Comparing Electoral Systems for the Massachusetts Legislature

MGGG Redistricting Lab

Contents

Contributors

Moon Duchin, Jack Gibson, Chanel Richardson, and Emarie De La Nuez contributed to the study design and execution and to the writing of this report. We thank the many members of the MGGG Redistricting Lab who participated in the development of the models and code used here.

This project was commissioned by New America, and we are grateful to the members of that organization— particularly Oscar Pocasangre, Maresa Strano, and Lee Drutman—for support and feedback. We also thank Nathan Lockwood for very helpful comments and conversations.

Version 2.0 | October 2024

1 Introduction

In this study, we have applied a suite of statistical models to test alternative systems of election in the setting of the Massachusetts state legislature, varying styles of voter behavior and other features of the electoral context. We present extensive results for *single transferable vote* (STV), which is a style of ranked choice voting with multi-member districts. Broadly, we find that STV is likely to secure representation that is in line with *support proportionality*, or the share of voters who support each slate of candidates. Our focus is on the election of candidates of choice of communities of color, but we also consider representation by political party.

1.1 Background

The election system in the United States is highly decentralized, allowing states and localities to enact their own voting procedures—as a result, voting methods can and do vary by state, county, and locality. Across that variety, one system is enduringly popular in American elections: plurality voting in single-member districts (PSMD), in which the one candidate who receives the most votes wins each district. This requires each jurisdiction to be subdivided into many territorial domains, raising the specter of abusive line-drawing practices known as *gerrymandering*. But worse—as numerous researchers have independently concluded—minority groups are systematically underrepresented by plurality election systems when single-member districts are blindly drawn, even without the intent to gerrymander.

Reformers have turned to other systems of election as mechanisms of change. Some systems allow voters to rank their preferences among the candidates, allowing support to transfer to lower-ranked choices when candidates are eliminated or pass the threshold of election with excess votes. The two main ranked choice voting systems are called IRV (instant runoff voting) and STV (single transferable vote). IRV, in particular, has considerable reform momentum around the United States. In 2022, Alaska joined Maine to become the second state employing IRV for statewide elections, and it has passed a first vote and awaits a second vote to confirm its adoption in Nevada. These join a large number of localities, especially across California, using IRV for mayoral races.[1] In local elections, Cambridge, Massachusetts has used STV to conduct city council and school board elections (magnitude 9 and 6, respectively) continuously since 1941. In 2023, Portland, Oregon became only the second sizeable city to adopt STV; it's fair to say that STV is not as far along but is getting a close look around the country. [2] See [5, Ch 20,21] for detailed information about how the systems work.

In Massachusetts, our primary focus is the state's bicameral legislative body, formally called the Massachusetts General Court, which is composed of a 40-member Senate and a 160-member House of Representatives. Historically, as in most states, White members of the legislature are present in shares that outpace population shares by a significant margin, while Black, Latino, and Asian residents are proportionally underrepresented [6]. In partisan terms, the recent composition of the state legislature leans heavily Democratic; Republicans hold just three to four seats in the Senate (7.5-10%) and twenty-five seats in the House (15.6%), despite pulling over 30% of the major-party vote in nearly every statewide election.[3]

[1] Voters in Memphis, TN approved IRV for their mayoral elections, but in 2022 the state legislature voted to ban it.

[2] Eastpointe, Michigan, adopted STV for one election, but then reverted to plurality when a consent decree expired.

[3] Peter Durant flipped a seat in Central Massachusetts in a special election in November 2023, raising the number of

Ultimately, the modeling presented here will allow us to assess how the move to an alternative system, such as STV, might have consequences for racial or partisan proportionality in either body.

1.2 Study design

The goal of the present study is to isolate the effects on representation as election conditions vary one at a time, holding voter preferences constant. We apply these questions in the Massachusetts Senate first, and then House, paying particular attention to the ability of communities of color to elect candidates of choice.

There are many possible systems of election to consider, as well as many possible formats for eliciting voter preferences. One popular reform option is to employ *rankings*—but it is important to remember that ranking is only a ballot style; several reasonable systems could then be applied to determine a winner set or consensus ranking once the ranked ballots have been tallied. In this project, we primarily focus on the system called single transferable vote, or STV for short. Below, we use "zone" as a succinct term for multi-member districts, "White" for non-Hispanic single-race White respondents on the Census, and "POC" (people of color) for all other respondents.

Models and methods. In prior work, MGGG has introduced three novel generative models for simulating voter rankings. These are known as **PL** (the Plackett-Luce or *impulsive voter* model, in which a voter selects from a pool of candidates sequentially, with likelihood in proportion to a fixed set of preference weights), **BT** (the Bradley-Terry or *deliberative voter* model, in which a voter weighs each pairwise comparison of candidates against preference weights), and **CS** (the Cambridge Sampler, in which ballot types are pulled from historical STV cast vote records in Cambridge, Massachusetts).[4]

In addition, we employ randomized algorithms to create alternative districting plans in which the PL/BT/CS models will be used to simulate elections. We will often use the term "zone" for these districts when they are larger and elect multiple representatives. To generate them, we use a graph algorithm known as *recombination*, which is surveyed in [3] and has been widely used in research and in redistricting litigation.[5] For the STV modeling, we generate large ensembles of alternatives and choose five plans from those, so that they are approximately independent, before simulating elections in each zone in each plan. (See §A.2 for details.) For the party list comparisons in §4, where no computationally costly ranking simulations are needed, we use a full ensemble of 100,000 alternative plans. A full replication repository is provided at [9].

Parameters. The results below are pulled from model runs that use specified choices and parameters to vary the conditions under which the election is held.

- **Structure**, or the system of election. In addition to STV, we've also modeled the status quo system of plurality in single member districts, party list voting, and IRV (*instant runoff voting*, which also uses ranked ballots but elects only one member per district).

Republicans in the Senate from three to four.

[4]The articles [1, 2] contain formal descriptions of the models and ideas about how they can be modified and tested. The white paper entitled *Modeling the Fair Representation Act* [4] is an earlier example of work using these models to study Congressional districting nationwide. In addition, MGGG maintains an open-source codebase called `VoteKit` to support experimentation with these models and other social choice research [7].

[5]The Python package called `GerryChain` offers an open-source implementation of recombination [8].

- **Magnitude**, or the number of representatives elected out of each district. Our models below use $m = 1$ (single-member districts), or $m = 4, 5$ (multi-member districts).
- **Ranking behavior**. Our model parameters allow us to vary whether voters rank their candidates impulsively (PL model) or deliberatively (BT model); we can also pull historical ranking data from Cambridge STV elections to get realistic real-world behavior—and in particular, the tendency of voters to truncate their ballots—incorporated into the model.
- **Bloc cohesion**, or the level of polarization: how much does each bloc stick to its own slate versus crossing over to support candidates from the other slate.
- **Candidate pool**. Who is able to run for office at all and appear on the ballot for election? This, of course, has a major impact on the representational outcomes. When the minority runs a large number of preferred candidates the models let assess the risks of vote splitting.
- **Candidate strength**, measuring the tendency of a voter bloc to agree on the order of their candidates. For instance, there may be a consensus strong candidate or there may be more variation of candidate order. This can help us model the impacts of coordinated campaigns of voter mobilization.
- Finally, **voter turnout** is a crucial variable that allows us to measure the effects if one bloc has a higher rate of voting participation than the other, for any combination of reasons.

Benchmarks. Let us start by setting our baseline expectation of how much representation to expect by candidates preferred by people of color in Massachusetts. We start by taking the POC share of the electorate to be roughly 30%, in line with the 29.57% share of voting age population (VAP) made up of non-White residents in the 2020 Census. (Later, we will consider a hypothesis of reduced turnout.) Our basic cohesion assumption will be W70/C80, meaning that White voters support their slate of preferred candidates with an overall 70% tendency, while POC voters have slightly higher 80% cohesion. This means combined support for POC-preferred candidates is made up of 80% of POC voters and 30% of White Voters.

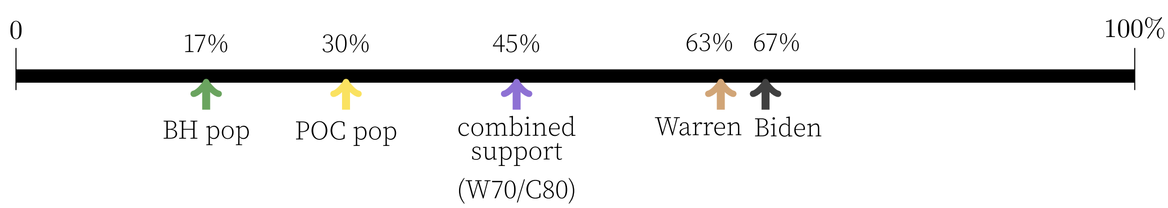

Figure 1. A series of benchmarks for proportional performance help us to contextualize the outcomes below: Black and Hispanic voting age population share (17%); combined POC VAP share (30%); combined POC & White support for the minority-preferred candidates under baseline polarization (45%); support for Warren (D) in Sen18 and Biden (D) in Pres20 (63% and 67%).

