Supreme Court of Ohio Clerk of Court - Filed October 22, 2021 - Case No. 2021-1193

IN THE SUPREME COURT OF OHIO

LEAGUE OF WOMEN VOTERS OF OHIO, *et al.*, *Relators*, *v.* OHIO REDISTRICTING COMMISSION, *et al.*, *Respondents*.	Case No. 2021-1193

AFFIDAVIT OF DR. KOSUKE IMAI

EVIDENCE OF RELATORS – DR. KOSUKE IMAI EXPERT REPORT

(*counsel listing on next page*)

Freda J. Levenson (0045916)
ACLU of Ohio Foundation, Inc.
4506 Chester Avenue
Cleveland, Ohio 44103
Tel: 614-586-1972 x 125
flevenson@acluohio.org

David J. Carey (0088787)
ACLU of Ohio Foundation, Inc.
1108 City Park Avenue, Suite 203
Columbus, OH 43206
(614) 586-1972 x2004
dcarey@acluohio.org

Alora Thomas (PHV 22010-2021)
Kelsey Miller*
Julie A. Ebenstein (PHV 25423-2021)
American Civil Liberties Union 125 Broad Street
New York, NY 10004
(212) 519-7866
athomas@aclu.org
jebenstein@aclu.org

Robert D. Fram (PHV 25414-2021)
Donald Brown** (PHV 25480-2021)
Joshua González (PHV 25424-2021)
Juliana Goldrosen (PHV 25193-2021)
David Denuyl (PHV 25452-2021)
COVINGTON & BURLING LLP
Salesforce Tower
415 Mission Street, Suite 5400
San Francisco, CA 94105-2533 (415) 591 6000
rfram@cov.com

DAVE YOST
OHIO ATTORNEY GENERAL
Michael J. Hendershot (0081842)
Michael A. Walton (0092201)
Bridget C. Coontz (0072919)
Julie M. Pfeiffer (0069762)
30 E. Broad Street Columbus, OH 43215
Tel: (614) 466-2872
Fax: (614) 728-7592
Michael.Hendershot@OhioAttorneyGeneral.gov
Michael.Walton@OhioAGO.gov
bridget.coontz@ohioago.gov
julie.pfeiffer@ohioago.gov

Counsel for Respondents
Governor Mike DeWine, Secretary of State Frank LaRose, and Auditor Keith Faber

W. Stuart Dornette (0002955)
Beth A. Bryan (0082076)
Philip D. Williamson (0097174)
TAFT STETTINIUS & HOLLISTER LLP
425 Walnut St., Suite 1800
Cincinnati, Ohio 45202-3957
T: (513) 381-2838
dornette@taftlaw.com
bryan@taftlaw.com
pwilliamson@taftlaw.com

Phillip J. Strach (PHV 25444-2021)
Thomas A. Farr (PHV 25461-2021)
John E. Branch, III (PHV 25460-2021) Alyssa M. Riggins (PHV 25441-2021) NELSON MULLINS RILEY & SCARBOROUGH LLP 4140 Parklake Ave., Suite 200
Raleigh, North Carolina 27612
phil.strach@nelsonmullins.com
tom.farr@nelsonmullins.com
john.branch@nelsonmullins.com
alyssa.riggins@nelsonmullins.com
T: (919) 329-3812

Counsel for Respondents
Senate President Matt Huffman and House Speaker Robert Cupp

James Smith*
Megan C. Keenan (PHV 25410-2021)
L. Brady Bender (PHV 25192-2021)
Alexander Thomson (PHV 25462-2021)
COVINGTON & BURLING LLP
One CityCenter
850 Tenth Street, NW
Washington, DC 20001-4956
(202) 662-6000
mkeenan@cov.com

Anupam Sharma (PHV 25418-2021)
James Hovard (PHV 25420-2021)
Yale Fu (PHV 25419-2021)
COVINGTON & BURLING LLP
3000 El Camino Real
5 Palo Alto Square, 10th Floor
Palo Alto, CA 94306-2112
(650) 632-4700
asharma@cov.com

Madison Arent*
COVINGTON & BURLING LLP
The New York Times Building
620 Eighth Avenue
New York, NY 10018-1405
(212) 841 1000
marent@cov.com

Counsel for Relators
League of Women Voters et al.
**Pro Hac Vice Motion Forthcoming*
***Pro Hac Vice Motion Pending*

John Gilligan (Ohio Bar No. 0024542)
Diane Menashe (Ohio Bar No. 0070305)
ICE MILLER LLP
250 West Street, Suite 700
Columbus, Ohio 43215
John.Gilligan@icemiller.com
Diane.Menashe@icemiller.com

Counsel for Respondents
Senator Vernon Sykes and
House Minority Leader Emilia Sykes

Erik J. Clark (Ohio Bar No. 0078732)
Ashley Merino (Ohio Bar No. 0096853)
ORGAN LAW LLP
1330 Dublin Road
Columbus, Ohio 43215
T: (614) 481-0900
F: (614) 481-0904
ejclark@organlegal.com
amerino@organlegal.com

Counsel for Respondent
Ohio Redistricting Commission

Affidavit of Kosuke Imai.pdf

DocVerify ID:	F28E1E3D-C0A8-4D4A-A864-16B12998E5F9
Created:	October 22, 2021 11:32:30 -8:00
Pages:	1
Remote Notary:	Yes / State: OH

This document is a DocVerify VeriVaulted protected version of the document named above. It was created by a notary or on the behalf of a notary, and it is also a DocVerify E-Sign document, which means this document was created for the purposes of Electronic Signatures and/or Electronic Notary. Tampered or altered documents can be easily verified and validated with the DocVerify veriCheck system. This remote online notarization involved the use of communication technology.

Go to www.docverify.com at any time to verify or validate the authenticity and integrity of this or any other DocVerify VeriVaulted document.

E-Signature Summary

E-Signature 1: Kosuke Imai (KI)
October 22, 2021 12:01:57 -8:00 [CE102AED8522] [140.247.116.184]
imai@harvard.edu (Principal)

E-Signature Notary: Theresa M Sabo (TMS)
October 22, 2021 12:01:57 -8:00 [D88C9F804C9B] [74.142.214.254]
tess.sabo@gmail.com
I, Theresa M Sabo, did witness the participants named above electronically sign this document.

DocVerify documents cannot be altered or tampered with in any way once they are protected by the DocVerify VeriVault System. Best viewed with Adobe Reader or Adobe Acrobat.
All visible electronic signatures contained in this document are symbolic representations of the persons signature, and not intended to be an accurate depiction of the persons actual signature as defined by various Acts and/or Laws.

Notarial act performed by audio-visual communication

IN THE SUPREME COURT OF OHIO

LEAGUE OF WOMEN VOTERS OF OHIO, et al., **Relators,** **v.** **OHIO REDISTRICTING COMMISSION, et al.,** **Respondents.**	Case No. 2021-1193 Original Action Pursuant to Ohio Const., Art. XI

AFFIDAVIT OF KOSUKE IMAI

Franklin County
/ss
State of Ohio

Now comes affiant Kosuke Imai, having been first duly cautioned and sworn, deposes and sates as follows:

1. I am over the age of 18 and fully competent to make this declaration. I have personal knowledge of the statements and facts contained herein.
2. For the purposes of this litigation, I have been asked by counsel for Relators to analyze relevant data and provide my expert opinions.
3. To that end, I have personally prepared the report attached to this affidavit as Exhibit A, and swear to its authenticity and to the faithfulness of the opinions expressed, and, to the best of my knowledge, the accuracy of the factual statements made therein.

FURTHER AFFIANT SAYETH NAUGHT

Executed on 10/22/2021,2021.

CE193AE0B02
Kosuke Imai
Signed on 2021/10/22 12:01:57 -8:00
Kosuke Imai

Sworn and subscribed before me this 10/22/2021 day of _____, 2021

NOTARY PUBLIC STATE OF OHIO
Theresa Michelle Sabo
Commission # 2016-re-619622
Electronic Notary Public
State of Ohio
My Comm Exp. Nov 28, 2021
Notary Stamp 2021/10/22 12:01:57 PST D88C9F804C9B

D88C9F804C9B
Signed on 2021/10/22 12:01:57 -8:00

F28E1E3D-C0A8-4D4A-A864-16B12998E5F9 --- 2021/10/22 11:32:30 -8:00 --- Remote Notary

EXHIBIT A

Table of Contents

I. Introduction and Scope of Work 3

II. Summary of Opinions 3

III. Qualifications, Experience, and Compensation 4

IV. Methodology 5

- A. Simulation Analysis 6
- B. Metrics Used to Measure Bias 7
- C. Description of Redistricting Simulation Software 9
- D. An Example Simulated Plan 9

V. Statewide Evaluation of the Enacted Plan 9

- A. Evaluation Using the Commission's Approach 10
- B. Evaluation Using the 13 Statewide Election Results 16

VI. Detailed Local Analysis of County Clusters 21

- A. Hamilton County 21
- B. Franklin County 25
- C. Cuyahoga, Summit, and Geauga Counties 27

VII. Appendix 30

- A. Introduction to Redistricting Simulation 30
- B. Incorporating Article XI Sections 3 and 4 into the Algorithm 32
- C. Implementation details 37
- D. An Example Simulated Plan 40
- E. Compliance with Section 6(C) 42
- F. Vote Share for Precincts 42
- G. References and Materials Considered 42

KI_001

I. INTRODUCTION AND SCOPE OF WORK

1. My name is Kosuke Imai, Ph.D., and I am a Professor in the Department of Government and the Department of Statistics at Harvard University. I specialize in the development of statistical methods for and their applications to social science research. I am also affiliated with Harvard's Institute for Quantitative Social Science.

2. I have been asked by counsel representing the Relators in this case to analyze relevant data and provide my expert opinions related to whether Ohio's recently enacted state legislative districting plan (hereafter the "enacted plan") meets the criteria in Article XI, Section 6 of Ohio's Constitution. More specifically, I have been asked:

- To statistically analyze the enacted plan's compliance with Article XI, Section 6(A) by comparing it against other alternative plans that are as or more compliant with other relevant requirements of Article XI.
- To statistically analyze the enacted plan's compliance with Article XI, Section 6(B) by comparing it against other alternative plans that are as or more compliant with other relevant requirements of Article XI.

II. SUMMARY OF OPINIONS

3. I simulated 5,000 hypothetical plans that are at least as compliant with Article XI as the enacted plan. The comparison of these simulated plans with the enacted plan yields the following findings:

- The enacted plan exhibits a significant partisan bias in favor of the Republican party. The magnitude of bias is much greater under the enacted plan than *any* of my 5,000 simulated plans, according to several standard metrics used in the academic literature.
- The enacted plan fails to meet the proportionality criteria of Section 6(B), making it almost certain for the Republican party to win disproportionately more seats relative to their statewide vote share. The degree of disproportionality is much greater under the enacted plan than *any* of my 5,000 simulated plans.

- In several counties including Hamilton, Franklin, and Cuyahoga, the enacted plan packs a disproportionately large number of Democratic voters in some districts while turning other districts into safe Republican seats.

III. QUALIFICATIONS, EXPERIENCE, AND COMPENSATION

4. I am trained as a political scientist (Ph.D. in 2003, Harvard) and a statistician (MA in 2002, Harvard). I have published more than 60 articles in peer reviewed journals, including premier political science journals (e.g., *American Journal of Political Science*, *American Political Science Review*, *Political Science*), statistics journals (e.g., *Biometrika*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society*), and general science journals (e.g., *Lancet*, *Nature Human Behavior*, *Science Advances*). My work has been widely cited across a diverse set of disciplines. For each of the past three years, Clarivate Analytics, which tracks citation counts in academic journals, has named me as a highly cited researcher in the cross-field category for producing "multiple highly cited papers that rank in the top 1% by citations for field and year in Web of Science."

5. I started my academic career at Princeton University, where I played a leading role in building interdisciplinary data science communities and programs on campus. I was the founding director of Princeton's Program in Statistics and Machine Learning from 2013 to 2017. In 2018, I moved to Harvard, where I am Professor jointly appointed in the Department of Government and the Department of Statistics, the first such appointment in the history of the university. Outside of universities, between 2017 and 2019, I served as the president of the Society for Political Methodology, a primary academic organization of more than one thousand researchers worldwide who conduct methodological research in political science. My introductory statistics textbook for social scientists, *Quantitative Social Science: An Introduction* (Princeton University Press, 2017), has been widely adopted at major research universities in the United States and beyond.

6. Computational social science is one of my major research areas. As part of this research agenda, I have developed simulation algorithms for evaluating legislative redistricting since the beginning of this emerging literature. At Harvard, I lead the Algorithm-Assisted Redistricting

Methodology (ALARM; https://alarm-redist.github.io/) Project, which studies how algorithms can be used to improve legislative redistricting practice and evaluation.

7. Back in 2014, along with Jonathan Mattingly's team at Duke, my collaborators and I were the first to use Monte Carlo algorithms to generate an ensemble of redistricting plans. Since then, my team has written several methodological articles on redistricting simulation algorithms (Fifield, Higgins, et al. 2020; Fifield, Imai, et al. 2020; McCartan and Imai 2020; Kenny et al. 2021).

8. I have also developed an open-source software package titled `redist` that allows researchers and policy makers to implement the cutting-edge simulation methods developed by us and others (Kenny et al. 2020). This software package can be installed for free on any personal computer with Windows, Mac, or Linux operating system. According to a website that tracks the download statistics of R packages, our software package has been downloaded more than 25,000 times since 2016 with an increasing download rate.[1]

9. In addition to redistricting simulation methods, I have also developed the methodology for ecological inference referenced in voting rights cases (Imai, Lu, and Strauss 2008; Imai and Khanna 2016). For example, my methodology for predicting individual's race using voter files and census data was extensively used in a recent decision by the Second Circuit Court of Appeals regarding a redistricting case (Docket No. 20-1668; Clerveaux *et al* v. East Ramapo Central School District).

10. A copy of my curriculum vitae is attached as Exhibit A.

11. I am being compensated at a rate of $450 per hour. My compensation does not depend in any way on the outcome of the case or on the opinions and testimony that I provide.

IV. METHODOLOGY

12. I conducted simulation analyses to evaluate the enacted plan's compliance with Sections 6(A) and 6(B). Redistricting simulation algorithms generate a representative sample of all possible plans under a specified set of criteria. This allows one to evaluate the properties of

1. https://ipub.com/dev-corner/apps/r-package-downloads/ (accessed on September 24, 2021)

a proposed plan by comparing them against those of the simulated plans. If the proposed plan unusually favors one party over another *when compared to* the ensemble of simulated plans, this serves as empirical evidence that the proposed plan is a partisan gerrymander. Furthermore, statistical theory allows us to quantify the degree to which the proposed plan is extreme relative to the ensemble of simulated plans in terms of partisan outcomes.

13. A primary advantage of the simulation-based approach, over the traditional methods, is its ability to account for the political and geographic features that are specific to each state, including spatial distribution of voters and configuration of administrative boundaries. Simulation methods can also incorporate each state's redistricting rules. These state-specific features limit the types of redistricting plans that can be drawn, making comparison across states difficult. The simulation-based approach therefore allows us to compare the enacted plan to a representative set of alternate districting plans subject to Ohio's administrative boundaries, political realities, and constitutional requirements. Appendix A provides a brief introduction to redistricting simulation.

A. Simulation Analysis

14. For the purposes of my analyses, I have assumed that the enacted plan is compliant with Sections 3 and 4. I have further ensured that all my simulated plans are equally or more compliant with Sections 3 and 4 than the enacted plan. My simulation procedure achieves this, in part, by exactly following many of the county-level decisions made by Respondents in creating the enacted plan. Appendix B provides detailed information about this process. For all simulations, I ensure districts fall within a 5% deviation from population parity, pursuant to Section 3(B)(1).

15. Section 6(A) states that no plan should be drawn "primarily to favor or disfavor a political party." One can ensure that a plan is compliant with this provision by drawing district boundaries in a way that does not favor or disfavor one political party. Accordingly, when instructing the algorithm to build districts, I apply a party-neutral constraint that places a smaller weight on the likelihood of creating districts that have vote shares for each party too far from 50%. The weight continuously increases as the two-party vote share of a district approaches a 50-50 split, which receives the greatest weight. Appendix C presents the exact formula of this constraint, which

mirrors the way other constraints are imposed on simulation algorithms (Herschlag et al. 2020a).

16. This constraint is designed to discourage "packing," which represents a common feature of gerrymandering (Owen and Grofman 1988; Best et al. 2018; Buzas and Warrington 2021). The boundaries of these packed districts are drawn so that they contain an excessive number of voters from one party, leading to that party disproportionately wasting votes (McGhee 2014; Stephanopoulos and McGhee 2015, 2018). Similarly, the constraint discourages "cracking" to the extent that a group of voters, which could form a majority in a district, is split into small groups across multiple districts.

17. This constraint is party-neutral, encouraging districts that maximize the voting power of each voter equally regardless of their partisanship. In other words, switching the party labels produces identical weights, and hence the same simulation results.

18. Lastly, in the generation of simulated plans, the algorithm does not use any of the partisan bias evaluation metrics discussed in Section B. Rather, such metrics are used to evaluate the resulting set of simulated plans once they are generated, in order to determine compliance with Section 6(A). The algorithm also does not use the proportionality criteria. Instead, I will use this criteria to evaluate the plan's compliance with Section 6(B) based on simulated plans. This separation between algorithmic constraints and evaluation metrics is critical in order to ensure fair evaluation of the enacted and simulated plans.

B. Metrics Used to Measure Bias

19. To measure compliance with Sections 6(A) and 6(B) in the set of simulated plans generated by the algorithm, the enacted plan, and the Democratic caucus plan, I apply a variety of metrics that are commonly used in the academic literature. These metrics are extensively discussed in Dr. Christopher Warshaw's affidavit, dated September 23, 2021, and the references therein. I have reviewed Dr. Warshaw's articulation of these metrics and they are consistent with my understanding, and appear to be applicable to the facts of this case.

20. To represent compliance with Section 6(A), I use the following partisan bias metrics whose definitions are discussed in Dr. Warshaw's affidavit and the references therein.

- Efficiency gap
- Mean-median gap
- Symmetry in the vote-seat curve across parties
- Declination

21. To measure compliance with Section 6(B), I use the proportionality metric, which is defined as the difference between the Republican seat share and the Republican vote share in statewide elections. According to the 13 statewide elections from 2012 to 2020 for which the election results are available at the precinct level (see Appendix G.1 for the list of these elections), the Republican vote share is 53.9% of the votes cast for two major parties when weighting each statewide contest equally. This percentage is essentially identical to the corresponding number (54%), which is reported by the Commission in its Article XI, Section 8(C)(2) Statement. This number reduces to 53.1% if I use the raw percentage of votes rather than the two-party votes. This suggests that my analysis based on two-party vote is more favorable to the enacted map when evaluating its compliance with Section 6(B) than if I used the raw percentage of votes. For each redistricting plan, I compute the average number of Republican seats won using these past statewide elections.

22. I compute the proportionality metric used to measure compliance with Section 6(B) as follows. First, consider the House of Representatives. Given a redistricting plan, I first determine likely winners of all districts based on the vote totals for each statewide election. This gives the total number of expected Republican seats won in each statewide election given the plan. I then average this number across all the statewide elections, arriving at the average number of seats Republican candidates are expected to win. Dividing this by the total number of House districts, which is 99, gives the average expected Republican seat share for the plan. Subtracting from this seat share the statewide Republican vote share for the election yields a measure of proportionality. The same procedure is applied to the Senate. The only difference is that the total number of Senate districts is 33 since the Ohio constitution requires each Senate district to consist of three House districts.

23. When this measure is positive, it means Republicans win more seats on average than their share of votes, and vice versa for Democrats when it is negative. Calculating the number of seats across elections is important, from both a legal and social scientific perspective: political scientists advocate evaluating redistricting plans by averaging across elections (Gelman and King 1994; Katz, King, and Rosenblatt 2020), and Section 6(B) of Article XI of the Ohio Constitution explicitly mandates evaluation on the basis of the statewide elections during the past 10 years.

C. Description of Redistricting Simulation Software

24. In my analysis, I use the open-source software package for redistricting analysis `redist` (Kenny et al. 2020), which implements a variety of redistricting simulation algorithms as well as other evaluation methods. My collaborators and I have written the code for this software package, so that other researchers and the general public can implement these state-of-the-art methods on their own. I supplement this package with code written primarily to account for the redistricting rules and criteria that are specific to Ohio. I conducted all of my analyses on a laptop. Indeed, all of my analysis code can be run on any personal computer once the required software packages, which are also freely available and open-source, are installed.

D. An Example Simulated Plan

25. Figure 21 of Appendix D shows a sample redistricting plan for the House generated using my algorithm. The plan scores the median value according to the proportionality measure described above. Republicans are expected to win an average of 58.9 seats under this simulated plan, using the 9 statewide election results from 2016, 2018, and 2020.

26. Similarly, Figure 22 of Appendix D shows a sample redistricting plan for the Senate generated using my algorithm. The plan also scores the median value according to the proportionality measure. Republicans are expected to win an average of 19.6 seats under this simulated plan, again using the 9 statewide election results from 2016, 2018, and 2020.

V. STATEWIDE EVALUATION OF THE ENACTED PLAN

27. Using the methodology described above, I evaluated the enacted plan's compliance with Article XI Sections 6(A) and 6(B). At the instruction of counsel for the Relators, I also

evaluated the compliance of the Democratic caucus plan, with Sections 6(A) and 6(B). Appendix G.1 provides the detailed information about data sources.

28. I simulated 5,000 alternative House of Representatives plans and 5,000 alternative Senate plans, using the simulation procedure described in Section IV. As explained in Appendix B, every simulated plan is at least as compliant with Sections 3 and 4 as the enacted plan, which I am assuming is compliant with those provisions for the purpose of this analysis. Appendix E also shows that the simulated plans are as compact as the enacted plan, pursuant to Section 6(c).

29. I can easily generate additional compliant plans by running the algorithm longer, but for the purpose of my analysis, 5,000 simulated plans will yield statistically precise conclusions. In other words, generating more than 5,000 plans, while possible, will not materially affect the conclusions of my analysis.

30. Below, I present the results of two evaluations based on different sets of statewide election results. First, I follow the Commission's approach and use a total of 9 statewide elections from 2016, 2018, and 2020 (see Section A). My analysis shows that the enacted plan has worse partisan bias and proportionality scores than any of my 5,000 simulated plans. Second, to give the Commission the benefit of the doubt, I repeat the same evaluation using a more complete set of statewide election results by adding the available election results from 2012 and 2014 (see Section B). I show that using this more complete set of statewide elections does not affect my substantive conclusions.

