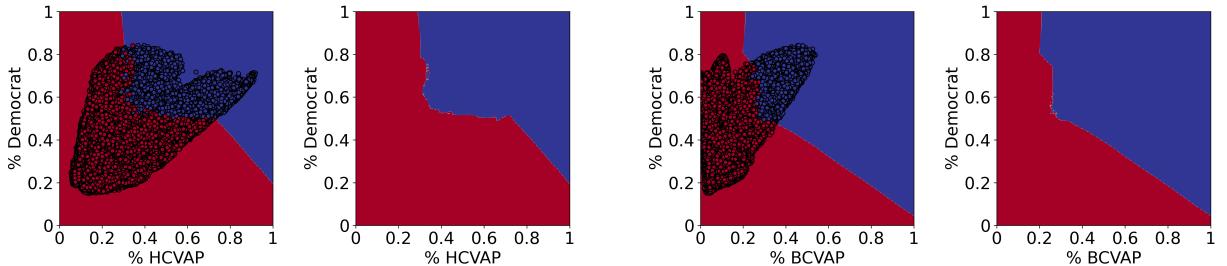


Figure 20: KNN classification fit to a yes/no label at the 60% effectiveness level for Latino and Black voters in Texas.

Next, we present the results of *decision-tree models* that find the best rectangular approximation

of the effectiveness zone. This would correspond to a rough rule of thumb such as *districts are likely to be effective if they are at least X% minority by CVAP and at least Y% Democratic*. The classification model essentially solves for the best  $X$  and  $Y$  by using a balanced accuracy metric. The accuracy achieved here is unsurprisingly lower than for KNN classifiers, because the form of the frontier is more constrained. The interesting findings here, shown in Figure 21, are that a Democratic share of nearly 50% together with about 30% HCVAP or about 23% BCVAP are the best threshold correlates of effectiveness. (Compare to the demographic thresholds of 45% HCVAP and 25% BCVAP used above in §5.4.)

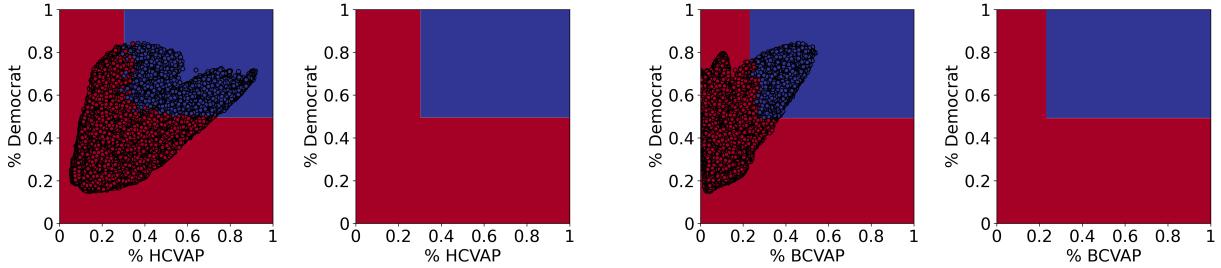


Figure 21: Decision trees fit to a yes/no label at the 60% effectiveness level for Latino and Black voters in Texas.

Finally, we include one more sample plot to test the hypothesis that district effectiveness is easier to classify if Latino and Black populations are combined. We find that the model outputs are less accurate than for the racial groups considered one at a time.