- POC population proportionality. POC coalition representation in direct proportion to the share of VAP would translate to just under 30% of the seats. This means 11.8 seats out of 40 for POC-preferred candidates in a 40-member Senate, or about 47.3 seats in a 160-member House.

- Black and Hispanic only. If we don't assume that all communities of color incline to support a common slate of preferred candidates, but restrict to Black and Hispanic voters only, then the share drops steeply to roughly 17% of VAP, which leads to proportional projections of about 6.8 seats out of 40 or 27.2 out of 160. Noting that the POC share of citizen voting age population is about 22% and that turnout rates for POC voters are low in Massachusetts compared to other states, this 17% benchmark can double as a more conservative estimate of the POC share of voters in legislative contests.[6]
- Support proportionality. In this study, our baseline assumption for group cohesion is that there are disjoint slates of candidates who can be identified as White-preferred or POC-preferred; 80% of POC voters support the POC slate, while 70% of White voters incline to the White-preferred slate. Thus if we take into account the combination of support for the POC slate—80% of POC voters are joined by 30% of White voters who cross over—then the level jumps to 44.8%, which is nearly 18 seats out of 40 or 71.7 seats out of 160.
- Party proportionality. Joe Biden received roughly 67% of the major-party vote in 2020, while Elizabeth Warren notched 63% against Geoff Diehl in the 2018 Senate contest. These would correspond to 26.8 and 25.2 seats out of 40, respectively, or 107.2 and 100.8 out of 160. Since these candidates were preferred by both POC and White voters alike, this should *overstate* support for the POC-preferred slate under conditions of polarization.

Taken together, these give us a set of benchmarks so that we can understand the model projections in the context of other possible outcomes.

2 Modeling STV for state Senate

2.1 State Senate, basic STV with 5-member districts

We now describe the basic model specifications, which create a starting point from which we will vary conditions one at a time in subsequent sections.

Structure and magnitude: STV with 4- and 5-member zones. **Ranking behavior:** we will not choose a single setting, but will compare the three ranking scenarios in each plot. **Bloc cohesion:** We start with W70/C80 cohesion (as described above) for the full POC voting bloc. **Candidate pool:** The state is divided into eight 5-member zones, and we assume 15 candidates run in each zone. We set the C/W candidate split among those 15 candidates to be proportional to the POC/W population in the zone, rounded to the nearest integer. For instance, if a zone has 55% POC share of voting age population, then we run our simulations with 8 POC-preferred candidates and 7 candidates preferred by the White voting bloc. **Voter turnout:** Rounding out our basic scenario, we start with each group turning out at the same rate as a share of voting age population.

It is important to note that these specifications are not chosen for maximal realism, but as a baseline against which we will test the effects of different variations, including lower cohesion, different composition of candidates, and a turnout gap.

[6]The Census Bureau releases estimates of registration and voting by race by state. See for instance `https://www.census.gov/data/tables/time-series/demo/voting-and-registration/p20-586.html`

How to read the model output summary plots. Figures 2-6 show Senate modeling outputs in a common style. The vertical stripes show the proportional number of seats for the benchmarks shown in Figure 1. (Note that the x axis does not extend all the way from zero to 40 seats.)

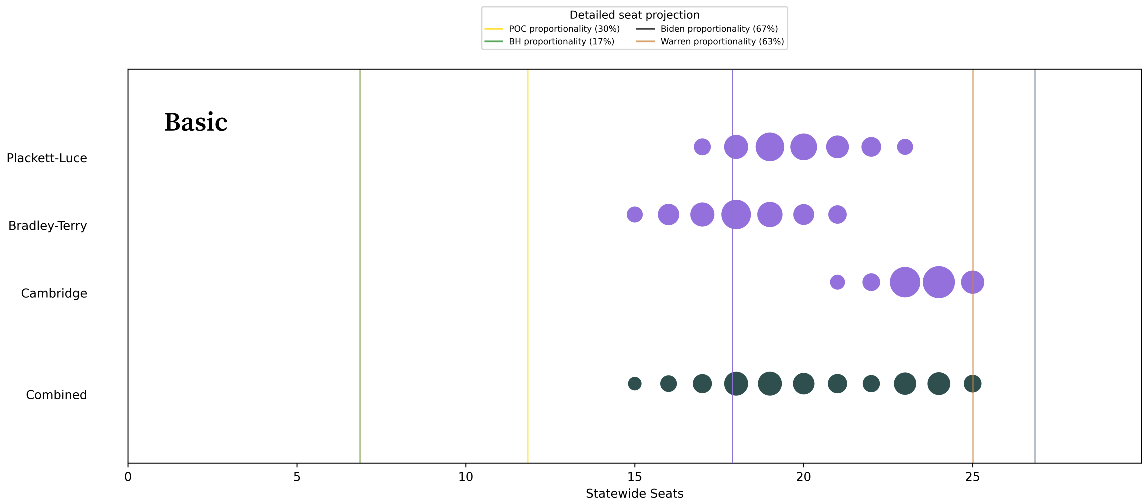

Figure 2. When we drew 5 random zoning plans and simulated 100 ranked elections in each zone, PL ranking behavior ("impulsively" selecting from the overall preference weights) led to 17-23 seats for POC-preferred candidates, BT behavior ("deliberatively" selected from preference weights) led to 15-21 seats, and following historical Cambridge ranking patterns gave 21-25 seats. That is most in line with support proportionality for these specifications—18 seats.

Each row of purple symbols represents runs with a different generative model: PL, BT, and CS, respectively. We drew five zoning plans for the state at random, and simulated 100 ranked elections in each zone with each model, creating 500 statewide seats outcomes in each of those rows. For instance, if a particular simulated contest in a particular map resulted in the election of 2, 0, 5, 3, 1, 2, 3, and 0 POC-preferred candidates across its eight zones, that would be recorded as 16 statewide seats for POC-preferred candidates overall. The size of the round symbols in a row corresponds to the frequency with which each outcome was observed with the runs from that behavior model; for instance, in the Plackett-Luce runs, 19 statewide seats for POC-preferred candidates was the most common outcome observed, which was slightly more frequent than the 20-seat outcome. All 1500 values are presented together in the Combined row (gray dots) as a summary. Figure 7 is similarly structured, but the red colored symbols are a reminder that this time the minority in question is partisan (Republicans) rather than racial and ethnic (people of color).

In many of the runs featured here, Cambridge-style voting gives quite different results from the PL and BT models; it is also the most prone to big swings as the conditions of election change. The primary for this is that PL and BT are set up so that every ballot is a complete ranking of all available candidates; because Cambridge sampler pulls from historical ballots, it can and frequently does generate truncated ballots, and even bullet votes (see also [1]). Exhausted ballots, in which votes can no longer transfer, blunt the proportionality tendencies of STV.

2.2 Isolating the role of turnout

When turnout for the POC coalition drops to just 2/3 the rate of White turnout, we expect the representation for POC-preferred candidates to suffer. The modeling reflects this, but the dropoff is modest, with 12-24 instead of 15-25 seats expected.

