IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF SOUTH CAROLINA
COLUMBIA DIVISION

THE SOUTH CAROLINA STATE CONFERENCE OF THE NAACP, et al.,

Plaintiffs,

v.

THOMAS C. ALEXANDER, et al.,

Defendants.

Case No. 3:21-cv-03302-MGL-TJH-RMG

SENATE DEFENDANTS' & HOUSE DEFENDANTS' MOTION IN LIMINE TO EXCLUDE THE TESTIMONY OF PLAINTIFFS' PUTATIVE EXPERT KOSUKE IMAI

The Senate Defendants and the House Defendants[1] respectfully move to exclude the testimony of Plaintiffs' putative expert Kosuke Imai on the ground that it fails to meet the demanding standards established by Federal Rule of Evidence 702 and Daubert v. Merrell Dow Pharmaceuticals., Inc., 509 U.S. 579 (1993). Dr. Imai lacks a "reliable foundation" for his proposed testimony regarding the Congressional Plan that Plaintiffs challenge in this action. Daubert, 509 U.S. at 597. Indeed, Dr. Imai committed the same fundamental error as the putative expert that this Court rejected ten years ago in Backus v. South Carolina: by his own admission, he "failed to consider all the traditional race-neutral principles that guide redistricting in South Carolina." Backus v. South Carolina, 857 F. Supp. 2d 553, 563 (D.S.C.), aff'd, 568 U.S. 801 (2012). Thus, his proposed testimony and evidence are "problematic," "incomplete," and "unconvincing," and he is "unable to provide the Court a reliable opinion that the General

[1] The Senate Defendants are Thomas C. Alexander, in his official capacity as President of the Senate, and Luke A. Rankin, in his official capacity as Chairman of the Senate Judiciary Committee. The House Defendants are James H. Lucas, in his official capacity as Speaker of the House of Representatives, Chris Murphy, in his official capacity as Chairman of the House of Representatives Judiciary Committee, and Wallace H. Jordan, in his official capacity as Chairman of the House of Representatives Elections Law Subcommittee. On May 12, 2022, James H. Lucas stepped down as Speaker of the South Carolina House of Representatives. The current Speaker of the House is Representative G. Murrell Smith, Jr.

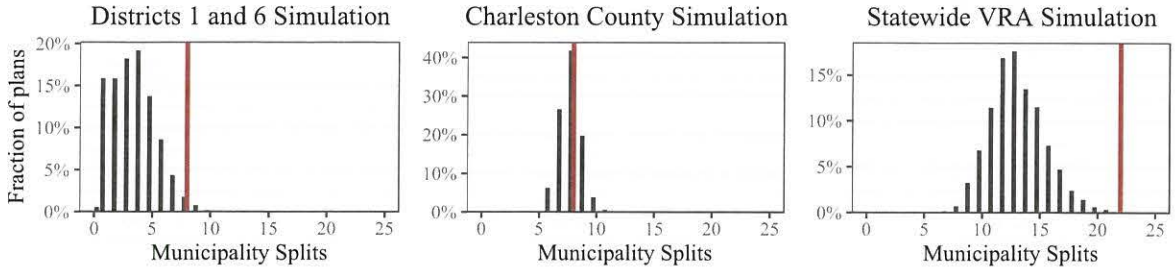


Figure 13: Municipality splits in simulation. The histogram shows the distribution of the number of split municipalities under the simulated plans while the red vertical line shows the enacted plan. On average, the simulated plans split fewer number of municipalities than the enacted plan.

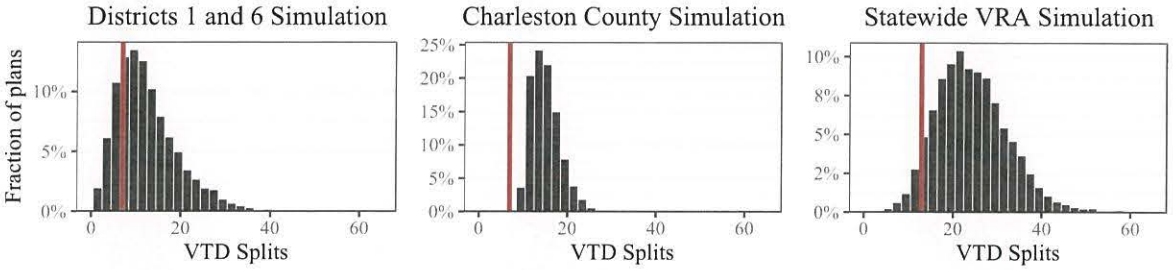


Figure 14: Precinct or Voting Tabulation District (VTD) splits in the simulation. The histogram shows the distribution of the number of split VTDs under the simulated plans while the red vertical line shows the enacted plan.

E. Municipality Splits of the Simulated Districts

60. Figure 13 shows that the number of municipalities split under the simulated plans (grey histograms) is no greater than that under the enacted plan (red vertical line).

F. Precinct Splits of the Simulated Districts

61. Figure 14 show that the number of split precincts or voting tabulation districts (VTDs) among the simulated plans (grey histogram) is generally compatible with that of the enacted plan (vertical red line) but tends to be somewhat higher on average. This is in part due to the fact that many municipalities split VTDs, implying that there often is a direct trade-off between municipality and precinct splits.

G. Data Sources

G.1. Data Acquisition

62. The 2020 Census Block shapefiles, 2020 Census Place shapefiles, total population by race and ethnicity, and voting age population by race and ethnicity directly were acquired from the Census FTP portal. In this report, when reporting the black voting age population, I count voters in the Census that are any-part black as black.

63. The VTD block assignment files and Census Place block assignment files were acquired from the Census website.

64. The incumbent addresses were acquired from the Redistricting Data Hub and subsequently modified based on public information and records (e.g., South Carolina State Election Commission filings, South Carolina property records) and input from plaintiffs' counsel. These addresses were then geocoded to census blocks.

65. The passed Congressional plan was acquired from the South Carolina House of Representatives Redistricting 2021 website.

66. The 2020 Census place block assignment files (for city and town boundaries) were obtained from the Census website.

G.2. Data Processing

67. For datasets that were on the 2020 census block level (total population, voting age population, VTD assignment, incumbent addresses, congressional district assignment, and census place assignment), these datasets were joined to the 2020 Census block shapefile.

G.3. Data Aggregation

68. The full block-level dataset was aggregated up to the level of the 2020 voting districts, taking into account (a) discontiguities in voting districts (b) splits of voting districts by the proposed Congressional plan and (c) splits of voting districts by cities and towns.

H. 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.*

Autry, Eric A., Daniel Carter, Gregory J. Herschlag, Zach Hunter, and Jonathan C. Mattingly. 2021. "Metropolized Multiscale Forest Recombination for Redistricting." *Multiscale Modeling & Simulation* 19 (4): 1885–1914.

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.

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

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.

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.

———. 2022. *redistmetrics: Redistricting Metrics.* https://CRAN.R-project.org/package=redistmetrics.

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

EXHIBIT A

Curriculum Vitae

Kosuke Imai

Curriculum Vitae

March 2022

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)

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 (2022).

2. *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, 2021).

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

4. *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).

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

Books

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

Llaudet, Elena, and Kosuke Imai. (forthcoming). *Data Analysis for Social Science: A Friendly and Practical Introduction.* Princeton University Press.

Refereed Journal Articles

1. Olivella, Santiago, Tyler Pratt, and Kosuke Imai. "Dynamic Stochastic Blockmodel Regression for Social Networks: Application to International Conflicts." *Journal of the American Statistical Association*, Forthcoming.

2. Fan, Jianqing, Kosuke Imai, Inbeom Lee, Han Liu, Yang Ning, and Xiaolin Yang. "Optimal Covariate Balancing Conditions in Propensity Score Estimation." *Journal of Business & Economic Statistics*, Forthcoming.

3. 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.

4. 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.

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

6. de la Cuesta, Brandon, Naoki Egami, and Kosuke Imai. (2022). "Experimental Design and Statistical Inference for Conjoint Analysis: The Essential Role of Population Distribution." *Political Analysis*, Vol. 30, No. 1 (January), pp. 19–45.

7. 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.

8. 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.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

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

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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.

30. 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.

31. 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.

32. 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)

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

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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).

40. 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.

41. 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.

42. 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.

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

44. 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.

45. 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)

46. 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)

47. 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.

48. 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.

49. 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.

50. 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)

51. 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.

52. 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.

53. 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.

54. 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.

55. 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.

56. 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.

57. 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.

58. 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.

59. 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.

60. 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.

61. 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.

62. 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.

63. 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.

64. 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.

65. 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.

66. 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.

67. 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.

68. 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.

69. 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.

70. 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. (2022). "Causal Diagrams and Social Science Research." *Probabilistic and Causal Inference: The Works of Judea Pearl.* Geffner, Hector and Dechter, Rina and Halpern, Joseph Y. (eds). Association for Computing Machinery and Morgan & Claypool, pp. 647–654.

2. 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.

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

4. 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.

5. 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.

6. 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).

7. 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.

8. 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).

9. 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.

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

11. 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.

12. 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.

13. 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.

14. 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.

15. 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. Ham, Dae Woong, Kosuke Imai, and Lucas Janson. "Using Machine Learning to Test Causal Hypotheses in Conjoint Analysis."

2. Goplerud, Max, Kosuke Imai, Nicole E. Pashley. "Estimating Heterogeneous Causal Effects of High-Dimensional Treatments: Application to Conjoint Analysis."

3. Malani, Anup, Phoebe Holtzman, Kosuke Imai, Cynthia Kinnan, Morgen Miller, Shailender Swaminathan, Alessandra Voena, Bartosz Woda, and Gabriella Conti. "Effect of Health Insurance in India: A Randomized Controlled Trial."

4. McCartan, Cory, Jacob Brown, and Kosuke Imai. "Measuring and Modeling Neighborhoods."

5. Ben-Michael, Eli, D. James Greiner, Kosuke Imai, and Zhichao Jiang. "Safe Policy Learning through Extrapolation: Application to Pre-trial Risk Assessment."

6. Tarr, Alexander and Kosuke Imai. "Estimating Average Treatment Effects with Support Vector Machines."

7. McCartan, Cory and Kosuke Imai. "Sequential Monte Carlo for Sampling Balanced and Compact Redistricting Plans."

8. Imai, Kosuke and Zhichao Jiang. "Principal Fairness for Human and Algorithmic Decision-Making."

9. 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."

10. Eshima, Shusei, Kosuke Imai, and Tomoya Sasaki. "Keyword Assisted Topic Models."

11. Tarr, Alexander, June Hwang, and Kosuke Imai. "Automated Coding of Political Campaign Advertisement Videos: An Empirical Validation Study."

12. Chan, K.C.G, K. Imai, S.C.P. Yam, Z. Zhang. "Efficient Nonparametric Estimation of Causal Mediation Effects."

13. Barber, Michael and Kosuke Imai. "Estimating Neighborhood Effects on Turnout from Geocoded Voter Registration Records."

14. 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 and Co-Principal Investigator

1. National Science Foundation (2022-2025). "Collaborative Research: Understanding the Evolution of Political Campaign Advertisements over the Last Century." (Accountable Institutions and Behavior Program, SES–2148928). Principal Investigator (with Michael Crespin and Bryce Dietrich) $538,484.

2. 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.

3. 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.

4. 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.

5. 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

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

7. 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.

8. 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

9. 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.

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

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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
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). To be Postdoctoral Fellow, Princeton Unviersity, followed by Assistant Professor, Department of Political Science, Boston University

4. Ambarish Chattopadhyay (Statistics). To be Postdoctoral Fellow, Stanford University
5. Shusei Eshima (Government)
6. Georgina Evans (Government). To be Research Scientist, Google Brain
7. Dae Woong Ham (Statistics)
8. Zeyang Jia (Statistics)
9. Christopher T. Kenny (Government)
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

Former Students

1. Michael Lingzhe Li (Ph.D. in 2021, Operations Research, MIT). Postdoctoral Fellow, MIT. To be Assistant Professor, Technology and Operations Management Unit, Harvard Business School
2. Alexander Tarr (Ph.D. in 2021, Department of Electrical and Computer Engineering, Princeton University; Dissertation Committee Chair)
3. Connor Jerzak (Ph.D. in 2021, Department of Government, Harvard University). Postdoctoral Fellow, Linkoping University. To be Assistant Professor, Department of Government, University of Texas, Austin
4. 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

5. Erik Wang (Ph.D. in 2020, Department of Politics, Princeton University). Assistant Professor, Department of Political and Social Change, Australian National University
6. Diana Stanescu (Ph.D. in 2020, Department of Politics, Princeton University). Postdoctoral Fellow, Stanford University
7. Nicole Pashley (Ph.D. in 2020, Department of Statistics, Harvard University). Assistant Professor, Department of Statistics, Rutgers University
8. Asya Magazinnik (Ph.D. in 2020, Department of Politics, Princeton University). Assistant Professor, Department of Political Science, Massachusetts Institute of Technology
9. Max Goplerud (Ph.D. in 2020, Department of Government, Harvard University). Assistant Professor, Department of Political Science, University of Pittsburgh
10. Naoki Egami (Ph.D. in 2020, Department of Politics, Princeton University; Dissertation Committee Chair). Assistant Professor, Department of Political Science, Columbia University
11. Brandon de la Cuesta (Ph.D. in 2019, Department of Politics, Princeton University). Postdoctoral Fellow, Center on Global Poverty and Development, Stanford University
12. Yang-Yang Zhou (Ph.D. in 2019, Department of Politics, Princeton University). Assistant Professor, Department of Political Science, University of British Columbia
13. Winston Chou (Ph.D. in 2019, Department of Politics, Princeton University). Senior Data Scientist at Apple
14. 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
15. Benjamin Fifield (Ph.D. in 2018, Department of Politics, Princeton University; Dissertation Committee Chair). Data Scientist, American Civil Liberties Union
16. Tyler Pratt. (Ph.D. in 2018, Department of Politics, Princeton University). Assistant Professor, Department of Political Science, Yale University
17. Romain Ferrali (Ph.D. in 2018, Department of Politics, Princeton University). Assistant Professor, Aix-Marseille School of Economics
18. Julia Morse (Ph.D. in 2017, Woodrow Wilson School, Princeton University). Assistant Professor, Department of Political Science, University of California, Santa Barbara
19. Yuki Shiraito (Ph.D. in 2017, Department of Politics, Princeton University; Dissertation Committee Chair). Assistant Professor, Department of Political Science, University of Michigan
20. Carlos Velasco Rivera (Ph.D. in 2016, Department of Politics, Princeton University). Research Scientist, Facebook
21. 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

22. Graeme Blair (Ph.D. in 2016, Department of Politics, Princeton University). Assistant Professor, University of California, Los Angeles

23. Jaquilyn R. Waddell Boie (Ph.D. in 2015, Department of Politics, Princeton University). Private consultant

24. Scott Abramson (Ph.D. in 2014, Department of Politics, Princeton University). Associate Professor, Department of Political Science, University of Rochester

25. Michael Barber (Ph.D. in 2014, Department of Politics, Princeton University). Associate Professor, Department of Political Science, Brigham Young University

26. In Song Kim (Ph.D. in 2014, Department of Politics, Princeton University). Associate Professor, Department of Political Science, Massachusetts Institute of Technology

27. Alex Ruder (Ph.D. in 2014, Department of Politics, Princeton University). Principal Advisor, Federal Reserve Bank of Atlanta

28. Meredith Wilf (Ph.D. in 2014, Department of Politics, Princeton University). Senior Director, Capital Rx

29. Will Bullock. (Ph.D. candidate, Department of Politics, Princeton University). Senior Researcher, Facebook

30. Teppei Yamamoto (Ph.D. in 2011, Department of Politics, Princeton University; Dissertation Committee Chair). Associate Professor, Department of Political Science, Massachusetts Institute of Technology

31. Dustin Tingley (Ph.D. in 2010, Department of Politics, Princeton University). Professor, Department of Government, Harvard University

32. Aaron Strauss (Ph.D. in 2009, Department of Politics, Princeton University). Former Executive Director, Analyst Institute

33. Samir Soneji (Ph.D. in 2008, Office of Population Research, Princeton University; Dissertation Committee Chair). Associate Professor, Department of Health Behavior at the Gillings School of Global Public Health, University of North Carolina, Chapel Hill

34. 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). Associate 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). Associate 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). Associate Professor, Departments of Political Science and Statistics, University of California, Los Angeles
12. Florian Hollenbach (Predoctoral Fellow, 2013–2014). Associate Professor, Department of International Economics, Government and Business at the Copenhagen Business School
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

Mmeber, Senior Lecturer Search Committee (2021-2022)

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 Review Committee member, Department of Political Science, University of Rochester (2022)

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.

