Chapter 4: Balancing Partisanship and Voting Rights Law in Michigan Legislative Maps

> A system cannot fail those it was never built to protect.
>
> Vann Newkirk

Note: This chapter is a slightly edited version of a white paper by the same name. This is joint work with Moon Duchin.

4.1 Introduction and background

In December 2023, Michigan's state House and Senate maps were struck down by three federal judges as unconstitutional racial gerrymanders in *Agee v. Benson*, sending the independent commission that created the maps (the Michigan Independent Citizens Redistricting Commission, or MICRC) back to the drawing board. The redrawn state House map was approved by the court on March 27, 2024 and the state Senate redraw is active at the time of writing, with the court's final sign-off given on July 26, 2024. It was widely assumed by observers that the requirements from the U.S. Constitution and the Voting Rights Act of

1965 around racial fairness sat in tension with the goals of partisan fairness that are indicated in the Michigan Constitution.[15] The main purpose of this report is to investigate that assumption.

Our general method will be to generate large collections, or "ensembles," of plausible districting plans at each level in Michigan (the 110-district state House, 38-district state Senate, and 13-district Congressional delegation) using a publicly available district-drawing algorithm.[16] We will construct a working definition that labels some districts as being "effective" for Black voters to elect their candidates of choice (§4.1.1) and will consider various metrics of partisan fairness, centering the efficiency gap (§4.1.2). By considering the relationship between VRA effectiveness and partisan fairness, we will learn about whether and how the fairness goals trade off against each other.

4.1.1 VRA effectiveness

At the outset, it is important to emphasize that the Voting Rights Act is not fundamentally about counting majority-Black districts; rather, it considers whether plaintiffs in a minority racial, ethnic, or language group (in this case Black voters) have an opportunity to elect candidates of choice. In fact, courts are increasingly likely to strike down maps that take racial data into account too centrally or without adequate justification—indeed, that was a key theme in the *Agee* lawsuit that invalidated the Michigan legislative maps.

To avoid these problems, the best practice is to use a measure that is based on the likelihood of minority-preferred candidates to be first *nominated* in the primary and then *elected* in the general election. The role of primary elections is ineliminably important. This is a problem in Michigan because there are essentially no primary elections for which the

[15] See legislature.mi.gov.

[16] A more detailed discussion of district generation is found in §A.1.

data is publicly available. This means that it is impossible to construct an effectiveness score based on confirming whether Black voters' candidates of choice would have won a proposed district in past statewide electoral contests. To work around this in the remedial phase of litigation, the commission contracted experts (Max Palmer and Lisa Handley) and provided them with primary data from selected counties near Detroit in certain districted elections. The Palmer–Handley team ran statistical inference methods to learn estimated turnout by race in Democratic primary contests in Oakland, Macomb, and Wayne counties, and they deemed a district to be effective for Black voters if their point estimates indicated that Black voters would have outnumbered White voters in recent Democratic primary elections. This is fairly crude—because it doesn't take cohesion and crossover voting into account—but it may be the best proxy available given the poor availability of primary data.[17] Unfortunately, we are not able to fully replicate this proxy because the commission has declined to make the Palmer–Handley dataset publicly available.

It is well established that in recent voting patterns, Black voters in the Detroit area have a strong preference for Democratic over Republican candidates. For our analysis, we therefore used the Black voting age population (BVAP) of a district and its share of major-party vote in the Presidential contest of 2016 that went to Biden (D) as opposed to Trump (R). If the BVAP is over a threshold, we assume that Black voters can likely get a candidate of choice nominated in the Democratic primary; if the Biden support is over a threshold, we infer that the Democratic candidate can likely be elected to office in the general election. We tested BVAP thresholds of 40%, 44%, 46%, and 50% together with Biden thresholds of 50%, 53%, 54%, and 55%. The sensitivity analysis in §A.2.1 shows that our main findings are robust to the choice of thresholds: larger numbers of districts effective for Black voters can easily be

[17]Cohesion is a measure of how likely Black voters are to vote for the same set of candidates, and crossover voting occurs when a voter defects from this set.

found while simultaneously passing strong tests of partisan fairness. All code can be found in our replication repository [40].

For the main part of the analysis, we will focus on an effectiveness score we call ”VRA 44/50”: a district is labeled effective if the share of Black residents among the voting age population is at least 44% (so that Black voters can nominate preferred candidates) and Biden won a simple majority with at least 50% of the major-party presidential vote (so that the nominated candidates can be elected to office).

A separate and widely misunderstood element of voting rights law is known as the Gingles 1 prong of a VRA case. To advance to trial, plaintiffs must provide a demonstration that it is possible to draw a "reasonably configured" districting plan with a larger number of majority-minority districts than are present in the challenged map. In this case, this means finding additional districts that have BVAP $>$ 50%. It is for this threshold test only that majority-BVAP districts are a legal requirement.

We perform a Gingles 1 analysis in order to set an upper bound for the number of districts that could be cited in a VRA claim. We find that while maintaining strong levels for other traditional districting criteria (like compactness, respect for county and municipal boundaries, and simplified efficiency gap) we can produce state Senate maps with 5 majority-BVAP districts and state House maps with 15 majority-BVAP districts (see Figure 4.1). This is greater by quite a margin than the number of districts likely to be effective for Black voters to elect candidates of choice in any of the maps considered by the courts, clearing the way for the rest of the VRA analysis contained here.