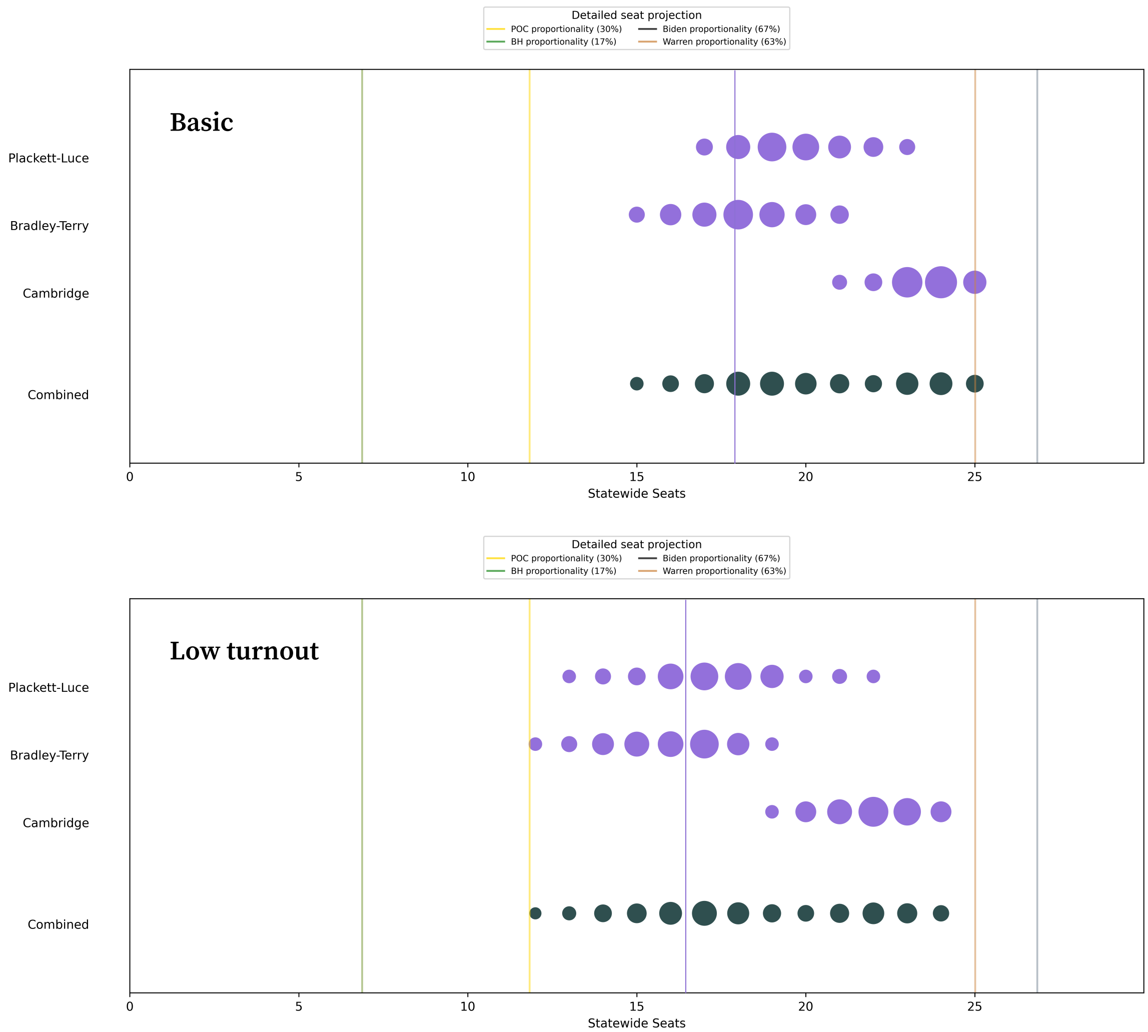

Figure 3. Top: basic scenario. Bottom: 2/3 turnout rate for POC. Since people of color make up nearly 30% of voting age population in Massachusetts, they would also constitute 30% of the electorate if turnout were equal across the blocs. If the turnout rate is 2/3 that of White voters, POC presence in the electorate would drop from 30 people out of 100 to 20 out of every 90, or just under 22%. (See A.1 for a discussion of demographics and turnout.) However, our assumptions include a substantial rate of crossover voting, which buffers the modeled electoral outcomes from displaying a calamitous drop.

2.3 Varying subgroup size and candidate pools

There is no question that the size and makeup of the candidate pool has a major qualitative impact on electoral outcomes: shutouts of minority preferences often occur in status quo plurality elections when there are many POC candidates in a primary, and supporters split their votes, allowing another candidate to get through with a small share of the overall vote. This famously occurred in the Detroit-area primary election of 2022 under the new Congressional maps in Michigan, allowing an eccentric Indian-born businessman to be elected with just 28.3% of the vote against eight Black candidates who collectively received the other 71.7%.[7] One of the main arguments in favor of ranked choice voting is that vote transfer can mitigate the risks of vote splitting.

To see this in sharp relief, first consider a scenario in which we drop the assumption that POC (non-White) voters have a common preferred slate of candidates; instead, we restrict the minority bloc to Massachusetts residents identifying as Black and/or Hispanic/Latino, moving Asian/AAPI, Native American, and Other-identified voters together with the White majority. This drops the population share of the highlighted minority from 30% to just 17% of voting age population. However, when we run an identical set of experiments to the basic runs above, changing only the population fraction, we see a very surprising outcome (Figure 4). Instead of losing electoral representation as their voter share drops by about half, the representation actually records a modest *gain*. This is because the choice to set up candidates in proportion to minority population now gives 3 minority-preferred candidates out of 15 in each zone, on average. This turns out to be extremely advantageous, allowing *all* of them to be elected in the Cambridge voting model. This is because the Cambridge data features a tendency for voters to abbreviate their ballots, with many voters choosing just one, two, or three options instead of ranking all available candidates, and it also features frequent alternation between slates by all voters. The combination of fewer candidates, shorter ballots, and frequent alternation produces a very sizeable advantage for a small group.

What this powerfully illustrates is it can be greatly advantageous for a group to have relatively *fewer* candidates. A smaller slate (as in Figure 4, middle) lets the minority group avoid the massive dropoff in performance that occurs when voters spread their support over the slate and ballots are exhausted, wasting their capacity to transfer (as in Figure 4, bottom).

2.4 Isolating the role of polarization and crossover voting

Most results in this study show POC-preferred candidates overperforming population proportionality, due to our model assumptions which don't posit solid bloc preferences, but rather allow for a significant rate of crossover vote support. Allowing crossover is realistic: POC voters can certainly rank White-preferred candidates highly; White voters can incline towards candidates who are generally POC-preferred; and all voters may intersperse candidates from various slates according to personal and political preferences of all kinds. Figure 5 shows the significant swings that result when the crossover patterns swing from one extreme to the other.

[7] See Ballotpedia, `ballotpedia.org/Michigan's_13th_Congressional_District_election,_2022`.

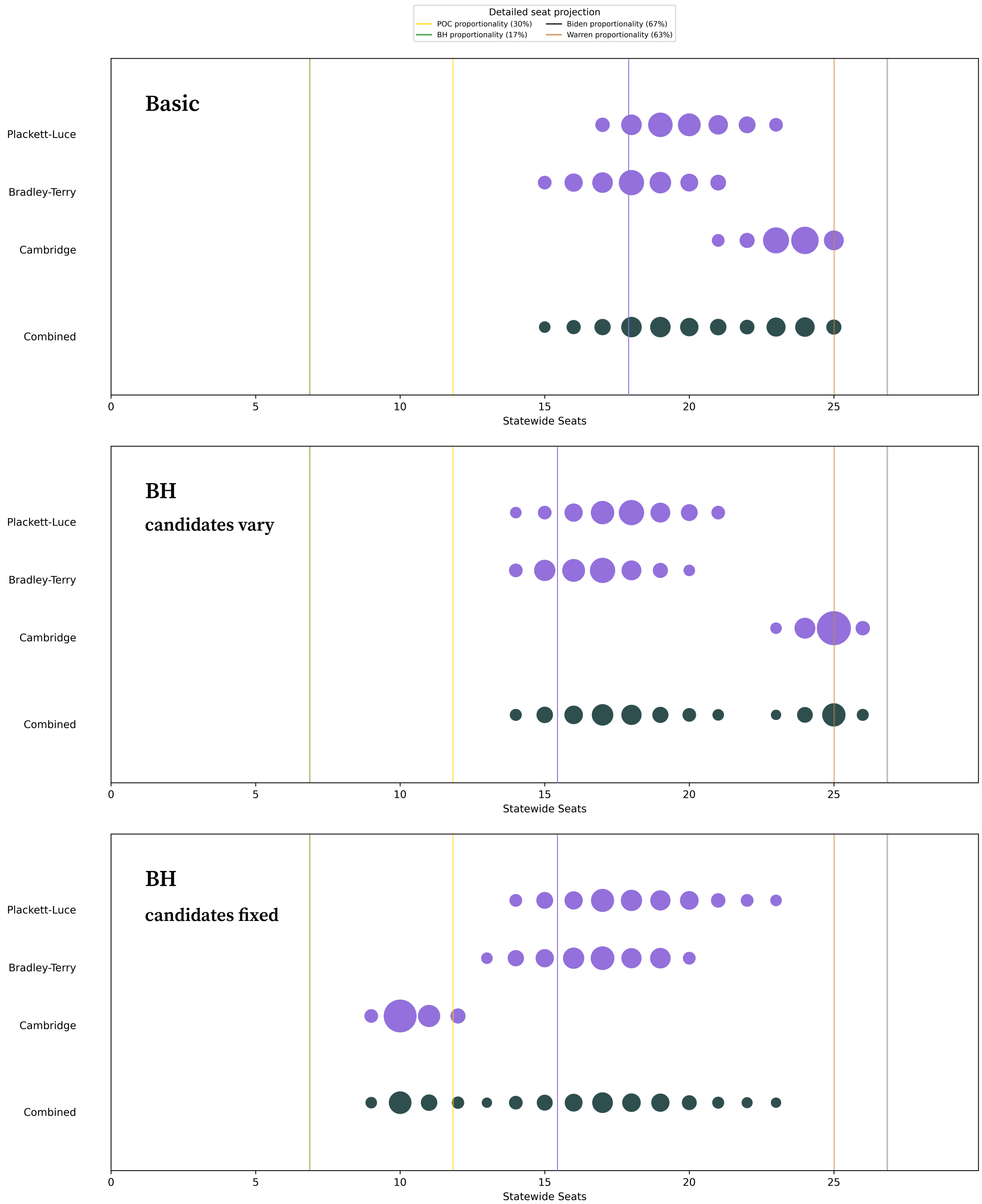

Figure 4. Top: basic scenario. Middle: the minority bloc drops from 30% to 17% when we restrict the POC population to Black and Hispanic voters (but not Asian and other non-White groups). However, under Cambridge voting patterns, representation actually *improves*, because we allowed the candidate slate to contract proportionally, which is very favorable for the minority. Bottom: by contrast, when a fixed seven candidates out of 15 are minority-preferred but the specs are otherwise identical to the middle plot, representation in the Cambridge model drops sharply.

