Analysis of partisan bias in the PA house plan proposed by the LRC (12/16/2021) for public review

John F Nagle

Introduction and Qualifications

I am a professor emeritus at Carnegie Mellon University in the Department of Physics and in the Department of Biological Sciences. My research there since 1967 obtains meaningful quantities from data in the fields of physics, biophysics, chemistry and biology, including data obtained from simulations and from experiments of my research group and others. Google Scholar reports over 24000 references to my over 200 publications.
https://scholar.google.com/scholar?hl=en&as_sdt=0%2C39&q=John+F+Nagle&btnG=

Since 2012 I have been extracting partisan bias from election data. I have written four peer reviewed papers in one of the most important journals that covers this subject. My most recent paper notes my connection with the DRA software which implements some of my methodology for obtaining partisan bias from districting plans.

Election Law Journal 20 (2021) 116-138 with A. Ramsay @ DRA
On Measuring Two-Party Partisan Bias in Unbalanced States

Election Law Journal 18 (2019) 63-77.
What Criteria Should Be Used for Redistricting Reform?

Election Law Journal 16, 196-209 (2017).
How competitive should a fair single member districting plan be?

Election Law Journal 14, 346-360 (2015)
Measures of Partisan Bias for Legislating Fair Elections

I initiated this report and I have not been compensated for it.

Summary

Measures of partisan bias are briefly reviewed. All the measures agree that the proposed house plan is biased in favor of the Republican party, although less so than the current plan. Simulations, when properly interpreted, support this conclusion.

Methodology substantiating the first conclusion regarding bias in the house map

There are many metrics for measuring partisan bias that have been devised by scholars. A good reason for the plenitude of metrics is the difficulty of evaluating bias in a state that leans far towards a single party.[1] However, when a state is nearly equally balanced between two major parties, the methodology is greatly simplified as will be shown. To substantiate taking advantage of this simplification, Table 1 shows that Pennsylvania is a well-balanced, essentially 50/50 state when considering the state house. The average two party vote percentage is shown by the blue number, in the last row and in the D 2-party column, to be 48.9% Democratic and therefore 51.1% Republican. (A similar 2-party vote of 49.0% was obtained for the legislative senate.) Table 1 also shows that the average percentage of Democratic seats was only 43% in the same period.[2]

Election Year	D vote all HDs	R vote all HDs	D vote 2-party %	D seats	R seats	D seats 2-party %
2020	3017689	3416942	46.9	90	113	44.3
2018	2568968	2075093	55.3	93	110	45.8
2016	2755058	2852921	49.1	82	121	40.4
2014	1408624	1825181	43.6	84	119	41.4
averages	2437585	2542534	**48.9**	**87**	**116**	**43.0**

Table 1. Elections since the last house reapportionment were chosen. The 2-party D percentages are obtained by dividing D votes by D+R votes. The website https://en.wikipedia.org/wiki/yyyy_Pennsylvania_House_of_Representatives_election provides votes and seats data in the above link where one replaces yyyy by the year.

Let us turn now to metrics of bias, ten of which are shown in Fig. 1.

[1] This is a matter that is discussed in detail in my 2021 paper.

[2] One expects a winner's bonus in single member district election systems of about R=2. (The winner's bonus is defined as the ratio of the percentage above 50% in seats divided by that percentage of the vote above 50%.) The winner's bonus for the current map should be flagged as a too large value, 6.4. This is consistent with the anti-majoritarian result in 2018 when the Democrats received considerably more than half the vote and considerably fewer than half the seats.

Metric		Description
• Proportional	2.08%	The simple deviation from proportionality using fractional seat shares
• Efficiency gap	2.23%	The relative two-party difference in wasted votes
• Gamma	2.18%	The fair difference in seats at the map-wide vote share
• Seats bias	2.18%	Half the difference in seats at 50% vote share
• Votes bias	1.31%	The excess votes required for half the seats
• Partisan bias	2.19%	The difference in seats between the map-wide vote share and the symmetrical counterfactua
• Global symmetry	2.71%	The overall symmetry of the seats-votes curve
• Partisan bias rating	71	The combined rating of seats bias & votes bias
• Declination	5.47°	A geometric measure of packing & cracking
• Mean–median	1.87%	The average vote share across all districts minus the median vote share

Figure 1. Screenshot from the Advanced section of DRA[3] that shows many metrics of partisan bias and their values when applied to the LRC proposed plan for the house. This screenshot uses the composite President 2016 & 2020 data which had 50.15% 2-party D vote share.