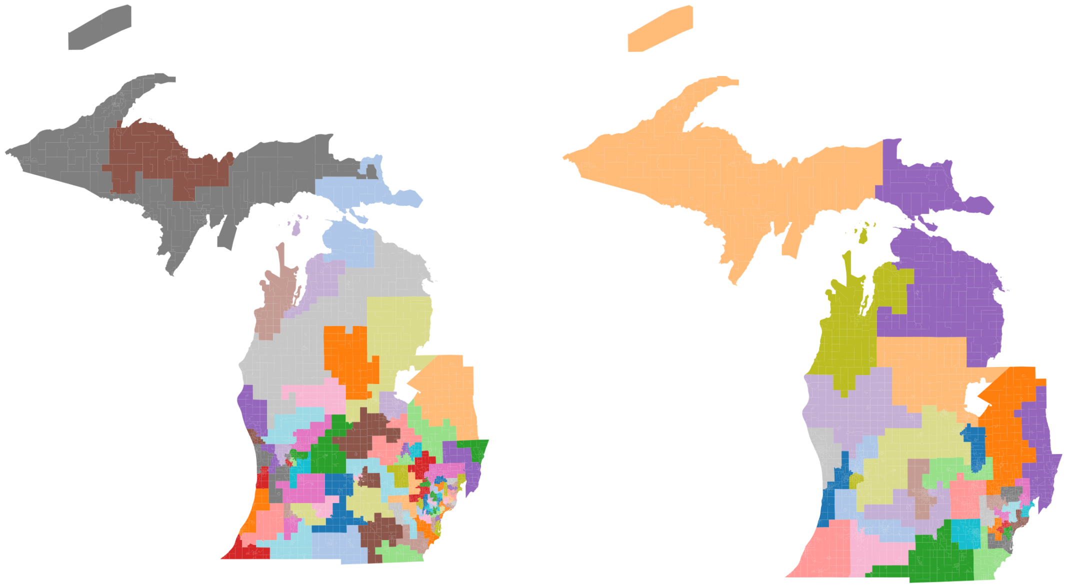

Figure 4.1: Left: a state House plan with 15 majority-BVAP districts. Right: a state Senate plan with 5 majority-BVAP districts. Both are comparable to or better than the plans considered by the court on the basis of compactness, respect for political boundaries, and simplified efficiency gap.

4.1.2 Partisan fairness metrics

Evaluating election-based scores

In order to measure the partisan bias of proposed maps, we turn to metrics of partisan fairness that were adopted by the MICRC for their compliance with state Constitutional requirements. To briefly explain the scores, we adopt the notation that S (or S_E) is the seat share and V (or V_E) is the vote share for a particular party under a particular election E.

To evaluate partisan fairness, we choose a score, a map, and a pattern of votes to evaluate. Typically, this means we need a number of D and R votes cast in every precinct, such as the votes cast in a single election, or a modeled election that is created as a combination of observed ones. The choice of election can be just as important as the choice of score in presenting a clear and interpretable picture of fairness. Many authors favor the use of statewide (exogenous) elections, because they are easily interpretable observations of voting behavior in real contests. Others prefer to use endogenous elections—Congressional elections to evaluate a Congressional map, legislative elections to evaluate legislative maps, and so on—but this requires employing techniques such as statistical regression that attempt to control for variable factors like candidate strength, incumbency, recency, and so on. It is quite uncommon to use endogenous elections in redistricting litigation, partially because they come with many user choices that make them more difficult to vet. Instead, a common approach in litigation is the use of an *election index*, consisting of a set of statewide elections averaged or blended together into a single, synthetic record.

We have prepared a dataset of statewide elections (see §A) for use with these scores. The MICRC was advised by Dr. Lisa Handley for its partisan fairness work, who created an election index but did not release information about its design. To attempt to replicate what was available to the commission, we present three variants on a blended election index in the supplemental material.[18] We will report serial scores in the body of this report, using interpretable statewide elections one at a time, and averaging the results; scores based on an averaged election are postponed to the supplement.

With every score we introduce, we will adopt the convention that the Republicans are the point-of-view party, which means that positive scores (> 0) will indicate an advantage for Democrats while negative scores (< 0) will indicate an advantage for Republicans.

[18] Out of five possible methods, we chose: method 0, which adds up all votes in a precinct across all elections, method 1, which does the same but equalizes statewide turnout across elections, and method 3, which averages vote shares in each precinct across elections.

Efficiency gap

One of the most commonly used partisan fairness metrics is known as the *efficiency gap*, as defined and explained by Stephanopoulos—McGhee in [52]. It has two main formulations, which we call the *original efficiency gap* (OEG) and the *simplified efficiency gap* (SEG).

Following the authors of [52], we say a vote is "wasted" if it is either a surplus vote for the winning party, or any vote for the losing party in a district. To compute OEG, sum up the wasted votes for each party over all the districts in a plan. The sum of wasted votes for party A, minus the sum of wasted votes for party B, divided by the total number of votes gives OEG, with a positive score indicating advantage for party B. The efficiency gap authors suggest that $|\text{OEG}| > .08$ flags a presumptive gerrymander, shifting the burden to the mapmaker to defend a map. A score of zero is considered ideal, because it means that each party has "wasted" an equal number of votes.

It is well known that the efficiency gap reduces to a simple formula in terms of seats vs. votes if every district has equal turnout: $\text{SEG} = S - 2V + 1/2$. Original efficiency gap differs from this simplified formula in the general case by a term depending on turnout disparities. Many authors have indicated that this simplified formula is preferable to the original formulation because it is directly tied to election outcomes, making it less gameable. The differences are usually small, but we will give SEG results in the main text and OEG results in the supplemental material.

Other metrics

- **Mean-median** If the Republican vote share per district is $V_1, V_2, \ldots, V_k$, then the mean-median score is simply the mean of those numbers minus the median. If the median (the middle district) is more favorable to Republicans than the mean (the average district)—making the score negative—this is thought to be sign of advantage for Republicans, so once again positive scores indicate Democratic advantage. Zero is considered ideal.

- **Lopsided margins** This is the average winning margin across all Republican-won districts minus the average winning margin across all Democratic-won districts. The idea is that gerrymandering for Democrats would lead to high Republican margins in a few districts (packing) and low Democratic margins in many districts (cracking), creating a positive score. Zero is considered ideal.

- **Mean disproportionality** The disproportionality in a single election is $V - S$, the difference between vote share and seat share. A positive result indicates that seat share has fallen short of vote share, which is disfavorable to the point-of-view party. Mean disproportionality simply averages this over the dataset of elections. It is near zero in two cases: either the map performs near-proportionally in each election, or the advantages to each party cancel out over time. Zero is considered ideal.

- **Competitive contests** We define a "competitive contest" in a single district to be one in which the vote share for each major party satisfies $.47 \leq S \leq .53$. Over our dataset, we have (number of districts) $\times$ (number of elections) district-level outcomes, and we count the number of competitive ones for a given map. Higher numbers of competitive contests are generally considered favorable, though many authors have noted that this priority can lead to significant seats/votes skews if there are many tight races and there is a slight swing in general public opinion.