Expert Reports

1. Milligan *et al. v.* Merrill *et al.* United States District Court for the Northern District of Alabama, Case No. 2:2021cv01530

2. League of Women Voters of Ohio *et al. v.* Ohio Redistricting Commission *et al.* The Supreme Court of Ohio, Case No. 2021–1449

3. League of Women Voters of Ohio *et al.* *v.* Ohio Redistricting Commission *et al.* The Supreme Court of Ohio, Case No. 2021–1193

4. League of Women Voters of Ohio *et al.* *v.* Frank LaRose *et al.* The Supreme Court of Ohio, Case No. 2022–0303

5. The South Carolina State Conference of the NAACP, *et al.* *v.* McMaster, *et al.* United States District Court for the District of South Carolina Columbia Division, Case No. 3-21-cv-03302-JMC-TJH-RMG

6. Benninghoff *v.* 2021 Legislative Reapportionment Commission. The Supreme Court of Pennsylvania, Case No. 11 MM 2022

7. Graham *et al.* *v.* Adams *et al.* Commonwealth of Kentucky Franklin Circuit Court Division, Case No. 22-CI-00047

Exhibit 2

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF SOUTH CAROLINA
COLUMBIA DIVISION

THE SOUTH CAROLINA STATE CONFERENCE OF THE NAACP, et al.,

Plaintiffs,

vs. CASE NO. 3:21-cv-03302-MGL-TJH-RMG

THOMAS C. ALEXANDER, et al.,

Defendants.

DEPOSITION OF: KOSUKE IMAI, PhD (Via VTC)
DATE: August 8, 2022
TIME: 11:04 a.m.
LOCATION: Cambridge, MA
TAKEN BY: Counsel for the Senate Defendants
REPORTED BY: SOLANGE RUIZ-URIBE, Court Reporter
Via Videoteleconference

APPEARANCES OF COUNSEL:
ATTORNEYS FOR THE PLAINTIFFS
THE SOUTH CAROLINA STATE CONFERENCE OF THE NAACP, ET AL.:
AMERICAN CIVIL LIBERTIES UNION
BY: ADRIEL I. CEPEDA
VIA VIDEOTELECONFERENCE
1730 M Street, NW, Suite 1800
Washington, DC 20036
(202) 457-0800
acepedaderieux@aclu.org
AND
AMERICAN CIVIL LIBERTIES UNION
BY: PATRICIA YAN
VIA VIDEOTELECONFERENCE
1730 M Street, NW, Suite 1800
Washington, DC 20036
(202) 457-0800
pyan@aclu.org
AND
NAACP LEGAL DEFENSE AND EDUCATIONAL FUND
BY: JOHN S. CUSICK
VIA VIDEOTELECONFERENCE
40 Rector Street, 5th Floor
New York, NY 10006
(212)-965-2200
jcusick@naacpldf.org
ATTORNEYS FOR THE DEFENDANTS
SENATE DEFENDANTS:

JONES DAY
BY: JOHN M. GORE
VIA VIDEOTELECONFERENCE
51 Louisiana Avenue, N.W.
Washington, D.C. 20001
(202) 879-3939
jmgore@jonesday.com

(appearances continued)

(appearances continued)

ATTORNEYS FOR THE DEFENDANT
HOUSE DEFENDANTS:

NEXSEN PRUET
BY: ANDREW MATHIAS
VIA VIDEOTELECONFERENCE
104 S. Main Street, Suite 900
Greenville, SC 29601
(864) 282-1195
amathias@nexsenpruet.com

ATTORNEYS FOR THE DEFENDANT
ELECTION COMMISSION DEFENDANTS:

BURR & FORMAN, LLP
BY: MICHAEL R. BURCHSTEAD
VIA VIDEOTELECONFERENCE
1221 Main Street, Suite 1800
Columbia, SC 29201
(803) 799-9800
mburchstead@burr.com

AND

BURR & FORMAN, LLP
BY: M. ELIZABETH CRUM
VIA VIDEOTELECONFERENCE
1221 Main Street, Suite 1800
Columbia, SC 29201
(803) 799-9800
lcrum@burr.com

ALSO PRESENT:
SAMANTHA FORAN, ACLU Intern

(INDEX AT REAR OF TRANSCRIPT)

KOSUKE IMAI, PhD,
after first being duly sworn, testified as follows:
EXAMINATION
BY MR. GORE:
Q. Good morning, Dr. Imai.
A. Good morning.
Q. Thank you for being here today. For the record will you say your full name and spell your last name, please.
A. Kosuke Imai, it's spelled I-M-A-I.
Q. Dr. Imai, my name is John Gore. I'm with the law firm of Jones Day. I represent the Senate defendants, Senate President Thomas Alexander and Senate Judiciary Committee Chairman, Luke Rankin, and we are here today in the Congressional redistricting case in South Carolina.
Do you understand that you are appearing in connection with that case?
A. Yes, I do.
Q. The court reporter has just placed you under oath and do you understand that means you are under an obligation to tell the truth today?
A. Yes, I do.
Q. Is there any reason you are unable to testify truthfully today?

A. No.

Q. The court reporter is here to take down your testimony and to do so she'll need verbal answers; yes and no. She can't record nods or shakes of the head. Does that make sense?

A. Okay, yes. That makes sense.

Q. And are you being represented by counsel here today?

A. Yes.

Q. And who is that?

A. Mr. Cepeda.

Q. So Mr. Cepeda may from time to time object to some of my questions, things like objection to form, and he's obviously entitled to do that. But unless he instructs you not to answer, you should go ahead and answer all of my questions.

A. Okay.

Q. Do you understand that?

A. That makes sense, yes.

Q. And most important rule today, if you don't understand a question please ask me to clarify. Can we have that agreement?

A. Yes, I will.

Q. And if you answer a question I'm going to assume that you understood it; is that fair?

A. That's fair.

Q. And of course, if you need to take a break for any reason just please let me know and we can take a break. We'll take periodic breaks throughout the day. My only request is that if a question is pending, you answer the question before we take the break; does that make sense?

A. Yes, that makes sense.

Q. Okay. Great. How many times have you been deposed, Dr. Imai?

A. I think I had once before.

Q. Okay. And do you remember which case that was?

A. I think it was a South Carolina house case.

Q. Okay. Dr. Imai, where are you today?

A. I'm in Cambridge, Massachusetts.

Q. And is anyone in the room with you?

A. No.

Q. What did you do to prepare for today's deposition?

A. I reviewed my own report and exhibits that were sent by Mr. Cepeda and I also had conversations with Mr. Cepeda.

Q. How many conversations have you had with

Mr. Cepeda?

A. I think we had maybe four times, four or five times before this -- before today.

Q. When did those conversations take place?

A. Most recently it took place on Friday and Saturday of last week and before that we had a conversation, if I recall correctly, I think the last week or the last couple weeks of July.

Q. How long did each of those conversations last?

A. I'd say an hour to two hours, one hour to two hours.

Q. Was anyone else present for those conversations?

A. I think that Ms. Yan may have been present in not all -- not all instances but one or two times.

Q. And apart from Mr. Cepeda and Ms. Yan have you discussed today's deposition with anyone else?

A. No.

Q. Do you have any documents with you today?

A. I have my own expert report which was sent to me, just printed out a copy of it, a paper copy of it as well as some other materials that was sent by the -- you know, the exhibit materials that were

sent by counsel, the South Carolina House Representative Redistricting Guideline copy as well as Senate Guideline copy and two reports by Mr. Sean Trende that were also sent to me by counsel as part of the exhibit, you know, the exhibit folder that I received. I just printed them up just in case that I need to reference them.

Q. Do any of your documents have handwritten notes on them?

A. No.

Q. We touched on this a little bit before but which documents did you review in preparing for today's deposition?

A. So I reviewed my own report. I also reviewed the reports by Mr. Sean Trende, two reports that were sent to me as part of the exhibit. I also reviewed the redistricting guidelines, the House and Senate Redistricting Guidelines that was also sent to me as part of the exhibit materials. And I also reviewed, you know, other materials that were sent to me as part of the exhibit, some of them brief, very briefly.

Q. At any time have you reviewed any of the other plaintiffs' expert reports in this case?

A. Can you repeat the question again? Sorry.

Q. Certainly. At any time have you reviewed any of the other plaintiffs' expert reports in this case?

A. I don't recall reviewing other experts reports, yeah.

Q. So have you reviewed any expert reports by Dr. Baodong Liu?

A. No.

Q. Moon Duchin?

A. No.

Q. Dr. Bagley?

A. No.

Q. Or Dr. Ragusa?

A. No.

Q. Thank you. Dr. Imai, I'm going to introduce as the first exhibit the document that was sent to you as tab one in the zip folder, if you want to turn to that.

A. Yeah.

(Defendant's Exhibit No. 1, SENATE DEFENDANTS' AMENDED NOTICE OF TAKING VIDEO CONFERENCE DEPOSITION OF KOSUKE IMAI, PHD, was marked for identification.)

THE WITNESS: Okay. I have it opened.

BY MR. GORE:

Q. And I'm going to display this as well as in Exhibit Share --

A. Okay.

Q. For the benefit of counsel and our court reporter. Dr. Imai, do you recognize this document?

A. Yes.

Q. What document is this?

A. This is the announcement about notice of deposition.

Q. And do you recall receiving this document?

A. Yes, I do.

Q. Now turn with me, if you will, to what I believe is the seventh page or so of this document.

A. Seventh page.

Q. The page has Exhibit A at the top.

A. Yes.

Q. And the letter A, B and C. Do you see that?

A. Yes, I see that.

Q. And this is a subpoena for you to produce documents in connection with your testimony. What did you do to collect documents listed here on Exhibit A?

A. For Exhibit A I listed all the books and treatises and articles and publication that were

used in my, you know, to formulate my opinions in this case in my own expert report. And so I asked counsel to simply submit those for Exhibit B and C. I was struggling during the time when this request came and so I asked counsel to prepare these because all these items were with counsel and, you know, they were cc'd on email.

And then I, upon my return from my travel I checked my inbox and to make sure that no item is missing, I sent all the list of items that I have to counsel to cross-check. And I confirmed that, all the materials, relevant materials being submitted.

Q. Dr. Imai, did you receive any data from plaintiffs' counsel in this case --

A. Yes.

Q. To form your report?

A. Yes, I did.

Q. And what data did you receive?

A. All the data that were used in my analysis which are listed in my expert report I received from counsel.

Q. And other than the data listed in your expert report did you receive any other data from plaintiffs' counsel?

A. There may have been included in the data that were shared with me that, you know, for example, there may have been variables that I did not use, I did not end up using in my analysis, but I don't recall exactly what were, you know, in the data shared but not included in my analysis.

Q. And whether you used the data or not, what types of data did plaintiffs' counsel provide you?

A. Data mostly from the census data, the shared files and population few years and racial information that I used. And then in addition there were also data about incumbent residence location which I used for my analysis.

Q. Did counsel provide you with any election data?

A. I don't think so but again, you know, I did not use them at all in my analysis. I cannot 100 percent certain that no election data were included in the -- in the data that were sent to me but I certainly did not look at them.

Q. Did counsel provide you any assumptions to rely on in your expert report?

A. We discussed the assumptions that I used in my analysis during conversations we had in the course of preparing my draft report and -- but yeah,

that's what I received, some feedback.

Q. And what assumptions did counsel provide you that you relied on in your expert report?

MR. CEPEDA: Objection; mischaracterizes the record.

MR. GORE: You can answer.

THE WITNESS: The assumptions were not directly provided. We may have discussed them during our conversations of draft reports but all the assumptions that I ended up imposing in my final report are my own decisions.

BY MR. GORE:

Q. And are those assumptions disclosed in your expert report?

A. Yes. All the assumptions are written in the expert report.

Q. Okay.

A. In the form of algorithm constraints, just to be clear.

Q. Thank you. Dr. Imai, I want to talk a little bit about how you were hired for this case. Do you recall who first contacted you about this case?

A. I think it was Ms. Yan but I might be wrong. It could be Mr. Cepeda but I think one of

the two people contacted me first.

Q. And do you have a formal engagement letter with the ACLU or the NAACP Legal Defense Fund?

A. Yes, I do.

Q. And who is that with?

A. I think it was with the ACLU.

Q. And other than this case are there any other cases where you have been retained by the ACLU?

A. Yes. Oh, would you like me to list them?

Q. I'll ask follow-up questions. Can you tell me what those cases are?

A. Okay. So it should be listed in my expert report but they are Alabama Congressional case and Ohio State House Congressional case. Let's see. And the South Carolina State House case. Yeah, I think that's -- that's it.

Q. How much are you charging for your work in this case?

A. So it's a hourly charge of $450.

Q. Do you have a cap on your fees for this case?

A. There may have been a cap in the agreement that needs to be discussed if I reach that cap but I don't recall specifics.

Q. Did anyone assist you in preparing your expert report?

A. Yes.

Q. Who is that?

A. So I have research assistants and there are three research assistants that were involved in preparing for my expert report. One is Shiro Kuriwaki. Would you like me to spell that for you?

Q. Please.

A. Okay. Shiro is S-H-I-R-O. And Kuriwaki is K-U-R-I-W-A-K-I. And Tyler Sinko. Sinko is S-I-N-K-O. And Kevin Wang. K-E -- oh, Kevin you know. Wang is W-A-N-G.

Q. What did Shiro do to assist you on your report?

A. He helped me, under my guidance, run simulations, make figures, et cetera.

Q. What did Tyler do to assist you on your expert report?

A. The same.

Q. How about Kevin?

A. The same.

Q. Has anyone conducted a peer review of your expert report?

A. I don't think so. Would you -- do you

mean the academic peer reviews?
Q. Yes.
A. No, no.
Q. Dr. Imai, I'm next going to pull up your expert report.
A. Okay.
Q. Which I believe is tab two in your binder.
A. Yes, I got this opened.
(Defendant's Exhibit No. 2, EXPERT REPORT OF KOSUKE IMAI, PH.D. DATED APRIL 4, 2022, was marked for identification.)
BY MR. GORE:
Q. I'm introducing this now as Exhibit Two and loading it into Exhibit Share. Dr. Imai, do you recognize this document?
A. Yes.
Q. What is this document?
A. This is a report I wrote.
Q. Does this expert report contain all of your opinions related to the Congressional plan litigation?
A. Yes.
Q. Do you intend to offer any opinions at trial that are not contained in your report?
A. No.

Q. Do you intend to supplement or amend your report?

A. Not right now.

Q. Did you prepare a rebuttal report in this case?

A. No, I didn't.

Q. Why not?

A. I reviewed the report by Mr. Trende and -- but he did not use any simulation analysis and, you know, my expertise is about redistricting simulation so I decided to not offer any opinions on that particular report.

Q. Dr. Imai, I'm going to turn to page 31 of your report.

A. Okay, page 31.

Q. Which is a copy of your CV.

A. Oh, okay.

Q. But I understand this may no longer be your current CV. Is this still your current CV?

A. No. This was prepared when I submitted my report back in April.

Q. And I'm going to pull up what I believe is your updated CV which unfortunately is not included in your zip file.

A. Oh, okay.

Q. But I will pull it up on Exhibit Share and I can share my screen with you if you can see it. Do you have access to Exhibit Share?

A. I don't have -- that's the software that --

Q. Yeah, it's part --

A. No.

Q. A browser.

A. Okay. I don't have that right now.

Q. Now, Dr. Imai, I have tried to share my screen with you.

A. Yeah.

Q. And it's small on your monitor. Can you see that?

A. I can see that. It's big now.

Q. And this I believe is your -- this is -- says Kosuke Imai CV June 2022; is that correct?

A. That's correct.

Q. And do you recognize this document?

A. Yes. I shared this with counsel.

Q. And what is this?

A. My CV.

Q. Is it your current CV?

A. Well, it's CV as of June, you know, as of June when I received that request.

Q. I've marked this as Exhibit Three for the record.

(Defendant's Exhibit No. 3, JUNE 2022 CURRICULUM VITAE OF KOSUKE IMAI, was marked for identification.)

BY MR. GORE:

Q. Dr. Imai, can you briefly describe your educational background.

A. Yes. I received bachelor of arts in liberal arts from University of Tokyo in 1998. I received a master's degree in statistics from Harvard in 2002 and I received PhD in political science from also Harvard in 2003.

Q. And as part of your education did you take any classes in legislative redistricting?

A. Oh, there is no such a class back then that I was aware of so I did not take it.

Q. And when did redistricting become part of your studies or scholarship?

A. I think it was around 2010.

Q. And how did you become interested in redistricting as a topic of scholarship?

A. You know, I was interested in methodology that's used to evaluate redistricting trends because my expertise is political methodology which is

greater study of algorithms and statistical methods for understanding political science. So this was one of the important topics in the field.

Q. And what is the focus of your legislative redistricting scholarship?

A. There are -- there are two components. One is the development and application of algorithms to evaluate the redistricting trends. And the second part is the method development for ecological inference which basically trying to inquire individual behavior from aggregate data.