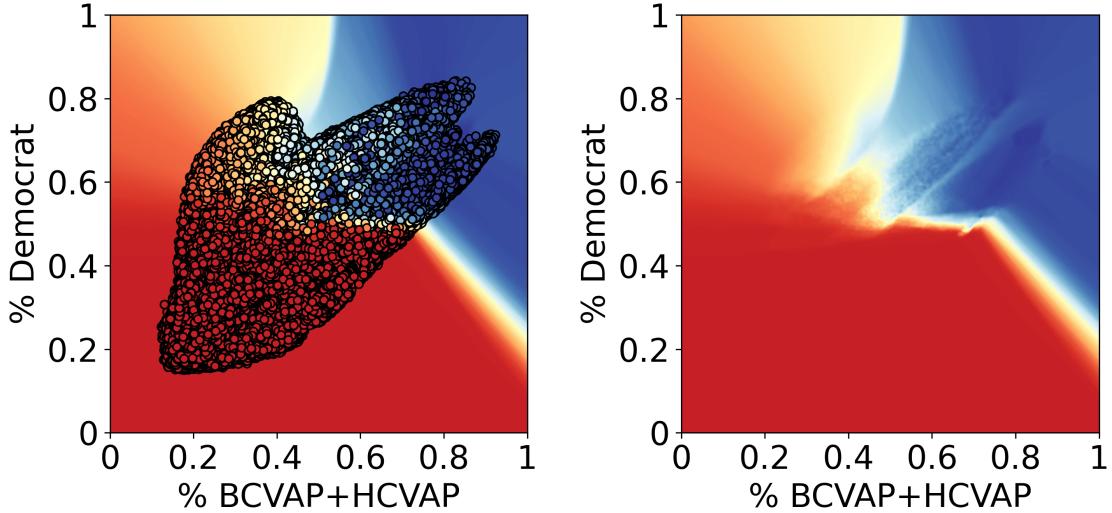


Figure 22: With Latino and Black population combined, it is even harder to describe the pattern of effectiveness. This finding holds up across various styles of regression and classification that we attempted.

## H Heuristic optimization

Our goal for heuristic optimization was to find a map with five Black-effective districts and ten or eleven Latino-effective districts, while ensuring that at least 15 are effective for at least one group. Here, the goal was not to draw a representative sample from the set of all valid plans, or even to see typical properties of especially good plans. Rather, we merely seek *examples* of interesting plans. To do this, we use the proposal-generation mechanism of ReCom and insert local-search optimization techniques with a specified objective function keyed to the effectiveness scores. Finding provable optima in a setting like this is NP-hard, and we emphasize that these short searches have almost certainly not located global optima. But we can still expect to find maps with good features in this way; the demonstration map featured in §7 was derived from a run of this local-search algorithm.

Suppose there are  $k$  districts in plan  $P$ , denoted  $P_1, \dots, P_k$ . Consider the piecewise function

$$g(x) = \begin{cases} 1 & x \geq .6 \\ 5x - 2 & .4 \leq x \leq .6 \\ 0 & x \leq .4 \end{cases}$$

One possible objective function  $F$  on the space of plans is defined by

$$F(P) = \sum_{i=0}^k g(s_L(P_i)) + \sum_{i=0}^k g(s_B(P_i)) - Ov(P),$$

where  $Ov$  is the number of effective districts for both Black and Latino voters and  $s_L(P_i)$  and  $s_B(P_i)$  are effectiveness scores of district  $P_i$  for Latino and Black voters, respectively.

The objective function incorporates Latino-effectiveness across all districts and Black-effectiveness across all districts, applying the function  $g$  to each so that scores less than 40% do not contribute to objective function value, the contribution rises quickly as effectiveness scores rise to 60%, and no additional contribution occurs after passing the 60% threshold. Finally, the overlap term is subtracted off to mitigate double-counting, since overlap contributes to both of the other terms.

For our optimization runs we tried various schemes to preferentially accept proposals with higher  $F$  values, such as by accepting lower-scoring proposals with a fixed probability, or by a probability based on the difference  $F(P) - F(Q)$  when proposing a move from map  $P$  to map  $Q$ .

We iterate steps of this procedure, following the *short bursts* method from [Cannon, Goldbloom-Helzner, Gupta, Matthews and Suwal, 2020](#). To use short bursts, we choose a burst length of  $b$  and run the chain normally in batches of  $b$  steps. The map with the highest score in each batch of  $b$  steps is used as the starting position for the subsequent batch.

We chose  $b = 50$  after experimenting with different values. After just a few thousand steps of this procedure, we are routinely finding maps with far more measured opportunity for minority voters than the levels seen in the current enacted plan. (See Figure 12 for a similar optimization run.) We make no claims to have pushed heuristic optimization to anywhere near its limits, and we welcome other approaches for finding interesting plans.