Table 4.1 highlights the enacted and invalidated maps (the scores for all benchmark maps can be found in Tables A.3, A.4, and A.5 in §A.3). Note that many of the scores are similar across the invalidated and remedial maps, and in particular the serial lopsided margins score is almost exactly the same across all maps (making it not very discerning). These scores are hard to interpret without an idea of what is typical for the state, so we have produced scatterplots showing the range of these scores over many possible plans for the House, Senate, and Congressional level in Figures 2, 3, and 4. It should be observed that the benchmark maps almost always lie closer to 0, the "ideal" scores, than any maps in the ensembles.

	Enacted HD 22	Motown Sound	Enacted SD 22	Crane	Enacted CD 22
SOEG	-0.025	-0.026	-0.018	-0.018	0.022
SSEG	-0.006	-0.007	0.003	0.003	0.033
SMM	-0.02	-0.021	-0.01	-0.015	-0.009
SLM	0.6	0.6	0.592	0.593	0.585
Disprop	0.304	0.295	0.429	0.429	0.844
OEG I0	-0.005	-0.005	0.002	0.003	0.049
SEG I0	0.014	0.014	0.02	0.02	0.056
MM I0	-0.022	-0.023	-0.012	-0.015	-0.008
LM I0	0.602	0.602	0.594	0.594	0.587
Disprop I0	0.043	0.043	0.049	0.049	0.086

Table 4.1: The baseline of partisan fairness found in the benchmark maps. Negative scores are Republican-favoring, while positive scores are Democratic-favoring, in the convention we have adopted for the partisan metrics.

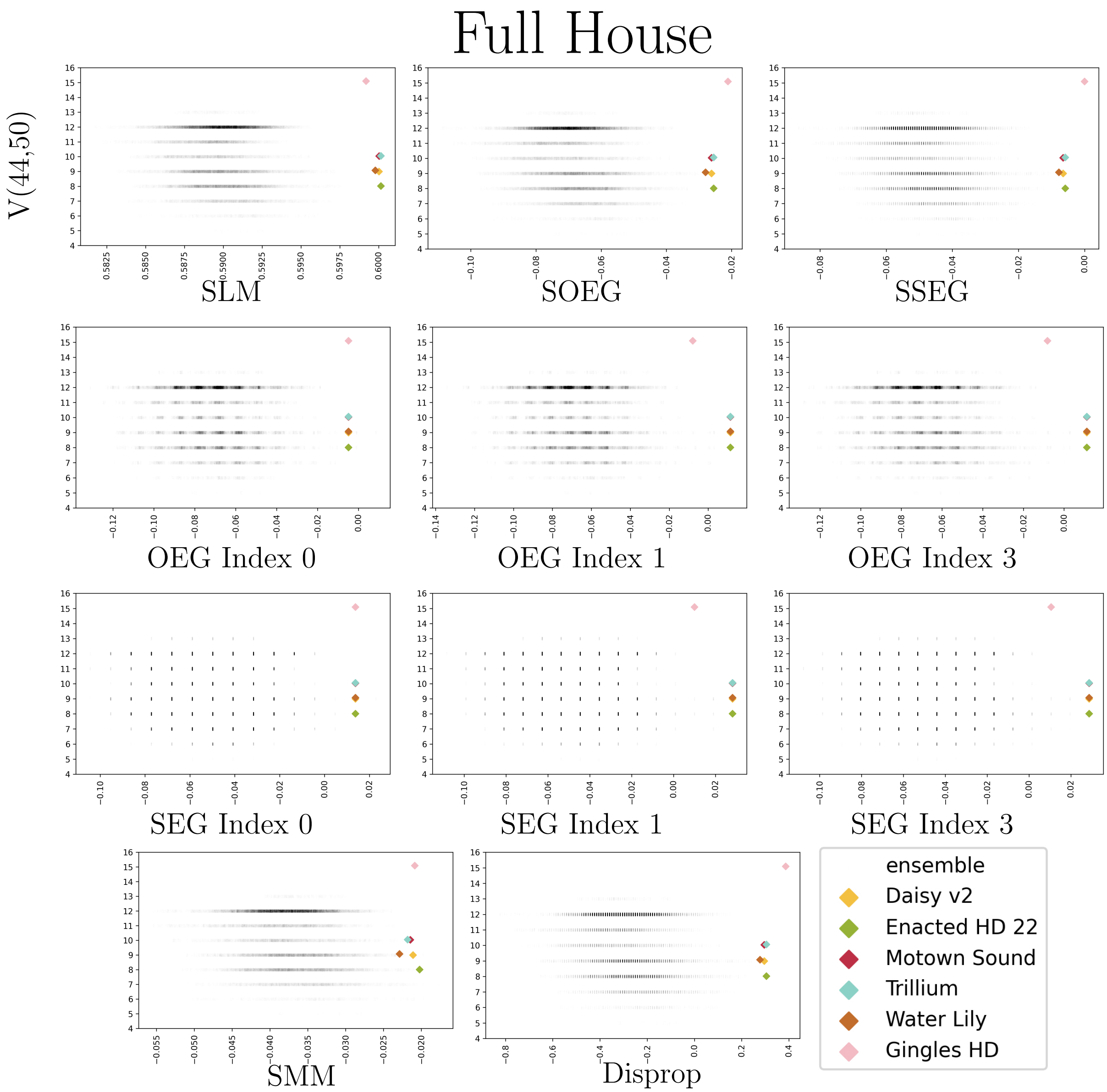

Figure 4.2: Scatterplots for 11 different measures of partisan bias compared to number of VRA effective districts in the state House. In the full House runs, we see that regardless of the choice of partisan score, the benchmark maps consistently have some of the highest scores found by the ensemble.