Q. Have you written any peer-reviewed publications about South Carolina?

A. No.

Q. Have you submitted any works for publication that were not published?

A. Can you repeat that question again?

Q. Sure. During your career --

A. Uh-huh.

Q. Have you submitted any articles, books or other written works for peer review that were rejected for publication and not published?

A. Oh, I see. Yes, yes.

Q. How many, how many of those?

A. Oh, you can count, you know, number of

publications I have but it's hard to count the number of rejections. Many.

Q. Any about redistricting?

A. Oh, yes.

Q. And what do you recall about any of those publications and why they were rejected?

A. I don't recall specifics of reasons for rejections, you know. It ranges from, you know, complaints about the algorithm itself or complaints about -- or criticism, I shouldn't say complaints -- criticism about applications, criticism about literature reviewed to, you know -- it ranges by one review to another. So I don't recall all the specifics.

Q. And were any of these draft articles about simulation analysis?

A. Oh, yes.

Q. And what do you recall about criticisms of those articles?

A. So for example, the paper I eventually published in Journal of Computational and Graphical Statistics, during that review process of that paper, there were criticism about how, you know, the full particular performance of the algorithm I was studying at that time. And there may have been also

criticisms about the interpretation of the empirical applications of that particular algorithm.

Q. And with respect to these draft articles, did any of them discuss or relate to the simulation analysis you used for your expert report in this case?

A. They are related but they are not the ones that I used for this analysis.

Q. What were the simulation analyses in your draft articles that were rejected for publication? Can you explain to me what those were?

A. Can you repeat that question again?

Q. Sure. I'd just like to understand the differences between those simulation analyses and the one you did in this report. So can you explain to me the differences or explain to me what those analyses were and I'll follow up by asking about your analysis in this case.

A. Okay. Just to be clear, rejection doesn't necessarily mean that the paper will never be published. So if -- it's an iterative process, so often it's very rare for paper in academic journals to be accepted when you submit and often it will be rejected and sometimes you will be asked to revise, to address the criticism and resubmit.

And eventually the paper may get accepted or the paper may be submitted to a different journal and then again through the review process you -- after it goes through the process eventually, you know, the paper gets accepted.

So it's a long process so the fact that the paper is rejected doesn't necessarily mean it will never be published, just to make sure that's understood.

And the difference between algorithm that I used which is the merge-split type algorithm that's explained in my expert report is different from the algorithms that were developed in the articles that I have. They are both, all these algorithms are part of the Monte Carlo methods, it's a broad class of mathematical algorithms that try to plan and regenerate representative sample from the particular populations. It's all related, part of the Monte Carlo family.

But the algorithm that I used which is the merge-split, broadly speaking, basically is a Markov chain Monte Carlo algorithm where you basically randomly choose two districts, adjacent districts, and you launch them and then you run the resplit in a specific way and then you retreat. So

that you start with a particular plan might be an enacted plan and then you'll randomly merge two adjacent districts and then randomly split them and that way you start modifying the plan.

The algorithm that I developed in my own article, there are two types. One is called Frep, which basically start modifying the boundaries of particular units by swapping certain units, say precincts, from one district to another in a specific way and then proceed.

The other one is called sequential Monte Carlo which basically starts from scratch so it's a blank slate and it starts creating one district at a time.

So these are the different types of algorithms that are all part of the family of Monte Carlo methods and for this case I used the more split-type algorithm.

Q. Thank you for all that. I'm going to scroll down on your CV to what I believe is the last page. It appears to be buffering.

A. Yeah.

Q. Let me do this a different way.

A. Yeah.

Q. Let's go back to your expert report if we

can.

A. Okay.

Q. Which was previously marked as Exhibit Two.

A. Right. I have Two, right.

Q. Yes.

A. Okay.

MR. CEPEDA: Tab two is the CV.

MR. GORE: I think tab one was the notice of deposition. Tab two is the report which contained an old version of the CV.

MR. CEPEDA: Yeah.

MR. GORE: And I've marked the new version of the CV as Exhibit Three, although it wasn't a tab. I'm trying to pull this back up if I can. That's back in Exhibit Share. Exhibit Two.

BY MR. GORE:

Q. And I want to go, Dr. Imai, if we might, to paragraph -- let's start with paragraph 13 on page 6.

A. Oh, so this is tab two?

Q. Yes, sir.

A. Thirteen. Okay page 13 of which paragraph?

Q. It's -- sorry, paragraph 13.

A. Oh, 13 you said.
Q. Yeah.
A. Thirteen, page 6. Okay. I have it here.
Q. Paragraph 13 mentions that your methodology for predicting individuals' race was used in a recent decision of the Second Circuit?
A. That's right.
Q. Were you an expert witness in that case?
A. No.
Q. What is your methodology for predicting individuals' race using voter files and census data?
A. So I focus on developing statistical methodology to use voter file information, information from voter files such as name and address of registered voters and tried to impute the race which is used in some of the ecological inference analysis.
Q. Is that the same as Bayesian Improved Surname Geocoding?
A. That's correct.
Q. Thank you. All right. Now, let's look at paragraph 14.
A. Okay.
Q. And I want to ask you if this is a complete list of all cases in which you have been an

expert witness?

A. So I have filed expert report in July, maybe it's three or four weeks ago for Jacksonville, Florida case which was not included when this CV was submitted back in March or April.

Q. And what -- is the Jacksonville case also a redistricting case?

A. Yes. It's a city council redistricting case.

Q. Other than the Jacksonville case does paragraph 14 list all cases in which you have appeared as an expert or submitted an expert report?

A. I think so. Except Ohio case has evolved and I'm involved in both state legislative case and Congressional case and sometimes I'm not 100 percent sure what the name of the case is. I think that has changed. But either way, I submitted reports to those two cases.

Q. Are all the cases in which you have been an expert redistricting cases?

A. Yes, these are all redistricting cases.

Q. And have you ever testified or submitted a report as an expert in any non-redistricting cases?

A. No.

Q. Have you ever been an expert on behalf of

a defendant in a redistricting case?

A. Yes.

Q. Which case was that?

A. I think the Supreme Court of Pennsylvania case, Pennsylvania State House redistricting case is -- I think it was a defendant because I testified as expert witness for the legislative, the apportionment commission.

Q. Will you identify all the cases in this paragraph that are racial gerrymandering cases.

A. Okay. So Alabama case is a racial gerrymandering case, Congressional and also South Carolina House, you know, South Carolina House, State House case is also racial gerrymandering case.

Q. Are the rest of the cases here in paragraph 14 partisan gerrymandering cases?

A. That's right. I think so.

Q. How about Jacksonville?

A. Oh, Jacksonville is a racial gerrymandering case. Thank you.

Q. So this case, Graham versus Adams in Kentucky, is a partisan gerrymandering case; is that right?

A. That's correct.

Q. Dr. Imai, are you a map drawer?

A. No.

Q. Have you ever been qualified as an expert in map drawing?

A. No.

Q. Have you ever been disqualified as an expert witness?

A. No, not that I know.

Q. And in any case, have some or all of your expert opinions been found inadmissible by a court?

A. Not that I know of.

Q. All right. I'd like to focus now on the summary of opinions which starts on the page 4. It's paragraph six to seven.

A. Okay. Hold on. Page 4, okay.

Q. And I believe your report analyzes the South Carolina Congressional map enacted earlier there year; is that correct?

A. That's correct.

Q. And I will generally refer to that map as the enacted plan; does that make sense to you?

A. That's how I refer to it so that works for me.

Q. Great. Does your report examine whether the enacted plan intentionally discriminates on the basis of race?

A. Not the part intentional, no.

Q. And so your report draws no conclusions about intent; is that right?

A. Right. That's correct. No opinion on intent.

Q. And your report -- does your report conclude that race predominated over traditional districting criteria in the enacted plan?

A. No.

Q. Why not?

A. This is a statistical analysis and a statistical evaluation so my analysis only addresses whether race played a significant role in determining the district boundaries in the enacted plan, not the legal conclusion suggested by the phrase, the predominant.

Q. Paragraph seven, starting at the sentence that carries over from page 4 to 5.

A. Uh-huh.

Q. Uses the word cracks and the word cracking. Do you see that?

A. Yes.

Q. How do you define cracking as that term is use in your report?

A. It's a simple definition that refers to

splitting particular geographical area and that's the definition. Splitting into multiple districts.

Q. So splitting a county, splitting any county in South Carolina, does that constitute cracking under your definition?

A. I wouldn't call it cracking a county. Usually I use cracking as particular group of orders for living, you know, specific geography, so not so much about administrative boundaries themselves.

Q. Does cracking occur any time a group of say black voters is divided into more than one district?

A. You may call that cracking. I think in my report you can, you know, think of cracking as just a splitting a group of voters who live in, you know, certain geographical area. It has no legal or any other meaning.

Q. So what does -- in your report is there a particular test for when cracking occurs or is it simply the splitting of voters, like you just said?

A. Splitting voters, you can test whether that's unusual. So you can test, use the statistical test, to see if particular splitting or cracking is unusual related to the simulated plans.

Q. Did you conduct an effectiveness analysis

on the enacted plan or in any of your simulation plans?

A. No, I did not.

Q. And did you analyze any election data in your report?

A. No, no.

Q. And did you analyze the political effect of any plan?

A. Would you clarify what you mean by political effect?

Q. Sure. Well, for example, did you consider whether districts were likely to elect a Republican or a Democrat?

A. Oh, no.

Q. And did you consider partisan performance in any district?

A. No.

Q. Now we talked a little bit before about the methodology you used, I believe you gave some detail on. And I want to understand. Let's turn to paragraph 17 on page 7 of your report.

A. Okay, hold on. Yes.

Q. I believe paragraph 17 says that you set to generate a representative sample of all possible plans that satisfy a specified set of criteria; is

that right?

A. That's correct.

Q. And how does your methodology ensure a representative sample?

A. Right. So this is -- the algorithm that I use is merge-split type algorithm and it is a type of Markov chain Monte Carlo algorithm and this algorithm is a target distribution which is specified by, in this case, specifying the set of constraints that are imposed in the algorithms and once those constraints are imposed the algorithm has a mathematical property that generates a representative sample from that distribution that you specified.

Q. And is there anything that you do in your methodology or through the algorithm to guarantee that the set of simulation plans you end up with is in fact a representative sample?

A. So the guarantee comes from mathematical property. So there is a mathematical theorem that gives mathematical guarantee. And in addition, we use standard diagnostics to ensure that the samples that we generate is following what in the field described as the best practice of using this type of algorithm.

Q. All right. Dr. Imai, I'm going to pull up what is Exhibit Four.

(Defendant's Exhibit No. 4, SEQUENTIAL MONTE CARLO FOR SAMPLING BALANCED AND COMPACT REDISTRICTING PLANS PAPER, was marked for identification.)

BY MR. GORE:

Q. Tab four in your binder and it will also be numbered as Exhibit Four.

A. Okay, tab four. Yes.

Q. Now I uploaded it in Exhibit Share as well. Dr. Imai, do you recognize this document?

A. Yes.

Q. What is this document?

A. This is a paper I've written on redistricting simulation.

Q. Has this paper been published?

A. Not yet.

Q. But you hope it will, right?

A. Yes.

Q. So I'd like to just walk through a couple of points that you make in this paper. So in the abstract --

A. Sure.

Q. Is that -- well, let me back up.

So this paper is about sequential

Monte Carlo method, right?

A. That's correct.

Q. And we talked about that earlier, that the sequential method starts from a blank slate.

A. That's right.

Q. Whereas the Markov chain method starts with an existing plan; is that correct?

A. That's correct.

Q. Are there any other differences between sequential Monte Carlo method and the Markov chain Monte Carlo method?

A. I mean, there are a lot of differences that, you know, in this application. Just so the most basic difference is this nature that sequential Monte Carlo starts from a blank slate and start building one district at a time. And then once you generate, you know, one -- so you basically generate multiple plans in parallel.

Whereas, Markov chain Monte Carlo will generate multiple plans, not in parallel, but in actually sequence. So it start generating multiple plans by modifying a plan, you know, sequentially. So even though sequential part is used for the sequential Monte Carlo, Markov chain is more independent. So it starts from the existing

plan and then generating different -- start generating different plans by modifying it.

Whereas the SMC is really about starting from a blank slate and start building one district at a time.

Q. Thank you. So I want to start with the third sentence in the abstract of this paper.

A. Okay.

Q. And I'm just going to read that out loud. It says: For successful application sampling methods must scale to large maps with many districts, incorporate realistic legal constraints and accurately and efficiently sample from a selected target distribution. Unfortunately, most existing methods struggle in at least one of these areas.

So my first question, Dr. Imai, did I read that correctly?

A. That's correct.

Q. Do you agree that simulation analysis must incorporate realistic legal constraints?

A. I agree.

Q. And the next sentence says that: Most existing methods struggle in at least one of these areas.

A. Uh-huh.

Q. And are the existing methods referred to in this sentence, Markov chain Monte Carlo methods?

A. Included is Markov chain Monte Carlo algorithm. That's not the only one, but yes.

Q. What other methods are included here?

A. There are enumeration algorithm, which simply enumerate all possible, you know, redistricting plans under certain set of constraints.

There are also something called constructive Monte Carlo methods which often start with a seed for each district. So you pick -- if you have six districts you pick six precincts, for example, and then start growing a district from each of those seed precincts.

So there are different type of algorithms out there other than the SMC and MCMC. And that sentence, the existing methods include all that.

Q. And I'm going to skip down a couple of sentences to a sentence that starts, because.

Because it draws many plans in parallel, the SMC algorithm can efficiently explore the relevant space of redistricting plans better

than the existing Markov chain Monte Carlo, MCMC algorithms that generate plans sequentially.

Did I read that correctly?

A. You did.

Q. And so is it your view that SMC is better than the Markov chain Monte Carlo algorithm?

A. So I don't want to make general statement like that because even though, you know, in the abstract that's one way to summarize how the sequential Monte Carlo improves upon the existing methods.

Depending on different cases, especially the purpose of analysis, you know, certain algorithm may do just fine and, you know, certain algorithm may be more suitable than other algorithms. So I don't want you to think of this statement applies generally to every single, you know, redistricting analysis algorithms on that.

Q. I'm going to scroll down to the next page which I think is page 2 of the document or the PDF, but it's page 1 in the footer. It's the page after the abstract.

A. Oh, okay. Okay.

Q. And then it's the fourth paragraph down, starts, MC algorithms.

A. Yes.

Q. MCMC algorithms, citation, can in theory sample from a specific target distribution and incorporate constraints through the use of an energy function. In practice, however, existing algorithms struggle to mix and traverse through a highly complex sampling space, making scale a bit difficult and accuracy hard to prove. Some of these algorithms make proposals by flipping precincts at the boundary of existing districts, rendering it difficult or even impossible to transition between points in the state space, especially as more constraints are imposed.

Did I read that correctly?

A. Yes, you did.

Q. And is that a valid criticism of MCM algorithms?

A. I think it's a valid general criticism of the MCMC algorithm but I would like to emphasize that depending on the specific use case, this particular, you know, criticism may not apply. So it should be evaluated based on the specific use case.

Q. And is that point made in this paragraph of the draft or elsewhere in the draft?

A. I don't recall all the sentences I've written but, you know, this is a general point that's being made, not a criticism of specific applications per se. So I think for any statistical analysis if you are criticizing a particular application of it, you would want to consider the specific factors that are relevant for that case.

You know, this is general weakness of the MCMC algorithms that are acknowledged in the field but that doesn't mean that, you know, that the application of it, all the applications of it are automatically invalid and it needs to be -- the validity of the application needs to be evaluated case by case.

Q. Dr. Imai, in light of these criticisms of MCMC algorithms, why did you use that approach here instead of SMC in your report?

A. Okay. There are several reasons. First, I'd like to point out that in the race-blind simulations that I have conducted, so this is referring to the first race-blind simulation which basically generates the simulated boundaries between districts one and six.

And then the second race-blind simulation which examines the District 1 within the

Charleston County, okay. So in these two race-blind simulations, because there is only two districts are in question, essentially the SMC and MCMC are very, very similar.

So if you recall, merge-split algorithm, they merge two districts and then spread them. And SMC starts from blank slate and then generate one district at a time.

In these cases where you only have two districts starting from the blank slate and merging and then splitting is essentially the same. So they are very, very similar in terms of the -- how the algorithm -- how the algorithm, you know, progresses.

And for the statewide analysis where I analyze all districts at a time, there are important differences, how the SMC handles that case as well as MCMC handles that case. But there I wanted to impose the constraint that the District 6 has this VRA percentage, BVAP percentage, between 45 and 50 percent, and so in order to specify, to tell the algorithm that District 6 should have that percentage, it's better to use MCMC algorithm where you start with existing primary states enacted plan, which has a District 6 and then generate the new

districts in sequence.