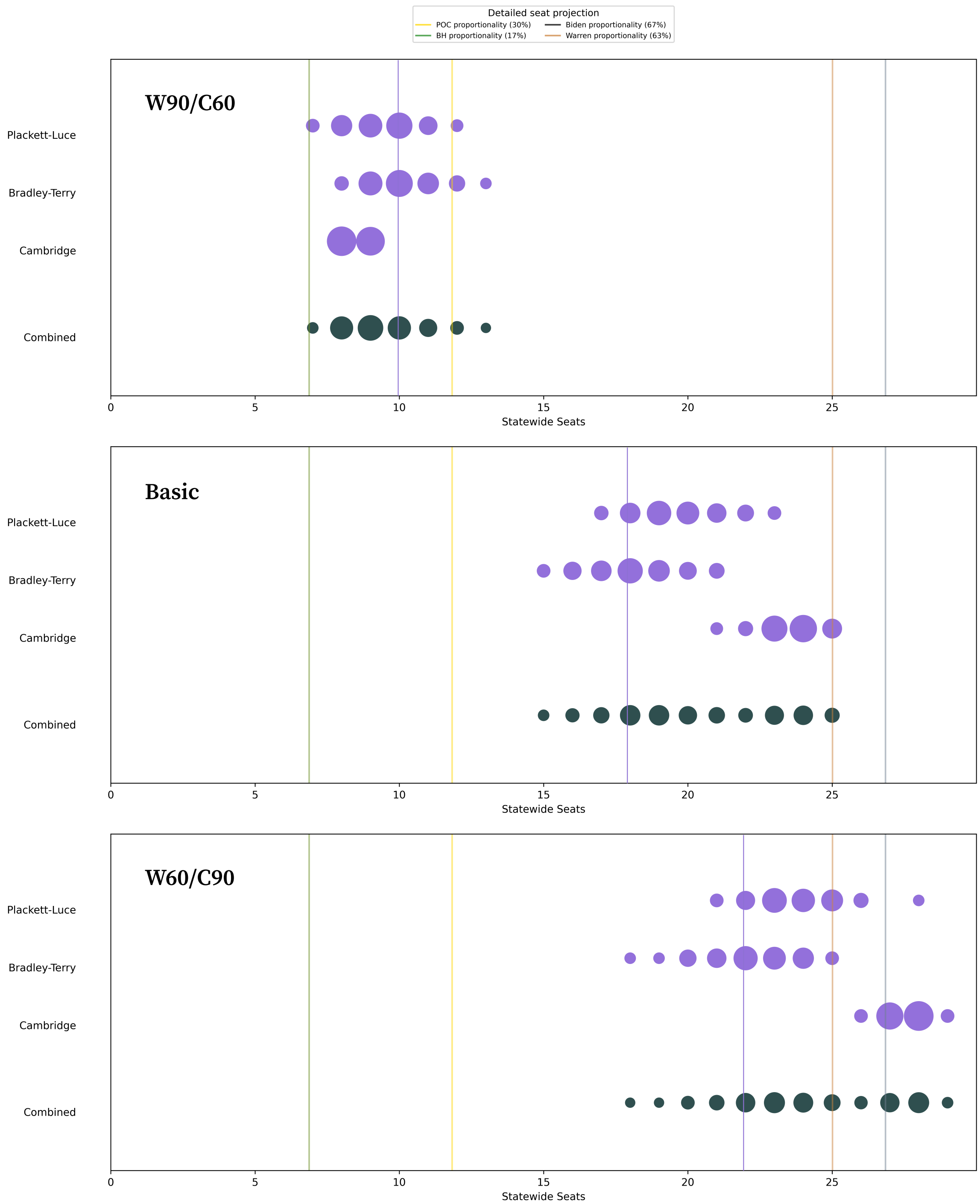

Figure 5. Top: White voters highly cohesive while POC voters are not (W90/C60). Middle: the basic scenario, repeated (W70/C80). Bottom: POC voters highly cohesive while White voters are not (W60/C90).

2.5 Magnitude effects

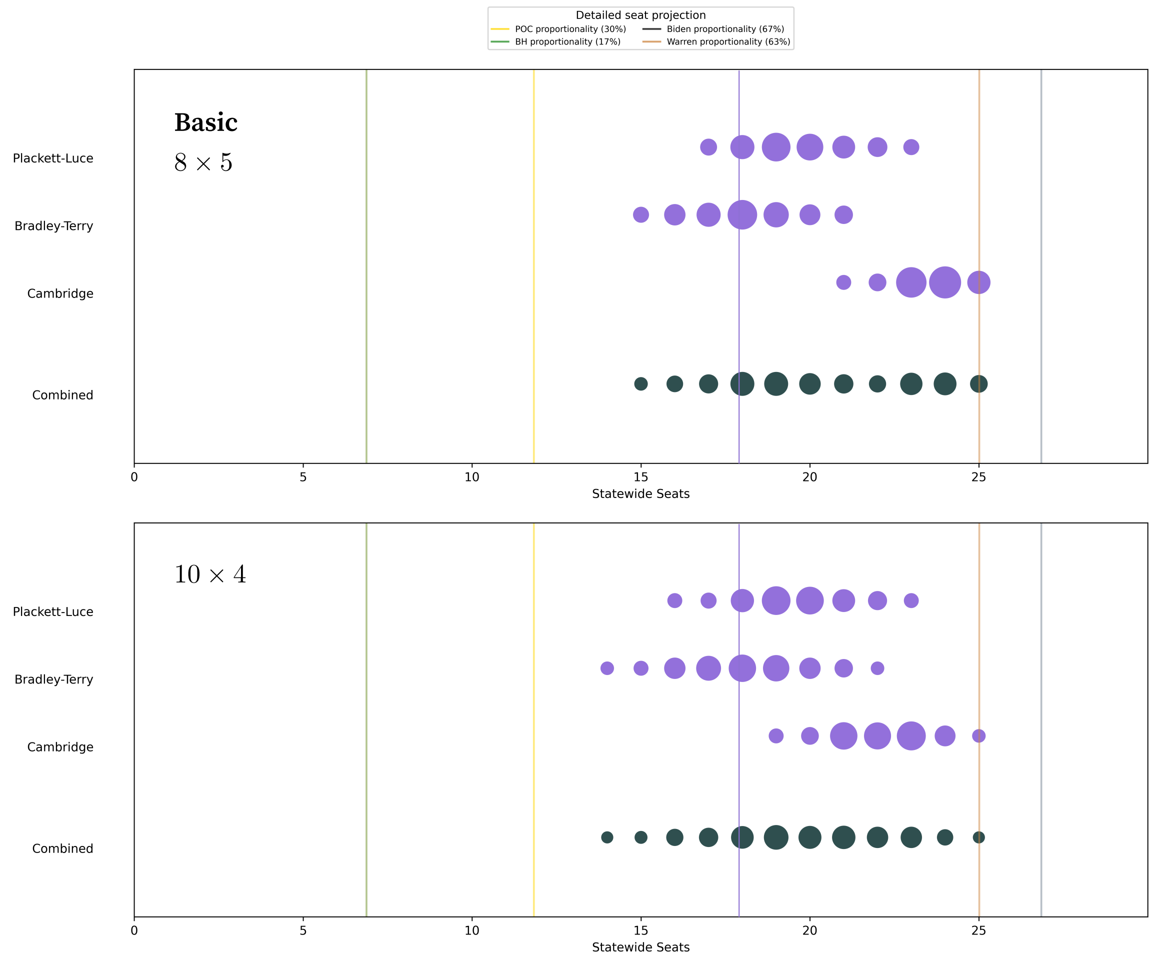

Figure 6. Top: basic, with 8×5 zones. Bottom: 10×4. The projected difference between eight 5-member zones and ten 4-member zones is quite modest.