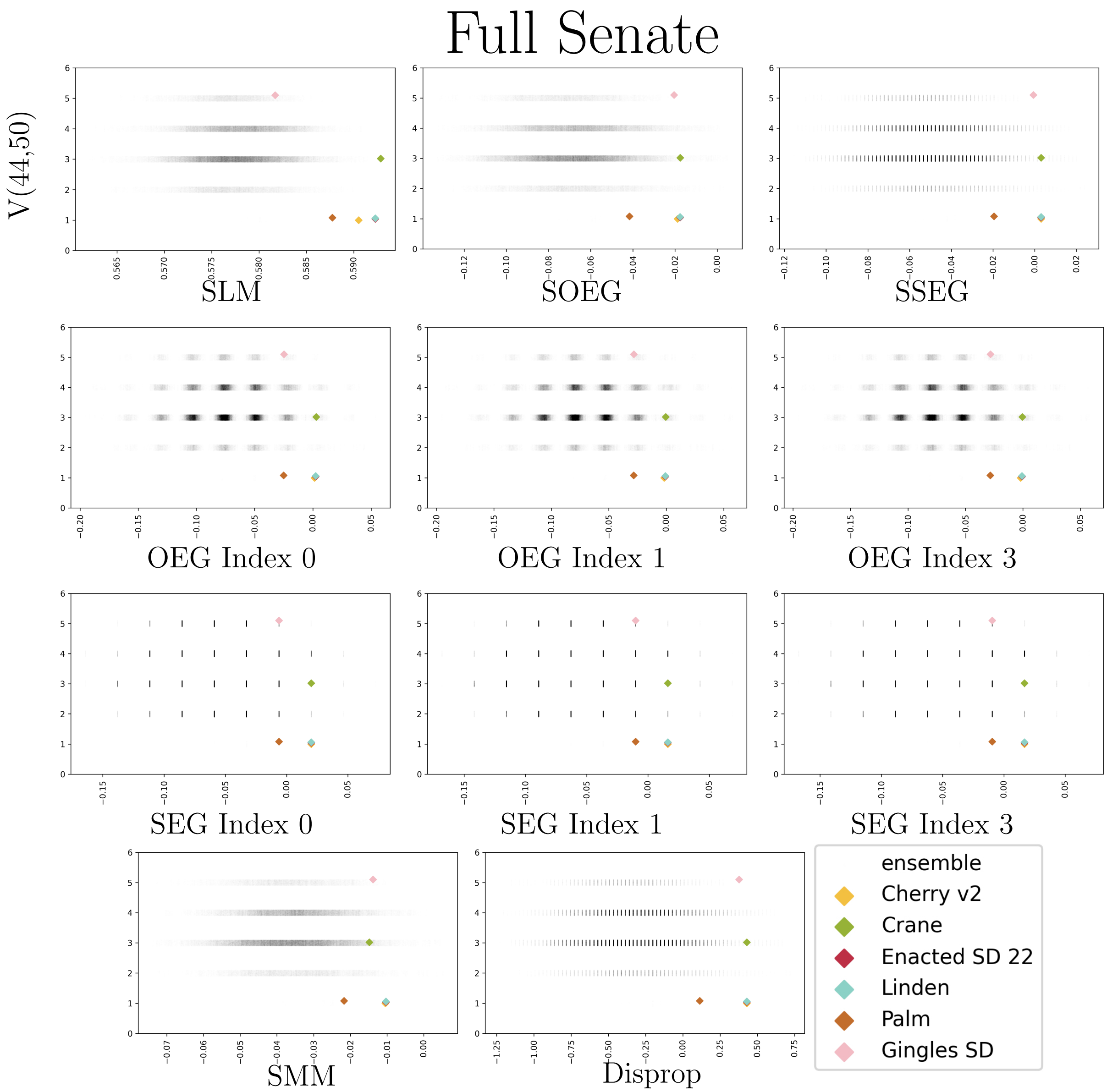

Figure 4.3: Scatterplots for 11 different measures of partisan bias compared to number of VRA effective districts in the state Senate. In the full Senate runs, we see that regardless of the choice of partisan score, the ensemble can find many maps that improve upon the partisanship score of the benchmark maps.