Whereas, if you start from a blank slate, it's a little bit more difficult to tell the algorithm District 6 should have BVAP proportion or percentage. So it makes sense for that statewide analysis to use the merge-split type of algorithm and for the sake of consistency, I use the merge-split algorithm in the race-blind simulation, whereas, the rest difference between the use of SMC and merge-split because there are only two districts that are involved.

Q. Could you have used the SMC algorithm for your statewide simulations?

A. It would have taken, you know, different type of constraints. I haven't -- you could try it. I haven't tried that.

Q. Is there any technical reason why that would not be possible?

A. I wouldn't say it's impossible because I haven't tried it, but as I explained, SMC starts from the blank slate and they build the district in sequence, one by one. And so to impose the VRA constraints, which is basically pairing the algorithm that the particular district should have certain range of BVAP proportion, it's much easier

to specify that constraint in the MSMC where you start with existing plan that has District 6.

Q. Dr. Imai, I'd like to go down to page -- it's page 3 of the draft article, it's the fourth page.

A. Okay.

Q. In what I have sent you since the abstract page is not numbered.

A. Okay.

Q. I just want to read some of this to you. So the second-to-last paragraph, the last sentence of that paragraph starts, however.

However, these methodological debates are also relevant for other cases where simulation algorithms have been extensively used by expert witnesses, citation, and highlight the difficulties in practically applying existing sampling algorithms to actual redistricting problems.

Did I read that correctly?

A. Yes.

Q. And do you stand by that statement about existing sampling algorithms?

A. I stand by it but again, I have not examined each application, you know, one by one, so I cannot say anything about this specific, you know,

case.

Q. And then I just want to read a couple more sentences. It says: First, the distributions that some of these algorithms sample from are not made explicit, leaving open the possibility that the generated ensemble is systematically different from the true set of all valid plans. Second, even when the distribution is known, MCMC algorithms used to sample from it may be prohibitively slow to mix and cannot yield a representative sample.

Did I read that correctly?

A. Yes, you did.

Q. So in certain cases MCMC algorithms don't yield representative samples; is that correct?

A. That's correct.

Q. In fact, in other cases or maybe even the same cases MCMC algorithms can generate ensemble plans or simulations that are systematically different from the true set of all valid plans; is that right?

A. That's certainly a possibility.

Q. I'd like to go back to the abstract page and ask you about the first footnote which is a star footnote where you thank a few people for input on the draft.

A. Yes.

Q. The first person is Ben Fifield. Do you see that?

A. Yes, I do.

Q. And who is Mr. Fifield?

A. He was ACLU staff but more importantly, he was my former student from Princeton when I taught there and he was also a collaborator on some of these papers, not this one, but some of these redistricting algorithm papers.

Q. And approximately how many papers have you collaborated with Mr. Fifield?

A. Oh, I think two, at least two papers that were published already and the software that I developed, he was also involved in that development at the earlier years, way back, you know, several years ago, I think, or maybe ten even.

Q. And do you know where Mr. Fifield works now?

A. I -- my understanding is that he left the ACLU and went to Facebook but I'm not 100 percent sure.

Q. Have you discussed this case with him?

A. Not about this case.

Q. How about your expert report, have you

discussed that with him?

A. No. Oh, just one clarification. So he was part of the -- some of the conversation I had with counsel. So I didn't discuss with him, you know, just with him alone about this case but he was present when I had a conversation about draft reports with Mr. Cepeda and other lawyers so...

Q. Did you have any conversations with Mr. Fifield about any draft maps or plans for South Carolina Congressional redistricting?

A. I don't think so.

Q. Did you receive any data from Mr. Fifield?

A. I did receive the data from Mr. Fifield through the email that -- where the -- my counsel was also copied and through the SharePoint side that they provided.

Q. Other than the data that was provided with counsel copied did you receive any other data from Mr. Fifield?

A. No.

Q. Did Mr. Fifield provide you any assumptions in this case?

A. No.

Q. I'm going to take down this exhibit for now and if I can figure out how to do it I'd like to

bring up your report again.

A. Okay.

MR. GORE: Adriel, can you see that in Exhibit Share?

MR. CEPEDA: No. I didn't have access when I tried. I'm sorry, I can't answer your question at the moment.

MR. GORE: I just want to make sure you're comfortable and you're able to follow along with all these.

MR. CEPEDA: I'm following along with the tabs so as long as you are saying where you are looking at, I'm there.

BY MR. GORE:

Q. Okay. So back to the expert report which is tab two.

A. Yes.

Q. And has previously been marked as Exhibit Two.

A. Right.

Q. I would ask you some questions about this but before I do so, Dr. Imai, did you run any simulation analysis on the benchmark plan?

A. I may have run it in the draft expert reports that were included in the draft expert

reports.

Q. And what was the result of that simulation analysis?

MR. CEPEDA: John, I'm going to object to the extent that the question calls for the substance of a draft under -- that would be protected under Rule 26B.

MR. GORE: Okay. Well, he said he ran the analysis but didn't report it. I would like to know why it's not in the report and if there is a reason for that.

MR. CEPEDA: Well, you are asking about a draft report which the rules explicitly protect.

MR. GORE: Well, I didn't ask about the substance of the draft report, I asked about the outcome of the analysis that your expert says he ran. We had this with Dr. Lieu last week. He had run some analyses that weren't in his report and he told me what the results of those analyses were.

MR. CEPEDA: All right.

BY MR. GORE:

Q. So I'm just going to ask you, Dr. Imai, what was the result of the simulation analysis you ran on the benchmark plan?

A. I don't recall the specific results.

Q. Did you get the same result from the simulation analysis on the benchmark plan that you got from the simulation analysis on the enacted plan?

A. Oh, it will be different. I mean, once you add different constraint it will be a different result.

Q. And did your analysis show that the benchmark plan was an outlier with respect to cracking of black voters?

MR. CEPEDA: Same objection as before.

BY MR. GORE:

Q. You can go ahead and answer.

A. I didn't evaluate benchmark plan. I may have used benchmark plan as parts of algorithm but I didn't evaluate directly the benchmark plan.

Q. So what analysis did you do on the benchmark plan?

MR. CEPEDA: Same objection.

BY MR. GORE:

Q. You can go ahead and answer.

A. I was asked by lawyers to -- by counsel to incorporate the benchmark plan in the simulation, not to evaluate the benchmark plan because my analysis focuses on the evaluation of enacted plan

so...

Q. Well, let me ask the question this way, did you run your Markov chain Monte Carlo simulation on the benchmark plan?

A. Can you clarify what you mean by, on the benchmark plan? I don't quite understand.

Q. Okay. Well, I read your report to say that you ran a Markov chain Monte Carlo simulation on the enacted plan or in relation to the enacted plan.

A. That's not quite correct.

Q. Okay.

A. Because when you run the simulation algorithm the enacted plan is not directly used.

Q. Okay.

A. You generate -- not directly used, right. You generate the simulated plan under a set of constraints. And then once you generate those simulated plans the next step is to compare those simulated plans with the enacted plan. And that's where enacted plan comes in.

Q. Okay. So thank you for that clarification. So the simulation plans take, as I understand it, the 2020 census data, for example, and other constraints which we'll talk about that

you programmed the algorithm to consider, right?

A. That's correct.

Q. And that generates a set of simulation plans; is that correct?

A. That's correct.

Q. And then you compare that to the enacted plan; is that correct?

A. That's right, that's right.

Q. Okay.

A. So that's where -- so the actual generation of simulated plans does not directly involve the enacted plan. In my -- the one that I, you know, the analysis that's presented in my report.

Q. Did you conduct any simulation using 2020 or 2010 census data which was the data used for the benchmark plan?

MR. CEPEDA: Same objection.

BY MR. GORE:

Q. You can answer.

A. So in my draft report I may have included the simulation results that you used the benchmark plan.

Q. And did you then compare those simulation results to the benchmark plan?

MR. CEPEDA: Same objection.
BY MR. GORE:
Q. You can answer.
A. I -- I don't compare with the benchmark plan because my goal -- I didn't because my goal is not to evaluate the benchmark plan.
Q. So when you --
A. I evaluated the enacted plan.
Q. I've got it. But when you ran the simulation set using the 2010 census data did you then compare that simulation set to the enacted plan?
MR. CEPEDA: Same objection.
BY MR. GORE:
Q. You can answer.
A. Can you repeat that question again? I'm not understanding.
Q. Certainly. So as I understand it, your analysis really has two steps. You generate a set of simulation plans and then you compare that set to the enacted plan and I believe what's in your report; is that correct?
A. That's in my report, yes, my final report.
Q. Okay. And I think in your report we just discussed that you generated some simulation sets

using 2020 census data --

A. That's right.

Q. Plus various constraints which we'll talk about later. And then you compared that simulation set to the enacted plan?

A. That's right.

Q. Okay. I just asked you if you ran the simulation set using 2010 census data.

A. I don't think so. The census data I've seen that, you know, the census data I used has been always 2020.

Q. Okay.

A. That's the data that matter.

Q. Okay. So did you -- let me ask the question again because I think the record may be unclear on this point.

Did you ever run a set for this case, this report, did you ever run a simulation set using 2010 census data?

MR. CEPEDA: Same objection.

BY MR. GORE:

Q. You can answer.

A. I mean, I may have run the simulation analysis that used the benchmark plan so to that extent, if you think of that as using the 2010

census data, then the answer is yes. But the population data that I used always come from 2020 census because that's the data that matters currently.

Q. So the simulations that you ran using the benchmark plan, how did those work?

MR. CEPEDA: Same objection.

BY MR. GORE:

Q. You can answer.

A. So as part of the, you know, draft expert report, I may have included a simulation analysis that, you know, imposes some similarity to the benchmark plan.

Q. Would that --

A. As additional constraint.

Q. Okay. And was that constraint an attempt to control for preservation of cores of the benchmark district?

MR. CEPEDA: Same objection.

BY MR. GORE:

Q. You can answer.

A. I was asked by counsel to perform that analysis and I normally don't, you know, use any plan including the previous plan in my redistricting analysis because, you know, it will contaminate

the -- it basically prevents me from isolating the role that race play and redrawing the boundaries of the districts in the enacted plan when you include some information, include any plan that for all the factors that affected it.

So that's the reason why I don't in general use the previous plans or any other plan to construct my -- directly construct my algorithm constraints.

But I recall that the draft expert reports may have included analysis that put some constraints with respect to the benchmark plan upon the request of counsel.

Q. Okay. And was that an effort to control for the shape and location of the benchmark districts and how that might have affected the enacted plan?

MR. CEPEDA: Same objection and also now calling for attorney-expert communication.

MR. GORE: Are you instructing him not to answer?

MR. CEPEDA: I'll let it but I'm lodging the objection.

MR. GORE: Thank you.

BY MR. GORE:

Q. Go ahead, Dr. Imai, you may answer.

A. I don't -- because it was a suggestion by counsel, I don't know exactly what the, you know, real goal of that particular analysis. But from my point of view that because it would prevent me from isolating the role the race plays in redistricting plan, I don't directly incorporate any plan in my algorithm constraints including the previous, you know, benchmark plan. And that's why in the final report it's not there.

Q. And you include other constraints in your simulation set, correct?

A. That's correct.

Q. And those other constraints such as contiguity or incumbency pairings don't prevent you from isolating the role of race, correct?

A. That's correct to the extent that those factors are controlled, right. Like I don't want to consider noncontiguous districts or the districts that have a large population deviation.

Q. So --

A. I would want to exclude those.

Q. Why does including or why did including the benchmark plan-related constraints prevent you from isolating the role that race played in the

enacted plan?

MR. CEPEDA: Object to form.

BY MR. GORE:

Q. You can answer.

A. Yeah, so the reason is that benchmark plan may be based on a set of unknown factors that I just don't know what factors went into create the benchmark plan and one of those factors could be race or it could be some other factors. And if I really want to isolate the role of the race I know what I include is unknown, you know, set of factors that may influence my results. Whereas, the population and contiguity, I know exactly what they are so I impose them.

Q. Did you --

A. I don't want noncontiguous districts.

Q. And do you have any reason to believe that the benchmark plan violated traditional districting principles?

MR. CEPEDA: Object to the form.

THE WITNESS: I don't evaluate benchmark plans so I don't form any opinion on benchmark plan.

BY MR. GORE:

Q. How is the benchmark plan-related constraint different from the other constraints you

imposed in your report?

A. As I said, the benchmark plan is perhaps created based on a number of factors that I don't know what they are because I don't evaluate the benchmark plan. My goal has been to evaluate the enacted plan so if I directly impose the benchmark plan in my algorithm as a constraint I would be inheriting a number of unknown factors that influence the benchmark plan which would prevent me from drawing any conclusions about the role race played in drawing the district in the enacted plan.

Q. Well, what if the map drawer had started with the benchmark plan when drawing the enacted plan, wouldn't that affect the set of plans from which the map drawer was drawing?

A. How the map drawer drew the enacted plan doesn't affect my analysis. My analysis is completely independent of exact intention or anything that the map drawer has had. The goal is just to only evaluate the enacted plan that resulted in the process.

Q. So it doesn't matter to your analysis whether the map drawer drew the enacted plan based on the benchmark plan; is that right?

A. I did not evaluate whether the map drawer,

you know, what information the map drawer used to draw enacted plan. My evaluation is really just about the final enacted plan, the characteristics of it and then specifically the role race played in the final product. And that's really the goal. It's just not my goal to understand how map drawer drew the enacted plan.

Q. So after you generated the simulation set that included the benchmark plan-related constraints, did you compare that set to the enacted plan?

MR. CEPEDA: Object to form; mischaracterizes testimony and asking for questions about the draft protected by Rule 26.

MR. GORE: Again, I'm not asking about the draft, I'm asking about the analysis.

BY MR. GORE:

Q. So Dr. Imai, I believe you can answer.

A. In the draft expert report the results may have been included but I don't recall specifics.

Q. And no such results are included in your final report; is that correct?

A. What do you mean by such results?

Q. You don't, your expert report doesn't contain a simulation set with a benchmark

plan-related constraint, right?

A. That's correct.

Q. Do you recall what the benchmark plan-related constraint was in the analysis that you ran that included that constraint?

MR. CEPEDA: Same objection as above.

THE WITNESS: There are several ways to do that but I don't recall which one I used.

BY MR. GORE:

Q. What are some of the possible ways to do that?

A. Like one possible way is to look at the geographical overlap. Another possible way is to look at the population overlap, although one has to be very careful about this because people have moved over the past ten years. So the fact that there is a number of people in a certain location may not necessarily, you know, reflect the people who've been living there for the last decade.

Q. Okay. So other than geographical overlap and population overlap, are there any other ways to control for the benchmark plan or any other benchmark plan-related constraints that would be possible in your simulation model?

A. So in my simulation model the incumbency

regions restraint is imposed so that also partially, you know, controls for the overlap, if you -- if you make sure that incumbency won't be paired in each district that will to some extent have some implications on the overlap with the benchmark plan.

In my race-blind analysis I basically only look at two districts so to the extent that the rest of the districts are untouched, they are the same as, you know, benchmark plan. And in my second race-blind simulation I freeze even the District One Six boundary outside of Charleston County, so to the extent that, you know, that part is following the boundary of -- under the benchmark plan or the enacted plan. The enacted plan is similar to the benchmark plan.

You know, there are implications in the overlap but I don't want to use the benchmark plan directly in my constraint because it inherits all sorts of factors that I don't know and basically my analysis.

Q. Okay. Is it possible that some of those factors the benchmark plan inherits are not related to race?

A. It's possible. Again, I didn't evaluate the benchmark plan, that was never the goal of my

analysis so I don't know what factors went into the benchmark plan and precisely for that reason I don't want to include it.

Q. And my final question, what -- on this for now, what was the benchmark plan-related constraint in your simulation analysis that you ran?

MR. CEPEDA: Object to form and object because it asks the question about the draft protected under Rule 26.

MR. GORE: Again, I'm not asking about the draft. I'm asking about the analysis you ran, Dr. Imai.

MR. CEPEDA: Dr. Imai has referenced that he conducted the analysis as part of a draft report several times.

MR. GORE: The fact that it's in a draft doesn't insulate it from discovery if it's an analysis he did.

BY MR. GORE:

Q. So Dr. Imai, go ahead.

MR. CEPEDA: We can disagree about that, John.

MR. GORE: I imagine we will. Thank you.

BY MR. GORE:

Q. Dr. Imai, can you answer the question?

A. Yes. So there are two possible ways to impose the overlap constraint with respect to the benchmark plan, as I explained to you, geography or population based. There might be some other ways but those are the two that were part of the software factors that I used in my own software case, and I don't recall which one I used. I'm sorry.

Q. And do you recall the results of comparing that simulation to the enacted plan?

MR. CEPEDA: Same objection.

THE WITNESS: I may -- again, the draft expert report may have included the comparison but I don't recall the specific results.