For a 40-member body, natural choices of zoning structure would be 8×5 and 10×4. An advantage for voters of the lower magnitudes would be the likelihood of seeing fewer candidates running per zone, reducing the information burden on voters.[8] From a technical point of view, the main difference is in the rounding effects. For instance, we could imagine that a minority group tends to reach a high enough threshold for 2 out of 5 seats, but not enough for 2 out of 4. If that effect is consistent enough, it could make the difference between 40% and 25% outcomes. As long as the population is heterogeneously distributed, however, rounding effects should sometimes incline in each direction and largely cancel out. That is consistent with what we see in Figure 6.

[8] In recent years, several advocacy groups have inclined against using $m = 4$ magnitude in their reform proposals, preferring $m = 3$ or $m = 5$. One reason is the assumption that voters might be discouraged by a tied 2-2 partisan outcome in a zone if their votes inclined more toward one party than the other. A counterpoint view is that an even outcome would be looked on favorably by the public in the case of a nearly even vote preference.

2.6 Partisan outcomes

For our final look at 8 × 5 STV modeling, we set aside the perspective of communities of color and turn to political parties. This time it is Republicans in the minority in Massachusetts.

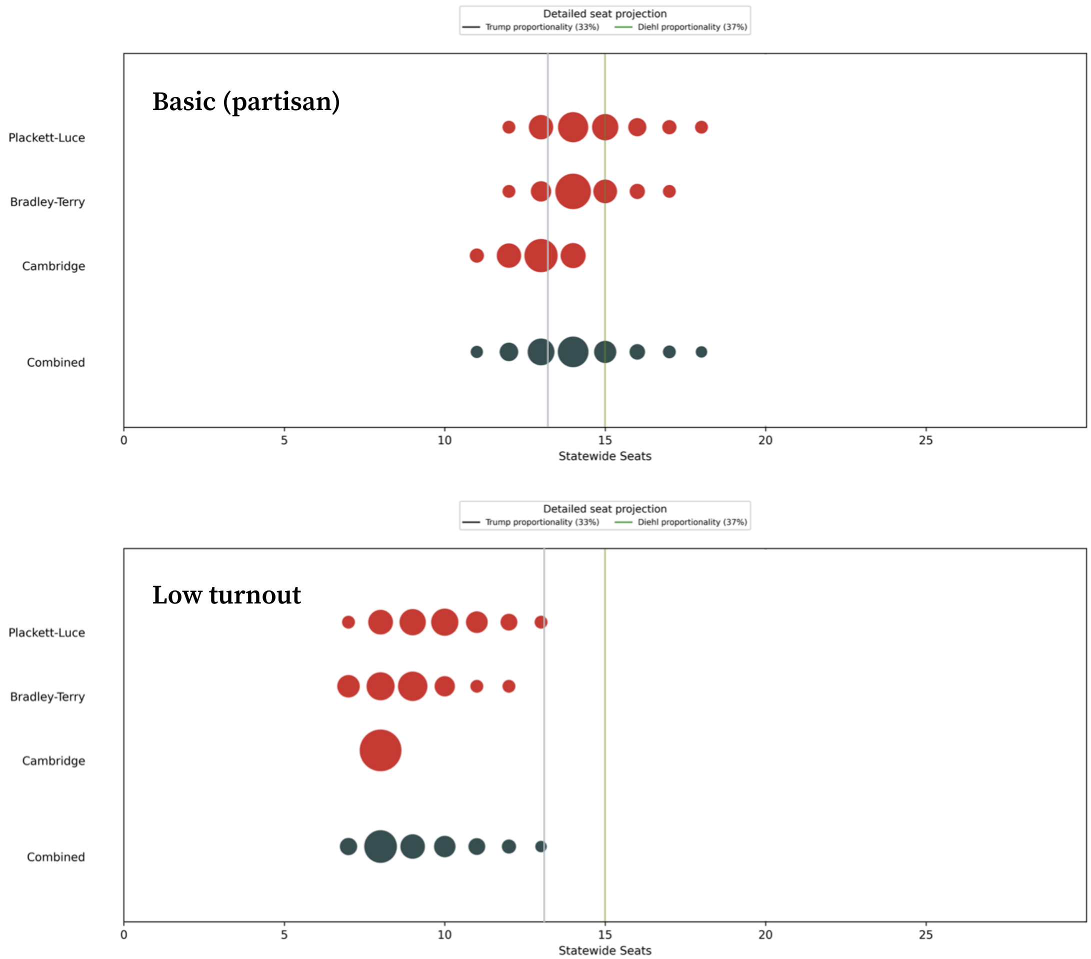

Figure 7. Top: equal turnout. Bottom: R turnout 2/3 of D. Projections for partisan STV elections for the 40-member Senate are shown here, using Sen18 votes to label Democrats and Republicans.

As a baseline for Republican performance, we use the 2018 U.S. Senate race between Elizabeth Warren (D) and Geoff Diehl (R); this was one of the stronger Republican years across recent presidential and senatorial contests, with Diehl receiving roughly 37% of the major-party vote. Diehl voters will serve as a baseline for Republican preference, and Warren voters mark the Democrats.

We set 90% D cohesion (that is, 9-to-1 preference for Democrats by Warren voters) and 80% R cohesion, with 15 candidates composed in proportion to D/R voters in each 5-member zone. This gives rough proportionality for Republicans in the simulations with equal turnout; this time, depressing the minority (R) turnout to 2/3 that of the majority takes a much bigger bite out of their electoral performance, because there is little crossover to cushion the losses.

3 Modeling STV for state House

Next, we repeat experiments for the 160-member state House. These are essentially identical to those presented above for the Senate, so we present them in a block.

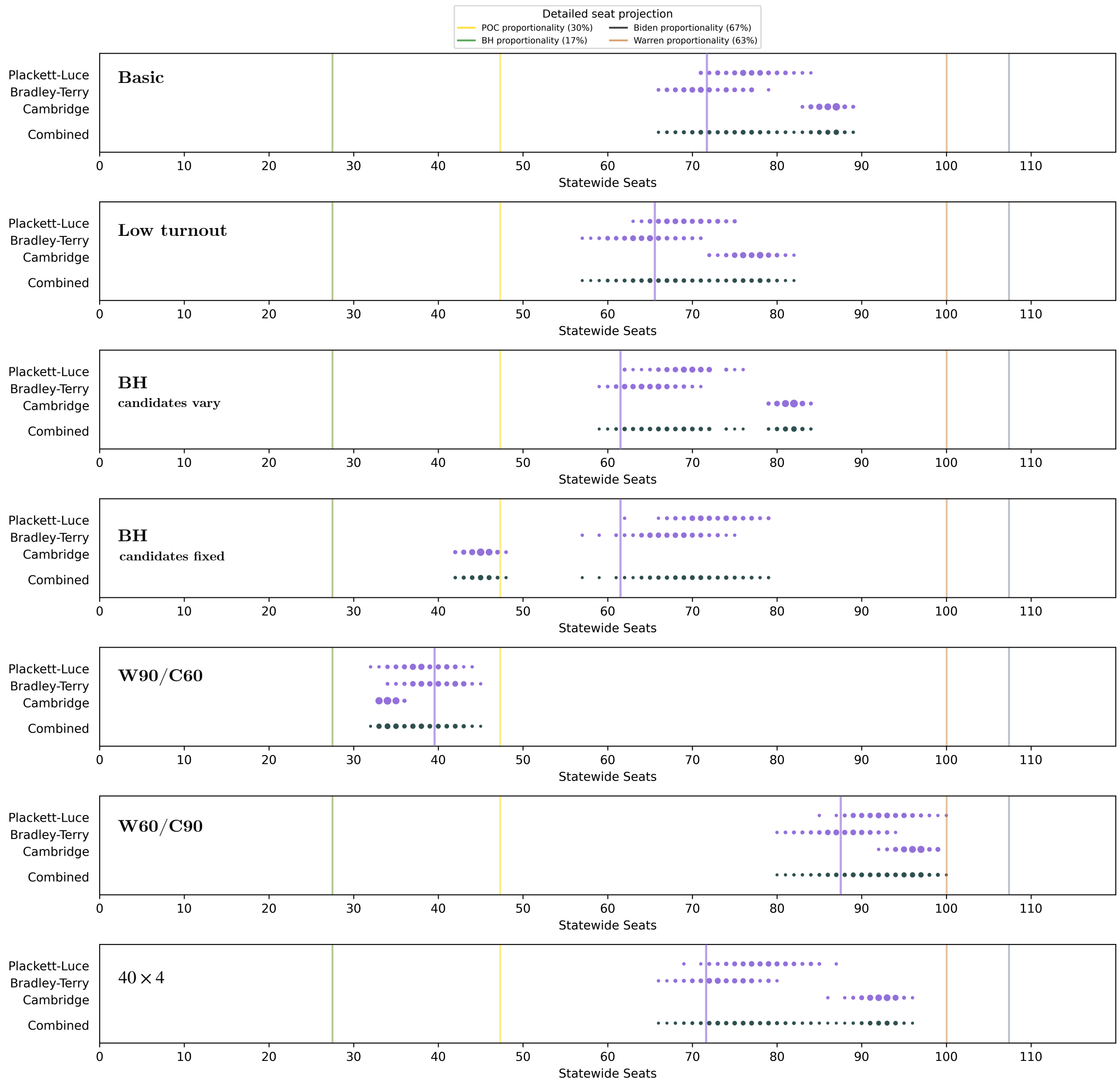

Figure 8. The baseline scenario is contrasted to alternatives varying turnout, candidate availability, polarization, and district magnitude. The combined level of voter support for POC-preferred candidates (purple line) moves as the scenario varies, with STV outcomes from simulated elections usually moving in a corresponding fashion.