A. Evaluation Using the Commission's Approach

31. I begin by evaluating the enacted plan's compliance with Sections 6(A) and 6(B), using the Commission's approach. In its Article XI, Section 8(C)(2) Statement, the Commission used only a total of 9 statewide elections from 2016, 2018, and 2020 to compute the expected Republican seat share under the enacted plan. This Commission's approach is not ideal given that Article XI, Section 6(B) states that the statewide voter preferences should be measured using the statewide election results during the last ten years. Nevertheless, I first follow the Commission's approach and evaluate the enacted plan's compliance using this particular subset of statewide elec-

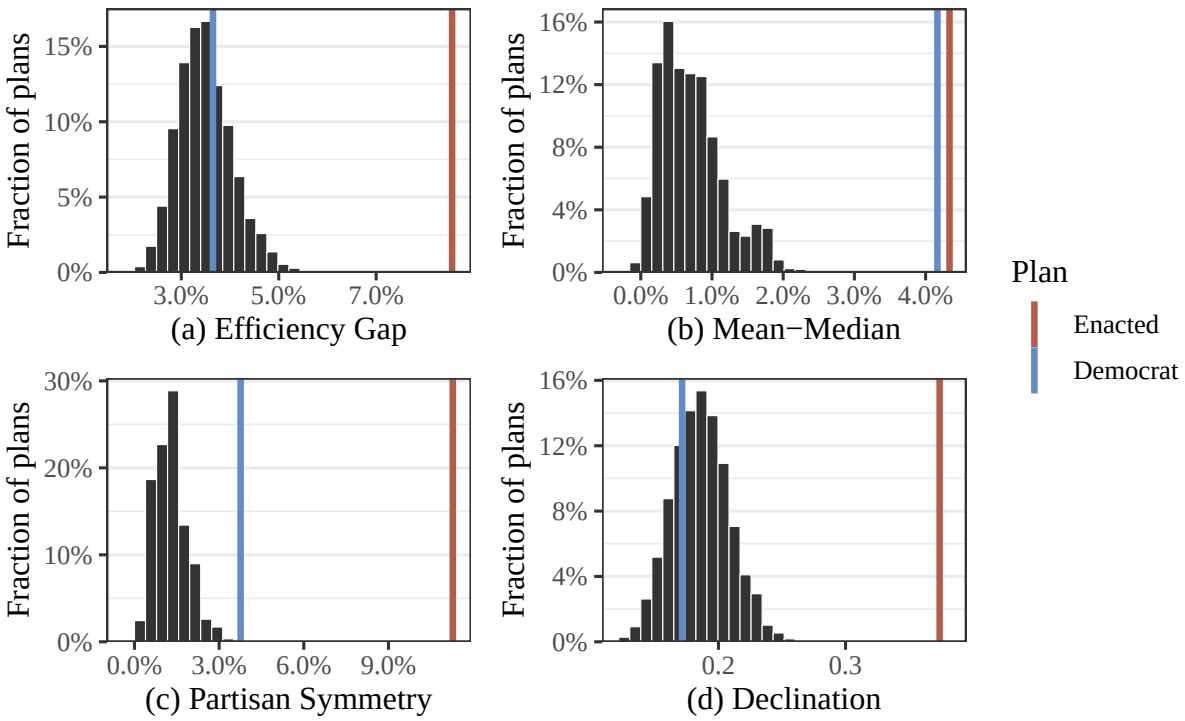

Figure 1: Four partisan bias measures calculated for the 5,000 simulated House redistricting plans computed by averaging across the 9 statewide elections from 2016 to 2020. Overlaid are the values for the enacted plan (red) and the Democratic caucus plan (blue). For each measure, larger values (towards the right) correspond to more Republican-favoring plans.

tion results.

A.1. Compliance with Section 6(A)

32. I first present the results regarding the enacted plan's compliance with Section 6(A) for the House (Figure 1) and Senate (Figure 2). We adjusted the sign of each metric so that a smaller value implies less partisan bias. Recall that the simulated plans follow several of the map-drawing decisions established by the enacted plan (see Appendix B). Despite this constraint, when compared to these simulated plans (black histogram), the enacted plan (red vertical line) is a clear outlier favoring the Republican party for both the House and Senate. Indeed, the enacted map is more biased than any of 5,000 simulated plans for all four partisan bias metrics I considered.

33. For the House, the efficiency gap, which captures both cracking and packing, is 8.6% for the enacted map, whereas the average efficiency gap for the simulated plans is only 3.4%.

This implies that the enacted plan wastes over 110,000 more Democratic votes on average than the simulated plans, and over 110,000 fewer Republican votes. As shown in Figure 1(a), the enacted map is a clear outlier according to this metric.

34. The mean-median gap is a measure of asymmetry in the distribution of votes across districts. The existence of packed districts may lead to a large mean-median gap. Figure 1(b) shows that in terms of the mean-median gap, the enacted plan is also a clear outlier relative to the simulated plans.

35. Partisan symmetry is based on the idea that each party should receive half of the seats if they each receive 50% of votes. Figure 1(c) shows that the enacted plan scores 11.3% on this metric while the simulated plans score 1.2%, on average. This suggests that under the enacted plan, the Republican party would gain roughly 22 more seats than the Democrats, for a hypothetical tied election. In contrast, the simulated plans would give only 2 more seats to the Republican party than the Democrats in the same situation. Again, the enacted plan is a clear outlier according to this metric.

36. Lastly, the declination represents another measure of asymmetry in the vote distribution. As shown in Figure 7(d), the enacted plan also scores worse on this metric than any of the 5,000 simulated plans.

37. The Democratic caucus plan (blue vertical line) scores better than the enacted plan across all partisan bias metrics with the exception of the mean-median metric, for which both plans perform poorly. In addition, this plan is an outlier for the mean-median and partisan symmetry metrics, while it does as well for the other two metrics as most of the simulated plans.

38. For the Senate, my simulation analysis uses the House districts of the enacted plan, which I found to be biased as shown above. Furthermore, as explained in Appendix B, the simulated plans follow additional map-drawing decisions established by the enacted plan. Despite this constraint, Figure 2 shows that the enacted plan is extreme relative to the simulated plans according to all four partisan bias metrics. For example, as shown in Figure 2(a), the efficiency gap of the enacted plan is 10.5% whereas the simulated plans score 3.5% on average for this metric. Like the

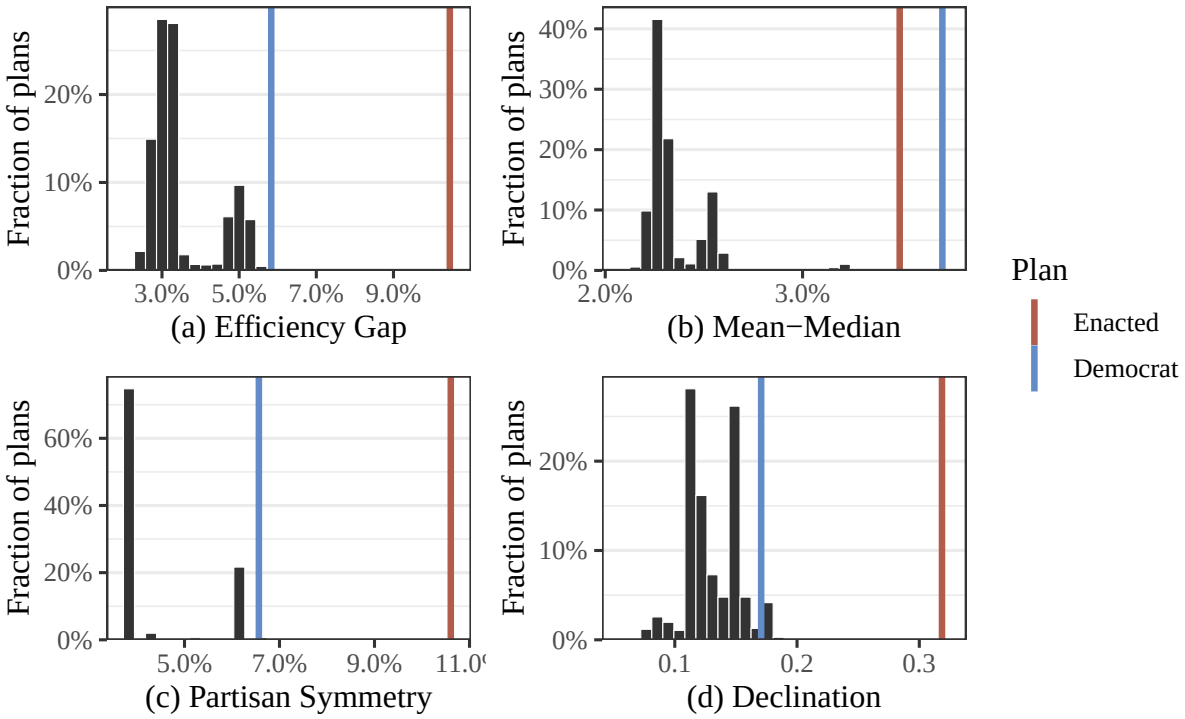

Figure 2: Four partisan bias measures calculated for the 5,000 simulated Senate redistricting plans computed by averaging across the 9 statewide elections from 2016 to 2020. Overlaid are the values for the enacted plan (red) and the Democratic caucus plan (blue). For each measure, larger values (towards the right) correspond to more Republican-favoring plans.

House, all of the 5,000 simulated plans have a lower (better) partisan bias score than the enacted plan across all four metrics considered here.

39. For the Senate, the Democratic caucus plan is also an outlier for all partisan bias metrics. But, it has better scores than the enacted plan with the exception of the mean-median metric.

A.2. Compliance with Section 6(B)

40. I next present the results regarding the plans' compliance with Section 6(B), using the Commission's approach. Section 6(B) states that "the statewide proportion of districts whose voters, based on statewide state and federal partisan general election results during the last ten years, favor each political party shall correspond closely to the statewide preferences of the voters of Ohio." Therefore, I use the proportionality metric to examine whether or not the statewide

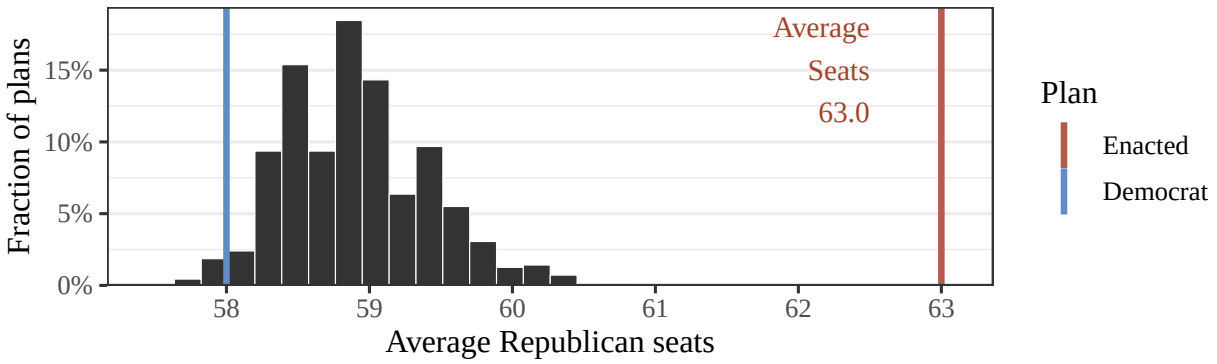

Figure 3: Average number of Repulican seats calculated for the 5,000 simulated House redistricting plans computed by averaging across the 9 statewide elections from 2016 to 2020. Overlaid are the values for the enacted plan (red) and the Democratic plan (blue).

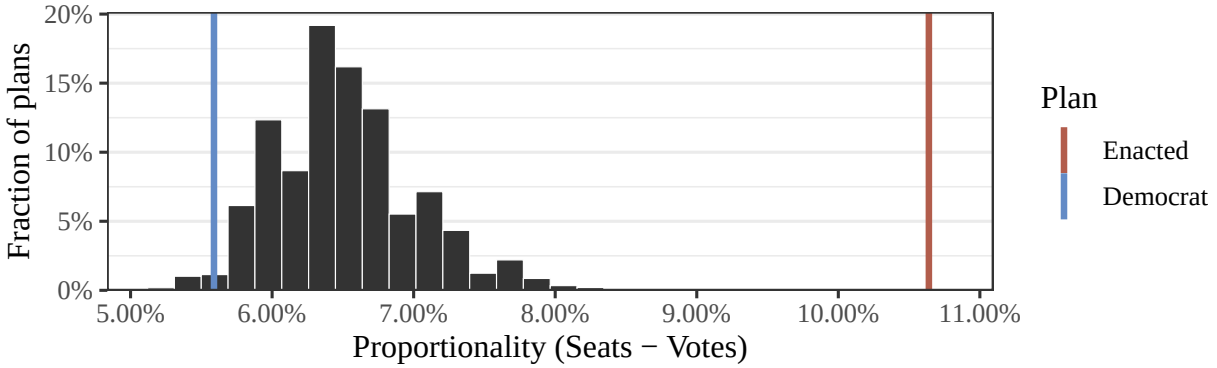

Figure 4: Corresponding proportionality measure calculated for the 5,000 simulated House redistricting plans computed by averaging across the 9 statewide elections from 2016 to 2020. Overlaid are the values for the enacted plan (red) and the Democratic plan (blue).

seat share of each party corresponds closely to its statewide vote share under each plan. As I show below, for both the House and Senate, the enacted plan is a clear outlier relative to the simulated plans. That is, although the simulated plans follow several of the map-drawing decisions established in the enacted plan, all of my 5,000 simulated plans are more compliant with Section 6(B) than the enacted plan.

41. For the House, Figure 3 shows that under the enacted plan, the Republican party is expected to win 63.0 seats, which is about 4 seats higher than the average simulated plan of 58.9 seats. None of my 5,000 simulated plans awards that many seats to Republicans. Under the Democratic caucus plan, the Republican party earns less seats than most of the simulated plans.

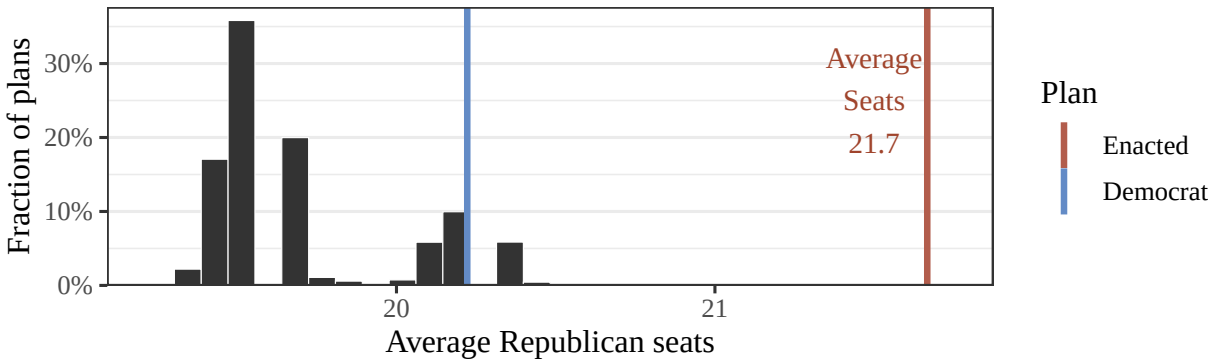

Figure 5: Average number of Repulican seats calculated for the 5,000 simulated Senate redistricting plans computed by averaging across the 9 statewide elections from 2016 to 2020. Overlaid are the values for the enacted plan (red) and the Democratic plan (blue).

42. This discrepancy is reflected in the proportionality metric, which is shown in Figure 4. A value of zero for this measure implies complete proportionality, while positive values indicate that Republicans win a larger share of seats than vote share, on average. A smaller value indicates a plan's better compliance with Section 6(B). The enacted plan has a proportionality score of 10.6%, implying that the Republican party would receive an average of 10.6% more seats under the enacted plan than under a proportional plan where the vote share is equal to the seat share. In contrast, under the simulated plans, the average proportionality score is only 6.5%. Indeed, all simulated plans score better than the enacted plan. It is worth noting that the Democratic caucus plan even outperforms most of the simulated plan.

43. For the Senate, the substantive conclusion is similar despite the fact that the simulated plans are based on the House districts of the enacted plan and follow several additional map-drawing decisions made by the Respondents. Figure 5 shows that the enacted plan favors the Republican party to a large degree and is a clear outlier. Under the enacted plan, the Republican party is expected to win 21.7 seats on average, which is much greater than any of my 5,000 simulated plans. On average, the simulated plans would award Republicans 19.7 seats, which is about 2 seats fewer than the enacted plan. The Democratic caucus plan awards fewer expected Republican seats than the enacted plan, but it tends to be more favorable to the Republican party than many of my simulated plans.

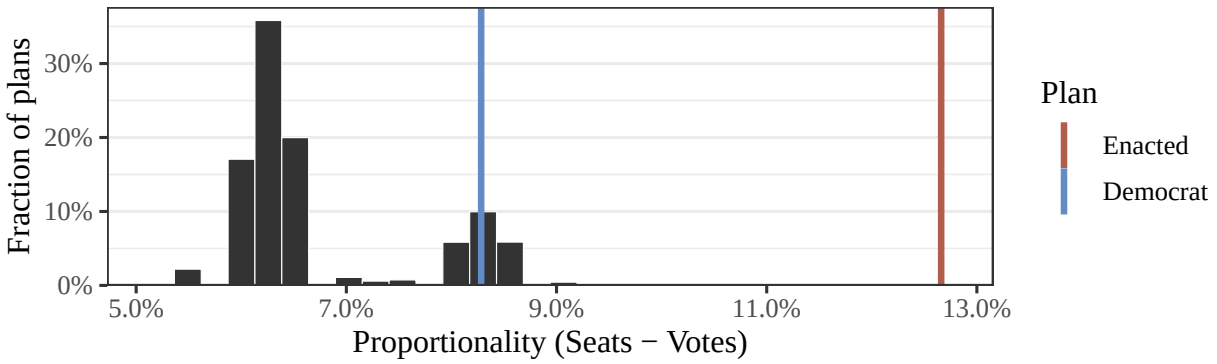

Figure 6: Corresponding proportionality measure calculated for the 5,000 simulated Senate redistricting plans computed by averaging across the 9 statewide elections from 2016 to 2020. Overlaid are the values for the enacted plan (red) and the Democratic plan (blue).

44. As for the proportionality criteria of Section 6(B), all of my 5,000 simulated Senate plans have smaller (better) proportionality scores than the enacted plan. The enacted plan has a deviation from proportionality that is nearly double the average simulated plan, giving Republicans 12.7% more seats on average above the proportional outcome. In contrast, the simulated plans would give Republicans only 6.7% more seats on average above the proportional outcome. The Democratic caucus plan performs better than the enacted plan but scores worse than most of my simulated plans.

B. Evaluation Using the 13 Statewide Election Results

45. To give the Commission the benefit of the doubt, I conducted an additional evaluation by supplementing these 9 elections with 4 additional statewide elections from 2012 and 2014 (see Appendix G.1 for the list of these 13 statewide elections). I show that the use of these additional statewide elections does not alter my substantive conclusions. My analysis demonstrates that regardless of which set of elections I use, for both the House and Senate, the enacted plan is a clear outlier relative to the simulated plans, according to all four partisan bias metrics. The enacted plan also has worse proportionality scores than any of the 5,000 simulated plans.

B.1. Compliance with Section 6(A)

46. For the House, the efficiency gap is 8.23% for the enacted map, whereas the average efficiency gap for the simulated plans is only 3.80%. This implies that the enacted plan wastes

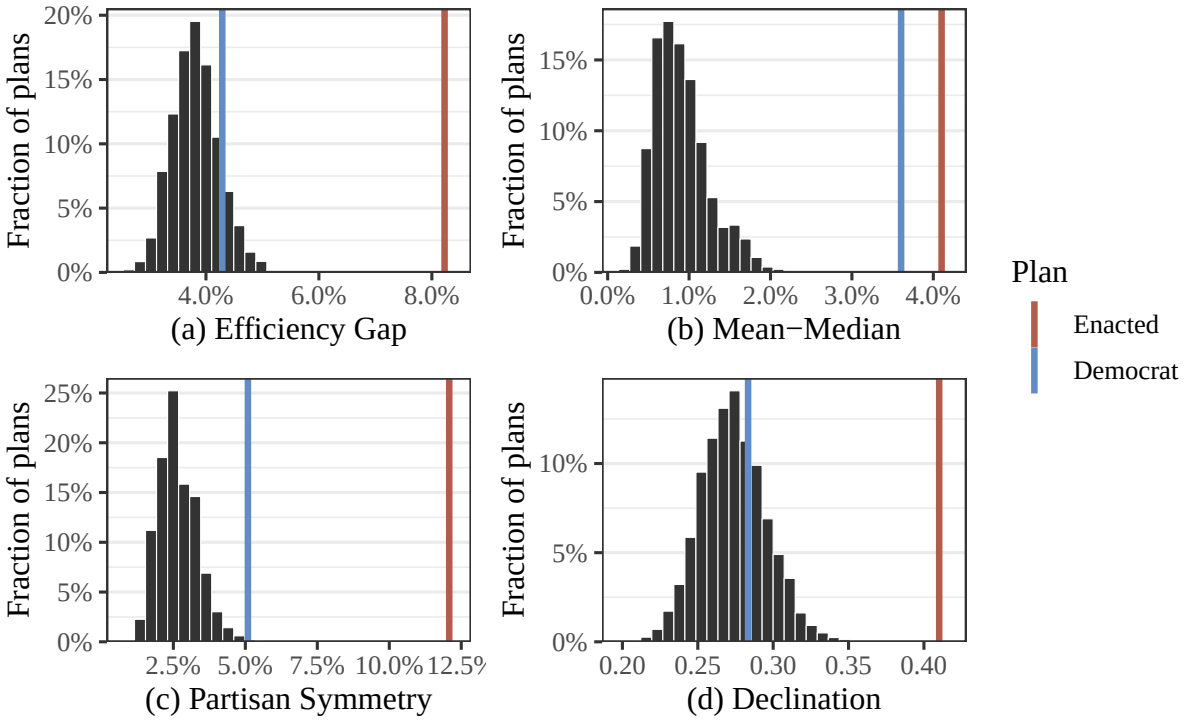

Figure 7: Four partisan bias measures calculated for the 5,000 simulated House of Representatives redistricting plans, using an alternative method of calculation. Overlaid are the values for the enacted plan (red) and the three comparison plans. For each measure, larger values (towards the right) correspond to more Republican-favoring plans.

over 100,000 more Democratic votes on average than the simulated plans, and over 100,000 fewer Republican votes. As shown in Figure 7(a), the enacted map is a clear outlier according to this metric. Figure 7(b) shows that in terms of the mean-median gap, the enacted plan is also extreme relative to the simulated plans.

47. In addition, Figure 7(c) shows that the enacted plan scores 12.1% on the partisan symmetry metric while the simulated plans score 2.6%, on average. This suggests that under the enacted plan, the Republican party would gain roughly 24 more seats than the Democrats, for a hypothetical tied election. Again, the enacted plan is a clear outlier according to this metric. Finally, as shown in Figure 7(d), the enacted plan also scores worse on the declination metric than any of the 5,000 simulated plans.