BY MR. GORE:

Q. And I want to ask you a few more questions about your report. So the simulation analysis that's in your report does not include the benchmark plan-related constraint, correct?

A. That's correct.

Q. Okay. Did you ever compare any of those simulations to, for example, any of the plans proposed by members of the public during the Congressional redistricting process?

A. No.

Q. So you didn't ever compare that to the

NAACP plan one, for example?
A. I didn't have access to those, yeah.
Q. So you didn't compare to NAACP Plan Two either; is that right?
A. No, I didn't do that.
Q. Or the League of Women Voters' plan?
A. No.
Q. Or the Harpootlian plan?
A. No.
Q. And so the analyses you ran were a local simulation of District 1 and 6; correct, was the first one?
A. Yeah. The first one, yeah.
Q. And the second one was a local simulation of Districts 1 and 6 limited to Charleston; is that correct?
A. Yeah, that's right. That's the second one.
Q. And the third is a statewide simulation; is that correct?
A. That's correct.
MR. GORE: At this point we have been going an hour and a half. Let's go off the record.
(A recess was taken.)
BY MR. GORE:

Q. So bear with me for on second if you would be so kind. Okay. I'd like to go to what's marked as tab three.

A. Tab three. Okay.

Q. In your zip.

A. Yes.

Q. I will go ahead and mark that as Exhibit Five.

(Defendant's Exhibit No. 5, RECOMBINATION: A FAMILY OF MARKOV CHAINS FOR REDISTRICTING ARTICLE, was marked for identification.)

BY MR. GORE:

Q. I have loaded it into Exhibit Share and now hopefully we are going to display. And I believe this is an article called Recombination: A Family of Markov Chains for Redistricting; is that right?

A. Yes.

Q. And I'm not even going to ask you to explain what those words mean. Thank you for confirming. I'd like to go to page 41 of this article, if we might.

A. Forty-one, okay. Hold on. Yes. Okay, yes.

Q. Thank you. And I believe there is a

sentence here that I would like to explore with you for a moment. So it's the first full sentence on the top of page 41, it starts, Comparator plans.

Comparator plans must be legally viable and pragmatically plausible to draw power from the conclusion that a proposed plan has very different properties.

Did I read that correctly?

A. Yes.

Q. And I believe this article was listed on your reliance list. Do you agree with that statement?

A. I have no opinion. I did not form opinion on this particular sentence.

Q. Okay. Let me ask it this way. Regardless of what the sentence says, do you agree that for simulation plans to tell you something about the enacted plan, the simulation plans have to be legally viable and pragmatically plausible?

A. I don't know what they mean by those words so I can't really form opinion on that particular sentence.

Q. Let me ask it this way. Forget about that sentence. I'm going to take that down.

A. Okay.

Q. Do you agree that for simulation plans to be instructive they have to comply with legal requirements for redistricting plans generally?

A. I disagree.

Q. Explain that, please.

A. Simulations can be used in many different purposes. So for example, you could see the impact of, you know, what would happen if you take out one particular requirement. And so depending on the goal of the analysis, a different set of constraints can be imposed.

And also, I'm not a lawyer so I don't really make judgment about whether those constraints, how they correspond to the legal requirements. They are informed by legal requirements but I don't make any judgment about the viability in the legal sense. The constraints are mathematical constraints and they are what they are. Nothing more, nothing less.

Q. So is it fair to say, Dr. Imai, that you did not analyze whether any of your simulation plans are legal?

A. I'm not a lawyer so my analysis does not draw any legal conclusions.

Q. Okay. And I just understand the scope of

your analysis.

A. Right.

Q. You didn't do anything to try to determine whether your plans were legal, correct?

A. Yeah. No, I didn't do that.

Q. Now, Dr. Imai, I believe your report mentions the South Carolina House and Senate redistricting criteria; is that right?

A. That's correct.

Q. So let's go to tab five of your binder.

A. Okay.

Q. Which is the House Redistricting Criteria.

A. All right. Tab five. Okay. House, yes. Okay.

Q. And I'm going to mark this as Exhibit Six.

(Defendant's Exhibit No. 6, SOUTH CAROLINA HOUSE OF REPRESENTATIVES JUDICIARY COMMITTEE REDISTRICTING AD HOC COMMITTEE 2021 GUIDELINES AND CRITERIA FOR CONGRESSIONAL AND LEGISLATIVE REDISTRICTING, was marked for identification.)

BY MR. GORE:

Q. And I hope I can figure out how to introduce it. Okay. Dr. Imai, do you recognize this document?

A. Yes.

Q. And what is this document?
A. This is the redistricting guidelines for South Carolina House of Representatives Redistricting Committee.
Q. And did you review this document as part of preparing your expert report?
A. Yes, I received this document from counsel.
Q. I'm going to take that down and I'm going to ask you about tab six in your binder.
A. Okay. It's open.
Q. Great. I'm going to mark this as Exhibit Seven.
(Defendant's Exhibit No. 7, 2021 REDISTRICTING GUIDELINES SOUTH CAROLINA SENATE JUDICIARY COMMITTEE REDISTRICTING SUBCOMMITTEE, was marked for identification.)
BY MR. GORE:
Q. And introduce it in Exhibit Share. Do you recognize this document?
A. Yes.
Q. And what is this document?
A. This is the redistricting guideline for Senate Judiciary Committee.
Q. And did you review this document?

A. Yes, I received it from counsel.
Q. Okay. We'll refer to both of those documents a little bit later.
A. Okay.
Q. I'd like to turn back to your expert report.
A. Okay.
Q. And ask you a few questions about that before we break. And I'd like to start with paragraph 20.
A. Okay.
Q. Which is on page 8.
A. Yes.
Q. And we may also talk about paragraph 22 which starts on page 9 and carries over to page 10.
A. Uh-huh.
Q. Both page 20 and page 22, excuse me, list as one of your criteria that all districts are geographically contiguous; is that right?
A. That's correct.
Q. And how did you -- how did you program the algorithm to guarantee that districts are contiguous?
A. Would you want the technical explanation?
Q. I guess I'd just like to understand.

A. Okay.

Q. Let me ask you this, does the algorithm consider water-to-water contiguity to be permissible?

A. Yes.

Q. How about point-to-point contiguity?

A. No.

Q. Are all of the districts in all of your simulation plans contiguous?

A. From that point of view, yes, from that definition, yes, based on that definition.

Q. Based on that definition?

A. Yes.

Q. So they are contiguous but don't use point-to-point contiguity; is that right?

A. That's correct, yes.

Q. Okay. And the next factor is, you say you program all relevant districts not to exceed an overall population deviation of plus or minus .1 percent; is that right?

A. That's correct.

Q. And that's 730 people, I think, approximately, according to footnote three; is that right?

A. That's my understanding, yes.

Q. You would agree with me though that this deviation is greater than what the House guidelines or the Senate guidelines permit, right?
A. Yes.
Q. And do you recall what the House guidelines and Senate guidelines say about equal population?
A. I don't recall exact language but...
Q. Do you recall what the requirement was for equal population in those guidelines?
A. So I don't recall exact language.
Q. Okay. That's no problem. It's not a memory test. So let's flip back to Exhibit Six, tab five which is the House redistricting criteria.
A. Okay, hold on. Tab five, okay, yes.
Q. All right. So this is the House criteria?
A. Uh-huh.
Q. And roman numeral four is equal population slash deviation?
A. Right.
Q. And letter B says: The number of persons in Congressional districts shall be as nearly equal in population as is practicable. The ideal population for Congressional districts shall be 731,204. In every case efforts shall be made to

achieve strict equality or produce the lowest overall range of deviation possible when taking into consideration geographic limitations.

Did I read that correctly?

A. That's correct.

Q. Okay. And now if you will turn to tab six which has been marked as Exhibit Seven, this is the Senate guidelines.

A. Right.

Q. Do you have that in front of you?

A. Yes.

Q. And roman numeral one says: Requirements of federal law. And A is population equality. Two is congressional districts. And it says, under the apportionment clause of Article One, Section Two of the U.S. Constitution: Any population deviation among congressional districts, no matter how small, must be justified through a showing that the specific deviation is required by legitimate redistricting policies such as making districts compact, respecting political subdivision boundaries, preserving the course of prior districts and avoiding contest between incumbent representatives. So that the state may avoid assuming this additional burden under federal law, a

congressional redistricting plan should not have population deviations greater than one person. Is that correct?

A. That's correct.

Q. So if we turn back to your expert report.

A. Uh-huh.

Q. Paragraph 20 and paragraph 22, you use the constraint of plus or minus 0.1 percent, correct?

A. That is right.

Q. And that's greater than what the -- certainly what the Senate guidelines permitted because Senate guidelines said the deviation should be no greater than one person; is that correct?

A. That's correct.

Q. Why did you use 0.1 percent as your constraint rather than one person?

A. The reason is that my simulation is based on the present level data.

Q. Can you elaborate on that?

A. So the precinct is much later greater than the census blocks which typically are used as units for actually doing the enacted plan and because a precinct has on average the population of a couple thousand often and there is a variation depending on the precincts you are looking at.

And so its appropriate to use the population deviation that corresponds to that particular size.

Q. Would it have been possible for you to program the algorithm to avoid a population deviation of greater than one person?

A. I think that would be difficult given the current state of algorithm development except you can always take a simulated plan that's generated using the precinct level data and modify the boundaries of the district to equalize the population. You know, you can do that.

But automatically generating at the census block level is -- is one -- there is no reason to do that and I can elaborate why there is no reason to do that. And two, even if you wanted to do it, it's not, you know, necessarily simple.

Q. And I think you said that you can take a plan generated through your method and equalize the population after the fact, correct?

A. Right, that's one possibility as opposed to generate directly sampling the plans as the census block level because there are many, many more census blocks than precincts in any given state. And I can explain why, you know, one should use

precincts rather than census blocks.

Q. Why don't you -- tell me why you used precincts as opposed to census block?

A. All right. So first thing that's very important is that I'm not trying to generate the plan that can be enacted, like I'm not trying to be automated map drawer. Like, my goal is to evaluate the characteristics of the enacted plan.

So in this case, I'm -- the goal is really to analyze the role that race played in drawing the district, the boundaries of districts in the enacted plan. So for that purpose, the equalizing the few hundred people -- and that's the maximum deviation -- so a lot of simulated plans have actually fewer deviation, it's not going to affect the substantive conclusions that I draw from the simulation analysis.

If you look at the results they are not different by a few hundred people, the difference is way, way diverse. So adjusting the population deviation is not going to change.

The other reason is that one of the things that guideline mentions is a preservation of the precincts and so it also makes sense to keep that simulation at the precinct level as opposed to

census block level.

Q. Thank you for all that. Did you in your report otherwise ever try to equalize the population on any of your simulation plans?

A. No.

Q. Or did you just take them as they were?

A. No, I did not try to equalize them.

MR. GORE: Let's go off the record for a minute.

(A recess was taken.)

BY MR. GORE:

Q. Dr. Imai, did you discuss your deposition with anyone during the break?

A. No.

Q. I'd like to go back to your expert report. At the break we were discussing paragraphs 20 and 22 where you list some of the principles that you used for your simulation. And we talked about contiguity and equal population.

The next few bullets in both of those paragraphs 20 and 22 say: No incumbent is paired with another incumbent. All the relevant districts are on average at least as compact as the enacted plan. The number of split counties is on average no greater than the corresponding number under the

enacted plan. And the number of split municipalities is on average no greater than the corresponding number under the enacted plan.

So I'd like to walk through each of these factors but before I do so, did you program constraints for each of these factors in your simulations?

A. Can you clarify what you mean by, did you program?

Q. Or did the algorithm operate subject to constraints for each of these factors?

A. So each of these factors, basically that you can instruct algorithm to discourage or encourage certain type of plans. So for example, for the no incumbent pairing constraint I encouraged algorithm not to sample the plans that paired incumbents in some districts.

Q. Let's turn to page 25 of your report.

A. Okay.

Q. I'd like to ask you about paragraph 57 on page 25.

A. Yes.

Q. Okay. You mention a couple of different kinds of constraints in this paragraph. I'd like to understand what they mean. First, what is a

hierarchical constraint?

A. Oh, hierarchical constraint is a constraint that, you can think of it as a hard constraint which limits the number of counties, county splits to basically like total number of districts minus one. So which in this case is six.

So this can be done in the process of splitting the merged district into two districts. So if you recall, the merged split algorithm merges two adjacent districts and then splits, merge district into two new districts. And when you do that you can make sure that you only split county boundary once which basically effectively limits the number of counties being split to six which is equal to the total number minus one.

So it's a hard constraint. So you can think of it as every simulated plan has this property.

Q. What is a soft constraint?

A. Soft constraint has a weight parameter and it encourages or discourages certain type of plans so it doesn't ensure that all the plans satisfy particular constraints. So for example, incumbency pairing avoidance is a soft constraint so some of the plan that we sample pairs incumbents, it's a

very small number. And what I do is basically drop those, you know, plans that don't satisfy the constraint afterwards.

But during the sampling process it's using a soft constraint which, you know, discourages the plans that pair incumbents in some districts.

Q. Okay. Paragraph 57 of your report assigns certain strengths to certain constraints. So for example, you have a default compactness of strength one and there are other numbers in here. Just, can you explain to me what the strength of the constraint is?

A. Right. So the strengths of the constraint is used for the soft constraint where the algorithm is instructed to discourage or encourage certain type of plans. And so that strength of the constraints basically tells us how strong that discouragement or encouragement is in the implementation of the algorithm. It's a parameter that controls that strength.

Q. So for each of these factors in paragraph 57, how did you choose the strength of the constraint?

A. So this is case-by-case, you know, choice so you, given this particular, each specific

analysis you choose these parameters such that a set of, you know, certain goals and criteria are being met. So for example,l if I want to ensure that incumbency is not paired in any district, if I choose the constraint that is too, you know, soft constraint that is too weak, I would end up with many plans that pair incumbents. So I increase it until I eliminate pretty much most of the incumbency avoidance. And of course, I have to do this while making sure that other properties of the simulated plans are satisfied as well.

Q. So let me ask you a question as an example. In the localized Charleston County simulation you raised the compactness constraint to 1.07, according to this paragraph. Why choose 1.07 as opposed to 1.2 or 1.35 or something else?

A. Uh-huh. So the default parameter value for this is one. There are mathematical reasons that particular default value is convenient for the sampling algorithm.

And if I wanted to make sure that the plans are more compact, so this is relative value so it's a reactive strength. And the reason why I raised it is because when I run it with constraint of strength one that's a default value, it's less

compact.

And in many cases, not all of the cases, many of the simulated plans are less compact than enacted plan so I wanted to make sure that the simulated plans are, you know, equally compact when compared to the enacted plan. That's why I raised that value, the encouraged, the simulation algorithm to sample more compact plan.

Q. How do you know that the precise strength you have selected is the right strength for that constraint in your model?

A. So you can basically run a simulation with different values and then see what the compactness of the simulated plans are.

Q. So in your model does changing the strength of the constraint change the sample of plans that the simulation generates?

A. That's correct. So changing the values of parameters will change the distribution, the target distribution. Hence it would change the output, simulated output.

Q. And would that be true if one of the factors, the strength of one factor was changed within the constraints?

A. Yeah. Like any factor that -- of the

constraint if you choose that it's -- it's empirical question, it's possible that it would change the simulated plans as well.

Q. Same if you change the strength or two or more of the constraints?

A. Yeah, that's correct.

Q. Or all of the constraints?

A. That's correct.

Q. Do the strengths of the constraints you selected approximate the constraints followed by the general assembly in the enacted plan?

A. No. In the sense that I don't -- my goal is not to -- you know, my constraints are informed by these guidelines and then traditional redistricting criteria. But they are not designed to mimic the way that the map drawers created these maps.

Q. So if the general assembly had assigned a different strength to these constraints it would have been working off of a different universe of potential plans, right?

A. Well, my understanding is the general assembly doesn't use the algorithm. So I'm not sure what do you mean by, if the general assembly changed the strength? Different constraints, strength

constraints.

Q. Yeah, that's a fair question. So would you agree that compliance with these principles requires tradeoffs or can require tradeoffs between various of the principles?

A. In some cases there are tradeoffs. In other cases the tradeoff is not very strong.

Q. And if the general assembly weighted certain factors more heavily than your model it would have come up with a different plan than your model, right?

A. I mean, again, I don't think that my model would inform how the general assembly came up with a plan but it's certainly the case that if you change the parameters of my model that would result in different simulated plans.

Q. Okay. So let's walk through each of these factors and some of the constraints that you used, if that's okay.

A. Sure.

Q. So incumbency pairing, I'm still in paragraph 57 on page 25, it looks like for the localized District 1 and 6 simulation, an incumbency pairing avoidance constraint of strength one was used.