The baseline 32×5 STV modeling for House uses W70/C80 cohesion and 15 candidates per zone, with the C and W slates present in proportion to POC population share in each zone. Next, we compare this to a low-turnout scenario with the same population base (POC 2/3 the rate of W).

When the minority is restricted to just Black and Hispanic communities instead of broader non-White POC, we observe the same effect as in the Senate, where outcomes depend heavily on the candidate pool in the Cambridge model (where voters are likely to vote incomplete ballots). Once again, having sharply reduced numbers of *candidates* can cause Black and Hispanic voters to overperform their population share.

Next, we vary the cohesion to extremes, with high cohesion for White voters, but low for POC voters (W90/C60) and the reverse (W60/C90). These are nearly the most unfavorable and favorable cohesion assumptions for the POC bloc, respectively, given that each group must favor its own slate by definition.

For the final comparison in Figure 8, we consider changing the district magnitude from five to four. The projected difference between 32 5-member zones and 40 4-member zones is modest, but the four-member zones perform better at this scale. And, crucially, the Senate itself is a 40-member body; this means that using single-member Senate districts as 4-member House districts is an eminently reasonable prospect.

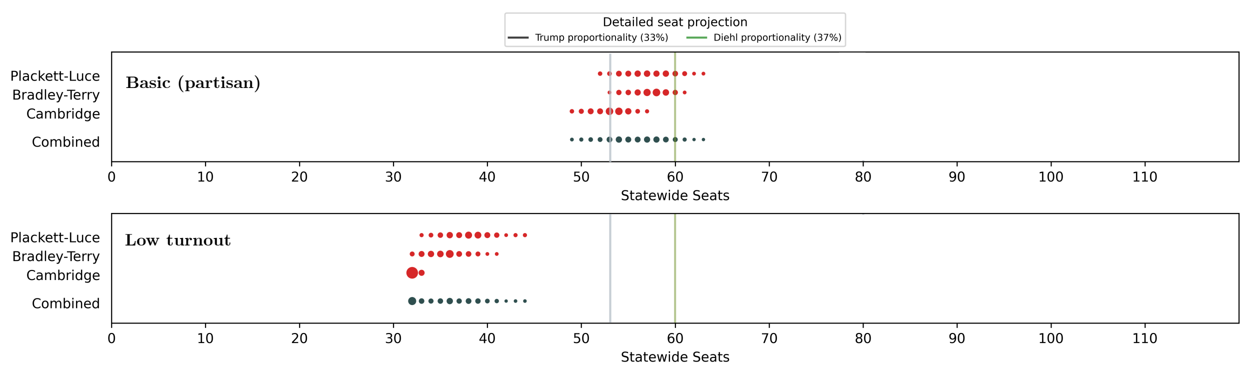

Figure 9. Projected outcomes of STV elections with partisan bloc voters, using Diehl's support in the Senate race of 2018 as the pattern of Republican voting.

Turning to partisan dimensions of representation in Figure 9, we again see a picture that is consonant with the earlier discussion in the Senate. With a baseline of 90% Democratic cohesion and 80% Republican cohesion—where those supporting Diehl over Warren are used to geographically locate Republicans—Republican electoral performance under 5-member zones would be well in line with their share of the population. But the geographic dispersion of Republicans (compared to people of color in the main analysis) means that they are especially susceptible to low turnout, which sees representation drop significantly. Even under these low turnout conditions, projected representation still exceeds the status quo (recently roughly 25 Republican members in the House).

4 Cross-system comparison: Plurality, party list, IRV, and STV

Having discussed plurality, IRV, and STV, we introduce one more system in order to set up a four-way comparison: **party list voting**, the system most widely used around the world to secure proportionality in governing bodies.

4.1 Introducing party list

Party list is an expressly proportional system in which the parties receive a number of seats as close as possible to their share of the vote in each zone; the seats are then filled by using a priority order that is either directly constructed by the party (*closed list*) or at least partially modifiable by voter preference (*open list*). Whether the list is open or closed does not matter for the overall party composition of elected officials, which only depends on how the vote share in each zone is rounded to a whole number of seats.

The plots in Figure 10 show the mechanism in action—whether in a Democratic-favoring contest like President 2020 (Biden/Trump) or a Republican-favoring contest like Governor 2018 (Baker/Gonzalez).

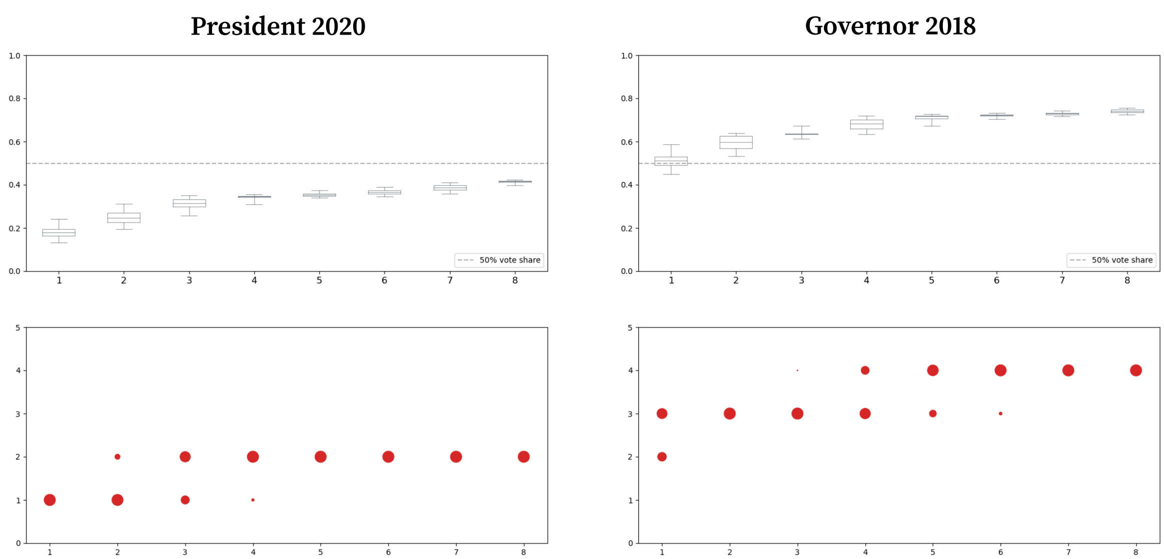

Figure 10. Top: box-and-whiskers plot showing the Republican vote share in each of the eight zones. In these plots, the districts are ordered from the lowest Republican share (numbered 1) to the highest (numbered 8). The range of outcomes is shown over 100,000 districting plans drawn at random using the *recombination* methods described in the supplemental material. Bottom: frequency of winning each number of seats out of the five available seats in each zone under party-list voting.

Note that in the Baker landslide election, more than half of the 100,000 districting plans have a Baker majority in *every one* of their eight zones. Accordingly, the most common outcomes are 3-4 R seats out of 5 in each zone, which sums to approximately 28 seats out of 40; compare the top plot in Figure 11, where this contest is shown in the yellow histogram.