48. For the House, the Democratic caucus plan (blue line) has better scores than the

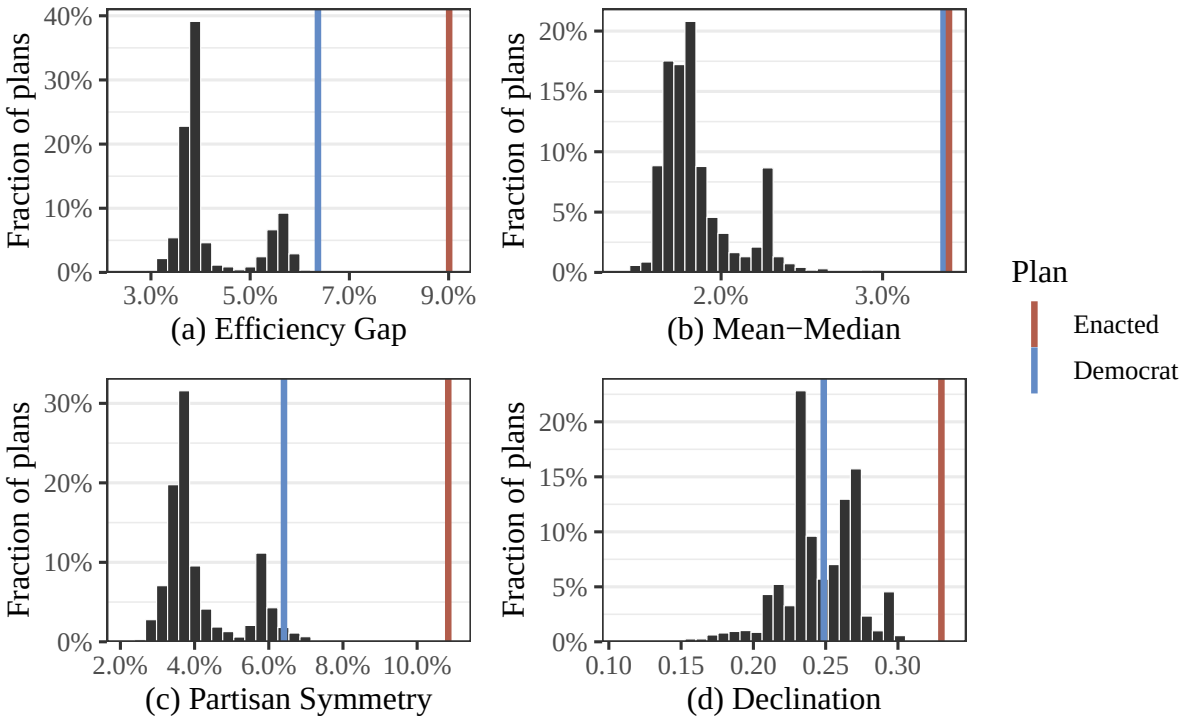

Figure 8: Four partisan bias measures calculated for the 5,000 simulated Senate redistricting plans computed by averaging across the 13 statewide elections, using an alternative method of calculation. Overlaid are the values for the enacted plan (red) and the Democratic caucus plan (blue). For each measure, larger values (towards the right) correspond to more Republican-favoring plans.

enacted plan for all four partisan bias metrics. Indeed, the Democratic caucus plan does as well for the efficiency gap and declination metrics as many of the simulated plans. Like the enacted plan, however, the Democratic caucus plan is an outlier for the mean-median and partisan symmetry metrics.

49. For the Senate, the results also remain essentially unaffected by the decision to use this more complete set of statewide election results. Although my simulated Senate plans are based on the House districts of the enacted plan, Figure 8 shows that the enacted plan is extreme relative to the simulated plans according to all four partisan bias metrics. For example, as shown in Figure 8(a), the efficiency gap of the enacted plan is 9.0% whereas the simulated plans score 3.9% on average for this metric. Like the House, all of the 5,000 simulated plans have a lower (better) partisan bias score than the enacted plan across all four metrics considered here.

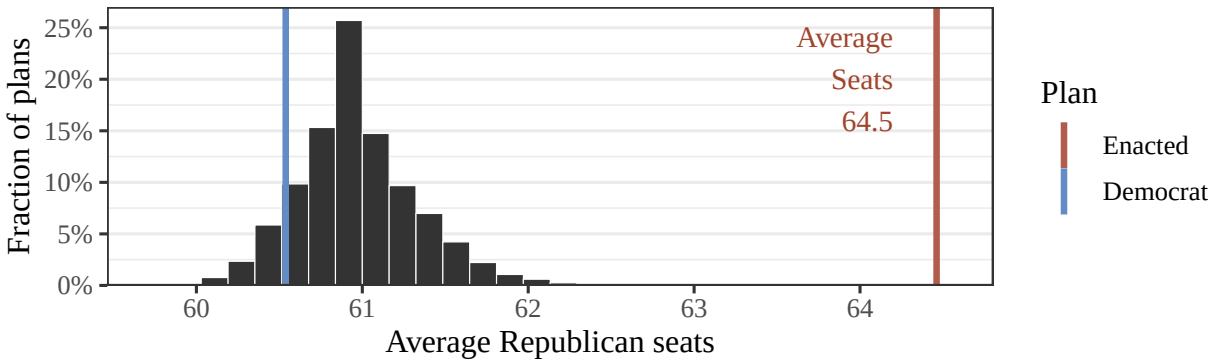

Figure 9: Average number of Repulican seats calculated for the 5,000 simulated House of Representatives redistricting plans, using an alternative method of calculation. Overlaid are the values for the enacted plan (red) and the three comparison plans.

50. For the Senate, the Democratic caucus plan is also an outlier for all the partisan metrics with the exception of declination. But, the Democratic caucus plan has better scores than the enacted plan though for the mean-median metric, both plans perform about the same.

B.2. Compliance with Section 6(B)

51. The results for the enacted plan's compliance with Section 6(B) also do not change when using this more complete set of statewide elections. For the House, across the simulated plans, Republicans are expected to earn 60.9 seats on average as shown in Figure 9. In comparison, under the enacted plan Republicans would earn an average of 64.5 seats, as indicated by the red vertical line. Thus, the enacted plan gives a roughly 4 seat advantage to Republicans on average when compared to the simulated plans. Indeed, none of the simulated plans came even close to awarding this many average seats to Republican candidates.

52. In terms of the proportionality criteria of Section 6(B), the enacted plan has an average proportionality score of about 0.11, implying that the Republican party would receive an average of 11% more seats under the enacted plan than under a proportional plan where the vote share is equal to the seat share. Again, all 5,000 simulated plans had smaller (better) proportionality scores. The enacted plan also achieves a worse proportionality score than the Democratic caucus plan, which, unlike the enacted plan, is not an outlier.

53. Under the Democratic caucus plan, the Republican party would be expected to win

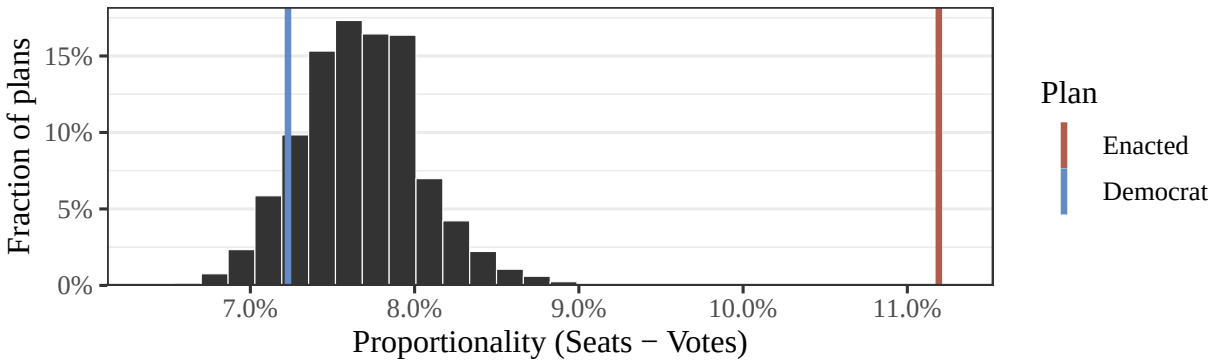

Figure 10: Corresponding proportionality measure calculated for the 5,000 simulated House of Representatives redistricting plans, using an alternative method of calculation. Overlaid are the values for the enacted plan (red) and the three comparison plans.

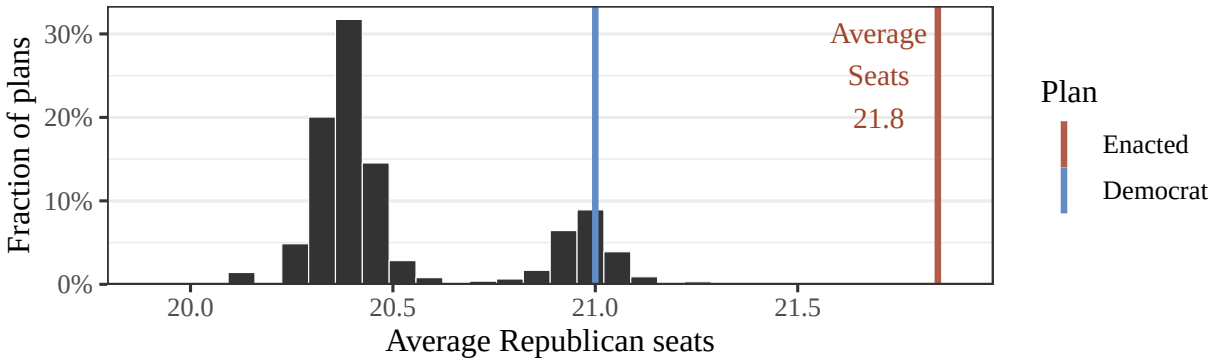

Figure 11: Average number of Repulican seats calculated for the 5,000 simulated Senate redistricting plans computed by averaging across the 13 statewide elections, using an alternative method of calculation. Overlaid are the values for the enacted plan (red) and the Democratic plan (blue).

about the same number of seats as many of the simulated plans. Accordingly, the Democratic caucus plan performs as well on the proportionality metric as many of the simulated plans.

54. For the Senate, the results also remain unaffected. Figure 11 shows that the enacted plan is the most favorable to the Republican party and is a clear outlier when compared to the simulated plans. Indeed, no simulated plan awards more seats to Republicans than the enacted plan. Republicans earn an average of 20.5 seats among the sampled plans, whereas the enacted map gives Republicans 21.8 seats on average.

55. As shown in Figure 12, the enacted plan has an average proportionality score of about 12.3%, which implies that the Republican party will receive about 12.3% more seats on

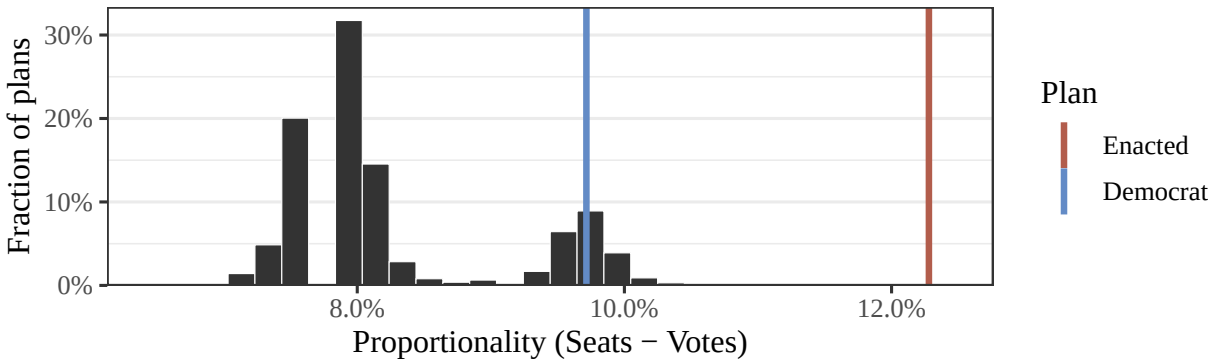

Figure 12: Corresponding proportionality measure calculated for the 5,000 simulated Senate redistricting plans computed by averaging across the 13 statewide elections, using an alternative method of calculation. Overlaid are the values for the enacted plan (red) and the Democratic plan (blue).

average under the enacted plan than under proportionality. As with the House simulations, all 5,000 simulated plans had better proportionality scores, with a mean proportionality score giving about 8.3% more seats on average to Republicans above the proportional outcome. The Democratic caucus plan has a better score than the enacted plan, though it has a worse score than most of the simulated plans.

VI. DETAILED LOCAL ANALYSIS OF COUNTY CLUSTERS

56. Partisan bias in the enacted plan is apparent not just in statewide summary statistics, as shown above, but also at the local level. To illustrate this, I performed a detailed analysis of the House and Senate districts in Hamilton, Franklin, and Cuyahoga-Summit-Geauga counties. My analysis of these counties shows that for both the House and Senate, the enacted plan packs a disproportionately large number of Democratic voters into some districts while turning other districts into Republican safe seats. The results shown in this section are based on the 13 statewide elections.

A. Hamilton County

A.1. House of Representatives

57. For the House districts, I began by calculating, for each precinct, the average two-party vote share of the district to which that precinct is assigned under the enacted plan. I also performed the same calculation under each simulated plan and then averaged these vote shares

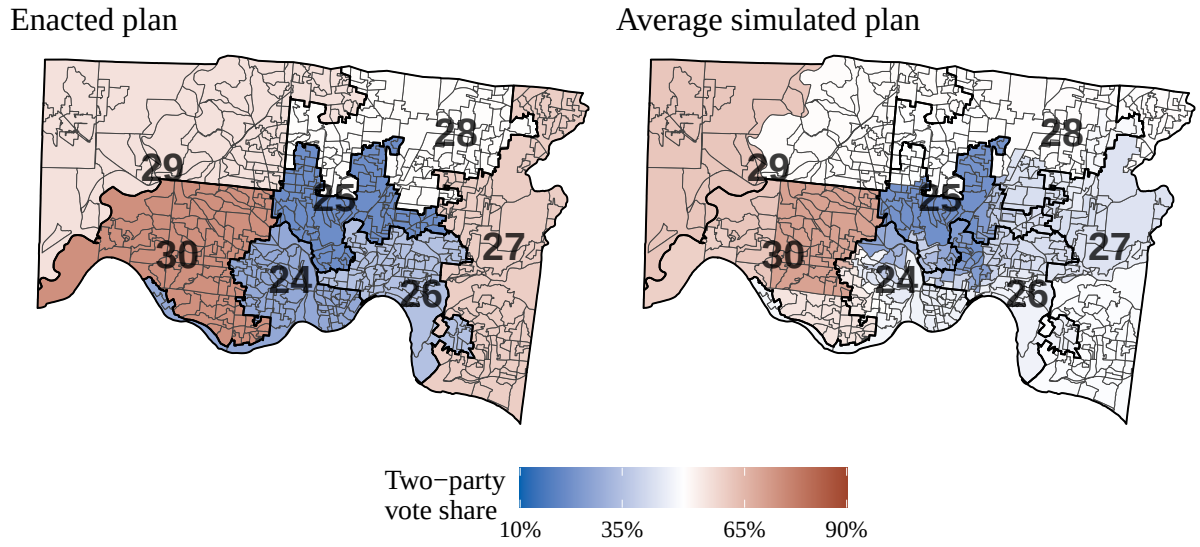

Figure 13: House districts in Hamilton county. The left and right maps show the average two-party vote share for each district under the enacted and average simulated plan, respectively. The enacted plan packs Democratic voters into districts 24, 25, and 26, turning districts 27, 29, and 30 into Republican safe seats. In contrast, under the average simulated plan, more voters live in competitive districts.

across all of the simulated plans. For example, precinct 061031AMM of Cincinnati lies within district 25 of the enacted map, which has an average Republican two-party vote share of 21.77%. However, the same precinct belongs to different districts in most of the simulated maps, each with their own Republican vote share. The average Republican vote share for the districts to which this precinct is assigned across all of the simulated plans is 38.92%, which is 17.16% higher than under the enacted plan. So, based on the representative set of simulated plans that have less partisan bias, precinct 061031AMM is packed into a more Democratic district under the enacted plan than would otherwise be expected.

58. Figure 13 shows the average vote share (averaged across the statewide contests) for each precinct under the enacted plan (left plot) and under the average simulated plan (right). Under the enacted plan, Democratic areas are packed into even-more Democratic districts, turning competitive and Republican-leaning areas into safe Republican seats. This is especially apparent along the southern border, with packed Democratic districts 24 and 26 allowing districts 27 and 30 to be shored up to safe Republican seats. In addition, more voters belong to competitive districts

under the average simulation plan than under the enacted plan. This is indicated by a much larger white area under the average simulated plan than under the enacted plan.

59. A closer look at each district reveals the packing and cracking of Democratic voters under the enacted map. For reference, I also include a map of two-party vote share for each precinct in Figure 24 of Appendix F. Consider enacted district 25 as an example. This district stretches into the Democratic-leaning area at its north west corner, making this district much more Democratic than the average simulated plan. In fact, most voters in this area would belong to competitive districts under the average simulation plan as indicated by its white color in the average simulated map. Similarly, the enacted plan packs district 24 with Democratic voters who, under the average simulated plan, would live in more competitive districts (again indicated by white color) under the average simulated plan. Yet another example is enacted district 29, which grabs a heavily Democratic area at its north east area. This cracking is possible without leading to a loss of Republican seat because the western side of this district is heavily Republican.

60. As a result, the enacted plan yields 3.3 Republican seats in Hamilton county, on average. Of the 5,000 simulated plans, more than 99.5% yield a lower average of Republican seats, with the average simulated plan leading to only 2.3 Republican seats. In other words, the enacted plan's packing of Democratic voters apparent in Figure 13 allows Republicans to gain an average of 1 seat in Hamilton County alone, out of 7 total.

A.2. Senate

61. My analysis reaches the same conclusion for the Senate. The enacted plan creates a total of 3 Senate districts out of 9 House districts in Hamilton and Warren counties. To be compliant with Sections 4(B)(1) and 4(B)(2), there are only 6 possible ways draw district boundaries from the House districts in the enacted plan (see Appendix B).

62. Figure 14 presents all of these plans along with the district-level average vote share under each plan. The enacted map (top left plot) packs a large number of Democratic voters into one district, which has 72.4% Democratic two-party vote share. At the same time, the enacted plan has two safe expected seats for Republicans with an average Democratic two-party vote share

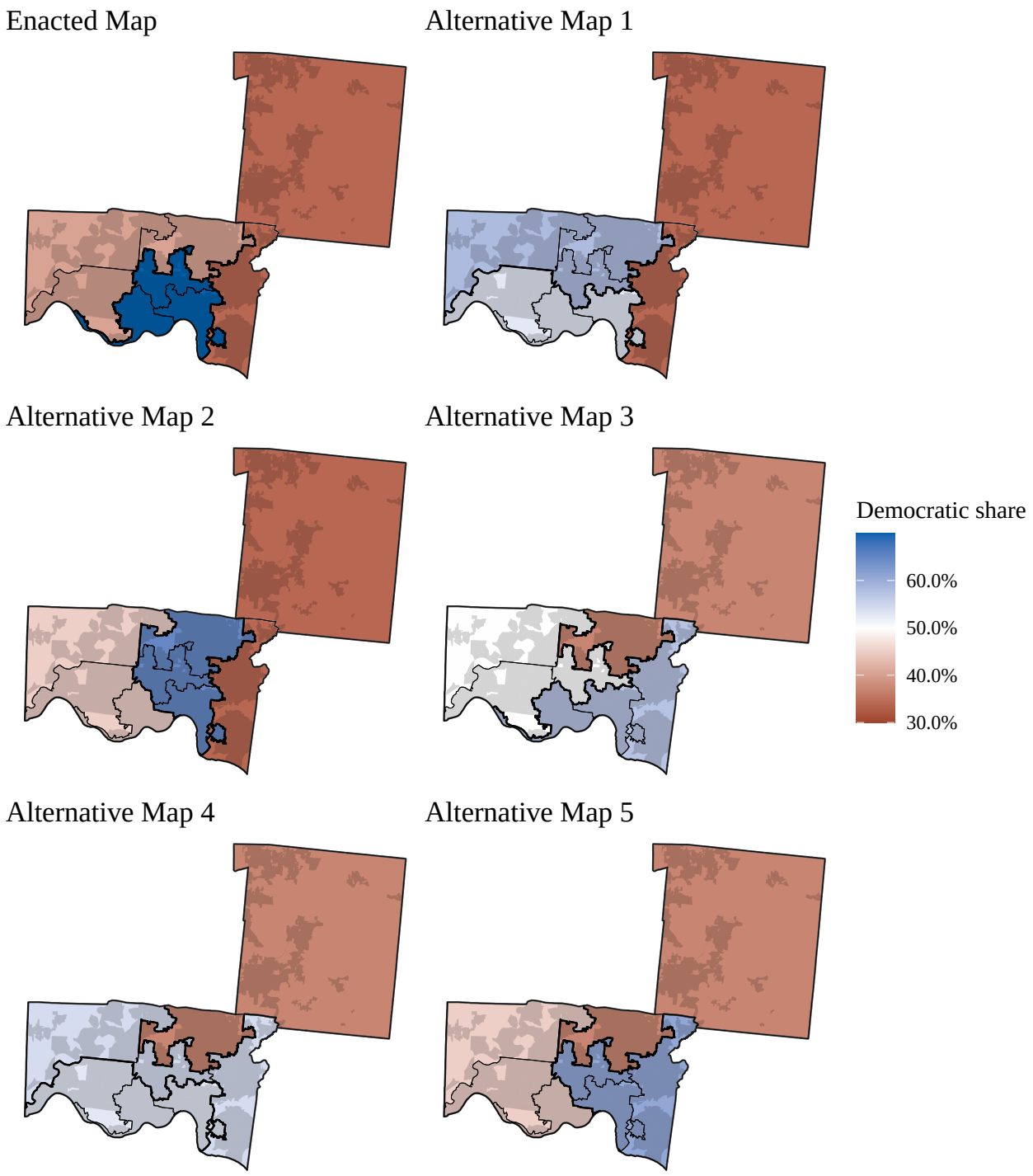

Figure 14: The 6 possible Senate districts in the Hamilton and Warren county cluster. The enacted plan is the top left plan. The enacted plan (top left) packs a disproportionately large number of Democratic voters into one district, creating two safe Republican districts. In contrast, the other plans create more competitive districts.

of 34.0% and 40.3%. In contrast, the other alternative plans do not have such a packed district. In particular, Alternative Map 3 (right middle plot) has one competitive district (Democratic vote-share of 49.9%) along with one Democratic (57.2%) and one Republican district (37.1%). This shows that the enacted plan unnecessarily packs Democratic voters into one district and is the most favorable to the Republican party among all possible plans in this area.

B. Franklin County

B.1. House of Representatives

63. Analogous to Figure 13, Figure 15 shows the average vote share (averaged across the statewide contests) for each precinct under the enacted plan (left plot) and under the average simulated plan (right plot) for Franklin county. Just like in Hamilton county, the enacted plan packs Democratic voters into a small number of districts (i.e., districts 1, 2, 3, and 7), allowing for the creation of two Republican seats in districts 10 and 12, and a third slightly Republican-leaning seat in district 4. For most of the areas of Franklin county which belong to Republican districts under the enacted plan, the average simulated plan would have placed them in more competitive or slightly Democratic-leaning districts.

64. This packing strategy can be seen clearly in the precinct-level vote shares as well, which are shown in Figure 25 of Appendix F. Districts 3 and 4 serve as illustrative examples. The boundary between the districts exactly follows the boudnary between the heavily-Democratic area around Columbus and the Republican-leaning area outside. A similar pattern is seen on the boundary of districts 4 and 9. The right plot of Figure 15 confirms that this boundary pattern is unusual, relative to the simulated plans: the average simulated district 4 is around five points more Democratic than the enacted district 4.

65. The net result of this packing is that the enacted plan yields 3.4 Republican seats in Franklin county, on average. Of the 5,000 simulated plans, all yield a lower average of Republican seats, with the average simulated plan leading to only 3.0 Republican seats. In other words, the enacted plan's packing of Democratic voters apparent in Figure 15 allows Republicans to gain an average of nearly half a seat in Franklin county, out of 12 total.