A. Uh-huh.
Q. Is that one value out of another number or what does that -- how do I understand that number?
A. Yes, so this is a reductive value so the reason you need a scale that tells you what one is, it really depends on the data that you are analyzing as well as other parameters that you are specifying.
So, you know, one doesn't nearly have intuitive meaning in the model, it's a relative strength so the higher this value is, you know, the stronger the constraint is and lower this value is, the weaker the constraint.
Q. And so --
A. Yeah.
Q. I'm sorry, I didn't mean to cut you off.
A. All this is meant to say that the zero means no constraint.
Q. So in the statewide VRA simulation --
A. Uh-huh.
Q. I notice that your incumbency pairing avoidance constraint had a strength of eight. Is that a stronger constraint than the constraint of one in the other simulation that we were just discussing?
A. So you can't really compare the statewide

simulation with the two-district simulation because now the data are very different and we have different number of districts and there are many other factors that are different.

So this number, say one or eight, needs to be understood within the particular simulation context as opposed to comparing that number across different simulation analysis.

Q. Okay. Thank you for that clarification. I think you mentioned before that none of your simulation plans pair incumbents; is that correct?

A. That's correct. But, you know, after dropping a small number of simulated plans that do pair incumbents, so because of the constraints that reduced the number of simulated plans that pair some incumbents, but in the final analysis simulated plans, the simulated plans that I used for the final analysis I dropped those small number of plans so none of those plans I used for the final analysis has incumbent pairing.

Q. So help me make sure that I got this right. Because the incumbency pairing constraint is a soft constraint, when you run the simulation, the simulation will generate some plans that pair incumbents somewhere?

A. Yeah.
Q. Did I have that right so far?
A. It may or may not.
Q. Okay.
A. It may or may not, yeah.
Q. Fair?
A. But it's possible.
Q. And to the extent it did so here, you just went through and eliminated those plans from your simulation set; is that right?
A. That's right. That's right. It was very small number, it was very trivial number.
Q. Yeah, not to hold you to this, but can you quantify it for me either in a number of plans or like a percentage that might have generated incumbency pairings?
A. I don't recall that.
Q. Okay. Now in each of your simulation plans you give the district number, the number for the incumbent's district in the enacted plan, correct?
A. That's correct because the simulated plan has no pairings so I can use the incumbency location to, you know, to level the district in the simulated plan.

Q. And that's true even though the simulated district often does not cover the same geographic area as the enacted district, right?

A. That's possible, certainly.

Q. Okay. Let's get back if we can, just briefly, to paragraph 23 of your report.

A. Twenty-three, okay.

Q. Which is on page 10 I think.

A. Okay. Hold on. Yes.

Q. And I'll just read the last sentence of that paragraph. It says: This renaming procedure allows me to compare each enacted district with a comparable simulated district even though the two districts often do not cover the same geographic area. Is that right?

A. You read it correctly.

Q. All right, thank you. So for example, in your plan, in your setup plan, your simulated plans, District 1 was the district where Nancy Mace resides; is that correct?

A. Right.

Q. Even if that district covered a different geographic area or some different geographic area than the enacted District 1; is that right?

A. Right, with varying degree.

Q. Okay. Thank you. So let's go back to page 25, paragraph 57 if we can and talk about some of the other constraints that you used.

A. Okay. Page, hold on. Okay.

Q. So for -- let's talk about split counties for a second. I think for the localized District 1 and 6 you used a split, a county split avoidance constraint of 0.4; is that right?

A. That's correct.

Q. And would you consider that a moderate strength for that constraint, a heavy strength? How would you -- help me understand that number a little better.

A. Yeah, again, it doesn't really have a scale so other than the relative scale so, you know, it's hard for me to characterize it that way.

Q. And so as a result of this constraint did your simulation sets generate any plans with more county splits than the enacted plan?

A. So the results of the county splits is in Figure 12. And so you see that, you know, many of my simulated plans split, you know, have fewer or equal number of county splits as enacted plan. There are a smaller number of maybe about 20 percent of the simulated plan has, you know, slightly more

county splits than the enacted plan.

Q. Thank you. I'm glad you sent me to Figure 12 because that was my next question.

A. Oh, okay.

Q. So on the right-hand side on Figure 12 there is a histogram for the statewide VRA simulation; is that right?

A. Right.

Q. So I'm having a little trouble reading these numbers but is each of these bars like a single number, like four or five or six or are they a range or?

A. Yeah.

Q. You have the whole number of county splits in the plan, right?

A. Yeah. It's a whole number even though it's a little bit, slightly shifted to the left.

Q. Okay.

A. Yeah.

Q. What's the minimum number of county splits in a seven district congressional plan with a plus or minus one population deviation?

A. A plus or minus -- I don't -- can you repeat the question again?

Q. Sure. Yeah. I'm asking about the -- I'm

asking you a background question and then I'll ask you about the histogram.

A. Oh.

Q. The enacted plan has seven districts, right?

A. Right.

Q. And in the -- and the Senate factors also limited the enacted plan to a plus or minus one person population deviation, right?

A. Right.

Q. So what would you expect to be the minimum number of county splits in a plan under those constraints or under those criteria?

A. If I do a simulation under that condition?

Q. Well, not -- either a simulation or just in the real world. I mean, before I think you mentioned something about six county splits and it seems to me that if you are drawing seven districts and you've got to get the district to plus or minus one person, then you are going to split six counties to do that, aren't you?

A. I can't really speak to a hypothetical.

Q. Okay.

A. So my analysis is based on, you know, present level data.

Q. Some of the plans in your statewide VRA simulation generate fewer than six county splits, right?

A. That's correct.

Q. Did you examine how a plan with seven districts can split only four or five counties?

A. Would you clarify what you mean by examine?

Q. Yeah. Do you have any explanation as to how a plan with seven districts splits only four or five counties?

A. I did not look at this, you know, the plan, simulated plan with this specific number of county splits.

Q. I believe your report says that you were calculating county splits, on average you were counting them so that -- or you programmed them so that on average it doesn't have more than an enacted plan. And I'm curious what you mean by average, what kind of average are you talking about?

A. Across simulated plans.

Q. Okay. So --

A. So 10,000 simulated plans.

Q. So if I average the number of county splits across the 10,000 simulated plans, you

program the algorithm so that that number would be the same or less than the number of county splits in the enacted plan; is that right?

A. Oh, I see. No. So what I did was basically choose the constraint parameter because that's the input into the algorithm such that, you know, when you look at the number of county splits on average, the simulated plans have fewer county splits than the enacted plan so some simulated plan may have than enacted plans.

So some simulated plans may have more county splits but many other simulated plans would have fewer county splits. It's not that directly told the algorithm to, you know, make sure an average number of county splits is fewer than enacted, it's more that I choose the parameters of the soft county split constraints in some cases such that the number of county splits is, after looking at the simulated plan, is on average fewer than the enacted plan.

Q. Okay. So just so I understand, and I appreciate your patience with me since I'm not familiar with your method, okay? But as I understand it, you are saying that each set of 10,000 plans that you have put forward in your

report, each set has on average fewer county splits per plan or as many as the enacted plan, right?

A. Right, that's correct.

Q. So within your set of 10,000 some may have more splits, some may have fewer, right?

A. Yes. That's exactly right.

Q. And I think Figure 12 shows that and the best way to see that is Districts 1 and 6 simulation?

A. Right.

Q. You've got the line for the enacted plan where it is and then you've got --

A. Yeah.

Q. Some plans from your simulation falling on either side of that; is that right?

A. That's right. So the goal was to put the simulated plans in about the same range or at least about the same range as the enacted plan.

Q. Okay. And is that also -- in turning how to split municipalities is that also how you calculated the average for split municipalities?

A. Yes, that's correct. That's Figure 13.

Q. And that's Figure 13 on page 27 of your report, right?

A. Uh-huh. That's right.

Q. So as I understand what happens, you assign a constraint strength and the algorithm generates a universe of plans and then you can go back and pair down that universe to 10,000 plans that enforces kind of your averages for county splits and municipalities and everything else; is that right?

A. Well, so you choose the parameters such that the sample plans, you know, have fewer or equal on average number of splits when compared to the enacted plan.

Q. So while we are talking about splits I have a couple of more questions.

A. Sure.

Q. Do some of your -- did you place --let me back up.

Let's take county splits as an example. Did you constrain the algorithm to split only the same counties as the enacted plan or can the algorithm draw plans that split different counties than the enacted plan?

A. For statewide simulation it could be a different counties that could be split. For the second race run simulation, the first one which is the District 1 and 6 simulation, it could also be a

different counties could be split.

For the second plan simulation where I focus on Charleston County you reach how the Charleston County was split is the focus of the analysis. And since the rest of the, you know, the districts are kept the same as the enacted plan they would split exactly the same place in county as the enacted plan. So it depends on which analysis that I've done.

Q. Okay. And could you have instructed the algorithm to split the same counties as the enacted plan only?

A. Right. I mean, it's possible to do that.

Q. And is there a reason you didn't do that?

A. So my constraints were guided, you know, in part based on these South Carolina redistricting guidelines and I don't recall that it specified the specific counties that need to be spread in that guideline, although I think there is provision that, you know, said something about dividing the number of county splits.

Q. Now, I'm going to ask you the same questions about the municipality splits. So do any of your simulated plans with different municipalities than the enacted plan?

A. Probably. I did not specify that certain municipalities should be split.

Q. Let's move to the next criteria which is compactness.

A. Okay.

Q. We talked a little bit already about paragraph 57 where there is a default compactness constraint of one and then there is a localized -- in a localized Charleston County simulation raised to 1.07; is that right?

A. Right, right.

Q. And again, your report says that all the relevant districts are on average at least as compact as the enacted plan?

A. Yeah.

Q. So what was your method for calculating that average?

A. Right. So this is Figures 10 and 11 so I used two different measures of compactness and, you know, so I specified the constraint and then generate the simulated plans and look at the compactness of districts using these two measures, Polsby-Popper and fractions of edges kept.

Q. How did you choose those measures?

A. So Polsby-Popper is perhaps most commonly

used measure in this type of analysis and fractions of edges kept is actually related, closely related to the -- the compactness that algorithm directly controls. So these measures are, you know, obviously related but they are different. And fraction of edges kept is something that algorithm directly controls or closely related to directly.

Q. Thank you for that. Are these mathematical measures of compactness?

A. Yes, they are.

Q. And again, on the averages, so within your each bucket of 10,000 simulated plans, is it that each bucket on average that districts are as compact as the corresponding district in the enacted plan, is that how you get the average?

A. So here I'm looking at the overall compactness. So Polsby-Popper is a major for each district, but I'm looking at the average of the districts so that's why there is one number for each limited plan. It's not like a comparison of the specific district and fraction of edges kept is a plan-wide measure to begin with. So in both cases I'm looking at the overall compactness, not like each district specific comparison of compactness of those.

Q. So if I can rephrase that, your average is the average of the plan like compactness?
A. Correct.
Q. Not any average of district specific compactness?
A. That's correct.
Q. Okay. Thank you. Dr. Imai, can you go back to tab five in your --
A. Okay, tab five. Yes, House guideline.
Q. Yes, that's the House redistricting guidelines?
A. Yes.
Q. And if you will scroll down with me I previously marked this as Exhibit Six. And on the second page of these guidelines --
A. Uh-huh.
Q. Roman numeral six is titled Compactness. Do you see that?
A. Yes.
Q. The second paragraph under that factor?
A. Uh-huh.
Q. It says: Compactness should be judged in part by the configuration of prior plans. Compactness should not be judge based on any mathematical, statistical or formula-based

calculation or determination.

Did I read that correctly?

A. You did.

Q. So did you do anything in your model to judge compactness by the configuration of prior plans?

A. Not directly.

Q. How about indirectly?

A. Well, to the extent that I tried to make sure that the overall compactness of my simulated plans is comparable to the enacted plan. So to the extent that the enacted plan does take this consideration into account, you know, they are related, but not directly. I didn't directly compare to the prior plans.

Q. And when you were -- to the extent that you were indirectly comparing, that was based on the mathematical measures, correct?

A. Yes, everything I do is mathematical.

Q. So it wasn't based on the shape of the district, for example, correct?

A. No. Shape in the mathematical sense.

Q. Fair enough. Very well said. And here the House says that compactness should not be judged based upon any mathematical calculation or

determination, correct?

A. That's -- you read it correctly.

Q. And Dr. Imai, if you would be so kind as to turn back to tab six with me which I previously marked as Exhibit Seven.

A. Senate Guidelines.

Q. And this is the Senate Guidelines.

A. Okay.

Q. And if you will go with me to page 2.

A. Okay.

Q. Roman numeral three is additional considerations?

A. Uh-huh.

Q. And F is district compactness?

A. Okay.

Q. And it reads: In determining the relative compactness of the district consideration should be given to geography, demography, communities of interest and the extent to which parts of the district are joined by roads, media outlets or other means for constituents to communicate effectively with each other and with their representative.

Did I read that correctly?

A. You did.

Q. And does the algorithm account for

geography, demography, communities of interest or the extent to which parts of the district are joined by roads, media outlets or other means for constituents to communicate effectively with each other?

A. So it depends on what we mean by geography, demography, community. So interest, because obviously the algorithm is using the state geography, it operates on the shape files and it uses population data to determine the size, population size of the district.

And in the statewide case, you know, analysis I take into account for the racial data. In terms of communities of interest, you know, I take into account the administrative boundaries like counties and municipalities. And so to the extent that those inputs I used are related to this type of consideration they are taken into account but, you know, the algorithm doesn't do anything more than that.

Q. And the Senate guidelines also don't mention use of a mathematical measure of compactness, correct?

A. Yes, I don't see that mentioned.

Q. Let's scroll up here, same page,

additional considerations on the Senate guidelines, letter B is constituent consistency and it lists: Preserving the cores of existing districts.

Did the algorithm consider preserving the cores of existing districts in generating plans?

A. So to the extent that, you know, I instructed the algorithm to avoid incumbents pairing and to the extent that my race plan simulations, for example, freezes, you know, all the districts other than Districts 1 and 6 and in the case of second race-blind simulation it freezes everything other than Charleston County.

So in that sense, you know, there are constraints that have implications of cores of existing districts, preservation.

Q. Did you --

A. But the analysis I presented in my final report did not directly use, you know, previous -- the benchmark plan.

Q. And so your analysis did not include a constraint for preserving the cores of districts, correct?

A. Not directly.

Q. And likewise, it did not include a constraint for keeping incumbents' residences in

districts with their core constituents, correct?

A. Yeah, incumbents weren't paired but there was no constraint that directly, you know, that needs a definition of what the core constituency of incumbents are. And that information was not available so I did not include that either.

Q. And as we discussed before, the districts in your simulation plans had the same numbers as districts in the enacted plan but may cover different geography; is that right?

A. That's correct, depending on, you know, this will change across analysis and, you know, I have three analyses. So first two analyses are probably much bigger overlap than statewide analysis, for example, but yeah.

Q. So for example, wouldn't that also mean that because the districts encompass different geography they encompass different populations and voters, correct?

A. That's correct, different people in different areas.

Q. And speaking with this page, communities of interest --

A. Uh-huh.

Q. Did you include any constraint for

communities of interest?

A. So again, only to the extent that, you know, things like administrative boundaries, like counties and municipalities overlap with these interest and to the extent that, you know, incumbent residence wasn't paired, but there is no definition of communities of interest available so I didn't use that.

Q. So there was no direct constraint on communities of interest, correct?

A. That's correct to the extent that --

Q. Okay.

A. Yeah, I don't have, you know, definitions of what these communities are.

Q. And so you didn't assign a strength to communities of interest, correct?

A. Right, because there is no mathematical, you know, geographical definition of communities of interest so I didn't assign that constraint directly to this.

Q. And so you also didn't assign a strength to preserving the course of existing districts, correct?

A. That's correct. For the reason that I explained that in order to isolate the role that

race played in determining the districts of enacted plan that I didn't want to include any plan including the benchmark plan.

Q. And similarly, you didn't assign a strength to keeping incumbents residences in districts with their core constituents, correct?

A. Right. So the weights are for just the avoidance of incumbent pairing and not with respect to their core constituents because they are not -- that definition was not available to me.

Q. Okay. Let's look down at letter E, minimizing divisions of voting precinct boundaries?

A. Uh-huh.

Q. Did you program a constraint in the algorithm for VTD splits or precinct splits?

A. Let's me double check. Yeah, I don't think so. It's no a listed in paragraph 57, which is not -- yeah.

Q. And I don't believe it's listed in paragraphs 20 or 22 either.

A. Yeah, I wanted to double check, yeah. I don't think I imposed that constraint.

Q. So let's go to -- can we go to figure 14 on page 27 of your report?

A. Yes.

Q. Wonderful. Thank you. So you do include here some histograms of VTD splits?
A. Uh-huh.
Q. Is that what I'm seeing in figure 14?
A. Yes.
Q. And it looks like the enacted plan performs better than most of the District 1 and six simulation plans on VTD splits; is that right?
A. That's correct.
Q. And the enacted plan appears to perform better than all the Charleston County simulation plans or at least the vast majority of them on VTD splits too, right?
A. Yes, that's correct.
Q. And I add that caveat because I can't tell based on this histogram whether there is -- Charleston County seems to be left to that line, right?
A. Yeah, sure.
Q. And then in terms of state wide VRA simulation the enacted plan again outperforms the vast majority of the plan from the statewide VRA simulations; is that right?
A. That's correct.
Q. And down here in paragraph F, the second

paragraph or I'm sorry the second sentence there in paragraph 61 under the heading F still on page 27?