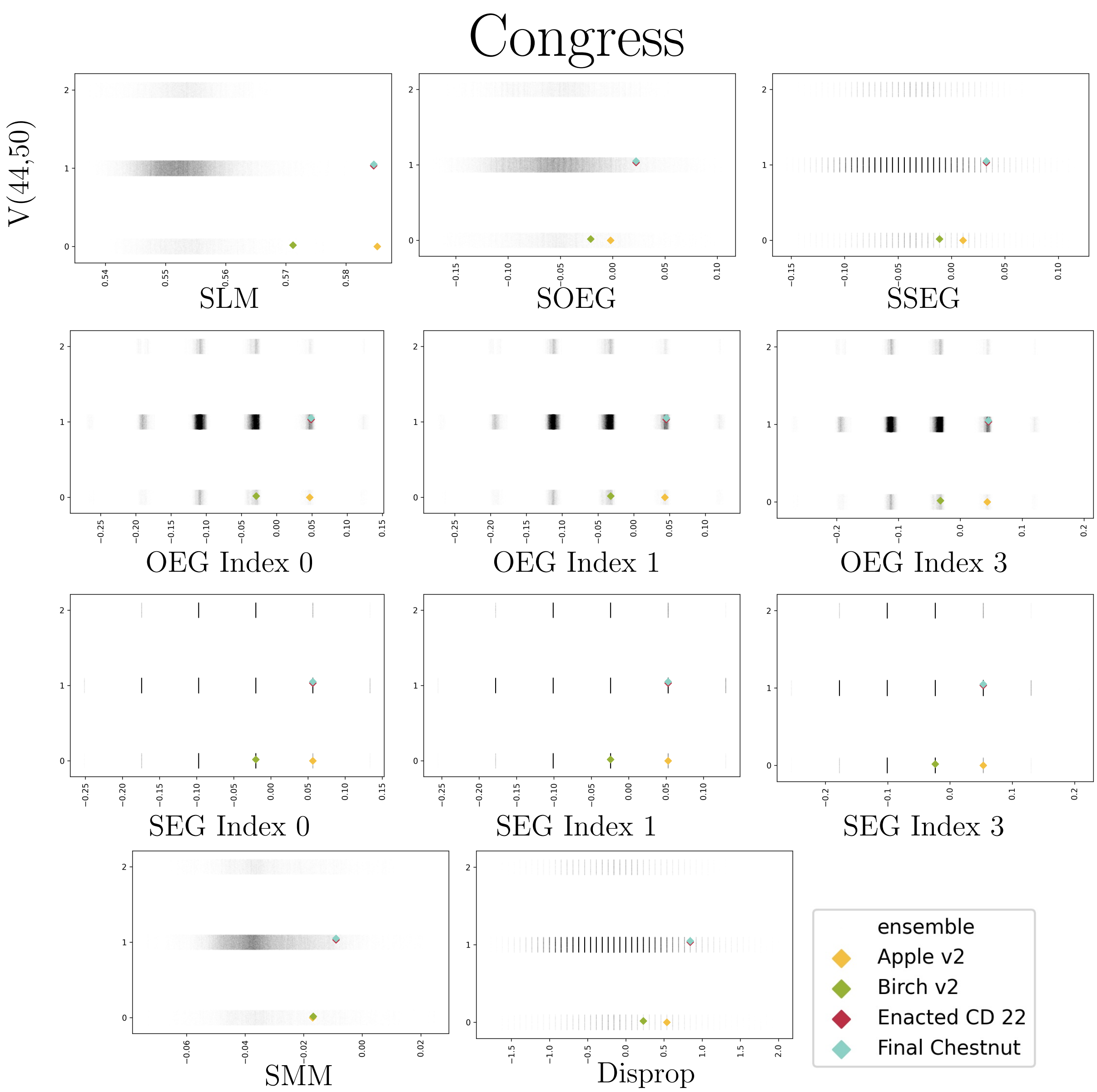

Figure 4.4: Scatterplots for 11 different measures of partisan bias compared to number of VRA effective districts in Congress. In the full Congressional runs, we see that regardless of the choice of partisan score, the ensemble can find many maps that improve upon the partisanship score of the benchmark maps.

4.2 Methods and Findings

In this section, we study the tradeoff between a highlighted partisan fairness metric—the (serial, simplified) efficiency gap—and the number of districts that meet the 44-50 threshold for effective opportunity for Black voters ($V(44, 50)$). (See §A.3 for other comparisons).

4.2.1 Methods

To probe the tradeoff, we generated ensembles of alternative maps for state House, state Senate, and Congress. In the House and Senate, we considered not only alternative maps for the whole state, but also limited re-draws, focused on the districts invalidated by the court.

We used six redistricting settings:

- State House: full state (110 districts), House 1 (13 districts), House 2 (15 districts)
- State Senate: full state (38 districts), Senate 1 (8 districts)
- Congress: full state (13 districts)

The limited redraws include the seven House districts and six Senate districts required to be remedied in the court order as well as adjacent districts that might reasonably be deemed necessary to redraw concurrently. (We note that the MICRC altered exactly the 15 districts in the House 2 set in its remedial map, named Motown Sound.) See Figure 4.5 to see where these appear in the state.

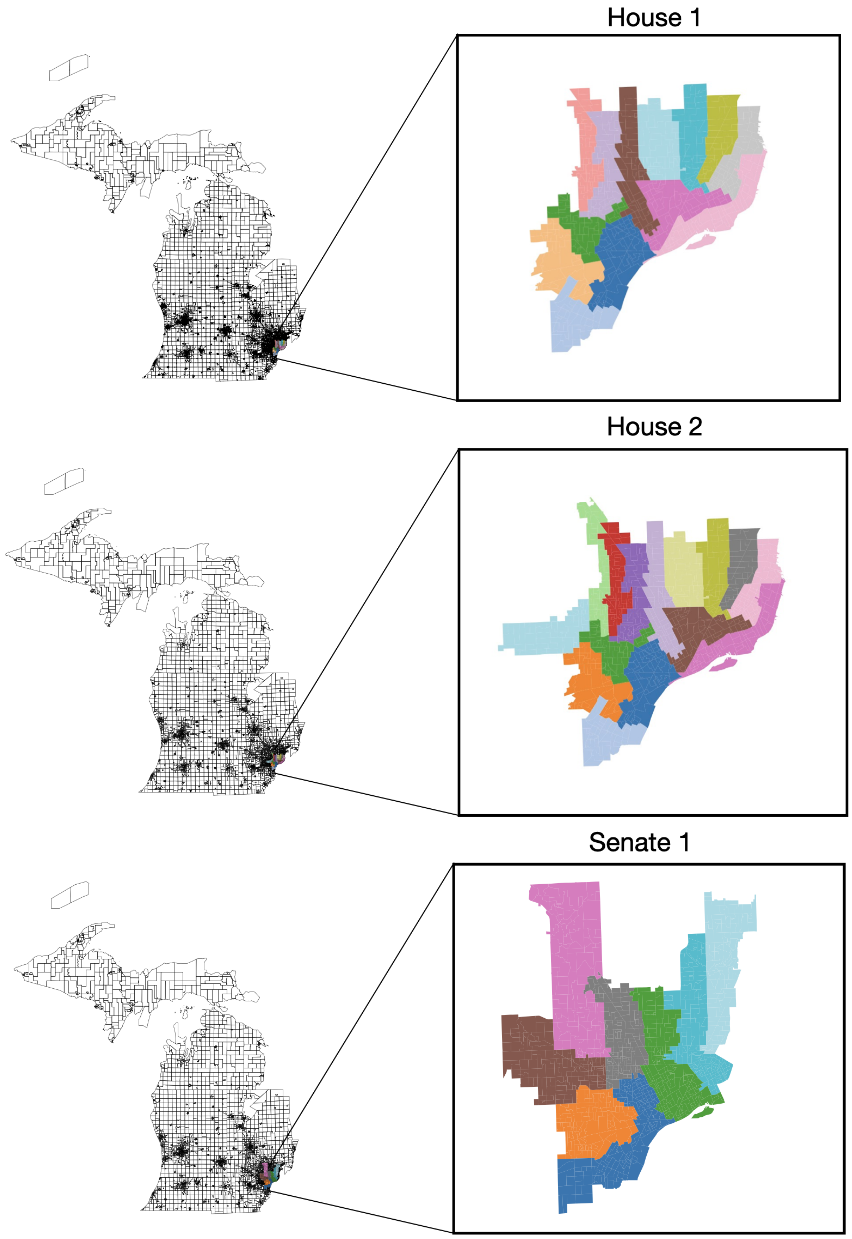

Figure 4.5: Maps showing the limited sets of districts being redrawn in the variants called House 1, House 2, and Senate 1. In House 1, these are invalidated districts 1, 7, 8, 10, 11, 12 and 14 and adjacent districts 2, 3, 4, 6, 9, and 13. In House 2, we also allow districts 5 and 16 to change. In Senate 1, SD 1, 3, 6, 8, 10, 11 must be redrawn, and SD 2 and 7 are reasonably deemed necessary to include.