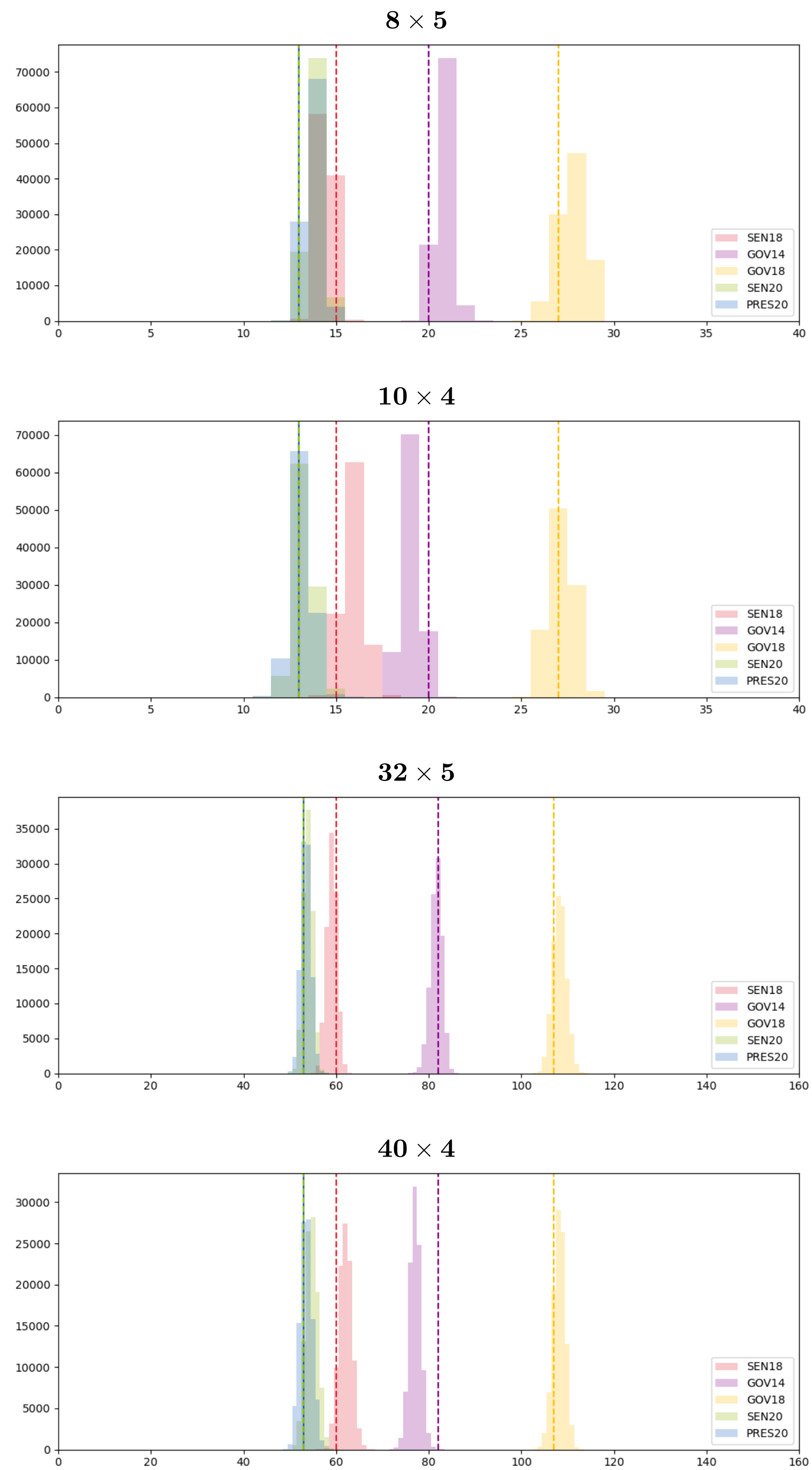

Figure 11. Party list histograms, showing the expected overall near-proportionality in each case. Rounding effects in four-member zones show up as slightly harmful overall for Republicans in the Gov14 pattern (purple), but helpful in the Sen18 pattern (red).

4.2 System compare

We now have the tools to compare four different systems functioning under corresponding electoral conditions. As before, we use Warren/Diehl support from the U.S. Senate contest in 2018 as a proxy for Democratic/Republican preference. This makes it easy to assess the plurality winner in each single-member district: we simply count whether there were more D or R voter. We do this across an ensemble of 100,000 districting plans to get the histograms at the top of Figures 12-13. Next, we run party list as above, with the five seats per zone awarded to D/R candidates in proportion to D/R votes. Next is an *open IRV*, in which 14 candidates run in every zone. In the 7R,7D version, the partisan split of candidates is fixed; in the variable version, the number of R candidates is in proportion to the R voters in that district. Voters rank their preferences using the PL, BT, and CS models used throughout this paper, and IRV is run to pick one winner in each zone. Finally, STV is run on exactly the same candidate and voter settings as IRV for a direct comparison.

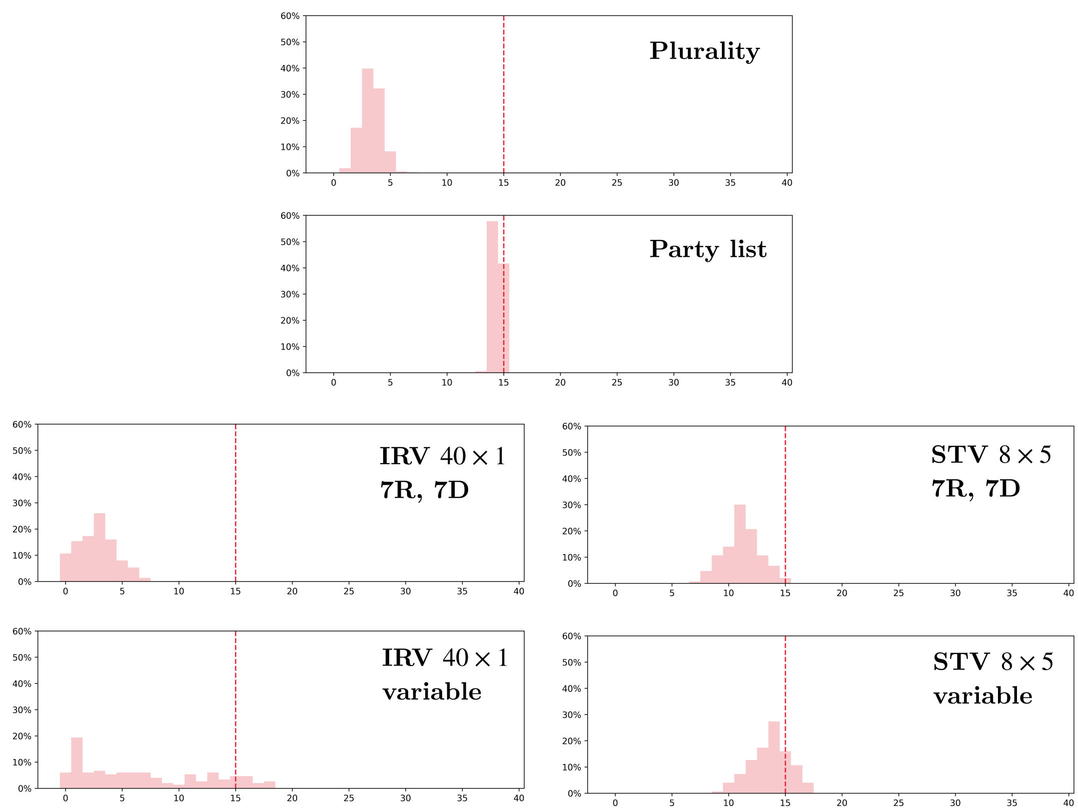

Figure 12. An apples-to-apples comparison of systems for the Senate under controlled electoral conditions. In each case we project the number of seats won by Republicans under the voting pattern for Geoff Diehl versus Elizabeth Warren in the 2018 Senate race.

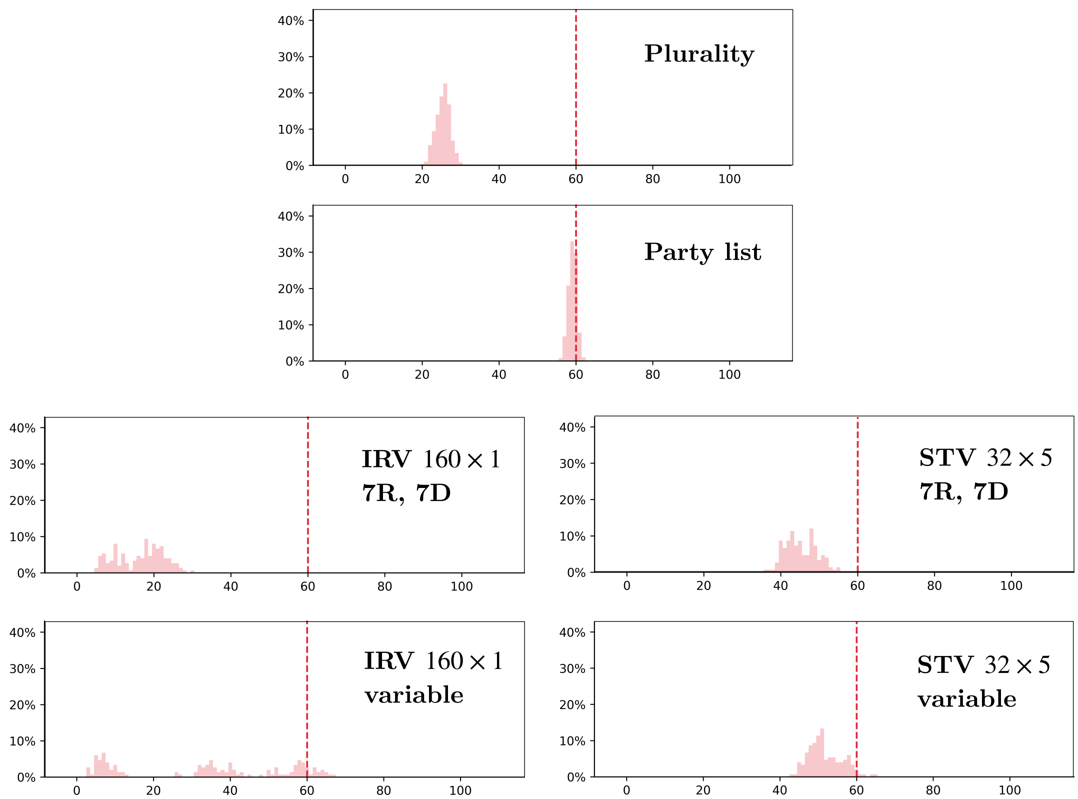

Figure 13. Similar comparison of systems for the House.