A. Yeah, I see that.

Q. You say: This is in part due to the fact that many municipalities split VTDs implying that there often is a direct tradeoff between municipality and precinct splits.

A. Uh-huh.

Q. Did I read that correctly?

A. Yes, you did.

Q. Can you explain what you mean by tradeoff between municipalities -- I'm sorry, municipality and precinct splits?

A. Yeah, so basically the precincts are not nested, in the case of South Carolina, are not necessarily nested in municipalities. So municipalities may split the VTD so if you want to reduce the number of municipality splits that may mean that you end up -- because a municipality cut to the precinct you may end up splitting the VTDs, the precincts.

Q. I've got it. So minimizing municipalities splits means that sometimes VTDs will be split; is that right?

A. That's right.

Q. Okay.

A. So if you are trying to keep the municipalities together you may end up splitting the VTDs whether as if you try to keep the VTDs together you may end up splitting municipalities.

Q. And the other thing I want to ask you about this sentence is you use VTDs in precinct?

A. Oh.

Q. Are you -- are you using those terms interchangeably?

A. Yes, yes, interchangeably.

Q. All right. I'd like to go back to paragraphs 20 and 22 back on page seven?

A. Okay, hold on.

Q. I guess. I'm sorry, I think maybe page eight.

A. Page 8, okay.

Q. In paragraph 20 it says: No race or partisan information was used and in paragraph 22 which is talking about your VRA set is says: No partisan information was used. So just to confirm, you didn't use partisan information in any of your simulations, correct?

A. No.

Q. So you didn't consider election data; is

that right?

A. Yeah, I didn't use it.

Q. And you didn't consider whether the district would enact -- would elect a Republican or a Democrat, right?

A. No, yeah.

Q. And you didn't --

A. That's right.

Q. And you didn't consider voter registration by party, correct?

A. No, yeah, I didn't consider that.

Q. Now, could you have instructed the algorithm to consider politics or added a politics constraint to the algorithm?

A. If the politics is, you know, what we mean by partisan mathematically formulated, yes.

Q. All right. So for example, could you have added a constraint to the algorithm that required that six of the seven districts elect a Republican rather than a Democrat?

A. Right, yeah, I could have done that.

Q. And could you have instructed the algorithm to require that District 1 have a higher percentage of the Trump vote share in 2020 than the benchmark District 1?

A. That's definitely possible as long as it's specified, you know, in a mathematical, operational way.

Q. And if you use some kind of constraint like this in partisan gerrymandering cases, right?

A. I -- I don't think so although, you know, it's -- I'm using my memory because in partisan gerrymandering case often is to see if you use a simulation without partisan information and then compare that nonpartisan baseline with enacted plan and look at the partisan, you know, how different they are with the partisan simulation. So in the simulation often you don't include the partisan information because you want to establish a nonpartisan baseline.

Q. I've got it, okay. Thank you for that clarification.

A. Uh-huh.

Q. But for these -- for this report at least you didn't include any kind of constraint like that, correct?

A. Right, for this report I did not use any partisan information.

Q. And why didn't you do that?

A. Because it was not clear from the

guideline how the partisan information should be incorporated, at least to me, so -- and I wasn't also instructed to do so either.

Q. Dr. Imai, do you have a view as to whether race and politics are highly correlated in South Carolina?

A. I have not analyzed that particular aspect so I don't have, you know, I don't have opinion on that particular question.

Q. And do you have an opinion on whether race rather than politics explains the enacted plan?

A. My analysis does not address that question so I have no opinion on that.

Q. All right. So now I can ask you some more questions about each of the simulations you conducted if that would be okay?

A. Sure.

Q. So we are still in your report and your first simulations, your localized simulations do not include a VRA constraint; is that right?

A. That's correct. It's race neutral.

Q. And why don't they include a VRA constraint?

A. Oh, because, you know, in order to assess whether race played any role in drawing the

boundaries of the enacted plan what you want to do as a first step is to, you know, create -- conduct a simulation analysis to have a race neutral baseline and then by comparing the race neutral baseline with the enacted plan you can see how race played a role in drawing boundaries, in this case between Districts 1 and 6 under the enacted plan.

Q. And in the statewide simulation you did include the VRA constraint, correct?

A. That's correct.

Q. And what -- I'm sorry, go ahead.

A. Yeah, that's correct.

Q. Okay. And is there a reason you didn't run a statewide simulation without the VRA constraint?

A. Oh, I see. The reason is that statewide simulation was there to address the possibility that you know, what I found in the race neutral simulation analysis is due to the VRA consideration. And that was the purpose of the statewide simulation analysis and so for that analysis I included the VRA constraint.

Q. Dr. Imai, let's go to -- let's talk first about your District 1 and District 6 localized simulation, if that's okay?

A. Sure.
Q. I'd like to start at page 11.
A. Okay.
Q. So as I understand the simulation you only altered the boundary between District 1 and 6, right?
A. That's correct.
Q. So the boundaries of Districts 2, 5 -- 2, 3, 4, 5 and 7 are all unchanged in this simulation?
A. Right, fixed under the enacted plan.
Q. Great. All right. So paragraph 27 and then you have a Figure 1 on page 12 --
A. Uh-huh.
Q. That shows precincts or VTDs by BVAP; is that right?
A. Right, that's correct.
Q. And you are using the total black voting age population by number rather than a black voting age population percentage, right?
A. That's correct.
Q. And so the black voting age population percentage in two districts with the same total black voting age population could be different, right?
A. Oh, yeah, that's true.

Q. And so moving either of those precincts in or out of the district could have a different net effect on the district's BVAP percentage, correct?

A. I'm not sure I understand the question.

Q. Sure. So let's -- let's spin this out. I think you agree with me there are two precincts -- let's take two precincts, each of which have 500 individuals -- black individuals of voting age population.

A. Uh-huh.

Q. One of those precincts -- let's say it only has 500 people in it so it's 100 percent BVAP precinct?

A. I see.

Q. The other precinct has a thousand people in it so it's a 50 percent BVAP precinct.

A. I see, I see.

Q. And moving those in between districts will have different effects on the districts' BVAP population, right?

A. Moving meaning -- what do you mean by moving?

Q. Well, either moving it in or out of the district.

A. Oh, I see.

Q. So if I have 100 percent BVAP precinct to a district that has a different effect on a district than if I have 50 percent BVAP?

A. Yeah.

Q. Okay.

A. So yeah, so the number -- BVAP population is the same but the proportion would be different, different depending on the total number of -- total population in that precinct, that's correct.

Q. Now, if I'm reading this correctly, Figure 1, on the right-hand side there -- you also have the -- this cool colored chart that shows how often the VTD was placed in District 1; is that right?

A. Uh-huh, right.

Q. And the darker numbers show that the VTD was more frequently placed in District 1?

A. That's right.

Q. In the lighter color, right?

A. That's correct.

Q. And so do you know as you sit here right now where Nancy Mace, the District 1 incumbent, lives?

A. Yes, I think near the border, right, the precinct border, the district border, if I remember

correctly.

Q. Yeah, I think she lives in Berkeley County if I'm not mistaken, but I don't know that for sure, but I'm pretty sure that's right. So of course, whatever precinct she lives in ends up in District 1 100 percent of the time?

A. That's right.

Q. Okay.

A. That's correct, yeah.

Q. All right. It looks like large parts of Charleston -- the city of Charleston and Charleston County end up at in District 1 very frequently, at least 90 percent of the time or more; is that right?

A. That's my finding.

Q. Okay. So in your simulated plans, Nancy Mace is frequently placed in a district with the City of Charleston or large portions of Charleston County; is that right?

A. That I think is consistent with my finding.

Q. Are you aware of any reason why the map drawer may not have wanted to include Nancy Mace in a district with the city of Charleston or a big portion of Charleston County?

A. No, I don't consider a map drawer's

intent.

Q. Did you consider the political effect on Nancy Mace's reelection chances --

A. No.

Q. Of Nancy Mace being placed with City of Charleston?

A. No.

Q. Or with a large portion of Charleston County?

A. No political information is used for my analysis.

Q. Okay. So in this analysis shown in Figure 1 --

A. Uh-huh.

Q. Does your analysis control at all for the benchmark district the VTD was in under the benchmark plan?

A. No, I don't use the benchmark plan for my analysis.

Q. And does it control for where the black voters and black individuals on this map live?

A. No, this is race-blind simulation so no racial information is used either.

Q. And in terms of the -- so we've talked about this, this was Districts 1 and 6; is that

correct?

A. That's correct.

Q. And I think on paragraph 29 you mentioned that the average difference in the BVAP proportion of District 1 between the enacted and race-blind simulated plans is about 5.8 percentage points; is that right?

A. Right, the difference, yes, average difference.

Q. And where is that 5.8 percentage point average coming -- where are those black voters coming from in your simulated plans?

A. Oh, I suspect they are coming from the City of Charleston.

Q. Okay. And so in your simulated plans --

A. And then maybe the City of North Charleston as well.

Q. So your simulated plans have 5.8 percentage points BVAP higher in District 1 --

A. Right.

Q. But they have a lower BVAP in some other district or districts as a result, correct?

A. Right. In this case, there is only two districts being analyzed so District 6 would be the one that's lower in terms of BVAP proportion under

the simulated plan compared to the enacted plan.

Q. So the 5.8 percentage points would mean that District 6's BVAP has dropped to something like 41.1 percent; is that right?

A. Yeah, it won't be exact because the population difference is there, but yeah, I think there is, you know, decrease in BVAP proportion of District 6 under the simulated plan when you compare that with enacted plan.

Q. So let me turn to the histogram on page 13 which is your Figure 2.

A. Page 13 -- oh, Figure 2, yes.

Q. So Figure 2 shows that some of the simulated plans have a District 1 BVAP proportion of up to 28 percent or maybe even 30 percent; is that right?

A. Yeah, close to 30 maybe.

Q. 30 percent, close to 30 would be about 12 to 13 percent higher in District 1 than in the enacted plan, right?

A. I don't recall exact -- yeah, no, that's right.

Q. Well, it's higher?

A. Yeah, I see the line. Yeah, I don't know the exact number but, yes.

Q. Yeah, I think in paragraph 29 it says 17.4 percent.

A. Okay.

Q. Okay. So if we have a plan in the simulation --

A. Yeah.

Q. At 29.4 percent that's 12 percent higher, right?

A. That's right.

Q. And that would mean that District 6's BVAP is 12 percent lower, right?

A. Right.

Q. But you didn't provide a histogram showing how enacted District 6's BVAP compares to the BVAP in your simulated plan, correct?

A. Right, because it would be, you know, a mirror image because I'm just working with two districts.

Q. So enacted District 6 has a higher BVAP on average than your simulated District 6s would have, correct?

A. For District 6, yeah.

Q. And in fact, it would be on average 5.8 percent or so higher --

A. Right.

Q. Once you account for population variation?
A. Yeah, in that range, that's correct.
Q. So in this simulation, black voters are just being moved between District 6 and District 1, correct?
A. Well, not just black voters but because I'm just doing the simulation between Districts 1 and 6 so that, you know, the voters are assigned to one of the two districts.
Q. And did you analyze the effect in District 6 of decreasing the BVAP, did you analyze the effect of that on black voters' ability to elect their candidates of choice?
A. Oh, I didn't use any partisan analysis in my report so no.
Q. Now, you didn't do a similar local simulation for Districts 2 and 6, correct?
A. Two and six, no.
Q. Why not?
A. Well, I started with the one and six because that boundary is the largest change that happened in the enacted plan when they -- you know, compared to the benchmark plan.
Q. And have you done any similar local simulation for Districts 2 and 6?

A. I don't think so.
Q. How about for Districts 5 and 6?
A. I don't think so.
Q. And why not for Districts 5 and 6?
A. Like I said, the largest change happened in district boundary in between one and six so it was natural for me to focus on that.
Q. So we were talking a minute ago about how the change in District 6's BVAP is the mirror image of the change in District 1's BVAP?
A. In this analysis?
Q. In this analysis, correct.
A. Right.
Q. Okay. So under your definition this analysis cracks black voters out of District 6, correct?
A. Well, it depends on which black voters you are looking at. If you look at black voters in, you know, Charleston, the City of Charleston, then the enacted plan cracks it because they divided these voters into Districts 1 and 6 and that's basically my finding for that analysis.
Q. In terms of -- I believe you said your definition of cracking is voters being split -- black voters living in a certain area are split into

separate districts; is that right?

A. That's right.

Q. So in your simulation plans black voters living in certain areas are split between Districts 6 and 1, correct?

A. Well, you -- if you give me a definition of group of black voters who live in a certain location I might be able to answer that question, but without that definition it's...

Q. Okay. Well, take Charleston for example.

A. Uh-huh.

Q. You said the enacted plan cracks voters in the City of Charleston but doesn't -- don't your simulation plans do the same thing?

A. I think more often the City of Charleston as a whole is included in the same district under my simulated plan.

Q. But some simulated plans split the City of Charleston.

A. Oh, yeah, of course. It's all statistical so it's a question of how frequently that occurs.

Q. And so those plans crack black voters, correct?

A. Well, I don't know whether there is such simulated plans and it could exist and if they do,

they do but I don't know whether any of my simulated plans have such, you know, specific split that you describe.

Q. But if the simulated plans do have that split then they crack black voters, right?

A. That particular plan does.

Q. And similarly, if a simulated plan cracks or splits North Charleston, for example, would that also be cracking under your definition?

A. If there is such a simulated plan then it is but we need to look at the distribution of simulated plans, not particular specific ones. That's sort of the whole point of statistical analysis is to look at the distribution as opposed to specific simulated plans.

Q. Right, but just conceptually speaking, a simulated plan that splits a black community in a defined location whether it's City of Charleston, North Charleston or somewhere else --

A. Uh-huh.

Q. Cracks that black community under your definition, right?

A. Right, but the -- the use of the term, cracked, does not have any legal meaning. It just means that the group of voters who living in a

certain geographical area are separated into two districts. So I just wanted to make sure that that's clear.

Q. Thank you. Yeah, I think you said that before.

A. Okay, okay.

Q. But thank you, I appreciate that.

A. All right.

Q. And there was not -- I didn't see anything in this section of your report discussing whether the simulated plans or how frequently the simulated plans split the City of Charleston. Is that in here?

A. I don't think so. Yeah, I don't think so.

Q. And similarly, it also doesn't include that kind of statistic or statement for North Charleston either I don't think; is that right?

A. Yeah, I don't think so.

Q. All right. Now, let's turn to your --

MR. GORE: Actually, go let's go off the record for just a moment.

THE WITNESS: Okay.

(A recess was taken.)

BY MR. GORE:

Q. All right. Let's talk about your

Charleston County simulation.

A. Okay.

Q. Which starts on page 13. And as I understand it, this is similar to the first simulation except with the further constraint that the boundaries of Districts 1 and 6 are the same except in Charleston County; is that right?

A. That's right.

Q. So you've frozen the enacted plan and all you are looking at is the boundary between Districts 1 and 6 in Charleston County; is that correct?

A. That's correct.

Q. And let's see. So pages 14 and 15 you again have a couple of histograms here.

A. Right.

Q. That show Charleston County BVAP in Figure 3; is that right?

A. Uh-huh.

Q. And again, the analysis here is total number of black voters rather than BVAP percentage; is that right?

A. Right, although in Figure 2 it's percentage.

Q. Right. And Figure 2 is a percentage for

District 1, correct?

A. Right, and here it's the BVAP in population who live in Charleston County.

Q. All right. So help me understand that because I had an understanding of figure four, so is figure four showing me something about Charleston County or district one as a whole?

A. Oh, this is among the black voters who live in Charleston County how many of them are assigned to District 1.

Q. Okay. I think I just confused us both. Are you talking about Figure 3 or Figure 4?

A. Figure 3.

Q. Figure 3, okay. Thank you. Okay. So Figure 3 is just Charleston County and it's BVAP total, not BVAP percentage, right?

A. Right, that's right.

Q. And again, in this simulation you didn't control for which benchmark district a voter was in, in the benchmark plan, correct?

A. No.

Q. And you also didn't control for which district the black voters lived in, in the benchmark plan, correct?

A. No, I didn't use benchmark plan.

Q. And I believe in paragraph 32 you said the difference on average, that District 1 has approximately 24,900 black voters on average in your simulation plans and that that's a difference of 9500 voters compared to the enacted plan; is that right?