All ensembles were made by an algorithm called *recombination*. Recombination leverages spanning trees for its main step; this ensures a high level of compactness without any need for additional weighting. We used a standard method (based on weighting graph edges) to encourage the preservation of counties and municipalities to an extent comparable to maps considered by the commission. We created both *neutral* ensembles and a *hill-climbing* variant designed to increase the number of VRA effective districts. We did not optimize for partisan fairness metrics, but instead looked to see whether driving up the number of VRA effective districts would have a deleterious impact on partisan metrics.

Figures 4.6, 4.7, and 4.8 show the relationship between the number of VRA effective districts and the serial SEG score in our six districting problems. The outputs show that additional effective districts can be added to the number found in the benchmark maps: at least one district at the Congressional level, four at the Senate level, and five at the House level.

In the Senate and Congressional ensembles, even without optimizing for the partisan score, we find many thousands of maps that simultaneously improve on the benchmark maps in VRA and partisan terms. (Recall that SEG = 0 is regarded as ideal, and that the commonly cited threshold for a presumptive gerrymander is .08.)

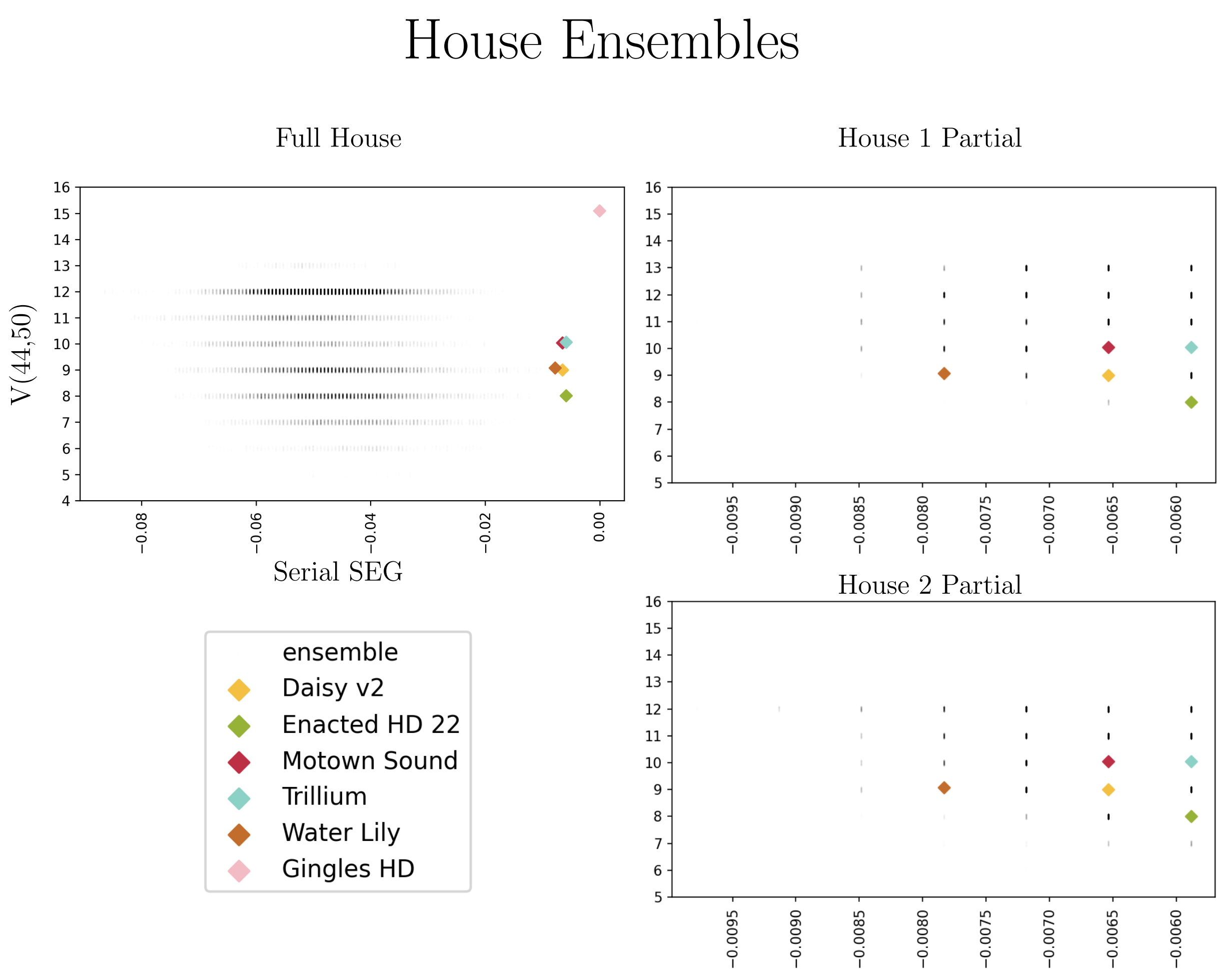

Figure 4.6: Scatterplots showing the relationship between efficiency gap and the number of VRA effective districts across all House ensembles, including both maps generated "neutrally" and those designed to increase effectiveness. Darker areas show more frequency. Note that the Gingles map does not appear on the partial scatterplots as it was optimized on the whole state.