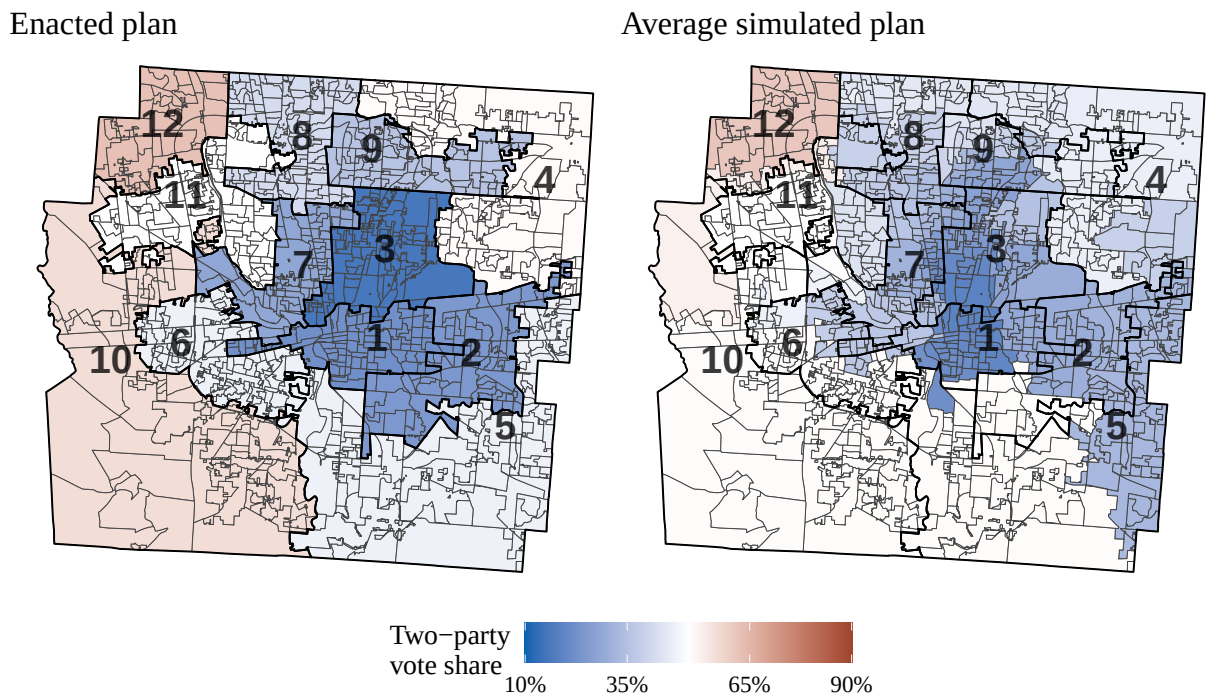

Figure 15: House districts in Franklin county. The left and right maps show the average two-party vote share for each district under the enacted and average simulated plan, respectively. The enacted plan packs Democratic voters into districts 1, 2, 3, and 7, turning districts 10 and 12 into Republican seats. In contrast, under the average simulated plan, more voters live in competitive districts.

B.2. Senate

66. For the Senate, as explained in Appendix B, my Senate analysis uses the House districts of the enacted plan. Since each Senate district consists of three House districts, the number of all possible Senate plans that satisfy Article XI Section 4(B) is relatively small. Thus, I used the algorithm of Fifield, Imai, et al. 2020 to enumerate all possible compliant plans. The algorithm found a total of 153 such compliant districting plans within this county cluster.

67. Panel (a) of Figure 16 presents each plan's two-party vote shares for the most Republican district (vertical axis) and the second most Republican district (horizontal axis). The plot clearly shows that the enacted plan, represented by the solid red square, chooses the combination of one safe Republican district and one competitive district. Panel (b) of the same figure shows that the enacted plan gives the best chance of electing two Republicans by packing the maximum

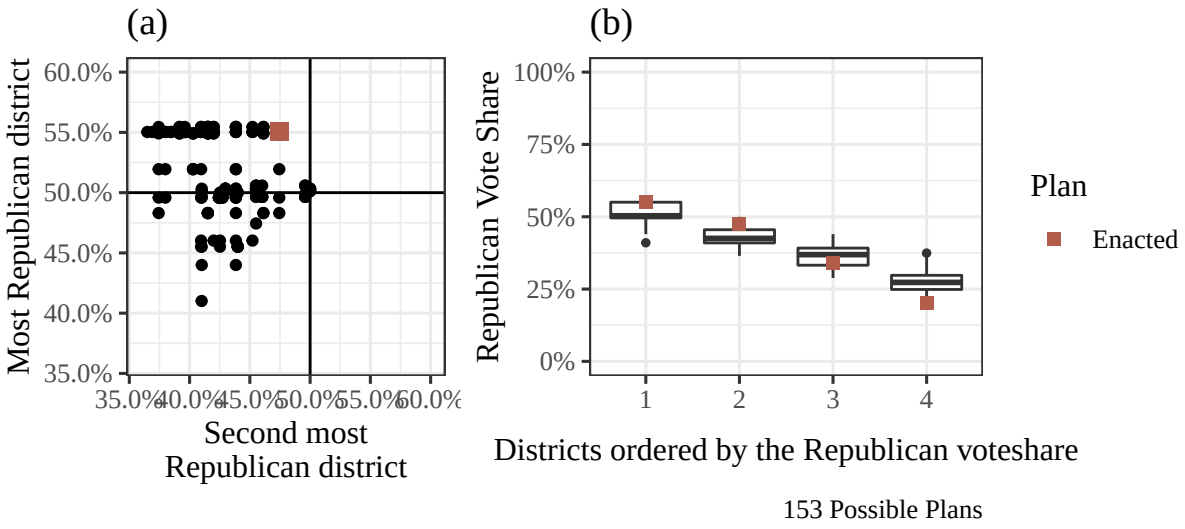

Figure 16: Comparison of simulated districts in Franklin and Union counties with the enacted districts. In panel (a), the vertical axis indicates the most Republican district and the horizontal axis indicates the next most Republican district. In panel (b), the districts are ordered horizontally by the Republican two-party vote share. The vertical axis indicates the Republican two-party vote share in that district.

number of Democratic voters into the most Democratic district. This shows that among all possible compliant plans in this county cluster, the enacted plan is the most favorable to the Republican party.

C. Cuyahoga, Summit, and Geauga Counties

C.1. House of Representatives

68. Figure 17 shows a similar pattern to Figures 13 and 15. The enacted plan creates additional Republican seats by concentrating Democrats and drawing district borders along partisan boundaries. In Cuyahoga, Summit, and Geauga counties, this is most apparent in districts 17 and 31, which under the simulated plans are generally more competitive or even Democratic-leaning, but which are Republican seats under the enacted plan.

69. This is achieved for enacted district 17 in part by having the boundary between districts 17 and 22 follow a partisan divide at a town boundary, as is visible at the precinct level in Figure 26 of Appendix F. In district 31, the enacted plan follows the western border of Akron exactly, and separates Akron proper from the towns of Norton and Barberton to its southwest.

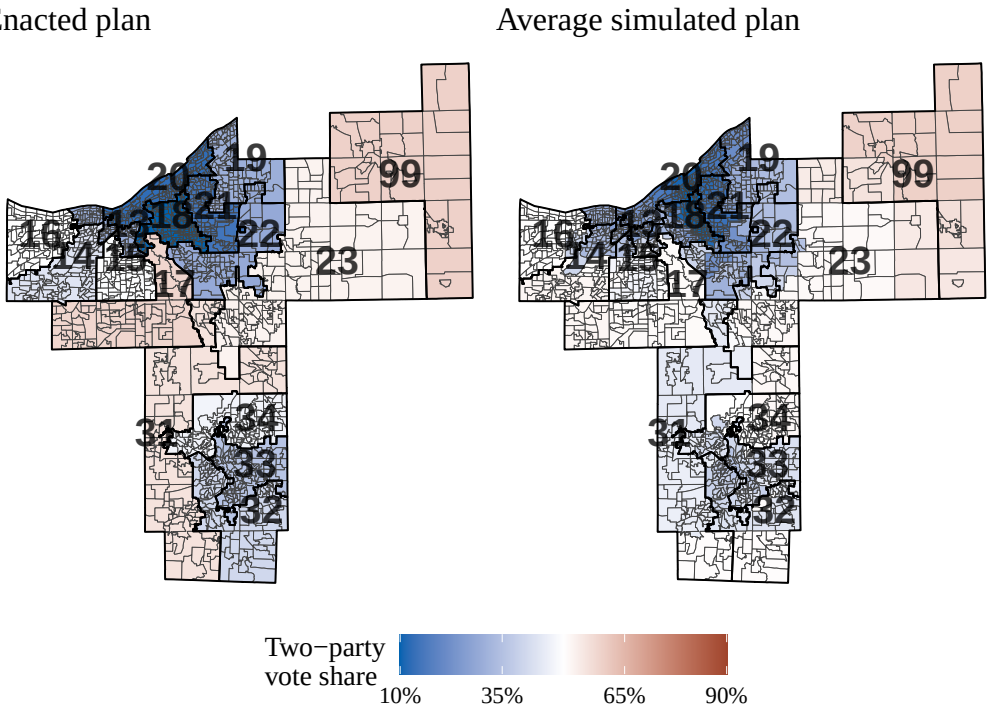

Figure 17: House districts in Cuyahoga, Summit, and Geauga counties. The left and right maps show the average two-party vote share for each district under the enacted and average simulated plan, respectively. The enacted plan packs Democratic voters in Cleveland districts, shoring up Republican vote shares in districts 17 and 31.

With the simulated plans, Norton and Barberton are more likely to be included with at least part of Akron, and consequently district 31 leans slightly Democratic.

70. In total, the enacted plan yields 6.3 Republican seats in these three counties, on average. Of the 5,000 simulated plans, all yield a lower average of Republican seats, with the average simulated plan leading to 5.4 Republican seats.

C.2. Senate

71. Like the Franklin county cluster, I used the enumeration algorithm to identify all possible compliant Senate plans within the Cuyahoga-Summit-Geauga county cluster. There are a total of 27 such plans in this case. Panel (a) of Figure 18 presents each plan's vote share for the most Republican district (vertical axis) and the second most Republican district (horizontal axis). The panel shows that the enacted plan chooses the districts, which are most favorable to the Republican party. Specifically, it chooses one safe district and one competitive district. Panel

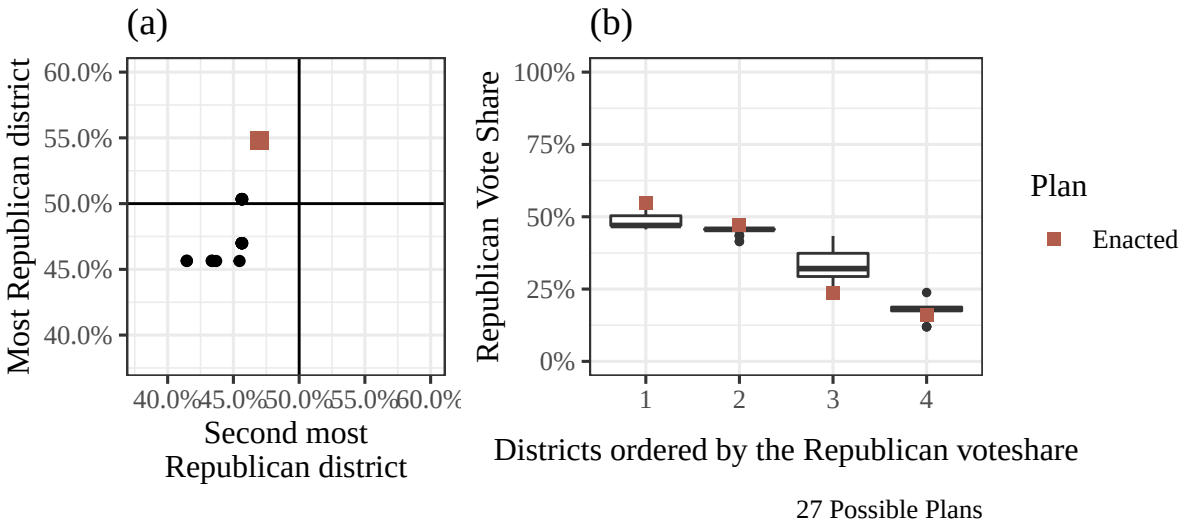

Figure 18: Comparison of simulated districts in Cuyahoga, Summit, and Geauga counties with the enacted districts. In panel (a), the vertical axis indicates the most Republican district and the horizontal axis indicates the next most Republican district. In panel (b), the districts are ordered horizontally by the Republican two-party vote share. The vertical axis indicates the Republican two-party vote share in that district.

(b) of the figure presents the Republican vote share across the districts that are ordered by the magnitude of their Republican vote shares. The enacted plan packs Democratic voters into the most Democratic districts, making the other two districts most Republican leaning possible. Again, among all compliant plans in this county cluster, the enacted plan is the most favorable to the Republican party.

VII. APPENDIX

A. Introduction to Redistricting Simulation

1. In recent years, redistricting simulation algorithms have played an increasingly important role in court cases involving redistricting plans. Simulation evidence has been presented to courts in Ohio and elsewhere, including Michigan, North Carolina, and Pennsylvania.[2]

2. Over the past several years, researchers have made major scientific advances to improve the theoretical properties and empirical performance of redistricting simulation algorithms. All of the state-of-the-art redistricting simulation algorithms belong to the family of Monte Carlo methods. They are based on random generation of spanning trees, which are mathematical objects in graph theory (DeFord, Duchin, and Solomon 2021). The use of these random spanning trees allows these state-of-the-art algorithms to efficiently sample a representative set of plans (Autry et al. 2020; Carter et al. 2019; McCartan and Imai 2020; Kenny et al. 2021). Algorithms developed earlier, which do not use random spanning trees and instead rely on incremental changes to district boundaries, are often not able to do so.

3. These algorithms are designed to sample plans from a specific probability distribution, which means that every legal redistricting plan has certain odds of being generated. The algorithms put as few restrictions as possible on these odds, except to ensure that, on average, the generated plans meet certain criteria. For example, the probabilities are set so that the generated plans reach a certain level of geographic compactness, on average. Other criteria, based on the state in question, may be fed into the algorithm by the researcher. In other words, this target distribution is based on the weakest assumption about the data under the specified constraints.

4. In addition, the algorithms ensure that all of the sampled plans (a) are geographically contiguous, and (b) have a population which deviates by no more than a specified amount

2. Declaration of Dr. Jonathan C. Mattingly, Common Cause v. Lewis (2019); Testimony of Dr. Jowei Chen, Common Cause v. Lewis (2019); Testimony of Dr. Pegden, Common Cause v. Lewis (2019); Expert Report of Jonathan Mattingly on the North Carolina State Legislature, Rucho v. Common Cause (2019); Expert Report of Jowei Chen, Rucho v. Common Cause (2019); Amicus Brief of Mathematicians, Law Professors, and Students in Support of Appellees and Affirmance, Rucho v. Common Cause (2019); Brief of Amici Curaiae Professors Wesley Pegden, Jonathan Rodden, and Samuel S.-H. Wang in Support of Appellees, Rucho v. Common Cause (2019); Intervenor's Memo, Ohio A. Philip Randolph Inst. et al. v. Larry Householder (2019); Expert Report of Jowei Chen, League of Women Voters of Michigan v. Benson (2019).

from a target population. These two guarantees are precisely those required by Article XI, § 03(B)(3) and § 03(B)(1), respectively.

5. There are two types of general Monte Carlo algorithms which generate redistricting plans with these guarantees and other properties: sequential Monte Carlo (SMC; Doucet, Freitas, and Gordon 2001) and Markov chain Monte Carlo (MCMC; Gilks, Richardson, and Spiegelhalter 1996) algorithms.

6. The SMC algorithm (McCartan and Imai 2020; Kenny et al. 2021) samples many redistricting plans in parallel, starting from a blank map. First, the algorithm draws a random spanning tree and removes an edge from it, creating a "split" in the map, which forms a new district. This process is repeated until the algorithm generates enough plans with just one district drawn. The algorithm calculates a weight for each plan in a specific way so that the algorithm yields a representative sample from the target probability distribution. Next, the algorithm selects one of the drawn plans at random. Plans with greater weights are more likely to be selected. The algorithm then draws another district using the same splitting procedure and calculates a new weight for each updated plan that comports with the target probability distribution. The whole process of random selection and drawing is repeated again and again, each time drawing one additional district on each plan. Once all districts are drawn, the algorithm yields a sample of maps representative of the target probability distribution.

7. The MCMC algorithms (Autry et al. 2020; Carter et al. 2019) also form districts by drawing a random spanning tree and splitting it. Unlike the SMC algorithm, however, these algorithms do not draw redistricting plans from scratch. Instead, the MCMC algorithms start with an existing plan and modify it, merging a random pair of districts and then splitting them a new way.

8. Diagnostic measures exist for both these algorithms which allow users to make sure the algorithms are functioning correctly and accurately. The original papers for these algorithms referenced above provide more detail on the algorithm specifics, empirical validation of their performance, and the appropriateness of the chosen target distribution.

B. Incorporating Article XI Sections 3 and 4 into the Algorithm

9. For the House of Representative plans, I follow the exact decisions made by Respondents under the enacted plan in creating clusters of counties, each of which contains a certain number of whole House districts. I simulate redistricting plans independently within each of these county clusters and combine them across the clusters to generate statewide plans.

10. For the Senate, my analysis is dependent on the House district boundaries in the enacted plan (Recall that a Senate district consists of exactly three House districts). I again follow the exact decisions made by Respondents in creating clusters of counties, each of which contains a certain number of whole Senate districts. Like the House of Representatives, I conduct a simulation analysis independently within each county cluster and then combine the results to generate statewide plans.

11. This process ensures that my simulated House and Senate plans are at least as compliant with Sections 3 and 4 as the enacted plan, which I am assuming is compliant with these provisions. I now explain this process in detail separately for the House and the Senate.

B.1. The House of Representatives

12. In drawing a redistricting plan for the House of Representatives, a multitude of constraints must be satisfied. We begin by classifying a total of 88 counties in Ohio into three categories based on their population according to Article XI Section 3(C) of the constitution: 3(C)(1), 3(C)(2), and 3(C)(3) counties, which are colored using green, blue, and yellow, respectively, in Figure 19.

13. There are a total of twenty-two 3(C)(1) counties. According to § 3(C)(1), each of these large counties should be "divided into as many house of representative districts as it has as it has whole ratios of representation." In addition, the article stipulates that "Any fraction of the population in excess of a whole ratio shall be a part of only one adjoining house of representatives district." There are many possible ways to choose the adjoining district when spilling over an excess fraction of the population from each of 3(C)(1) county into neighboring counties. The enacted map makes certain choices about how to allocate excess population from 3(C)(1) counties

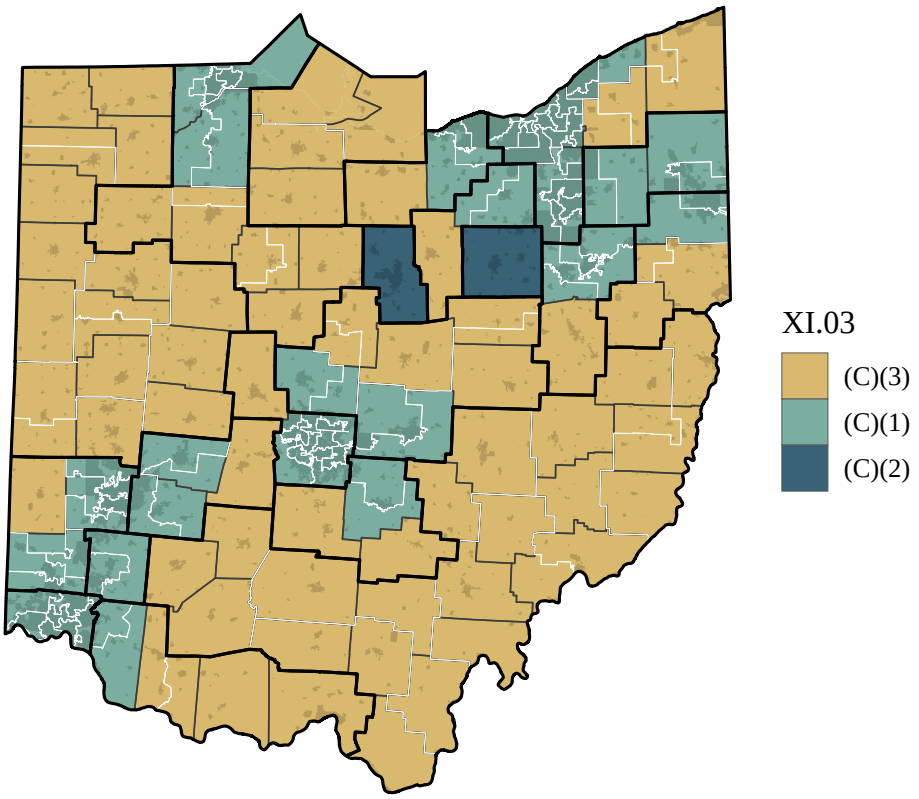

Figure 19: Ohio counties, colored by the subsection of Article XI.03 which they are subject to. Gray lines are county borders, and white lines are the district borders of the plan enacted by Respondents. Thick black lines demarcate independent county clusters used in simulation.

into neighboring counties. We follow these decisions of the enacted plan by starting with each 3(C)(1) county and selecting the minimal set of adjacent counties that contain whole districts in the enacted plan. These minimal sets of adjacent counties that contain whole districts sometimes include counties smaller than the ratio of representation, and we ensure that each of these counties is not split more than once, as required by § 3(C)(3). This results in 18 non-overlapping clusters of counties, as shown in Table 1. These clusters are demarcated in Figure 19 using the solid black boundary lines.

14. These clusters are determined by starting with each 3(C)(1) county and selecting the minimal set of adjacent counties so that no district in the enacted plan crossed their borders. For example, according to the enacted plan, all seven districts in Hamilton county lie entirely within the county, so Hamilton county is its own cluster. In contrast, in the enacted plan, one of the districts in Lorain county spills into Huron county (but goes no further), and so Lorain and Huron

Table 1: The clusters of counties that contain whole districts according to the enacted plan.

Counties	Districts
Franklin and Union	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12
Cuyahoga, Summit, Lake, Geauga, and Ashtabula	13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 31, 32, 33, 34, 56, 57, and 99
Hamilton	24, 25, 26, 27, 28, 29, and 30
Butler, Montgomery, and Preble	35, 36, 37, 38, 39, 44, 45, and 46
Lucas, Wood, Hancock, Putnam, Wyandot, Crawford, and Marion	40, 41, 42, 43, 76, 83, and 87
Stark and Tuscarawas	47, 48, 49, and 50
Portage and Trumbull	64, 65, and 72
Lorain and Huron	51, 52, and 53
Warren	54 and 55
Mahoning, Columbiana, and Carroll	58, 59, and 79
Licking, Delaware, Morrow, Knox, Holmes, and Coshocton	60, 61, 68, 69, and 98
Clermont, Brown, Adams, and Scioto	62, 63, and 90
Fairfield, Pickaway, and Hocking	73 and 74
Medina and Ashland	66 and 67
Clark, Greene, and Madison	70, 71, and 75
Williams, Fulton, Defiance, Henry, Paulding, Van Wert, Mercer, Allen, Auglaize, Hardin, Logan, Champaign, Shelby, Darke, and Miami	80, 81, 82, 84, 85, and 86
Ottawa, Erie, Sandusky, and Seneca	88 and 89
Clinton, Fayette, Highland, Ross, Pike, Vinton, Jackson, Lawrence, Gallia, Meigs, Athens, Perry, Morgan, Washington, Monroe, Noble, Belmont, Jefferson, Harrison, Guernsey, and Muskingum	91, 92, 93, 94, 95, 96, and 97

form a cluster.

15. In addition, there are two 3(C)(2) counties—Richland and Wayne—whose population falls between 95% and 105% of the target population. The enacted plan complies with § 3(C)(2) and assigns one district to each of these two counties. My analysis treats these two counties in the same way, and therefore no simulation is required.

16. Lastly, under the enacted plan, the remainder of the state (i.e., the entire state minus two 3(C)(2) counties and 19 clusters) is divided into three contiguous sets of counties, which consist of a subset of 3(C)(3) counties (see Figure 19). The list of counties that belong to each of these remaining clusters is given in the final three rows of Table 1. Per § 3(C)(3), these counties should not be split more than once. Occasionally, the algorithm will by chance split one of these counties more than once. I discard these simulations, leaving only those which are fully compliant with § 3(C)(3).