Brief descriptions of the measures are provided in Fig. 1 and more extensive definitions are given in the information tabs in the DRA Advanced section. A positive value of a metric means that the plan is biased in favor of the GOP. Notice that all ten metrics have positive values when applied to the LRC proposed house plan. Next, notice that the values for Proportional, Efficiency Gap, Gamma, Seats Bias, and Partisan Bias are nearly the same; that is because these metrics become identical when the 2-party vote is 50/50. It is therefore convenient and appropriate to focus on only one of those metrics. The seats bias gives a number from which one estimates how many seats would be expected on average[4] when the 2-party vote is 50% each. Then, the

[3] DRA (Dave's Redistricting App 2020) has the largest variety of partisan bias metrics of the various software packages. Incidentally, I am the inventor of the gamma and the global symmetry metrics.

[4] This assumes that the quality of the candidates and incumbency advantage is equal when averaged over many districts and many elections. Guessing near term outcomes based on knowledge of incumbents is not a valid way to estimate the intrinsic bias of a plan. Bias

value of 2.18% in Fig. 1 estimates the number of Republican seats to be 203(0.5 + 0.0218) = 105.9 and the number of Democratic seats to be 203(0.5 - 0.0218) = 97.1 when averaged over many elections and candidates.

The DRA software allows one to choose different election data. The DRA default data base is a composite average over all recent statewide elections. This composite includes two landslide elections in 2018 that give it 52.46% 2-party D vote. Table 2 shows that the plan then gives a majority of D seats, as it should for such a substantial D majority vote. However, the seats bias is nearly the same as for the President 16&20 data; both data sets give only 97 D seats for 50% of the vote. Partisan bias is also revealed by the votes bias in the last column of Table 2; 1.22% votes bias means that Democrats would have to obtain 51.22% of the vote to obtain half the seats on average.

Election Data	Vote V%	D Seats @V%	Seats Bias	D Seats @ V =50%	Votes Bias
DRA Composite	52.46	105.6	2.20	97.0	1.22
President 16&20	50.15	97.6	2.18	97.1	1.31
President 20	50.60	101.5	1.06	99.3	0.61
Att General 20	52.33	105.3	0.30	100.9	0.16
President 16	49.62	92.6	3.53	94.3	2.15
Senate 16	49.25	85.3	5.97	89.4	2.73
Senate 18	56.57	118.3	2.55	96.3	1.27
Governor 18	58.67	130.0	2.02	97.4	0.90
average of all	52.46	104.5	2.48	96.5	1.29
standard deviation	3.44	14.1	1.71	3.5	0.82

Table 2. Estimates of bias in the LRC proposed plan using different election data sets. For different election data in column one, column two gives the statewide D vote share. Column three gives the number of D seats at that vote share using the DRA seats/votes curve; rounding these numbers to integers gives the same number as obtained by simply counting the winning party in each district and summing. Subsequent columns give the seats bias, the estimated D seats at 50% 2-party vote share, ending with the votes bias. The final two rows give the average and the standard deviation, respectively, of the previous rows.

Table 2 also shows estimates of bias from other statewide elections. While there are substantial deviations from the average for specific elections, such as Attorney General 2020 and

should estimate how level is the playing field, not the prowess of the players or the resources of the team.

Senate 2016, all estimates of seats and votes bias favor the GOP. Bias in favor of the GOP is further indicated by the election data for President 16&20 which gives the anti-majoritarian result that fewer than 50% D seats would be obtained for greater than 50% D 2-party vote.

Interpretation of simulations

I turn here to criticize a recent solicited report by Dr. Michael Barber that has been entered into the LRC record.[5] The following Table is excerpted from Dr. Barber's report.