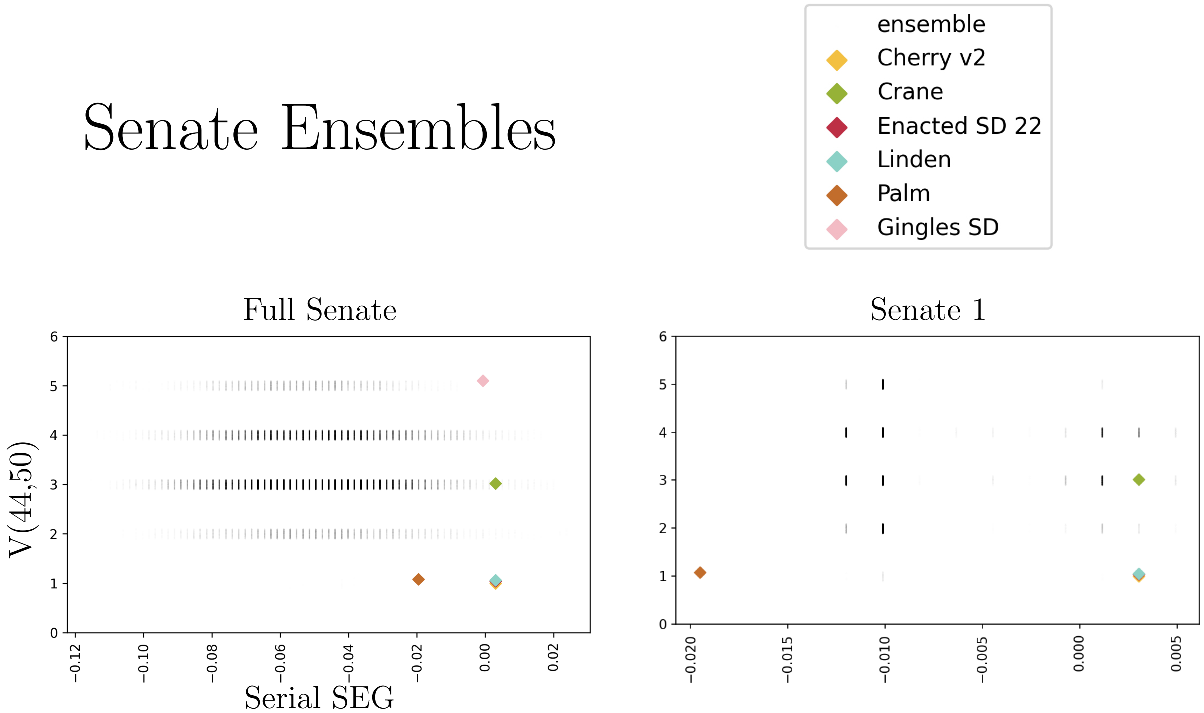

Figure 4.7: Scatterplots showing the relationship between efficiency gap and the number of VRA effective districts across all Senate ensembles, including both maps generated "neutrally" and those designed to increase effectiveness. Darker areas show more frequency. Note that the Gingles map does not appear on the partial scatterplots as it was optimized on the whole state.

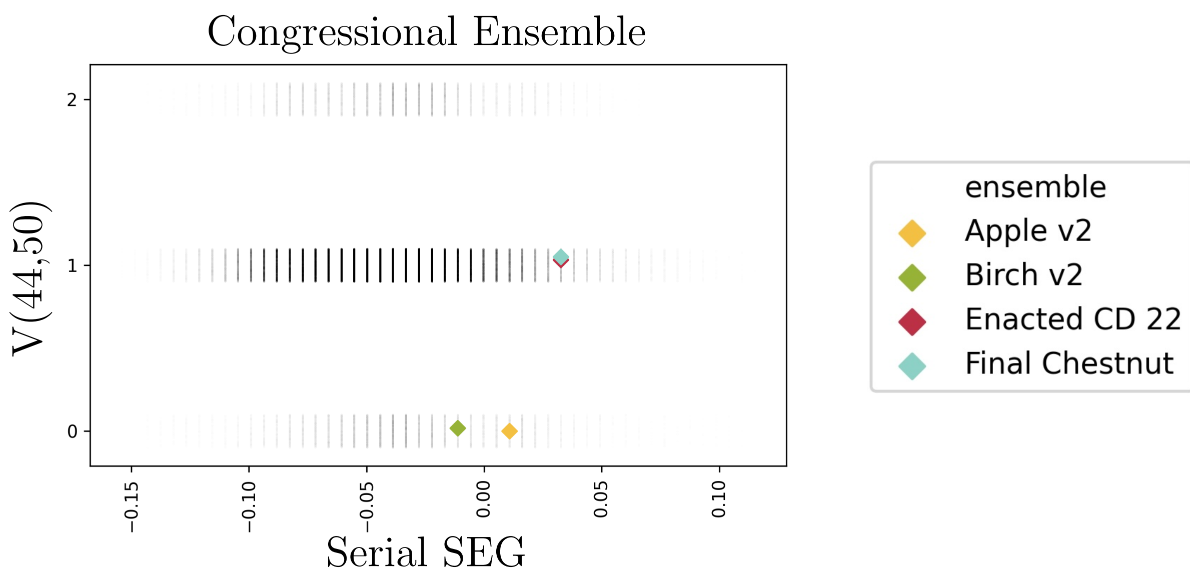

Figure 4.8: Scatterplots showing the relationship between efficiency gap and the number of VRA effective districts for the Congressional ensembles, including both maps generated "neutrally" and those designed to increase effectiveness. Darker areas show more frequency.

In contrast, in the full-state House ensemble, the best SEG scores fall a little bit short of the benchmark map. However, by targeting the redraw to the limited regions House 1 and House 2, we can maintain an SEG score identical to the benchmark plan while significantly increasing the number of VRA effective districts.

4.2.2 Findings for State House

Table 4.2 shows us some basic statistics about the 2022 enacted House map and our ensembles. Table 4.3 gives these statistics for the benchmark maps. In the enacted map, by the VRA 44/50 definition, there are eight VRA effective districts. In the full state ensemble, we achieve a range of 5 to 12 VRA effective districts and in House 1 and House 2 a range of 8 to 12 and 6 to 12 VRA effective districts respectively. This difference in the lower bound indicates that there are some VRA effective districts that can be found in the unscrambled portion of the state.

	2022 Enacted HD Map	HD Full Run	HD 1 Scramble Run	HD 2 Scramble Run
# Districts Scrambled	0	110	13	15
# Districts Fixed	110	0	97	95
# Counties Split by Scramble	N/A	N/A	(3, 3)	(3, 3)
# Munis Split by Scramble	N/A	N/A	(2, 15)	(2, 18)
# Counties Split by Full Map	48	(40, 67)	(48, 48)	(48, 48)
# Munis Split by Full Map	87	(81, 193)	(85, 94)	(84, 93)
# VRA Effective (44, 50)	8	(5, 12)	(4, 9) + 4	(4, 10) + 2
# VRA Effective (40, 50)	13	(6, 13)	(4, 9) + 5	(6, 10) + 3
# VRA Effective (40, 48)	13	(7, 13)	(4, 9) + 5	(5, 10) + 3
Max # of VRA (44, 50) While SEG Serial Beats State	8	–	12	12
Max # of VRA (44, 50) While $\lvert SSEG \rvert \leq .04$	8	11	12	12

Table 4.2: Statistics from state House ensembles and the enacted map. Numbers in parentheses denote a range of values (inclusive) achieved by the ensemble. Note that Michigan has 83 counties and 1520 municipalities (munis).