17. The enacted plan has no violation of § 3(C)(1). To ensure perfect compliance with this provision, I instruct the algorithm to follow the enacted plan and avoid creating districts that cross certain county boundaries. These boundaries are borders between Delaware and Licking, Delaware and Knox, Licking and Knox, Butler and Montgomery, Greene and Clark, Geauga and Cuyahoga, Lake and Cuyahoga, Summit and Cuyahoga, and Geauga and Lake counties. Preserving these boundaries is needed to guarantee that my simulated plans do not violate § 3(C)(1), and make the same choice as the enacted plan in terms of county splits.

18. Another important set of choices is which municipalities or townships to split, pursuant to § 3(D)(2) and § 3(D)(3). I ensured that the simulated plans complied with § 3(D)(2) and § 3(D)(3) as much as or more than the enacted plan by instructing the algorithm to avoid splitting any municipalities or townships smaller than the ratio of representation, except for those split by Respondents in the enacted plan. There are at least eleven instances in which the enacted plan splits municipalities or townships. They are the cities of Cleveland, Columbus, Cincinnati, Toledo, Akron, Dayton, Solon, and New Albany (the largest contiguous portion lying within Franklin county), and the townships of Jackson (in Franklin County), Copley, and Nimishillen. The algo-

Table 2: The clusters of counties that are consistent with the enacted plan. These clusters avoid violations of XI.04.

Districts	Counties
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	Franklin, Union
35, 36, 37, 38, 39, 80	Montgomery, Butler*, Preble, Miami*, Darke*
24, 25, 26, 27, 28, 29, 30, 54, 55	Hamilton, Warren
13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 31	Cuyahoga, Summit*, Geauga*
32, 33, 34, 40, 41, 42, 44, 45, 46, 47, 48, 49, 51, 52, 53, 56, 57, 99, 64, 65, 72, 70, 71, 75	Summit*, Lucas*, Butler*, Lorain, Huron, Lake, Ashtabula*, Trumbull, Portage, Clark, Greene, Madison
43, 50, 58, 59, 60, 61, 62, 63, 66, 67, 68, 69, 73, 74, 76, 77, 78, 79, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98	All remaining counties and partial counties

rithm is allowed to split these municipalities or townships along the specific district lines adopted in the enacted plan. None of these municipalities or townships are between 50% and 100% of ratio of representation and therefore do not violate § 3(D)(2).

B.2. The Senate

19. Like my analysis of the enacted plan for the House of Representatives described above, I follow many of the decisions made by Respondents in creating the enacted plan for the Senate. I begin my analysis of the enacted Senate plan by using the enacted House plan (recall that each Senate district should consist of exactly three House districts).

20. Given the enacted House plan, I consider the restrictions the Ohio constitution imposes on the construction of Senate districts. Specifically, § 4(B)(1) states that a large county, which contains at least one whole Senate ratio of representation, should contain as many whole Senate districts as possible, and any excess fraction should be part of only one adjoining Senate district. In addition, § 4(B)(2) demands that a small county, which contains less than one Senate ratio of representation but more than one House ratio of representation, should not be split into multiple Senate districts.

21. As done for my House analysis, I follow the exact decisions made by Respondents in creating the cluster of counties that contain a certain number of whole Senate districts without spilling into an adjacent county. Table 2 presents the list of such county clusters used in the enacted plan along with their Senate districts. These clusters are colored in Figure 20. We conduct separate simulation analyses within each of the following county clusters—Franklin (red), Cuyahoga-Summit-Geauga (CSG; yellow), Hamilton (purple), Montgomery-Butler-Preble-Miami-Darke (MBPMD; orange). In the figure, the "Determined" county clusters (dark blue) refer to the House districts which can only be in one Senate district to be compliant. No simulation is necessary for any of these "Determined" clusters because we follow the enacted Senate district that was adopted. Finally, the "Remainder" county cluster (white) represents the rest of counties that need not be grouped to be compliant with the Section 4 constraints. Like other county clusters, we conduct separate simulations within this cluster.

C. Implementation details

22. In my analysis, I use the SMC algorithm for several reasons. First, unlike the MCMC algorithms, the SMC algorithm generates nearly independent samples, leading to a diverse set of redistricting plans that satisfy the specified constraints. Second, the SMC algorithm avoids splitting political subdivision boundaries where possible, an important consideration in the case of Ohio. Third, the SMC algorithm continues to perform accurately in large states with many districts, a critical feature for the Ohio House of Representatives districts.

23. The mathematical function I used to discourage packed districts mirrors the way other constraints are imposed on simulation algorithms (e.g., Herschlag et al. 2020a) and is given by $C(|x_d - 0.5||x_r - 0.5|)^p$ where x_d and x_r represent the two-party vote share for Democrats and Republican (averaged across the statewide elections used in my analysis), and C is a parameter controlling the strength of the constraint. This mathematical function is completely symmetric between the two parties—switching the party labels produces the exact same value. The values of $p = 0.15$ (House) and $p = 1.5$ (Senate) were selected for the exponent based on my experience implementing similar constraints for the Voting Rights Act compliance, and by simulation experi-

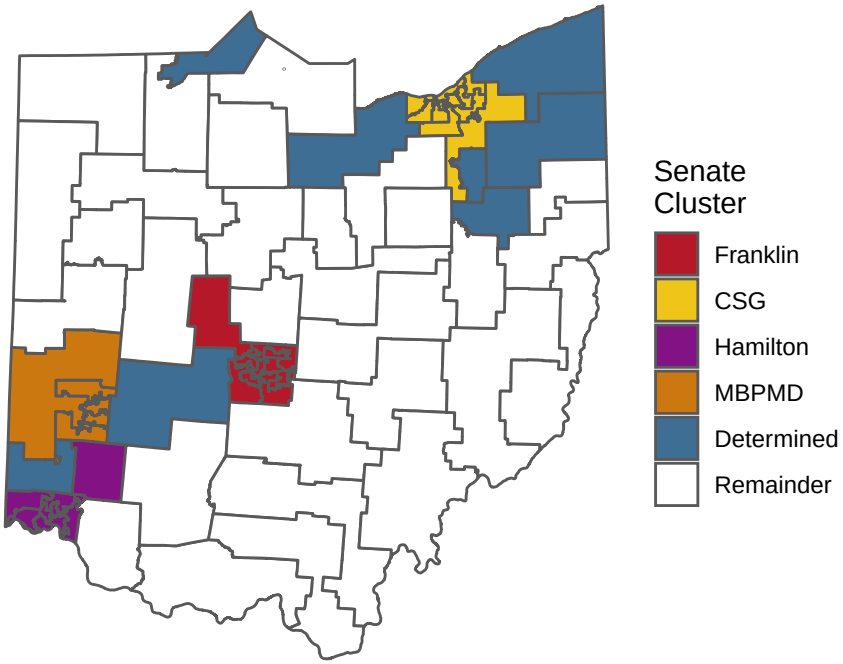

Figure 20: County clusters for the Senate implied by the decisions made to create the enacted House plan ensuring that no violations of Article XI Section 4(B)(1) or 4(B)(2). 'Determined' refers to the clusters, which there is only one compliant districting, whereas 'Remainder' refers to the rest of counties that need not be grouped to comply with the Section 4 constraints.

ments on this data. As a result, it is impossible for this constraint to favor one party over another. Note that for the Senate, removing this additional constraint yields substantively similar results.

24. I allowed the value of C to vary between 5 and 100 for each cluster simulation. Variance across clusters is necessary because each cluster has a different number and configuration of districts, and these affect how well the constraint function binds. Within the 5 to 100 range, I chose the maximum value which still maintained the accuracy of the algorithm, according to several diagnostic measures. Specifically, I increased the value of C in increments of 5, until either the resampling efficiency at any stage of the iteration fell below 1%, or the diversity of the sample, as measured by the pairwise variation of information distance between 100 randomly selected plans, was below 0.35–0.40. More detail about these diagnostic measures may be found in the

original SMC algorithm paper (McCartan and Imai 2020).

C.1. The House of Representatives

25. For the House plans, I run the algorithm independently within each county cluster and then combine the results to obtain a statewide plan. Thus, my analysis will examine how each cluster can be divided into the fixed number of districts in different ways, and how this drawing process affects each plan's compliance with Sections 6(A) and 6(B).

26. In Hamilton county, I ensured that there be one district whose majority of voting age population identify themselves in any part as Black. I made this decision based on the affidavit of Dr. Lisa Handley, which I reviewed. To accomplish this, I used a Voting Rights Act constraint and tuned it so that at least 75% of simulated plans in Hamilton county had one such majority-minority district (MMD). This constraint may be written mathematically as $\sqrt{\max(x_b - 0.51, 0)}$, where x_b is the share of a district's VAP that is Black. This is a common way to formulate the VRA constraint (Herschlag et al. 2020b).

27. Because this county uses both partisan bias and VRA constraints, which interact with one another, I employed a different rule in selecting the value of C for Hamilton county. I first adjusted the strength of the VRA constraint until at least 75% of simulated plans had one or more MMDs. Then, I increased the value of C in increments of 5 until the diversity of the sample reached 0.2. After generating redistricting plans in Hamilton county, I discarded the simulated plans that do not have at least one such MMD so that my simulated plans are perfectly compliant with this requirement.

C.2. The Senate

28. Simulating the Senate plans proceeds similarly, using the House districts of the enacted plan rather than precincts as geographical units. Simulating redistricting plans independently within each of these county clusters ensures that the combined statewide plans are in compliance with § 4(B)(1) and § 4(B)(2). After conducting a simulation analysis within each county cluster, I then combine the simulated plans from each cluster to create statewide plans. As with the House district simulation approach, I sample districts using 5% population bounds in accordance with

§ 3(B)(1). This guarantees that all 3 district plans are achievable in terms of the total statewide population. I also apply our party-neutral constraint, increasing its strength incrementally until the stopping criteria is met, as done in the House simulation. Per instruction of counsel for the Relators, I do not impose a VRA constraint.

D. An Example Simulated Plan

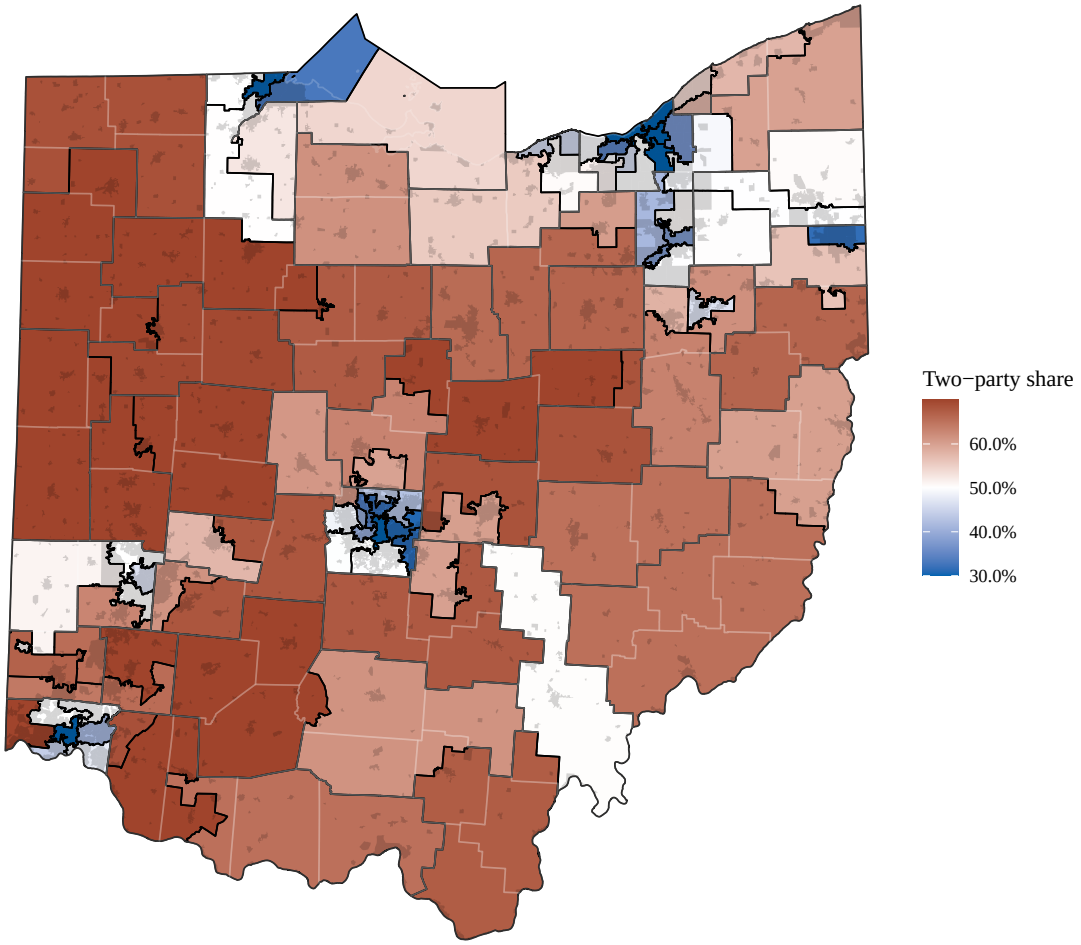

Figure 21: An example simulated redistricting plan for the House, with districts colored by their average two-party vote share.

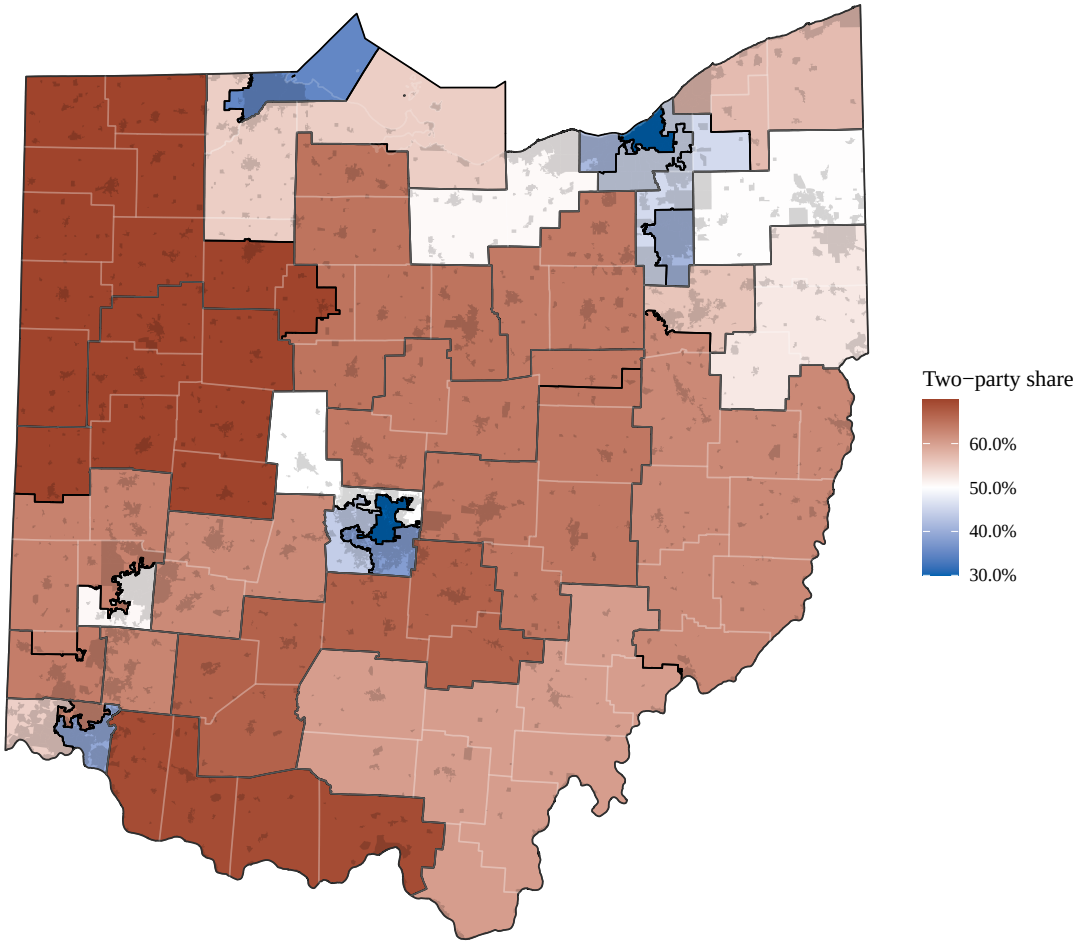

Figure 22: An example simulated redistricting plan for the Senate, with districts colored by their average two-party vote share.

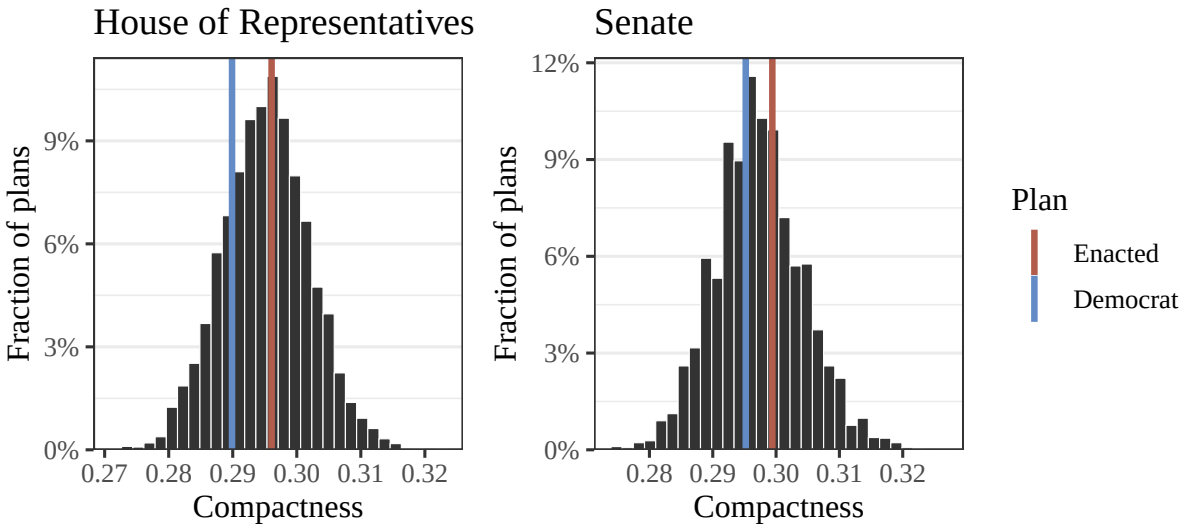

Figure 23: Polsby–Popper compactness scores for the simulated redistricting plans. Overlaid are scores for the enacted (red) and the Democratic caucus plan (blue). Larger values indicate more compact districts.

E. Compliance with Section 6(C)

29. The results in Section V show that the simulated plans and the Democratic caucus plan are much more compliant with Sections 6(A) and 6(B) than the enacted plan. I now show that this superior compliance is achieved without sacrificing compliance with Section 6(C), which requires districts to be compact. I use the Polsby–Popper score, a commonly-used quantitative measure of district compactness (Polsby and Popper 1991).

30. Figure 23 shows that the enacted plan and the Democratic caucus plan are both as compact as the simulated plans, on average. The result clearly implies that it is possible to be more compliant with Sections 6(A) and 6(B) without sacrificing the compliance with Section 6(C).

F. Vote Share for Precincts

31. Figure 24 presents the two-party vote share for precincts of Hamilton county. Figure 25 presents the two-party vote share for precincts of Franklin county. Figure 26 presents the two-party vote share for precincts of Cuyahoga, Summit, and Geauga Counties.

G. References and Materials Considered

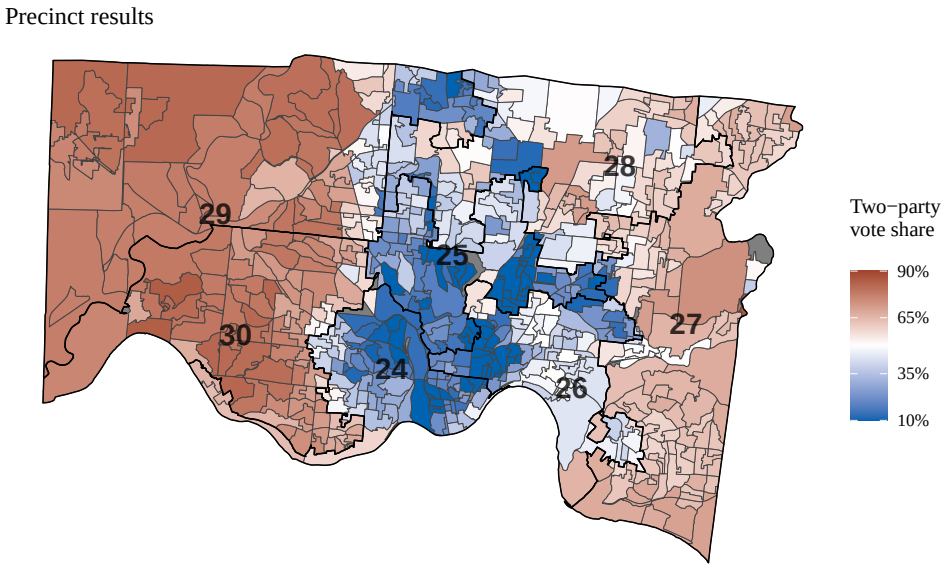

Figure 24: Vote shares for the precincts of Hamilton county.

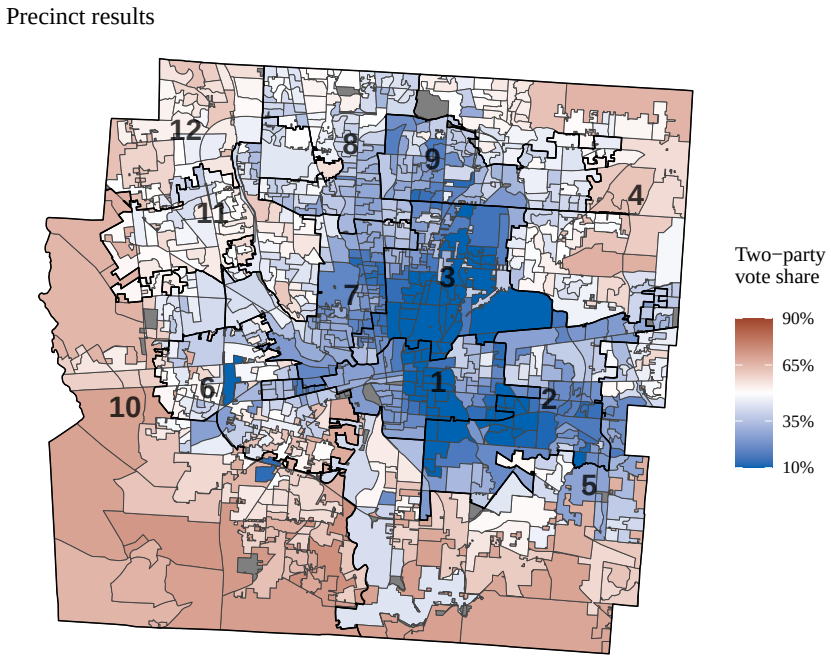

Figure 25: Vote shares for the precincts of Franklin county.