Table 3: Comparison of Seat Composition Under Different Elections/Indices

	Commission Plan	
Election Indices:	Number D Districts	Number R Districts
DRA index	105	98
Barber Replication of DRA Index	105	98
Barber 2012-2020 index	107	96
Barber 2014-2020 index	105	98
Barber 2020 index	104	99

It is important that this report acknowledges, in the two rows above the bold black line, that DRA is a valid tool to obtain the number of districts from a plan. What this table and this report does not reveal is that the DRA composite index has a 52.46% D 2-party vote.[6] As I emphasize in my Table 2, 105 D districts are even fewer than what should be obtained with such a vote majority.[7]

---

[5] Michael Barber, Report on Proposed Redistricting Plan from the Pennsylvania Legislative Reapportionment Commission.

[6] The last three lines in Barber's table give similar results for different voting data, but again with no indication of the 2-party vote.

[7] My Table 2 even gives more, 106, seats to Democrats because DRA accounts for the obvious fact that competitive districts should be counted as fractions for each party instead of using simple plurality as was done in Barber's report. See my 2019 and 2021 papers for a discussion of this improvement. Apparently, Dr. Barber agrees because his text on p. 49 also says that the

A popular enterprise is to use a computer algorithm to draw many plans. Barber purports that the LRC proposed plan is biased against the GOP because it yields 8-10 more Democratic seats than is obtained by averaging the ensemble of his simulations. But all Barber's simulated averages would give the antimajoritarian result that fewer than 50% of the two-party vote would give the GOP more than half the seats.[8]

The fallacy of averaging the ensemble of simulations can be revealed by an analogy. A professional basketball coach could consider 1000 people who know how to play the game and then randomly choose an average one to play center. That is like choosing a plan from many simulated plans in the middle of the ensemble of simulated plans. Or the coach could hire LeBron James. That is like picking the LRC proposed plan.

Barber's simulation does illustrate an important fact, namely, that the political geography of PA favors the GOP, and that is because Democrats are relatively more packed in Philadelphia, Pittsburgh and the southeast.[9] The LRC proposed house plan largely alleviates this geopolitical packing bias, but not enough to bias the plan against the GOP.

Given the political geography of PA, fairer plans would likely be found in the tail of the distribution of all plans if avoidance of partisan bias were not included in the code. A new paper (Becker et al. *Election Law Journal*, 2021, **20**, 407-441) from a simulation group that generates hundreds of thousands of plans has made the point that one should not idealize choosing a plan from the center of a distribution (see especially p. 412), and that people ultimately have to do redistricting.

"DRA index predicts 106 Democratic leaning seats." However, these are small differences that do not affect the broader discussion in the text.

[8] It should also be noted that Barber's Table 1 shows that his simulations on average are not as compact as, and split more counties than, the LRC plan.

[9] This well-known fact was mentioned by the LRC chair in his introduction to the 12/16 LRC hearing. Interestingly, the actual extent of this geopolitical bias is much larger for Barber's simulations than it was for the peer-reviewed simulations of Chen and Cottrell, *Elect Stud*, 2016, **44**, 329-340. It is also much larger than in the STATEMENT TO PENNSYLVANIA REAPPORTIONMENT COMMISION REGARDING PROPOSED HOUSE PRELIMINARY PLAN by Kosuke Imai.

Conclusions

The analysis in this memo rests on the principle that, a party that obtains the same number of votes as another party, should obtain, on average over many elections, the same number of seats as the other party.[10,11] The employed DRA methodology estimates the number of seats with small enough uncertainties that it is clear that the proposed LRC house plan is not biased against the GOP, but is instead biased by about 2% in its favor, likely due to the difficulty of overcoming the geopolitical bias of the state. Reported simulations confirm this geopolitical bias; they do not show bias in favor of Democrats for the LRC proposed house plan of 12/16/2021.

John F. Nagle
Carnegie Mellon University
nagle@cmu.edu
http://lipid.phys.cmu.edu
http://lipid.phys.cmu.edu/nagle for districting research

---

[10] This is not the same as proportional representation which requires proportionality for all vote shares. See my 2021 and 2017 papers for a thorough discussion of this distinction.

[11] It is often asserted that fairness should only be concerned with following some procedure that is blind to carefully estimated outcomes. That is not a sound prescription for business or other human endeavors. Its application to districting has been branded the myth of non-partisan cartography by political scientists. (Taylor, P.J. and G. Gudgin. 1976. The Myth of Non-Partisan Cartography: A Study of Electoral Biases in the English Boundary Commission's Redistribution for 1955–1970. Urban Studies 13: 13–25.)