Table 4.2 also shows the relationship between VRA effective districts and SEG scores. We find that in the full state ensemble, we cannot improve on the number of VRA districts

	HD 2022 Enacted Map	HD Remedial Map (Motown Sound) VTD	HD Daisy 2	HD Trillium	HD Water Lily
Split Counties	48	48	48	48	48
Split Municipalities	87	86	88	86	86
# VRA Effective (44, 50)	8	10	9	10	9
# VRA Effective (40, 50)	13	12	11	11	10
# VRA Effective (40, 48)	13	12	11	11	10
SSEG	-0.006	-0.007	-0.007	-0.006	-0.008

Table 4.3: Various scores for state House benchmark maps.

while also improving on the enacted map's SEG score. The enacted map had the smallest SEG score out of all of the maps generated, which can be seen in Figure 4.6. In the House 1 and House 2 partial scrambles, we can find maps that have 12 VRA effective districts and a smaller SEG score than the enacted map.

Rather than using the enacted map's SEG score as a benchmark, we can also ask if we can increase the number of VRA districts while having SEG magnitude less than .04. In the full run, we can achieve 11 VRA districts while having SEG magnitude less than .04. In either of the partial runs, we can achieve 12 VRA districts while having SEG magnitude less than .04. Thus, while there may be a tradeoff between VRA effectiveness and the SEG threshold of the enacted map, there is no tradeoff between VRA effectiveness and a small SEG score.

4.2.3 Findings for State Senate

Table 4.4 shows us some basic statistics about the enacted Senate map and our ensembles. Table 4.5 gives these statistic for the benchmark maps. In the enacted map, by the VRA 44/50 definition, there is one VRA effective district. In the full state ensemble, we achieve a range of 1 to 4 VRA effective districts and in Senate 1 a range of 1 to 5 VRA effective districts.

	SD 2022 Enacted Map	SD Full Run	SD 1 Scramble Run
# Districts Scrambled	0	38	8
# Districts Fixed	38	0	30
# Counties Split by Scramble	N/A	N/A	(3, 3)
# Munis Split by Scramble	N/A	N/A	(1, 14)
# Counties Split by Full Map	31	(22, 51)	(31, 31)
# Munis Split by Full Map	29	(8, 122)	(24, 35)
# VRA Effective (44, 50)	1	(1, 4)	(1, 5) + 0
# VRA Effective (40, 50)	4	(2, 5)	(1, 5) + 0
# VRA Effective (40, 48)	4	(2, 4)	(1, 5) + 0
Max # of VRA (0.44, 0.5) While SEG Serial Beats State	0	4	4
Max # of VRA (0.44, 0.5) While |SEG Serial| <.04	1	4	5

Table 4.4: Statistics from state Senate ensembles and the enacted map. Numbers in parentheses denote a range of values (inclusive) achieved by the ensemble.

	SD 2022 Enacted Map	SD Cherry v2	SD Linden	SD Palm
Split Counties	31	28	31	27
Split Municipalities	29	28	29	30
# VRA Effective (44, 50)	1	1	1	1
# VRA Effective (40, 50)	4	4	4	4
# VRA Effective (40, 48)	4	4	4	4
SSEG	0.003	0.003	0.003	-0.02

Table 4.5: Various scores for benchmark state Senate maps.

Table 4.4 also shows the relationship between VRA effective districts and SEG scores. We find that in the full state ensemble, we can find 4 VRA effective districts while improving the SEG score. This is in contrast to the House level, where we could not improve the SEG score of the enacted map. In the Senate 1 partial scramble, we can find maps that have 4 VRA effective districts and a smaller SEG score than the enacted map.

Rather than using the enacted map's SEG score as a benchmark, we can also ask if we can increase the number of VRA districts while having SEG magnitude less than .04. In the full run, we can achieve 4 VRA districts while having SEG magnitude less than .04. In the partial run, we can achieve 5. Thus, there is no tradeoff between VRA effectiveness and SEG score at the Senate level.

4.2.4 Findings for Congress

Even though the Congressional maps were not challenged in *Agee v. Benson*, we perform the same analysis on that level of redistricting for completeness.

Table 4.6 shows us some basic statistics about the enacted House map and our ensembles. Table 4.7 gives these statistic for the benchmark maps. In the enacted map, by the VRA 44/50 definition, there is one VRA effective districts. In the full state ensemble, we achieve a range of 0 to 2 VRA effective districts.

	CD Enacted Map	CD Full Run
# Districts Scrambled	0	13
# Districts Fixed	13	0
# Counties Split by Full Map	14	(5, 41)
# Munis Split by Full Map	12	(3, 75)
# VRA Effective (44, 50)	1	(0, 2)
# VRA Effective (40, 50)	2	(0, 2)
# VRA Effective (40, 48)	2	(0, 2)
Max # of VRA (0.44, 0.5) While SEG Serial Beats State	0	2
Max # of VRA (0.44, 0.5) While $\|$SEG Serial$\|$ $<.04$	1	2

Table 4.6: Statistics from the Congressional ensemble and the enacted map. Numbers in parentheses denote a range of values (inclusive) achieved by the ensemble.

Table 4.6 also shows the relationship between VRA effective districts and SEG scores. We find that in the full state ensemble, we can find 2 VRA effective districts while improving the SEG score. This is in contrast to the House level, where we could not improve the SEG score of from the enacted map.

	CD Enacted Map	CD Apple v2	CD Birch v2	CD Chestnut
Split Counties	14	18	13	14
Split Municipalities	12	10	10	12
# VRA Effective (44, 50)	1	0	0	1
# VRA Effective (40, 50)	2	2	2	2
# VRA Effective (40, 48)	2	2	2	2
SSEG	0.033	0.011	-0.011	0.033

Table 4.7: Various scores for Congressional benchmark maps.

Rather than using the enacted map's SEG score as a benchmark, we can also ask if we can increase the number of VRA districts while having SEG magnitude less than .04. In the full run, we can achieve 2 VRA districts while having SEG magnitude less than .04. Thus, there is no tradeoff between VRA effectiveness and SEG score at the Congressional level.

4.3 Conclusion

We find that overall, there is little to no tradeoff between VRA effectiveness and partisanship. Only in the case of the full House scramble did we find it hard to improve upon partisanship as measured by SEG. However even in that scenario, there were many maps that improved upon VRA effectiveness while maintaining the same degree of partisanship, and almost all maps found by the ensemble had SEG magnitude less than .08.