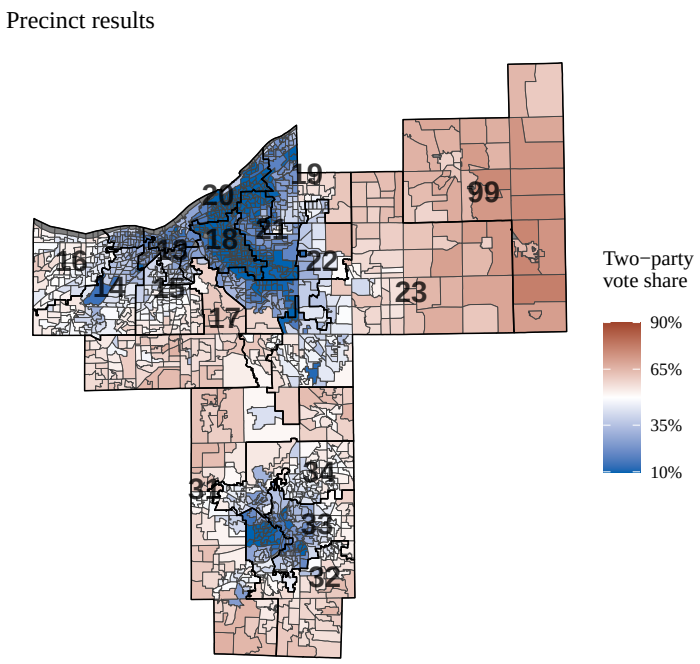

Figure 26: Vote shares for the precincts of Cuyahoga, Summit, and Geauga counties.

G.1. Data Sources

Data Aquisition

- I analyze a total of 13 statewide elections: US President (2012, 2016, 2020), US Senate (2012, 2016, 2018), Secretary of State (2014, 2018), Governor (2014, 2018), Attorney General (2018), Treasurer (2018), Auditor (2018)

- The 2016, 2018, and 2020 precinct-level shapefiles were acquired from the Voting and Election Science Team at the University of Florida and Wichita State University. This data is publicly available on the Harvard Dataverse, an online repository of social science data. Those shapefiles were joined to precinct-level election returns from the Ohio Secretary of State's office, which had been processed and cleaned by OpenElections.

- The 2012 and 2014 election returns pro-rated to the 2010 VTD level were acquired from Bill Cooper. Counsel has informed that Bill Cooper provided the following description of the data: The 2012 results are disaggregated to the block level (based on block centroids)

from the statewide 2012 precinct file. The 2014 results are based on a geocoding of about 3.15 million voters who cast ballots in Nov. 2014. These addresses were matched to census blocks and the blocks were aggregated to the precinct level. These virtual precincts were next matched to the 2014 election results and then disaggregated back to the block level, with block-level matches. When aggregated to the congressional level, the differences are measured in the tenths of a percent for House contests. As a final step, these datasets were aggregated from the block-level to the 2010 VTD level. Finally, it is important to note that there is a 2% to 3% undercount statewide for all votes cast in the 2014 election.

- Given the missing votes for the 2014 contests in Lorain County, the VTD-level totals in that county were approximated using the official precinct 2014 returns. First, after identifying the township, city, or village of each 2014 precinct, the official precinct-level returns were aggregated up to that level. Those municipality-level returns were then disaggregated for each candidate down to the VTDs in each municipality, proportionally to the vote counts for the candidate running for the same office and party in the 2018 midterm cycle.

- The 2020 Census Block shapefiles, total population by race and ethnicity, and voting age population by race and ethnicity were obtained directly from the Census FTP portal.

- The 2020 Census place block assignment files (for city and village boundaries), VTD block assignment files, lower general assembly district block assignment files, and upper general assembly district block assignment files were obtained from the Census website.

- The 2020 Census county subdivision shapefiles (for Ohio township boundaries) were obtained from the Census website.

- The enacted plan data and the House and Senate Democratic Caucuses plan data were obtained from the Ohio Redistricting Commission website, as block assignment files.

Data Processing

- The datasets that were on the 2020 census block level (total population, voting age population, Census place assignment, VTD assignment, lower GA district assignment, upper GA district assignment, Democratic proposed plans, enacted plans) were joined to the 2020 Census block shapefile.

- The datasets that were not on the level of the census block (2016, 2018, and 2020 election returns – precinct; 2012 and 2014 election returns – 2010 VTD) were disaggregated down to the 2020 census block level. Then, the resulting data were joined to the 2020 Census block shapefile.

- For the 2020 Census county subdivision shapefile, each 2020 Census block was assigned to its corresponding county subdivision assignment by overlaying the county subdivision shapefile onto the 2020 Census blocks.

- Given that some of Ohio's voting districts are geographically discontiguous, the separate discontiguous pieces of each voting district were identified.

Data Aggregation

- The full block-level dataset was aggregated up to the level of the 2020 voting districts, taking into account (a) discontiguous voting districts and (b) splits of voting districts by upper and lower General Assembly plans.

- The final municipality ID was constructed on the aggregated dataset. Where a VTD belonged to a village or a city, the municipality ID took the value of that village or city. Otherwise, it took the value of the county subdivision of the VTD. Then, discontiguous municipalities or townships were identified, and assigned to unique identifiers. The final muncipality ID concatenates the original municipality ID, the identifier for each discontiguous piece, and a county identifier, so that it identifies a unique contiguous piece of a municipality within a given county.

G.2. References

Autry, Eric, Daniel Carter, Gregory Herschlag, Zach Hunter, and Jonathan Mattingly. 2020. "Multi-scale merge-split Markov chain Monte Carlo for Redistricting." *arXiv preprint arXiv:2008.08054.*

Best, Robin E., Shawn J. Donahue, Jonathan Krasno, Daniel B. Magleby, and Michael D. McDonald. 2018. "Considering the Prospects for Establishing a Packing Gerrymandering Standard." *Election Law Journal: Rules, Politics, and Policy* 17 (1): 1–20.

Buzas, Jeffrey S., and Gregory S. Warrington. 2021. "Simulated Packing and Cracking." *Election Law Journal,* https://doi.org/10.1089/elj.2020.0673.

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 preprint arXiv:1911.01503.*

DeFord, Daryl, Moon Duchin, and Justin Solomon. 2021. "Recombination: A Family of Markov Chains for Redistricting." Https://hdsr.mitpress.mit.edu/pub/1ds8ptxu, *Harvard Data Science Review* (March 31, 2021). https://doi.org/10.1162/99608f92.eb30390f. https://hdsr.mitpress.mit.edu/pub/1ds8ptxu.

Doucet, Arnaud, Nando de Freitas, and Neil Gordon. 2001. *Sequential Monte Carlo methods in practice.* New York: Springer.

Fifield, Benjamin, Michael Higgins, Kosuke Imai, and Alexander Tarr. 2020. "Automated Redistricting Simulation Using Markov Chain Monte Carlo." *Journal of Computational and Graphical Statistics* 29 (4): 715–728.

Fifield, Benjamin, Kosuke Imai, Jun Kawahara, and Christopher T Kenny. 2020. "The essential role of empirical validation in legislative redistricting simulation." *Statistics and Public Policy* 7 (1): 52–68.

Gelman, Andrew, and Gary King. 1994. "A unified method of evaluating electoral systems and redistricting plans." *American Journal of Political Science,* 514–554.

Gilks, Walter R., Sylvia Richardson, and David J. Spiegelhalter. 1996. *Markov chain Monte Carlo in Practice.* London: Chapman & Hall.

Herschlag, Gregory, Han Sung Kang, Justin Luo, Christy Vaughn Graves, Sachet Bangia, Robert Ravier, and Jonathan C Mattingly. 2020a. "Quantifying Gerrymandering in North Carolina" [in eng]. *Statistics and Public Policy* 7 (1): 30–38. ISSN: 2330-443X.

———. 2020b. "Quantifying gerrymandering in North Carolina: Supplemental Appendix." *Statistics and Public Policy* 7 (1): 30–38.

Imai, Kosuke, and Kabir Khanna. 2016. "Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Record." *Political Analysis* 24 (2): 263–272.

Imai, Kosuke, Ying Lu, and Aaron Strauss. 2008. "Bayesian and Likelihood Inference for 2×2 Ecological Tables: An Incomplete Data Approach." *Political Analysis* 16 (1): 41–69.

Katz, Jonathan N, Gary King, and Elizabeth Rosenblatt. 2020. "Theoretical foundations and empirical evaluations of partisan fairness in district-based democracies." *American Political Science Review* 114 (1): 164–178.

Kenny, Christopher T., Shiro Kuriwaki, Cory McCartan, Evan Rosenman, Tyler Simko, and Kosuke Imai. 2021. "The Use of Differential Privacy for Census Data and its Impact on Redistricting: The Case of the 2020 U.S. Census." *Science Advances* 7, no. 41 (October): 1–17.

Kenny, Christopher T., Cory McCartan, Benjamin Fifield, and Kosuke Imai. 2020. *redist: Computational Algorithms for Redistricting Simulation.* https://CRAN.R-project.org/package=redist.

McCartan, Cory, and Kosuke Imai. 2020. "Sequential Monte Carlo for sampling balanced and compact redistricting plans." *arXiv preprint arXiv:2008.06131.*

McGhee, Eric. 2014. "Measuring Partisan Bias in Single-Member District Electoral Systems." *Legislative Studies Quarterly* 39 (1): 55–85.

Owen, Guillermo, and Bernard Grofman. 1988. "Optimal Partisan Gerrymandering." *Political Geography Quarterly* 7 (1): 5–22.

Polsby, Daniel D, and Robert D Popper. 1991. "The third criterion: Compactness as a procedural safeguard against partisan gerrymandering." *Yale Law & Policy Review* 9 (2): 301–353.

Stephanopoulos, Nicholas O., and Eric M. McGhee. 2015. "Partisan Gerrymandering and the Efficiency Gap." *University of Chicago Law Review* 82 (2): 831–900.

———. 2018. "The Measure of a Metric: The Debate over Quantifying Partisan Gerrymandering." *Stanford Law Review* 70:1503–1568.

Exhibit A of Expert Report

Kosuke Imai

Curriculum Vitae

October 2021

Contact Information

1737 Cambridge Street
Institute for Quantitative Social Science
Harvard University
Cambridge, MA 02138

Phone: 617-384-6778
Email: Imai@Harvard.Edu
URL: https://imai.fas.harvard.edu

Education

Ph.D. in Political Science, Harvard University (1999–2003)

A.M. in Statistics, Harvard University (2000–2002)

B.A. in Liberal Arts, The University of Tokyo (1994–1998)

Positions

Professor, Department of Government and Department of Statistics, Harvard University (2018 – present)

Professor, Department of Politics and Center for Statistics and Machine Learning, Princeton University (2013 – 2018)

Founding Director, Program in Statistics and Machine Learning (2013 – 2017)

Professor of Visiting Status, Graduate Schools of Law and Politics, The University of Tokyo (2016 – present)

Associate Professor, Department of Politics, Princeton University (2012 – 2013)

Assistant Professor, Department of Politics, Princeton University (2004 – 2012)

Visiting Researcher, Faculty of Economics, The University of Tokyo (August, 2006)

Instructor, Department of Politics, Princeton University (2003 – 2004)

KI_050

Honors and Awards

1. Invited to read "Experimental Evaluation of Computer-Assisted Human Decision-Making: Application to Pretrial Risk Assessment Instrument." before the Royal Statistical Society Research Section, London (2021).

2. *Excellence in Mentoring Award*, awarded by the Society for Political Methodology (2021).

3. *Statistical Software Award* for developing statistical software that makes a significant research contribution, for "fastLink: Fast Probabilistic Record Linkage," awarded by the Society for Political Methodology (2021).

4. *Highly Cited Researcher* (cross-field category) for "production of multiple highly cited papers that rank in the top 1% by citations for field and year in Web of Science," awarded by Clarivate Analytics (2018, 2019, 2020).

5. *President*, The Society for Political Methodology (2017–2019). *Vice President and President-elect* (2015–2017).

6. *Elected Fellow*, The Society for Political Methodology (2017).

7. *The Nils Petter Gleditsch Article of the Year Award* (2017), awarded by *Journal of Peace Research*.

8. *Statistical Software Award* for developing statistical software that makes a significant research contribution, for "mediation: R Package for Causal Mediation Analysis," awarded by the Society for Political Methodology (2015).

9. *Outstanding Reviewer Award* for Journal of Educational and Behavioral Statistics, given by the American Educational Research Association (2014).

10. *The Stanley Kelley, Jr. Teaching Award*, given by the Department of Politics, Princeton University (2013).

11. *Pi Sigma Alpha Award* for the best paper presented at the 2012 Midwest Political Science Association annual meeting, for "Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan," awarded by the Midwest Political Science Association (2013).

12. Invited to read "Experimental Designs for Identifying Causal Mechanisms" before the Royal Statistical Society Research Section, London (2012).

13. Inaugural recipient of the *Emerging Scholar Award* for a young scholar making exceptional contributions to political methodology who is within ten years of their terminal degree, awarded by the Society for Political Methodology (2011).

14. *Political Analysis Editors' Choice Award* for an article providing an especially significant contribution to political methodology, for "Estimation of Heterogeneous Treatment Effects from Randomized Experiments, with Application to the Optimal Planning of the Get-out-the-vote Campaign," awarded by the Society for Political Methodology and Oxford University Press (2011).

15. *Tom Ten Have Memorial Award* for the best poster presented at the 2011 Atlantic Causal Inference Conference, for "Identifying Treatment Effect Heterogeneity through Optimal Classification and Variable Selection," awarded by the Departments of Biostatistics and Statistics, University of Michigan (2011).

16. Nominated for the *Graduate Mentoring Award*, The McGraw Center for Teaching and Learning, Princeton University (2010, 2011).

17. *New Hot Paper*, for the most-cited paper in the field of Economics & Business in the last two months among papers published in the last year, for "Misunderstandings among Experimentalists and Observationalists about Causal Inference," named by Thomson Reuters' ScienceWatch (2009).

18. *Warren Miller Prize* for the best article published in *Political Analysis*, for "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," awarded by the Society for Political Methodology and Oxford University Press (2008).

19. *Fast Breaking Paper* for the article with the largest percentage increase in citations among those in the top 1% of total citations across the social sciences in the last two years, for "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," named by Thomson Reuters' ScienceWatch (2008).

20. *Pharmacoepidemiology and Drug Safety Outstanding Reviewer Recognition* (2008).

21. *Miyake Award* for the best political science article published in 2005, for "Do Get-Out-The-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments," awarded by the Japanese Political Science Association (2006).

22. *Toppan Prize* for the best dissertation in political science, for *Essays on Political Methodology*, awarded by Harvard University (2004). Also, nominated for American Political Science Association E.E. Schattschneider Award for the best doctoral dissertation in the field of American government and politics.

Publications in English

Book

Imai, Kosuke. (2017). *Quantitative Social Science: An Introduction.* Princeton University Press. Translated into Japanese (2018), Chinese (2020), and Korean (2021).

Stata version (2021) with Lori D. Bougher.

Tidyverse version (forthcoming) with Nora Webb Williams

Refereed Journal Articles

1. Imai, Kosuke, Zhichao Jiang, D. James Greiner, Ryan Halen, and Sooahn Shin. "Experimental Evaluation of Computer-Assisted Human Decision-Making: Application to Pretrial Risk Assessment Instrument." (with discussion) *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, Forthcoming. To be read before the Royal Statistical Society.

2. Imai, Kosuke, In Song Kim, and Erik Wang. "Matching Methods for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science*, Forthcoming.

3. Imai, Kosuke and Michael Lingzhi Li. "Experimental Evaluation of Individualized Treatment Rules." *Journal of the American Statistical Association*, Forthcoming.

4. de la Cuesta, Brandon, Naoki Egami, and Kosuke Imai."Experimental Design and Statistical Inference for Conjoint Analysis: The Essential Role of Population Distribution." *Political Analysis*, Forthcoming.

5. Kenny, Christopher T., Shiro Kuriwaki, Cory McCartan, Evan Rosenman, Tyler Simko, and Kosuke Imai. (2021). "The Use of Differential Privacy for Census Data and its Impact on Redistricting: The Case of the 2020 U.S. Census." *Science Advances*, Vol. 7, No. 7 (October), pp. 1-17.

6. Imai, Kosuke and James Lo. (2021). " Robustness of Empirical Evidence for the Democratic Peace: A Nonparametric Sensitivity Analysis." *International Organization*, Vol. 75, No. 3 (Summer), pp. 901–919.

7. Imai, Kosuke, Zhichao Jiang, and Anup Malani. (2021). "Causal Inference with Interference and Noncompliance in the Two-Stage Randomized Experiments." *Journal of the American Statistical Association*, Vol. 116, No. 534, pp. 632-644.

8. Imai, Kosuke, and In Song Kim. (2021). "On the Use of Two-way Fixed Effects Regression Models for Causal Inference with Panel Data." *Political Analysis*, Vol. 29, No. 3 (July), pp. 405–415.

9. Imai, Kosuke and Zhichao Jiang. (2020). "Identification and Sensitivity Analysis of Contagion Effects with Randomized Placebo-Controlled Trials." *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, Vol. 183, No. 4 (October), pp. 1637–1657.

10. Fifield, Benjamin, Michael Higgins, Kosuke Imai, and Alexander Tarr. (2020). "Automated Redistricting Simulation Using Markov Chain Monte Carlo." *Journal of Computational and Graphical Statistics*, Vol. 29, No. 4, pp. 715–728.

11. Fifield, Benjamin, Kosuke Imai, Jun Kawahara, and Christopher T. Kenny. (2020). "The Essential Role of Empirical Validation in Legislative Redistricting Simulation." *Statistics and Public Policy*, Vol. 7, No 1, pp. 52–68.

12. Ning, Yang, Sida Peng, and Kosuke Imai. (2020). "Robust Estimation of Causal Effects via High-Dimensional Covariate Balancing Propensity Score." *Biometrika*, Vol. 107, No. 3 (September), pp. 533—554.

13. Chou, Winston, Kosuke Imai, and Bryn Rosenfeld. (2020). "Sensitive Survey Questions with Auxiliary Information." *Sociological Methods & Research*, Vol. 49, No. 2 (May), pp. 418–454.

14. Imai, Kosuke, Gary King, and Carlos Velasco Rivera. (2020). "Do Nonpartisan Programmatic Policies Have Partisan Electoral Effects? Evidence from Two Large Scale Randomized Experiments." *Journal of Politics*, Vol. 82, No. 2 (April), pp. 714–730.

15. Zhao, Shandong, David A. van Dyk, and Kosuke Imai. (2020). "Propensity-Score Based Methods for Causal Inference in Observational Studies with Non-Binary Treatments." *Statistical Methods in Medical Research*, Vol. 29, No. 3 (March), pp. 709–727.

16. Lyall, Jason, Yang-Yang Zhou, and Kosuke Imai. (2020). "Can Economic Assistance Shape Combatant Support in Wartime? Experimental Evidence from Afghanistan." *American Political Science Review*, Vol. 114, No. 1 (February), pp. 126–143.

17. Kim, In Song, Steven Liao, and Kosuke Imai. (2020). "Measuring Trade Profile with Granular Product-level Trade Data." *American Journal of Political Science*, Vol. 64, No. 1 (January), pp. 102-117.

18. Enamorado, Ted and Kosuke Imai. (2019). "Validating Self-reported Turnout by Linking Public Opinion Surveys with Administrative Records." *Public Opinion Quarterly*, Vol. 83, No. 4 (Winter), pp. 723—748.

19. Blair, Graeme, Winston Chou, and Kosuke Imai. (2019). "List Experiments with Measurement Error." *Political Analysis*, Vol. 27, No. 4 (October), pp. 455–480.

20. Egami, Naoki, and Kosuke Imai. "Causal Interaction in Factorial Experiments: Application to Conjoint Analysis." *Journal of the American Statistical Association*, Vol. 114, No. 526 (June), pp. 529-540.

21. Enamorado, Ted, Benjamin Fifield, and Kosuke Imai. (2019). "Using a Probabilistic Model to Assist Merging of Large-scale Administrative Records." *American Political Science Review*, Vol. 113, No. 2 (May), pp. 353–371.

22. Imai, Kosuke and In Song Kim. (2019) "When Should We Use Linear Fixed Effects Regression Models for Causal Inference with Longitudinal Data?." *American Journal of Political Science*, Vol. 63, No. 2 (April), pp. 467–490.

23. Imai, Kosuke, and Zhichao Jiang. (2018). "A Sensitivity Analysis for Missing Outcomes Due to Truncation-by-Death under the Matched-Pairs Design." *Statistics in Medicine*, Vol. 37, No. 20 (September), pp. 2907–2922.

24. Fong, Christian, Chad Hazlett, and Kosuke Imai. (2018). "Covariate Balancing Propensity Score for a Continuous Treatment: Application to the Efficacy of Political Advertisements." *Annals of Applied Statistics*, Vol. 12, No. 1, pp. 156–177.

25. Hirose, Kentaro, Kosuke Imai, and Jason Lyall. (2017). "Can Civilian Attitudes Predict Insurgent Violence?: Ideology and Insurgent Tactical Choice in Civil War" *Journal of Peace Research*, Vol. 51, No. 1 (January), pp. 47–63.

26. Imai, Kosuke, James Lo, and Jonathan Olmsted. (2016). "Fast Estimation of Ideal Points with Massive Data." *American Political Science Review*, Vol. 110, No. 4 (December), pp. 631–656.

27. Rosenfeld, Bryn, Kosuke Imai, and Jacob Shapiro. (2016). "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions." *American Journal of Political Science*, Vol. 60, No. 3 (July), pp. 783–802.

28. Imai, Kosuke and Kabir Khanna. (2016). "Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Record." *Political Analysis*, Vol. 24, No. 2 (Spring), pp. 263–272.

29. Blair, Graeme, Kosuke Imai, and Yang-Yang Zhou. (2015). "Design and Analysis of the Randomized Response Technique." *Journal of the American Statistical Association*, Vol. 110, No. 511 (September), pp. 1304–1319.

30. Imai, Kosuke and Marc Ratkovic. (2015). "Robust Estimation of Inverse Probability Weights for Marginal Structural Models." *Journal of the American Statistical Association*, Vol. 110, No. 511 (September), pp. 1013–1023. (lead article)

31. Lyall, Jason, Yuki Shiraito, and Kosuke Imai. (2015). "Coethnic Bias and Wartime Informing." *Journal of Politics*, Vol. 77, No. 3 (July), pp. 833–848.

32. Imai, Kosuke, Bethany Park, and Kenneth Greene. (2015). "Using the Predicted Responses from List Experiments as Explanatory Variables in Regression Models." *Political Analysis*, Vol. 23, No. 2 (Spring), pp. 180–196. Translated in Portuguese and Reprinted in *Revista Debates* Vol. 9, No 1.

33. Blair, Graeme, Kosuke Imai, and Jason Lyall. (2014). "Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan." *American Journal of Political Science*, Vol. 58, No. 4 (October), pp. 1043–1063.

34. Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. (2014). "mediation: R Package for Causal Mediation Analysis." *Journal of Statistical Software*, Vol. 59, No. 5 (August), pp. 1–38.

35. Imai, Kosuke and Marc Ratkovic. (2014). "Covariate Balancing Propensity Score." *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, Vol. 76, No. 1 (January), pp. 243–263.

36. Lyall, Jason, Graeme Blair, and Kosuke Imai. (2013). "Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan." *American Political Science Review*, Vol. 107, No. 4 (November), pp. 679-705. Winner of the Pi Sigma Alpha Award.

37. Imai, Kosuke and Teppei Yamamoto. (2013). "Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments." *Political Analysis*, Vol. 21, No. 2 (Spring), pp. 141–171. (lead article).

38. Imai, Kosuke and Marc Ratkovic. (2013). "Estimating Treatment Effect Heterogeneity in Randomized Program Evaluation." *Annals of Applied Statistics*, Vol. 7, No. 1 (March), pp. 443–470. Winner of the Tom Ten Have Memorial Award. Reprinted in *Advances in Political Methodology*, R. Franzese, Jr. ed., Edward Elger, 2017.

39. Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. (2013). "Experimental Designs for Identifying Causal Mechanisms."(with discussions) *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, Vol. 176, No. 1 (January), pp. 5–51. (lead article) Read before the Royal Statistical Society, March 2012.

40. Imai, Kosuke, and Dustin Tingley. (2012). "A Statistical Method for Empirical Testing of Competing Theories." *American Journal of Political Science*, Vol. 56, No. 1 (January), pp. 218–236.

41. Blair, Graeme, and Kosuke Imai. (2012). "Statistical Analysis of List Experiments." *Political Analysis*, Vol. 20, No. 1 (Winter), pp. 47–77.

42. Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. (2011). "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review*, Vol. 105, No. 4 (November), pp. 765–789. Reprinted in *Advances in Political Methodology*, R. Franzese, Jr. ed., Edward Elger, 2017.

43. Bullock, Will, Kosuke Imai, and Jacob N. Shapiro. (2011). "Statistical Analysis of Endorsement Experiments: Measuring Support for Militant Groups in Pakistan." *Political Analysis*, Vol. 19, No. 4 (Autumn), pp. 363–384. (lead article)

44. Imai, Kosuke. (2011). "Multivariate Regression Analysis for the Item Count Technique." *Journal of the American Statistical Association*, Vol. 106, No. 494 (June), pp. 407–416. (featured article)

45. Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth Stuart. (2011). "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software*, Vol. 42 (Special Volume on Political Methodology), No. 8 (June), pp. 1–28.

46. Imai, Kosuke, Ying Lu, and Aaron Strauss. (2011). "eco: R Package for Ecological Inference in 2×2 Tables." *Journal of Statistical Software*, Vol. 42 (Special Volume on Political Methodology), No. 5 (June), pp. 1–23.

47. Imai, Kosuke and Aaron Strauss. (2011). "Estimation of Heterogeneous Treatment Effects from Randomized Experiments, with Application to the Optimal Planning of the Get-out-the-vote Campaign." *Political Analysis*, Vol. 19, No. 1 (Winter), pp. 1–19. (lead article) Winner of the Political Analysis Editors' Choice Award.

48. Imai, Kosuke, Luke Keele, and Dustin Tingley. (2010). "A General Approach to Causal Mediation Analysis." *Psychological Methods*, Vol. 15, No. 4 (December), pp. 309–334. (lead article)

49. Imai, Kosuke and Teppei Yamamoto. (2010). "Causal Inference with Differential Measurement Error: Nonparametric Identification and Sensitivity Analysis." *American Journal of Political Science*, Vol. 54, No. 2 (April), pp. 543–560.

50. Imai, Kosuke, Luke Keele, and Teppei Yamamoto. (2010). "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science*, Vol. 25, No. 1 (February), pp. 51–71.

51. King, Gary, Emmanuela Gakidou, Kosuke Imai, Jason Lakin, Ryan T. Moore, Clayton Nall, Nirmala Ravishankar, Manett Vargas, Martha María Téllez-Rojo, Juan Eugenio Hernández Ávila, Mauricio Hernández Ávila, and Héctor Hernández Llamas. (2009). "Public Policy for the Poor? A Randomized Ten-Month Evaluation of the Mexican Universal Health Insurance Program." (with a comment) *The Lancet*, Vol. 373, No. 9673 (April), pp. 1447–1454.

52. Imai, Kosuke, Gary King, and Clayton Nall. (2009). "The Essential Role of Pair Matching in Cluster-Randomized Experiments, with Application to the Mexican Universal Health

Insurance Evaluation." (with discussions) *Statistical Science*, Vol. 24, No. 1 (February), pp. 29–53.

53. Imai, Kosuke. (2009). "Statistical Analysis of Randomized Experiments with Nonignorable Missing Binary Outcomes: An Application to a Voting Experiment." *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Vol. 58, No. 1 (February), pp. 83–104.

54. Imai, Kosuke, Gary King, and Olivia Lau. (2008). "Toward A Common Framework of Statistical Analysis and Development." *Journal of Computational and Graphical Statistics*, Vol. 17, No. 4 (December), pp. 892–913.

55. Imai, Kosuke. (2008). "Variance Identification and Efficiency Analysis in Experiments under the Matched-Pair Design." *Statistics in Medicine*, Vol. 27, No. 4 (October), pp. 4857–4873.

56. Ho, Daniel E., and Kosuke Imai. (2008). "Estimating Causal Effects of Ballot Order from a Randomized Natural Experiment: California Alphabet Lottery, 1978–2002." *Public Opinion Quarterly*, Vol. 72, No. 2 (Summer), pp. 216–240.

57. Imai, Kosuke, Gary King, and Elizabeth A. Stuart. (2008). "Misunderstandings among Experimentalists and Observationalists: Balance Test Fallacies in Causal Inference." *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, Vol. 171, No. 2 (April), pp. 481–502. Reprinted in *Field Experiments and their Critics*, D. Teele ed., New Haven: Yale University Press, 2013.

58. Imai, Kosuke, Ying Lu, and Aaron Strauss. (2008). "Bayesian and Likelihood Ecological Inference for 2×2 Tables: An Incomplete Data Approach." *Political Analysis*, Vol. 16, No. 1 (Winter), pp. 41–69.

59. Imai, Kosuke. (2008). "Sharp Bounds on the Causal Effects in Randomized Experiments with "Truncation-by-Death"." *Statistics & Probability Letters*, Vol. 78, No. 2 (February), pp. 144–149.

60. Imai, Kosuke and Samir Soneji. (2007). "On the Estimation of Disability-Free Life Expectancy: Sullivan's Method and Its Extension." *Journal of the American Statistical Association*, Vol. 102, No. 480 (December), pp. 1199–1211.

61. Horiuchi, Yusaku, Kosuke Imai, and Naoko Taniguchi. (2007). "Designing and Analyzing Randomized Experiments: Application to a Japanese Election Survey Experiment." *American Journal of Political Science*, Vol. 51, No. 3 (July), pp. 669–687.

62. Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. (2007). "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis*, Vol. 15, No. 3 (Summer), pp. 199–236. (lead article) Winner of the Warren Miller Prize.

63. Ho, Daniel E., and Kosuke Imai. (2006). "Randomization Inference with Natural Experiments: An Analysis of Ballot Effects in the 2003 California Recall Election." *Journal of the American Statistical Association*, Vol. 101, No. 475 (September), pp. 888–900.

64. Imai, Kosuke, and David A. van Dyk. (2005). "MNP: R Package for Fitting the Multinomial Probit Model." *Journal of Statistical Software*, Vol. 14, No. 3 (May), pp. 1–32. abstract reprinted in *Journal of Computational and Graphical Statistics* (2005) Vol. 14, No. 3 (September), p. 747.

65. Imai, Kosuke. (2005). "Do Get-Out-The-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments." *American Political Science Review*, Vol. 99, No. 2 (May), pp. 283–300.

66. Imai, Kosuke, and David A. van Dyk. (2005). "A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation." *Journal of Econometrics*, Vol. 124, No. 2 (February), pp. 311–334.

67. Imai, Kosuke, and David A. van Dyk. (2004). "Causal Inference With General Treatment Regimes: Generalizing the Propensity Score." *Journal of the American Statistical Association*, Vol. 99, No. 467 (September), pp. 854–866.

68. Imai, Kosuke, and Gary King. (2004). "Did Illegal Overseas Absentee Ballots Decide the 2000 U.S. Presidential Election?" *Perspectives on Politics*, Vol. 2, No. 3 (September), pp. 537–549. Our analysis is a part of *The New York Times* article, "How Bush Took Florida: Mining the Overseas Absentee Vote" By David Barstow and Don van Natta Jr. July 15, 2001, Page 1, Column 1.

Invited Contributions

1. Imai, Kosuke, and Zhichao Jiang. (2019). "Comment: The Challenges of Multiple Causes." *Journal of the American Statistical Association*, Vol. 114, No. 528, pp. 1605—1610.

2. Benjamin, Daniel J., *et al.* (2018). "Redefine Statistical Significance." *Nature Human Behaviour*, Vol. 2, No. 1, pp. 6–10.

3. de la Cuesta, Brandon and Kosuke Imai. (2016). "Misunderstandings about the Regression Discontinuity Design in the Study of Close Elections." *Annual Review of Political Science*, Vol. 19, pp. 375–396.

4. Imai, Kosuke (2016). "Book Review of *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* by Guido W. Imbens and Donald B. Rubin." *Journal of the American Statistical Association*, Vol. 111, No. 515, pp. 1365–1366.

5. Imai, Kosuke, Bethany Park, and Kenneth F. Greene. (2015). "Usando as respostas previsíveis da abordagem list-experiments como variaveis explicativás em modelos de regressão." *Revista Debates*, Vol. 9, No. 1, pp. 121–151. First printed in *Political Analysis*, Vol. 23, No. 2 (Spring).

6. Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. (2014). "Comment on Pearl: Practical Implications of Theoretical Results for Causal Mediation Analysis." *Psychological Methods*, Vol. 19, No. 4 (December), pp. 482–487.

7. Imai, Kosuke, Gary King, and Elizabeth A. Stuart. (2014). "Misunderstandings among Experimentalists and Observationalists: Balance Test Fallacies in Causal Inference." in *Field Experiments and their Critics: Essays on the Uses and Abuses of Experimentation*

in the Social Sciences, D. L. Teele ed., New Haven: Yale University Press, pp. 196–227. First printed in *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, Vol. 171, No. 2 (April).

8. Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. (2013). "Reply to Discussions of "Experimental Designs for Identifying Causal Mechanisms"." *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, Vol. 173, No. 1 (January), pp. 46–49.

9. Imai, Kosuke. (2012). "Comments: Improving Weighting Methods for Causal Mediation Analysis." *Journal of Research on Educational Effectiveness*, Vol. 5, No. 3, pp. 293–295.

10. Imai, Kosuke. (2011). "Introduction to the Virtual Issue: Past and Future Research Agenda on Causal Inference." *Political Analysis*, Virtual Issue: Causal Inference and Political Methodology.

11. Imai, Kosuke, Booil Jo, and Elizabeth A. Stuart. (2011). "Commentary: Using Potential Outcomes to Understand Causal Mediation Analysis." *Multivariate Behavioral Research*, Vol. 46, No. 5, pp. 842–854.

12. Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. (2010). "Causal Mediation Analysis Using R," in *Advances in Social Science Research Using R*, H. D. Vinod (ed.), New York: Springer (Lecture Notes in Statistics), pp. 129–154.

13. Imai, Kosuke, Gary King, and Clayton Nall. (2009). "Rejoinder: Matched Pairs and the Future of Cluster-Randomized Experiments." *Statistical Science*, Vol. 24, No. 1 (February), pp. 65–72.

14. Imai, Kosuke. (2003). "Review of Jeff Gill's *Bayesian Methods: A Social and Behavioral Sciences Approach*," *The Political Methodologist*, Vol. 11 No. 1, 9–10.

Refereed Conference Proceedings

1. Svyatkovskiy, Alexey, Kosuke Imai, Mary Kroeger, and Yuki Shiraito. (2016). "Large-scale text processing pipeline with Apache Spark," *IEEE International Conference on Big Data*, Washington, DC, pp. 3928-3935.

Other Publications and Manuscripts

1. Goldstein, Daniel, Kosuke Imai, Anja S. Göritz, and Peter M. Gollwitzer. (2008). "Nudging Turnout: Mere Measurement and Implementation Planning of Intentions to Vote."

2. Ho, Daniel E. and Kosuke Imai. (2004). " The Impact of Partisan Electoral Regulation: Ballot Effects from the California Alphabet Lottery, 1978–2002." Princeton Law & Public Affairs Paper No. 04-001; Harvard Public Law Working Paper No. 89.

3. Imai, Kosuke. (2003). "Essays on Political Methodology," *Ph.D. Thesis.* Department of Government, Harvard University.

4. Imai, Kosuke, and Jeremy M. Weinstein. (2000). "Measuring the Economic Impact of Civil War," Working Paper Series No. 51, Center for International Development, Harvard University.

Selected Manuscripts

1. Ben-Michael, Eli, D. James Greiner, Kosuke Imai, and Zhichao Jiang. "Safe Policy Learning through Extrapolation: Application to Pre-trial Risk Assessment."
2. Tarr, Alexander and Kosuke Imai. "Estimating Average Treatment Effects with Support Vector Machines."
3. McCartan, Cory and Kosuke Imai. "Sequential Monte Carlo for Sampling Balanced and Compact Redistricting Plans."
4. Imai, Kosuke and Zhichao Jiang. "Principal Fairness for Human and Algorithmic Decision-Making."
5. Papadogeorgou, Georgia, Kosuke Imai, Jason Lyall, and Fan Li. "Causal Inference with Spatio-temporal Data: Estimating the Effects of Airstrikes on Insurgent Violence in Iraq."
6. Eshima, Shusei, Kosuke Imai, and Tomoya Sasaki. "Keyword Assisted Topic Models."
7. Tarr, Alexander, June Hwang, and Kosuke Imai. "Automated Coding of Political Campaign Advertisement Videos: An Empirical Validation Study."
8. Olivella, Santiago, Tyler Pratt, and Kosuke Imai. "Dynamic Stochastic Blockmodel Regression for Social Networks: Application to International Conflicts."
9. Chan, K.C.G, K. Imai, S.C.P. Yam, Z. Zhang. "Efficient Nonparametric Estimation of Causal Mediation Effects."
10. Fan, Jianqing, Kosuke Imai, Han Liu, Yang Ning, and Xiaolin Yang. "Improving Covariate Balancing Propensity Score: A Doubly Robust and Efficient Approach."
11. Barber, Michael and Kosuke Imai. "Estimating Neighborhood Effects on Turnout from Geocoded Voter Registration Records."
12. Hirano, Shigeo, Kosuke Imai, Yuki Shiraito, and Masaki Taniguchi. "Policy Positions in Mixed Member Electoral Systems: Evidence from Japan."

Publications in Japanese

1. Imai, Kosuke. (2007). "Keiryō Seijigaku niokeru Ingateki Suiron (Causal Inference in Quantitative Political Science)." *Leviathan*, Vol. 40, Spring, pp. 224–233.
2. Horiuchi, Yusaku, Kosuke Imai, and Naoko Taniguchi. (2005). "Seisaku Jyōhō to Tōhyō Sanka: Field Jikken ni yoru Kensyō (Policy Information and Voter Participation: A Field Experiment)." *Nenpō Seijigaku (The Annals of the Japanese Political Science Association)*, 2005–I, pp. 161–180.
3. Taniguchi, Naoko, Yusaku Horiuchi, and Kosuke Imai. (2004). "Seitō Saito no Etsuran ha Tohyō Kōdō ni Eikyō Suruka? (Does Visiting Political Party Websites Influence Voting Behavior?)" *Nikkei Research Report*, Vol. IV, pp. 16–19.

Statistical Software

1. Eshima, Shusei, Kosuke Imai, and Tomoya Sasaki. "Keyword Assisted Topic Models." The Comprehensive R Archive Network and GitHub. 2020.

2. Li, Michael Lingzhi and Kosuke Imai. "evalITR: Evaluating Individualized Treatment Rules." available through The Comprehensive R Archive Network and GitHub. 2020.

3. Egami, Naoki, Brandon de la Cuesta, and Kosuke Imai. "factorEx: Design and Analysis for Factorial Experiments." available through The Comprehensive R Archive Network and GitHub. 2019.

4. Kim, In Song, Erik Wang, Adam Rauh, and Kosuke Imai. "PanelMatch: Matching Methods for Causal Inference with Time-Series Cross-Section Data." available through GitHub. 2018.

5. Olivella, Santiago, Adeline Lo, Tyler Pratt, and Kosuke Imai. "NetMix: Mixed-membership Regression Stochastic Blockmodel for Networks." available through CRAN and Github. 2019.

6. Enamorado, Ted, Benjamin Fifield, and Kosuke Imai. "fastLink: Fast Probabilistic Record Linkage." available through The Comprehensive R Archive Network and GitHub. Winner of the Statistical Software Award. 2017.

7. Khanna, Kabir, and Kosuke Imai. "wru: Who Are You? Bayesian Predictions of Racial Category Using Surname and Geolocation." available through The Comprehensive R Archive Network and GitHub. 2015.

8. Fifield, Benjamin, Christopher T. Kenny, Cory McCartan, and Kosuke Imai. "redist: Markov Chain Monte Carlo Methods for Redistricting Simulation." available through The Comprehensive R Archive Network and GitHub. 2015.

9. Imai, Kosuke, James Lo, and Jonathan Olmsted. "emIRT: EM Algorithms for Estimating Item Response Theory Models." available through The Comprehensive R Archive Network. 2015.

10. Blair, Graeme, Yang-Yang Zhou, and Kosuke Imai. "rr: Statistical Methods for the Randomized Response Technique." available through The Comprehensive R Archive Network and GitHub. 2015.

11. Fong, Christian, Marc Ratkovic, and Kosuke Imai. "CBPS: R Package for Covariate Balancing Propensity Score." available through The Comprehensive R Archive Network and GitHub. 2012.

12. Egami, Naoki, Marc Ratkovic, and Kosuke Imai. "FindIt: R Package for Finding Heterogeneous Treatment Effects." available through The Comprehensive R Archive Network and GitHub. 2012.

13. Kim, In Song, and Kosuke Imai. "wfe: Weighted Linear Fixed Effects Regression Models for Causal Inference." available through The Comprehensive R Archive Network. 2011.

14. Shiraito, Yuki, and Kosuke Imai. "endorse: R Package for Analyzing Endorsement Experiments." available through The Comprehensive R Archive Network and GitHub. 2012.

15. Blair, Graeme, and Kosuke Imai. "list: Statistical Methods for the Item Count Technique and List Experiments." available through The Comprehensive R Archive Network and GitHub. 2011.

16. Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. "mediation: R Package for Causal Mediation Analysis." available through The Comprehensive R Archive Network and GitHub. 2009. Winner of the Statistical Software Award. Reviewed in *Journal of Educational and Behavioral Statistics*.

17. Imai, Kosuke. "experiment: R Package for Designing and Analyzing Randomized Experiments." available through The Comprehensive R Archive Network. 2007.

18. Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth Stuart. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." available through The Comprehensive R Archive Network and GitHub. 2005.

19. Imai, Kosuke, Ying Lu, and Aaron Strauss. "eco: Ecological Inference in 2×2 Tables." available through The Comprehensive R Archive Network and GitHub. 2004.

20. Imai, Kosuke, and David A. van Dyk. "MNP: R Package for Fitting the Multinomial Probit Model." available through The Comprehensive R Archive Network and GitHub. 2004.

21. Imai, Kosuke, Gary King, and Olivia Lau. "Zelig: Everyone's Statistical Software." available through The Comprehensive R Archive Network. 2004.

External Research Grants

Principal Investigator

1. National Science Foundation (2021–2024). "Collaborative Research: Causal Inference with Spatio-Temporal Data on Human Dynamics in Conflict Settings." (Algorithm for Threat Detection Program; DMS-2124463). Principal Investigator (with Georgia Papadogeorgou and Jason Lyall) $485,340.

2. National Science Foundation (2021–2023). "Evaluating the Impacts of Machine Learning Algorithms on Human Decisions." (Methodology, Measurement, and Statistics Program; SES-2051196). Principal Investigator (with D. James Greiner and Zhichao Jiang) $330,000.

3. Cisco Systems, Inc. (2020–2022). "Evaluating the Impacts of Algorithmic Recommendations on the Fairness of Human Decisions." (Ethics in AI; CG# 2370386) Principal Investigator (with D. James Greiner and Zhichao Jiang) $110,085.

4. The Alfred P. Sloan Foundation (2020–2022). "Causal Inference with Complex Treatment Regimes: Design, Identification, Estimation, and Heterogeneity." (Economics Program; 2020–13946) Co-Principal Investigator (with Francesca Dominici and Jose Zubizarreta) $996,299

5. Facebook Research Grant (2018). $25,000.

6. National Science Foundation (2016–2021). "Collaborative Conference Proposal: Support for Conferences and Mentoring of Women and Underrepresented Groups in Political Methodology." (Methodology, Measurement and Statistics and Political Science Programs; SES–1628102) Principal Investigator (with Jeffrey Lewis) $312,322. Supplement (SES–1831370) $60,000.

7. The United States Agency for International Development (2015–2017). "Unemployment and Insurgent Violence in Afghanistan: Evidence from the Community Development Program." (AID–OAA–A–12–00096) Principal Investigator (with Jason Lyall) $188,037

8. The United States Institute of Peace (2015–2016). "Assessing the Links between Economic Interventions and Stability: An impact evaluation of vocational and skills training in Kandahar, Afghanistan," Principal Investigator (with David Haines, Jon Kurtz, and Jason Lyall) $144,494.

9. Amazon Web Services in Education Research Grant (2014). Principal Investigator (with Graeme Blair and Carlos Velasco Rivera) $3,000.

10. Development Bank of Latin America (CAF) (2013). "The Origins of Citizen Support for Narcos: An Empirical Investigation," Principal Investigator (with Graeme Blair, Fabiana Machado, and Carlos Velasco Rivera). $15,000.

11. The International Growth Centre (2011–2013). "Poverty, Militancy, and Citizen Demands in Natural Resource-Rich Regions: Randomized Evaluation of the Oil Profits Dividend Plan for the Niger Delta" (RA–2010–12–013). Principal Investigator (with Graeme Blair). $117,116.

12. National Science Foundation, (2009–2012). "Statistical Analysis of Causal Mechanisms: Identification, Inference, and Sensitivity Analysis," (Methodology, Measurement, and Statistics Program and Political Science Program; SES–0918968). Principal Investigator. $97,574.

13. National Science Foundation, (2009–2011). "Collaborative Research: The Measurement and Identification of Media Priming Effects in Political Science," (Methodology, Measurement, and Statistics Program and Political Science Program; SES–0849715). Principal Investigator (with Nicholas Valentino). $317,126.

14. National Science Foundation, (2008–2009). "New Statistical Methods for Randomized Experiments in Political Science and Public Policy," (Political Science Program; SES–0752050). Principal Investigator. $52,565.

15. National Science Foundation, (2006–2009). "Collaborative Research: Generalized Propensity Score Methods," (Methodology, Measurement and Statistics Program; SES–0550873). Principal Investigator (with Donald B. Rubin and David A. van Dyk). $460,000.

16. The Telecommunications Advancement Foundation, (2004). "Analyzing the Effects of Party Webpages on Political Opinions and Voting Behavior," Principal Investigator (with Naoko Taniguchi and Yusaku Horiuchi). $12,000.

Adviser and Statistical Consultant

1. National Science Foundation (2016–2017). "Doctoral Dissertation Research: Crossing Africa's Arbitrary Borders: How Refugees Shape National Boundaries by Challenging Them." (Political Science Program, SES–1560636). Principal Investigator and Adviser for Co-PI Yang-Yang Zhou's Dissertation Research. $18,900.

2. Institute of Education Sciences (2012–2014). "Academic and Behavioral Consequences of Visible Security Measures in Schools" (R305A120181). Statistical Consultant (Emily Tanner-Smith, Principal Investigator). $351,228.

3. National Science Foundation (2013–2014). "Doctoral Dissertation Research: Open Trade for Sale: Lobbying by Productive Exporting Firm" (Political Science Program, SES–1264090). Principal Investigator and Adviser for Co-PI In Song Kim's Dissertation Research. $22,540.

4. National Science Foundation (2012–2013). "Doctoral Dissertation Research: The Politics of Location in Resource Rent Distribution and the Projection of Power in Africa" (Political Science Program, SES–1260754). Principal Investigator and Adviser for Co-PI Graeme Blair's Dissertation Research. $17,640.

Invited Short Courses and Outreach Lectures

1. Short Course on Causal Inference and Statistics – Department of Political Science, Rice University, 2009; Institute of Political Science, Academia Sinica, 2014.