Figures 12-13 highlight some very surprising observations when we juxtapose the results. First, note that at the time experiments were conducted, the actual number of Republicans in the legislature is 3 in the Senate and 25 in the House—right in line with the plurality projections made by the model. This should increase confidence in the model.

STV electing five representatives per district gives nearly proportional outcomes, close to the performance of party list despite the use of geographical districts and a geography in Massachusetts that is famously unfavorable to Republicans.[9] On the other hand, the performance of "open IRV" is comparable to the status quo plurality system—and in fact less favorable for Republicans—when the candidate pool is balanced. Switching to a variable candidate pool, where few Republican candidates contest seats in very "blue" areas, has an extremely large effect and can even help IRV reach proportionality under some of the simulated elections.

It is important to recognize that this form of IRV, with no primary election and a single "jungle-style" general with many candidates, is not identical to reform proposals by advocacy groups in Massachusetts, or in any other location to our knowledge. We feature it here to highlight that identical

[9] One caveat to keep in mind is that the Cambridge data comes from non-partisan municipal elections and was coded by the race of the candidate, not by party. We employ it to get realistic variety in the number of candidates ranked and the alternation between candidate types on realistic voter ballots, but its justification is clearly diminished when mapping onto partisan legislative elections.

ranking behavior, identical voter blocs, and a similar ranked choice mechanism can have widely different outcomes based on whether a single-winner or multi-winner format is adopted. Primaries and restricted ballots may still be necessary to prevent fence-outs in IRV; the system alone will not suffice to ensure rough proportionality. By contrast, STV is much less sensitive to these elements of system design.

5 Conclusion

The tools of statistical modeling enable comparisons across electoral systems, while varying assumptions about voting conditions one at a time. This gives users access to robust projections that can inform reform efforts with realistic trends and caveats.

Modeling studies must be interpreted with caution, as their outputs are far from providing guarantees of future performance. These techniques can only show how outcomes vary if some factors are held constant while others change; the models do not have the capacity to predict endogenous factors, such as the impact on voter mobilization brought on by system change itself, or the likely changes to campaign strategies and candidate availability that would ensue. This kind of applied modeling is best regarded as a tool to tease out structural elements of systems: central tendencies, on one hand, and combinations of hypotheses that jar the projected outcomes, on the other.

The broad findings under STV in the two chambers of the state legislature (§2-3) indicate that STV is likely to provide representation of choice for communities of color in proportion to the *combined support* for their slates by minority and majority voters—in sharp contrast to the status quo, where the representation falls far short of the actual level of combined support. These proportional outcomes are observed in our experiments across a wide range of variations in voter behavior, cohesion, candidate pools, candidate strength, district magnitudes, and the choice of boundary lines for zones.

In order to highlight that this proportional tendency is far from automatic, we present a comparison from the Republican point of view in §4. This shows that the mere adoption of ranked choice formats for voters does not secure this outcome—our hypothetical "open IRV" system gives projections that can lag as far behind proportionality as status quo plurality elections do, depending on the detailed conditions. This squares with an obvious fact about single-member elections: no clever choice of system can give a bloc of voters their "fair share" of just one person. Multi-member districts, paired with a ranked choice system, combine to produce better proportionality both in each district and in the legislature as a whole.

References

[1] Gerdus Benadè, Ruth Buck, Moon Duchin, Dara Gold, and Thomas Weighill, *Ranked Choice Voting and Proportional Representation*.
Preprint available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3778021.

[2] Gerdus Benadè, Chris Donnay, Moon Duchin, and Thomas Weighill, *Proportionality for Rankings*. Preprint available at https://mggg.org/PRVTP.

[3] Daryl DeFord, Moon Duchin, and Justin Solomon, *Recombination: A Family of Markov Chains for Redistricting*. Harvard Data Science Review **3** (2021).

[4] Moon Duchin and Anthony Pizzimenti, *Modeling the Fair Representation Act*.
White paper at https://mggg.org/FRA-Report and https://mggg.org/FRA-Supplement.

[5] Moon Duchin and Olivia Walch (Eds.), Political Geometry: Rethinking Redistricting in the U.S. with Math, Law, and Everything In Between, Birkhäuser, 2022.

[6] Peter Levine, Benjamin Forman, and Laurel Bliss, *MassForward: Advancing Democratic Innovation and Electoral Reform in Massachusetts*.
White paper at https://massincmain.wpenginepowered.com/wp-content/uploads/2019/11/MassForward.pdf.

[7] MGGG Redistricting Lab, *VoteKit Python Package*, http://github.com/mggg/VoteKit. Accessed Nov 21, 2023.

[8] MGGG Redistricting Lab, *GerryChain Python Package*, http://github.com/mggg/GerryChain. Accessed Nov 21, 2023.

[9] MGGG Redistricting Lab, *Massachusetts RCV Replication Repo*, https://github.com/mggg/Massachusetts-RCV. Accessed Mar 17, 2024.

A Additional notes on methodology

A.1 Demographics, registration, and turnout

This study mostly uses voting age population (VAP) as the basis for demographic numbers. VAP is found in the Decennial Census release, which is the authoritative source of data for redistricting. Citizen voting age population, or CVAP, is only provided in the American Community Survey, which was compromised in 2020 by collection challenges related to COVID-19. If we use the 5-year rolling average from the period 2017-2021, then the POC shares are roughly 29% of total population; 26% of voting age population; and 22% of citizen voting age population. Thus the scenario in Figure 3, where the minority makes up 22% of the electorate, is also in proportion to this estimate of CVAP.

The U.S. Census Bureau releases periodic estimates of voting and registration rates broken down by age, sex, race/ethnicity, and various other social and economic variables. For instance, November 2022 estimates can be found here. According to those estimates, those residents identified as non-Hispanic single-race white (which we call White in the shorthand of this report) registered to vote at a rate of 74% of VAP, while all other Massachusetts residents (called POC) were registered at a rate of 44% of VAP. In terms of who actually voted in November 2022, the Bureau estimates that White voter turnout was just under 60% of VAP, while POC turnout was 28% of VAP. These estimates place the POC share of the November 2022 electorate at about 16%.

This is partially attributable to the unusually uncompetitive legislative and congressional electoral environment, with large numbers of uncontested races and very few people of color on the ballot. (See, e.g., [6].) This gives us reason to believe that a transition of election systems that unsettles the patterns of candidacy could result in a substantial shift in participation.

A.2 Generating random zoning plans

Recalling that *zones* is the term used here for multimember districts, we now describe the randomized algorithm used to create zoning plans for Massachusetts for use in the studies presented above. Plans were generated using an open-source software package called GerryChain [8], implementing a *recombination* algorithm that was developed for use in redistricting research and litigation. This algorithm uses a Markov chain (random walk) process to iteratively transform plans by fusing a pair of districts into a double-district ($2\times$ ideal size), then halving that into two new districts at random to replace the ones that were fused. The mechanism used to split the double-district is to generate a random spanning tree, then search for an edge whose deletion leaves two population-balanced components. This replaces the previous pair of districts with a new pair. This process of making successive changes to districting plans, one pair of districts at a time, can then be repeated thousands or hundreds of thousands of times to produce a large and diverse collection of maps, called an *ensemble* of alternative plans. See DeFord–Duchin–Solomon [3] for details.

Because two successive plans in a recombination chain differ in only two districts, it may take dozens or hundreds of steps before all or nearly all districts have been replaced. This means that if a small sample of plans is required, they should be chosen with substantial spacing from a chain of modifications. For instance, to choose five plans for each STV simulation study above, we take 100,000 steps and select every 20,000th plan. Since standard heuristics indicate good sample quality, this produces approximately independent draws from the sampling distribution.