2. Short Course on Causal Inference and Identification, The Empirical Implications of Theoretical Models (EITM) Summer Institute – Harris School of Public Policy, University of Chicago, 2011; Department of Politics, Princeton University, 2012.

3. Short Course on Causal Mediation Analysis – Summer Graduate Seminar, Institute of Statistical Mathematics, Tokyo Japan, 2010; Society for Research on Educational Effectiveness Conference, Washington DC, Fall 2011, Spring 2012, Spring 2015; Inter-American Development Bank, 2012; Center for Education Research, University of Wisconsin, Madison, 2012; Bobst Center for Peace and Justice, Princeton University, 2014; Graduate School of Education, University of Pennsylvania, 2014; EITM Summer Institute, Duke University, 2014; Center for Lifespan Psychology, Max Planck Institute for Human Development, 2015; School of Communication Research, University of Amsterdam, 2015; Uppsala University, 2016

4. Short Course on Covariate Balancing Propensity Score – Society for Research on Educational Effectiveness Conference, Washington DC, Spring 2013; Uppsala University, 2016

5. Short Course on Matching Methods for Causal Inference – Institute of Behavioral Science, University of Colorado, Boulder, 2009; Department of Political Science, Duke University, 2013.

6. Lecture on Statistics and Social Sciences – New Jersey Japanese School, 2011, 2016; Kaisei Academy, 2012, 2014; Princeton University Wilson College, 2012; University of Tokyo, 2014

Selected Presentations

1. Distinguished speaker, Harvard College Summer Program for Undergraduates in Data Science, 2021.
2. Keynote speaker, Kansas-Western Missouri Chapter of the American Statistical Association, 2021.
3. Invited plenary panelist, Association for Computing Machinery Conference on Fairness, Accountability, and Transparency (ACM FAccT) 2021.
4. Keynote speaker, Taiwan Political Science Association, 2020.
5. Keynote speaker, Boston Japanese Researchers Forum, Massachusetts Institute of Technology, 2020.
6. Keynote speaker, Causal Mediation Analysis Training Workshop, Mailman School of Public Health, Columbia University, 2020.
7. Keynote speaker, Special Workshop on Evidence-based Policy Making. World Economic Forum, Centre for the Fourth Industrial Revolution, Japan, 2020.
8. Distinguished speaker, Institute for Data, Systems, and Society. Massachusetts Institute of Technology, 2019.
9. Keynote speaker, The Harvard Experimental Political Science Graduate Student Conference, Harvard University, 2019.
10. Invited speaker, Beyond Curve Fitting: Causation, Counterfactuals, and Imagination-based AI. Association for the Advancement of Artificial Intelligence, Spring Symposium, Stanford University, 2019.
11. Inaugural speaker, Causal Inference Seminar, Departments of Biostatistics and Statistics, Boston University, 2019.
12. Keynote speaker, The Second Latin American Political Methodology Meeting, Universidad de los Andes (Department of Political Science), 2018.
13. Keynote speaker, The First Latin American Political Methodology Meeting, Pontifical Catholic University of Chile (Department of Political Science), 2017.
14. Keynote speaker, Workshop on Uncovering Causal Mechanisms, University of Munich (Department of Economics), 2016.
15. Keynote speaker, The National Quality Registry Research Conference, Stockholm, 2016.
16. Keynote speaker, The UK-Causal Inference Meeting, University of Bristol (School of Mathematics), 2015.
17. Keynote speaker, The UP-STAT Conference, the Upstate Chapters of the American Statistical Association, 2015.
18. Keynote speaker, The Winter Conference in Statistics, Swedish Statistical Society and Umeå University (Department of Mathematics and Mathematical Statistics), 2015.

19. Inaugural invited speaker, The International Methods Colloquium, Rice University, 2015.
20. Invited speaker, The International Meeting on Experimental and Behavioral Social Sciences, University of Oxford (Nuffield College), 2014.
21. Keynote speaker, The Annual Conference of Australian Society for Quantitative Political Science, University of Sydney, 2013.
22. Keynote speaker, The Graduate Student Conference on Experiments in Interactive Decision Making, Princeton University. 2008.

Conferences Organized

1. The Asian Political Methodology Meetings (January 2014, 2015, 2016, 2017, 2018; co-organizer)
2. The Experimental Research Workshop (September 2012; co-organizer)
3. The 12th World Meeting of the International Society for Bayesian Analysis (June 2012; a member of the organizing committee)
4. Conference on Causal Inference and the Study of Conflict and State Building (May 2012; organizer)
5. The 28th Annual Society for Political Methodology Summer Meeting (July 2011; host)
6. Conference on New Methodologies and their Applications in Comparative Politics and International Relations (February 2011; co-organizer)

Teaching

Courses Taught at Harvard

1. Stat 286/Gov 2003 Causal Inference (formally Stat 186/Gov 2002): introduction to causal inference
2. Gov 2003 Topics in Quantitative Methodology: causal inference, applied Bayesian statistics, machine learning

Courses Taught at Princeton

1. POL 245 Visualizing Data: exploratory data analysis, graphical statistics, data visualization
2. POL 345 Quantitative Analysis and Politics: a first course in quantitative social science
3. POL 451 Statistical Methods in Political Science: basic probability and statistical theory, their applications in the social sciences
4. POL 502 Mathematics for Political Science: real analysis, linear algebra, calculus
5. POL 571 Quantitative Analysis I: probability theory, statistical theory, linear models
6. POL 572 Quantitative Analysis II: intermediate applied statistics

KI_066

7. POL 573 Quantitative Analysis III: advanced applied statistics
8. POL 574 Quantitative Analysis IV: advanced applied statistics with various topics including Bayesian statistics and causal inference
9. Reading Courses: basic mathematical probability and statistics, applied bayesian statistics, spatial statistics

Advising

Current Students

1. Soubhik Barari (Government)
2. Adam Breuer (Computer Science and Government). To be Assistant Professor, Department of Government and Department of Computer Science, Dartmouth College
3. Jacob Brown (Government)
4. Ambarish Chattopadhyay (Statistics)
5. Shusei Eshima (Government)
6. Georgina Evans (Government)
7. Dae Woong Ham (Statistics)
8. Christopher T. Kenny (Government)
9. Michael Lingzhe Li (MIT, Operations Research Center)
10. Jialu Li (Government)
11. Cory McCartan (Statistics)
12. Sayumi Miyano (Princeton, Politics)
13. Sun Young Park (Government)
14. Casey Petroff (Political Economy and Government)
15. Averell Schmidt (Kennedy School)
16. Sooahn Shin (Government)
17. Tyler Simko (Government)
18. Soichiro Yamauchi (Government)
19. Yi Zhang (Statistics)

Current Postdocs

1. Eli Ben-Michael
2. Evan Rosenman

KI_067

Former Students

1. Alexander Tarr (Ph.D. in 2021, Department of Electrical and Computer Engineering, Princeton University; Dissertation Committee Chair)

2. Connor Jerzak (Ph.D. in 2021, Department of Government, Harvard University). Postdoctoral Fellow. To be Assistant Professor, Department of Government, University of Texas, Austin

3. Shiro Kuriwaki (Ph.D. in 2021, Department of Government, Harvard University). Postdoctoral Fellow, Stanford University. To be Assistant Professor, Department of Political Science, Yale University

4. Diana Stanescu (Ph.D. in 2020, Department of Politics, Princeton University). Postdoctoral Fellow, U.S.-Japan Program, Harvard University

5. Erik Wang (Ph.D. in 2020, Department of Politics, Princeton University). Assistant Professor, Department of Political and Social Change, Australian National University

6. Asya Magazinnik (Ph.D. in 2020, Department of Politics, Princeton University). Assistant Professor, Department of Political Science, Massachusetts Institute of Technology

7. Max Goplerud (Ph.D. in 2020, Department of Government, Harvard University). Assistant Professor, Department of Political Science, University of Pittsburgh

8. Nicole Pashley (Ph.D. in 2020, Department of Statistics, Harvard University). Assistant Professor, Department of Statistics, Rutgers University

9. Naoki Egami (Ph.D. in 2020, Department of Politics, Princeton University; Dissertation Committee Chair). Assistant Professor, Department of Political Science, Columbia University

10. Brandon de la Cuesta (Ph.D. in 2019, Department of Politics, Princeton University). Postdoctoral Fellow, Center on Global Poverty and Development, Stanford University

11. Yang-Yang Zhou (Ph.D. in 2019, Department of Politics, Princeton University). Assistant Professor, Department of Political Science, University of British Columbia

12. Winston Chou (Ph.D. in 2019, Department of Politics, Princeton University). Senior Data Scientist at Apple

13. Ted Enamorado (Ph.D. in 2019, Department of Politics, Princeton University; Dissertation Committee Chair). Assistant Professor, Department of Political Science, Washington University in St. Louis

14. Benjamin Fifield (Ph.D. in 2018, Department of Politics, Princeton University; Dissertation Committee Chair). Data Scientist, American Civil Liberties Union

15. Tyler Pratt. (Ph.D. in 2018, Department of Politics, Princeton University). Assistant Professor, Department of Political Science, Yale University

16. Romain Ferrali (Ph.D. in 2018, Department of Politics, Princeton University). Postdoctoral Fellow, New York University, Abu Dhabi

KI_068

17. Julia Morse (Ph.D. in 2017, Woodrow Wilson School, Princeton University). Assistant Professor, Department of Political Science, University of California, Santa Barbara
18. Yuki Shiraito (Ph.D. in 2017, Department of Politics, Princeton University; Dissertation Committee Chair). Assistant Professor, Department of Political Science, University of Michigan
19. Carlos Velasco Rivera (Ph.D. in 2016, Department of Politics, Princeton University). Research Scientist, Facebook
20. Gabriel Lopez Moctezuma (Ph.D. in 2016, Department of Politics, Princeton University). Assistant Professor, Division of the Humanities and Social Sciences, California Institute of Technology
21. Graeme Blair (Ph.D. in 2016, Department of Politics, Princeton University). Assistant Professor, University of California, Los Angeles
22. Jaquilyn R. Waddell Boie (Ph.D. in 2015, Department of Politics, Princeton University). Private consultant
23. Scott Abramson (Ph.D. in 2014, Department of Politics, Princeton University). Associate Professor, Department of Political Science, University of Rochester
24. Michael Barber (Ph.D. in 2014, Department of Politics, Princeton University). Associate Professor, Department of Political Science, Brigham Young University
25. In Song Kim (Ph.D. in 2014, Department of Politics, Princeton University). Associate Professor, Department of Political Science, Massachusetts Institute of Technology
26. Alex Ruder (Ph.D. in 2014, Department of Politics, Princeton University). Senior Community Economic Development Advisor, Federal Reserve Bank of Atlanta
27. Meredith Wilf (Ph.D. in 2014, Department of Politics, Princeton University). Assistant Professor, Graduate School of Public and International Affairs, University of Pittsburgh
28. Will Bullock. (Ph.D. candidate, Department of Politics, Princeton University). Senior Researcher, Facebook
29. Teppei Yamamoto (Ph.D. in 2011, Department of Politics, Princeton University; Dissertation Committee Chair). Associate Professor, Department of Political Science, Massachusetts Institute of Technology
30. Dustin Tingley (Ph.D. in 2010, Department of Politics, Princeton University). Professor, Department of Government, Harvard University
31. Aaron Strauss (Ph.D. in 2009, Department of Politics, Princeton University). Executive Director, Analyst Institute
32. Samir Soneji (Ph.D. in 2008, Office of Population Research, Princeton University; Dissertation Committee Chair). Associate Professor, Dartmouth Institute for Health Policy & Clinical Practice, Geisel School of Medicine, Dartmouth College
33. Ying Lu (Ph.D. in 2005, Woodrow Wilson School, Princeton University; Dissertation Committee Chair). Associate Professor, Steinhardt School of Culture, Education, and Human Development, New York University

Former Predocs and Postdocs

1. Zhichao Jiang (Postdoctoral Fellow, 2016–2019). Assistant Professor, Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst
2. Adeline Lo (Postdoctoral Fellow, 2016–2019). Assistant Professor, Department of Political Science, University of Wisconsin, Madison
3. Yunkyu Sohn (Postdoctoral Fellow, 2016–2018). Assistant Professor, School of Political Science and Economics, Waseda University
4. Xiaolin Yang (Postdoctoral Fellow, 2015–2017). Research Scientist, Amazon
5. Santiago Olivella (Postdoctoral Fellow, 2015–2016). Assistant Professor, Department of Political Science, University of North Carolina
6. Drew Dimmery (Predoctoral Fellow, 2015–2016). Research Scientist, Facebook
7. James Lo (Postdoctoral Fellow, 2014–2016). Assistant Professor, Department of Political Science, University of Southern California
8. Steven Liao (Predoctoral Fellow, 2014–2015). Assistant Professor, Department of Political Science, University of California, Riverside
9. Michael Higgins (Postdoctoral Fellow, 2013–2015). Assistant Professor, Department of Statistics, Kansas State University
10. Kentaro Hirose (Postdoctoral Fellow, 2012–2015). Assistant Professor, Waseda Institute for Advanced Studies
11. Chad Hazlett (Predoctoral Fellow, 2013–2014). Assistant Professor, Departments of Political Science and Statistics, University of California, Los Angeles
12. Florian Hollenbach (Predoctoral Fellow, 2013–2014). Assistant Professor, Department of Political Science, Texas A&M University
13. Marc Ratkovic (Predoctoral and Postdoctoral Fellow, 2010–2012). Assistant Professor, Department of Politics, Princeton University

Editorial and Referee Service

Co-editor for *Journal of Causal Inference* (2014 – present)

Associate editor for *American Journal of Political Science* (2014 – 2019), *Journal of Business & Economic Statistics* (2015 – 2024), *Journal of Causal Inference* (2011 – 2014), *Journal of Experimental Political Science* (2013 – 2017), *Observational Studies* (2014 – present), *Political Analysis* (2014 – 2017).

Editorial board member for *Asian Journal of Comparative Politics* (2014 – present), *Journal of Educational and Behavioral Statistics* (2011 – present), *Journal of Politics* (2007 – 2008, 2019–2020), *Journal of Research on Educational Effectiveness* (2014 – 2016), *Political Analysis* (2010 – 2013), *Political Science Research and Methods* (2019 – present).

Guest editor for *Political Analysis* virtual issue on causal inference (2011).

Referee for *ACM Computing Surveys*, *American Economic Journal: Applied Economics*, *American Economic Review: Insights*, *American Journal of Epidemiology*, *American Journal of Evaluation*, *American Journal of Political Science*, *American Political Science Review*, *American Politics Research*, *American Sociological Review*, *Annals of Applied Statistics*, *Annals of Statistics*, *Annals of the Institute of Statistical Mathematics*, *Biometrics*, *Biometrika*, *Biostatistics*, *BMC Medical Research Methodology*, *British Journal of Mathematical and Statistical Psychology*, *British Journal of Political Science*, *Canadian Journal of Statistics*, *Chapman & Hall/CRC Press*, *Child Development*, *Communications for Statistical Applications and Methods*, *Computational Statistics and Data Analysis*, *Electoral Studies*, *Econometrica*, *Econometrics*, *Empirical Economics*, *Environmental Management*, *Epidemiology*, *European Union Politics*, *IEEE Transactions on Information Theory*, *International Journal of Biostatistics*, *International Journal of Epidemiology*, *International Journal of Public Opinion Research*, *International Migration Review*, John Wiley & Sons, *Journal of Applied Econometrics*, *Journal of Applied Statistics*, *Journal of Biopharmaceutical Statistics*, *Journal of Business and Economic Statistics*, *Journal of Causal Inference*, *Journal of Computational and Graphical Statistics*, *Journal of Conflict Resolution*, *Journal of Consulting and Clinical Psychology*, *Journal of Econometrics*, *Journal of Educational and Behavioral Statistics*, *Journal of Empirical Legal Studies*, *Journal of Multivariate Analysis*, *Journal of Official Statistics*, *Journal of Peace Research*, *Journal of Politics*, *Journal of Research on Educational Effectiveness*,*Journal of Statistical Planning and Inference*, *Journal of Statistical Software*, *Journal of Survey Statistics and Methodology*, *Journal of the American Statistical Association (Case Studies and Applications; Theory and Methods)*, *Journal of the Japanese and International Economies*, *Journal of the Japan Statistical Society*, *Journal of the Royal Statistical Society (Series A; Series B; Series C)*, *Law & Social Inquiry*, *Legislative Studies Quarterly*, *Management Science*, Multivariate Behavioral Research, National Science Foundation (Economics; Methodology, Measurement, and Statistics; Political Science), Natural Sciences and Engineering Research Council of Canada, *Nature Machine Intelligence*, *NeuroImage*, *Osteoporosis International*, *Oxford Bulletin of Economics and Statistics*, *Pharmaceutical Statistics*, *Pharmacoepidemiology and Drug Safety*, *PLOS One*, *Policy and Internet*, *Political Analysis*, *Political Behavior*, *Political Communication*, *Political Research Quarterly*, *Political Science Research and Methods*, *Population Health Metrics*, *Population Studies*, *Prevention Science*, *Proceedings of the National Academy of Sciences*, Princeton University Press, *Psychological Methods*, *Psychometrika*, *Public Opinion Quarterly*, *Quarterly Journal of Economics*, *Quarterly Journal of Political Science*, *Review of Economics and Statistics*, Routledge, Sage Publications, *Scandinavian Journal of Statistics*, *Science*, Sloan Foundation, Springer, *Sociological Methodology*, *Sociological Methods & Research*, *Statistical Methodology*, *Statistical Methods and Applications*, *Statistical Methods in Medical Research*, *Statistical Science*, *Statistica Sinica*, *Statistics & Probability Letters*, *Statistics in Medicine*, *Systems Biology*, U.S.-Israel Binational Science Foundation, *Value in Health*, *World Politics*.

University and Departmental Committees

Harvard University

Department of Government

Member, Curriculum and Educational Policy Committee (2020–2021)

Member, Second-year Progress Committee (2019–2020)

Member, Graduate Placement Committee (2019–2020)

Member, Graduate Admissions Committee (2018–2019)

Member, Graduate Poster Session Committee (2018–2019)

Department of Statistics

Chair, Senior Faculty Search Committee (2021–2022)

Member, Junior Faculty Search Committee (2018–2019)

Member, Second-year Progress Committee (2018–2019, 2020–2021)

Princeton University

University

Executive Committee Member, Program in Statistics and Machine Learning (2013–2018)

Executive Committee Member, Committee for Statistical Studies (2011-2018)

Member, Organizing Committee, Retreat on Data and Information Science at Princeton (2016)

Member, Council of the Princeton University Community (2015)

Member, Search Committee for the Dean of College (2015)

Member, Committee on the Library and Computing (2013–2016)

Member, Committee on the Fund for Experimental Social Science (2013–2018)

Member, Personally Identifiable Research Data Group (2012–2018)

Member, Research Computing Advisory Group (2013–2018)

Member, Task Force on Statistics and Machine Learning (2014–2015)

Department of Politics

Chair, Department Committee on Research and Computing (2012–2018)

Chair, Formal and Quantitative Methods Junior Search Committee (2012–2013, 2014–2015, 2016–2017)

Chair, Reappointment Committee (2015–2016)

Member, Diversity Initiative Committee (2014–2015)

Member, American Politics Junior Search Committee (2012–2014)

Member, Department Chair's Advisory Committee (2010–2013, 2015–2016)

Member, Department Priority Committee (2012–2013, 2014–2015, 2016–2017)

Member, Formal and Quantitative Methods Curriculum Committee (2005–2006)

Member, Formal and Quantitative Methods Junior Search Committee (2009–2010, 2015–2016)

Member, Formal and Quantitative Methods Postdoc Search Committee (2009–2018)

Member, Graduate Admissions Committee (2012–2013)

Member, Reappointment Committee (2014–2016)

Member, Space Committee (2014–2016)

Member, Undergraduate Curriculum Committee (2014–2015)

Member, Undergraduate Exam Committee (2007–2008)

Member, Undergraduate Thesis Prize Committee (2005–2006, 2008–2011)

Center for Statistics and Machine Learning

Executive Committee Member (2016–2018)

Member, Search Committee (2015–2017)

Services to the Profession

National Academies of Sciences, Engineering, and Medicine

Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Panel on the Review and Evaluation of the 2014 Survey of Income and Program Participation Content and Design (2014–2017)

National Science Foundation

Proposal Review Panel (2020)

The Society for Political Methodology

President (2017–2019)

Vice President and President Elect (2015–2017)

Annual Meeting Committee, Chair (2011)

Career Award Committee (2015–2017)

Program Committee for Annual Meeting (2012), Chair (2011)

Graduate Student Selection Committee for the Annual Meeting (2005), Chair (2011)

Miller Prize Selection Committee (2010–2011)

Statistical Software Award Committee (2009–2010)

Emerging Scholar Award Committee (2013)

American Statistical Association

Journal of Educational and Behavioral Statistics Management Committee (2016 – present)

Others

External Expert, Department of Methodology, London School of Economics and Political Science (2017)

Memberships

American Political Science Association; American Statistical Association; Midwest Political Science Association; The Society for Political Methodology.

CERTIFICATE OF SERVICE

I, Freda J. Levenson, hereby certify that on October 22, 2021, I caused a true and correct copy of the following documents to be served by email upon the counsel listed below:

1. **Affidavit of Dr. Kosuke Imai**
2. **Exhibit A - Dr. Kosuke Imai Expert Report (pages 1 - 73)**

DAVE YOST
OHIO ATTORNEY GENERAL
Bridget C. Coontz (0072919)
Julie M. Pfeiffer (0069762)
Michael A. Walton (0092201)
Michael J. Hendershot (0081842)
30 E. Broad St.
Columbus, OH 43215
Tel: (614) 466-2872
Fax: (614) 728-7592
bridget.coontz@ohioago.gov
julie.pfeiffer@ohioago.gov
michael.walton@ohioago.gov
michael.hendershot@ohioago.gov

Counsel for Respondents Governor Mike DeWine,
Secretary of State Frank LaRose, and
Auditor Keith Faber

W. Stuart Dornette (0002955)
Beth A. Bryan (0082076)
Philip D. Williamson (0097174)
TAFT STETTINIUS & HOLLISTER LLP
425 Walnut St., Suite 1800
Cincinnati, OH 45202
Tel: (513) 381-2838
dornette@taftlaw.com
bryan@taftlaw.com
pwilliamson@taftlaw.com

Phillip J. Strach (PHV 25444-2021)
Thomas A. Farr (PHV 25461-2021)

John E. Branch (PHV 25460-2021)
Alyssa M. Riggings (PHV 25441-2021)
Greg McGuire (PHV 25483-2021)
NELSON MULLINS RILEY & SCARBOROUGH LLP
4140 Parklake Ave., Ste. 200
Raleigh, NC 27612
phil.strach@nelsonmullins.com
tom.farr@nelsonmullins.com
john.branch@nelsonmullins.com
alyssa.riggins@nelsonmullins.com
greg.mcguire@nelsonmullins.com
Tel: (919) 329-3812

Counsel for Respondents
Senate President Matt Huffman and
House Speaker Robert Cupp

John Gilligan (0024542)
Diane Menashe (0070305)
ICE MILLER LLP
250 West St., Ste., 700
Columbus, OH 43215
john.gilligan@icemiller.com
diane.menashe@icemiller.com

Counsel for Respondents
Senator Vernon Sykes and
House Minority Leader Emilia Sykes

Erik J. Clark (0078732)
Ashley Merino (0096853)
ORGAN LAW LLP
1330 Dublin Rd.
Columbus, OH 43215
Tel: (614) 481-0900
Fax: (614) 481-0904

Counsel for Respondent Ohio Redistricting Commission

/s/ Freda J. Levenson