A. Yeah, that's correct.

Q. And 9500 voters, do you know what that translates into in terms of District 1's BVAP percentage?

A. Oh, you just have to divide by the population.

Q. By the population, okay. So if I told you that I tried to do this math last night and I think it's 1.3 percent, does that sound about right to you?

A. Yes, I suppose. Yeah, it's more than 1 percent. Yeah, that sounds right. I trust you.

Q. Okay. Well, that's fine. We'll -- you can take my word for it, you don't have to agree with me, but you have no reason to dispute that it might be 1.3 percent?

A. Right.

Q. 1.29 or whatever it rounds off to?

A. Sure.

Q. And so that would mean that under the simulation plans the BVAP in District 1 is 1.3 percent higher on average but in District 6 it's 1.3 percent lower on average; is that right?

A. Right, in the simulation analysis, that's correct, yeah.

Q. And have you done any effect of this analysis on this 1.3 percent change in these two districts' BVAP?

A. No, I have not done any partisan analysis.

Q. All right. Let's go to page 26 in your report.

A. Twenty-six, okay.

Q. I'd like to look at Figure 10.

A. Okay. Figure 10. Okay.

Q. As I understand it, Figure 10 shows the Polsby-Popper compactness scores for the various simulations and the enacted plan; is that right?

A. Right, that's correct.

Q. And on Polsby-Popper metric a higher number means more compact, correct?

A. That's correct, yeah.

Q. So I want to look at the Charleston County simulation table you have here in Figure 10 or a histogram rather than a table.

A. Yeah.

Q. So you froze all of the other district boundaries in the enacted plan except for Charleston County between one and six?

A. Uh-huh.

Q. It appears here that the enacted plan performs better on average than the vast majority of plans under the Charleston County simulation on the Polsby-Popper; is that right?

A. I wouldn't characterize it that way. It would be, it was in, you know, within statistical differences. It's not significantly different.

Q. But the --

A. They are essentially in the same range.

Q. Okay. But the majority of the simulated plans on average are to the left of that line, right?

A. It looks like it but, again, I would not take that difference, you know, to be statistically meaningful and I would characterize it as they are basic -- essentially in the same range.

Q. And do you have any view on what accounts for that difference?

A. Can you clarify the question?

Q. Well, I'm just curious, it's more for my

academic interest.

A. Uh-huh.

Q. You know, if there is something about the model that would explain this difference as to why the simulation --

A. Oh, I see.

Q. I mean, on average -- given the other constraints that you programmed and other things that you did.

A. Yeah, I mean, compactness, you know, there are different measures and each measure reflects different aspects of geographical shape of districts and, you know, the algorithm controls the fraction of edges kept which is the one in Figure 11 so that is the one that sort of much more closely related to what algorithm does, just a mathematical formulation.

And so Polsby-Popper is related to that but not the same thing so, you know, I would characterize the results on the Polsby-Popper to be essentially in the same range, simulated plan and enacted plan essentially the same. For the fraction of edges kept simulated plans is maybe a little bit better but it's not that, you know, huge difference. That would be my characterization.

Q. Okay. And just so I'm clear, this histogram in Figure 10, the Charleston County simulation average Polsby-Popper score, is that showing the averages for Districts 1 and 6 or is that showing the average for the whole plan?

A. For Districts 1 and 6.

Q. Okay, thank you.

A. Oh, wait. Actually, I don't recall so...

Q. Okay.

A. So either way, the other districts are frozen so it won't affect the difference --

Q. Right, okay.

A. Between the two, yeah.

Q. Yeah, so in relative terms it would be the same, right?

A. Yeah, yeah, that's right. And that's what matters is that, you know, the differences as opposed to the number itself.

Q. Other than Charleston County, are you aware of any other counties that are split between Districts 1 and 6 in the enacted plan?

A. Yeah, I think I analyzed Richland and Sumter in the statewide analysis.

Q. Are those counties split between Districts 1 and 6?

A. Oh, one and six, I'm sorry. I thought any county that's split by District 6. So I -- yeah, I misunderstood your question.

Q. Would you like me to rephrase it again?

A. Yeah, yeah, yeah.

Q. Yeah. So other than Charleston County are you aware of any other counties that are split between District 1 and District 6 in the enacted plan?

A. Yeah, I know there are others. I don't recall exactly which one but...

Q. And did you conduct a similar local simulation analysis of any of those counties?

A. Oh, no.

Q. So you didn't do one for Dorchester County; is that correct?

A. No, I didn't do that.

Q. Or Colleton County?

A. No.

Q. Or Jasper County?

A. No.

MR. GORE: So let's go off the record again.

(A recess was taken.)

BY MR. GORE:

Q. Dr. Imai, did you discuss your deposition with anyone on the break?

A. No.

Q. All right. Let's move to your statewide simulation analysis with the VRA constraint.

A. Okay.

Q. And in the statewide simulation unlike the local simulations you don't freeze any districts, correct?

A. That's correct.

Q. And you add the additional criterion of your VRA constraint, correct?

A. Uh-huh, that's correct.

Q. And that criterion was to keep the overall BVAP proportion of District 6 between 45 percent and 50 percent, correct?

A. That's right.

Q. And in your simulation that district would have been Congressman Clyburn's district since he is the representative of District 6 in the enacted plan, right?

A. That's right.

Q. And why did you add this criterion or constraint?

A. Oh, because in the statewide simulation as

we discussed the district boundary that's been simulated is just between Districts 1 and 6 so as we discussed, you know, the increase in BVAP in District 1 would decrease the BVAP proportion of District 6 because there is only two districts that's being considered in the analysis.

So I wanted to make sure that when, you know, when this BVAP for District 6 is at the certain level in enacted plan if that consideration was incorporated in that simulation analysis my results are still robust to that change.

So by considering other districts it allows me to keep the BVAP proportion of District 6 at the same level, you know, in comparison to the two district analysis where increasing the BVAP in District 1 would automatically just by construction decreases the BVAP proportion of District 6.

Q. And are you aware of whether the General Assembly had a goal to keep District 6's BVAP between 45 percent and 50 percent?

A. No, I don't take into account for any intent of General Assembly in my analysis.

Q. And did -- was this 45 percent to 50 percent District 6 BVAP constraint anywhere in the House or Senate Redistricting Guidelines?

A. Not this specific number, no.
Q. Okay. How did you select this specific number?
A. Oh, just the same range as the one in the enacted plan, which I think is 46-point some percent so.
Q. So you chose this range because it approximately straddles the actual BVAP percentage in enacted District 6?
A. That's correct, yeah.
Q. And do you believe that a district drawn between 45 and 50 percent BVAP complies with the Voting Rights Act?
A. I'm not a lawyer so I can't -- I don't have, you know, legal opinion on that.
Q. Do you have a nonlegal opinion on that?
MR. CEPEDA: Object to form.
BY MR. GORE:
Q. Do you have any opinion on that?
A. The question is legal so I have no opinion on that legal question.
Q. I followed up with nonlegal, but then I said, do you have any opinion on that one way or the other?
A. Well, if the question is legal I don't

have any opinion on that.

Q. All right. Let's go back to page 26 of your report, paragraph 57.

A. Okay. Wait, can you say page number again?

Q. I'm sorry, page 25.

A. Oh, 25.

Q. Page 25, paragraph 57.

A. Twenty-five, 57, okay. All right. Okay.

Q. Okay. So we are back to the strength of the constraint?

A. Uh-huh.

Q. And then the statewide VRA simulation it says: A custom constraint that penalizes plans in which District 6's BVAP is outside the range of 0.45 to 0.5, this constraint is given a strength of eight.

Did I read that correctly?

A. Yes.

Q. So what does it mean to have a custom constraint that penalizes certain plans?

A. Oh, I see. So the custom just means that it's very specific to this particular analysis because you are choosing a specific district, District 6, and choosing a specific range of BVAP so

that's all it means. It's just a constraint but it's sort of a specific constraint used for this particular analysis as opposed to generic, you know, constraint such as compactness or county splits.

And again, the strength of it is all relative so eight itself doesn't really mean anything and it's basically there to ensure that, you know, most of the simulated plan don't have BVAP of District 6 going outside of that range.

Q. Is that a hierarchical constraint or a soft constraint?

A. No, that's a soft constraint.

Q. So within your final set of 10,000 statewide VRA simulation plans, do any place District 6 BVAP outside of that range?

A. So again, similar to the incumbency pairing, there are sometimes that went below 45 percent, a very small number, but that was removed from the final 10,000 simulation plans.

Q. Similarly, were there any that went above 50 percent?

A. I don't think so although I don't recall the detail, but I don't think so. It tends to be lower.

Q. So would you agree that keeping the

overall BVAP proportion of District 6 between 45 and 50 percent was a significant factor in the statewide simulation plans?

A. Can you clarify what you mean by significant factor?

Q. Yeah. I mean, well, I want to know if it's a significant factor. Was it a determinant as to --

Well, first of all, it sounds like every plan in your set of 10,000 satisfies that criteria, correct?

A. Right, right.

Q. And any plan that didn't satisfy that criterion is excluded from your set of 10,000 statewide simulation plans; is that right?

A. That's right. That's right.

Q. So no plan outside of that range could have been included in your set of 10,000 simulation plans, correct?

A. That's correct.

Q. So you within your analysis didn't compromise on that factor, correct?

A. That's right. Compromise meaning -- yeah, that's correct, that constraint was imposed.

Q. And I believe your report says that race

is a significant factor in the enacted plan. And so isn't it also true that this criterion is a significant factor in your statewide simulation?

MR. CEPEDA: Object to form.

THE WITNESS: Right. So my conclusion of the report says for the statewide simulation analysis race was a significant factor beyond the purpose of, you know, keeping this District 6's BVAP in that range.

So the finding that I observed, basically the difference between enacted plan and simulated plan in terms of racial, you konw, factors that cannot be explained by the possibility of keeping District 6 BVAP in this range, that's what I mean. So that's the purpose of the statewide analysis is to see what I found in the local analysis can be explained by the fact that perhaps District 6's BVAP had to be higher because in the localized analysis, you know, if the BVAP increases in District 1 automatically that will reduce the BVAP of District 6.

But in the statewide simulation that's not necessarily the case because I'll make sure that the District 6 BVAP is kept in the same range as the one in the enacted plan. So then the question is, if I

keep -- place that constraint, do I still observe race being -- playing a significant role in determining district boundaries between one and six and then elsewhere.

Q. But in terms -- I'm asking about how the simulation plans are drawn?

A. Uh-huh.

Q. What you described to me makes it sound like this BVAP range in District 6 was a significant factor in how those plans were drawn, correct?

A. It is a significant constraint in the sense that if I don't have that particular constraint the results would have been different, but my conclusion basically says even if you place that constraint the race played a significant role beyond that particular constraint.

Q. Well, and isn't it also a significant constraint because any plan that didn't satisfy that constraint was excluded from your set of 10,000 plans, right?

A. That's right. So if I don't impose this particular constraint the results would have been different sometimes the order -- the analysis be incorporated and so on.

Q. And so this VRA -- satisfying the VRA

constraint was the price of admission for a plan to be part of your 10,000 plans in the simulation set, right?

A. What do you mean by the price of admission?

MR. CEPEDA: Object to form.

BY MR. GORE:

Q. That's a colloquialism. But let me just say it this way. Only plans that satisfied that criterion were included in your set of 10,000, correct?

A. That's correct.

Q. Okay.

A. But there are plans that may satisfy that constraint but still were excluded for other reasons.

Q. Sure, because you have other constraints programmed?

A. Yeah, other constraints, yeah, okay. I just wanted to make sure.

Q. I think I understand now.

A. That's not the only -- that's not the -- like, you know, constraint that dominates everything else.

Q. Did you examine whether race was a

significant factor in any of your statewide simulation plans?

A. Can you repeat that question again?

Q. Sure. Did you consider or examine whether race was a significant factor in any of your statewide simulation plans?

A. Right, that's what I did.

Q. Well, help me understand that a little bit. I think that you said that you considered -- you used the simulation plan to determine whether the enact -- race was a significant factor in the enacted plan?

A. Oh, simulated plan.

Q. Between the simulated plans did you look at any of those plans to determine whether race was a significant factor in any of those plans?

A. Okay. I use the simulated plan to evaluate the enacted plan, not the other way around so...

Q. Right. But did you compare any simulated plan to the set of simulated plans to determine whether any of those plans --

A. Oh, no, I did not evaluate the simulated plans.

Q. Okay. And did you conduct any of that

kind of analysis with respect to any of the plans submitted by members of the public?

A. No.

Q. So you didn't do it for the NAACP Plan One?

A. No.

Q. Or Plan Two?

A. I don't even know what these plans are so...

Q. Okay.

A. Yeah.

Q. You didn't do it for the League of Women Voters' plan?

A. No.

Q. Or Harpootlian plan?

A. No.

Q. All right. Let's flip back to pages -- maybe start around page 15.

A. Okay.

Q. In the statewide simulation here.

A. Okay, 15. Okay, hold on.

Q. Starting on page -- paragraph 35.

A. Okay.

Q. So here, again, you are analyzing the district boundary in Charleston County, correct?

A. That's right.
Q. And that's -- the district boundary running through there is Districts 1 and 6, correct?
A. Uh-huh, that's right.
Q. And it says here -- so again, in this analysis -- let's flip over to Figure 5.
A. Yeah.
Q. Okay. And Figure 5 you have again the cool colored chart that shows the likelihood of a VTD or precinct being placed in District 1; is that right?
A. That's correct.
Q. And again, this is the total number -- do you have -- do you show -- I guess you don't show anything about BVAP here, right? This is just likelihood of being placed in District 1; is that right?
A. That's correct.
Q. And again, we are seeing that in your model Nancy Mace's portion of Berkeley is being placed very frequently with City of Charleston or a large swath of Charleston County; is that right?
A. Again, I don't have a precise location of her region so I think that's probably right.
Q. Okay. And again, in this analysis you

didn't control for which district the VTD was in the benchmark plan, correct?

A. Right, I did not use benchmark plan.

Q. Now, I think in paragraph 35 you say: On average the BVAP proportion of District 1 under the enacted plan is 6.5 percentage points, 4.5 standard deviations of the simulated distribution lower than the corresponding number under the simulated plans.

A. Uh-huh.

Q. Did I read that correctly?

A. Yes.

Q. So where -- in the simulated plans where is that 6.5 percentage points on average coming from?

A. Yeah, I think my guess, again, I don't have exact numbers but, you know, it's from the City of Charleston and perhaps North Charleston.

Q. And so when it's being moved into District 1 or placed in District 1 under your simulation plans it's necessarily being placed outside of other districts, right?

A. Right.

Q. So it may be coming from District 6, maybe in some of the plans it's coming from District 7 even; is that possible?

A. That's possible. Probably from, you know, District 6 in this case.

Q. Yeah, but probably District 6 because District 6 has black voters and District 7 doesn't have as many, right?

A. Right. That's correct and also District 6 is the part that will be, you know, assigned with high probability to District 1 in the Charleston County area.

Q. And so again, in your simulation, statewide simulation --

A. Uh-huh.

Q. Those plans also may crack black voters when they split communities where black voters live just like you localized simulations we discussed before, correct?

MR. CEPEDA: Objection; mischaracterizes testimony.

THE WITNESS: What do you mean by those plans? The simulated plans?

BY MR. GORE:

Q. Yeah, so I think earlier we talked about your definition of cracking.

A. Uh-huh.

Q. And how in changing the line between

District 1 and District 6 various black communities could be cracked in any individual simulation plan?

MR. CEPEDA: Object to form.

BY MR. GORE:

Q. Do you recall that?

MR. CEPEDA: Object to form.

THE WITNESS: It's not any individual -- it's just a description of how the district boundary, you know, runs through a specific group of voters who live in the same area.

BY MR. GORE:

Q. Right. And yeah, so I'm using your -- I'm trying to use your definition of cracking.

A. Okay.

Q. And I think you said that the enacted plan cracks black voters between Districts 1 and 6?

A. Uh-huh.

Q. Because of how the line runs through Charleston or may or may not run through Charleston, correct?

A. Right, right.

Q. And some of the simulation plans also may crack black voters because of how the lines run through certain communities, correct?

A. Right, but my finding is that the

simulated plan tend to keep those black voters who are, you know, cracked by enacted plan tend to keep them together.

Q. Correct, but they could possibly -- it could possibly -- the simulation plans could crack that community of black voters or some other community of black voters somewhere else, correct?

A. It's possible but unless you tell me which area of black voters, I cannot tell whether the simulated plan or enacted plan cracks that community of voters.

Q. And did you review any of the public testimony or comment in the legislative record for this plan?

A. No.

Q. And did you review any statements by legislators?

A. No.

Q. So paragraph 36 says: The way in which the enacted plan splits Charleston County by displaying a disproportionately large number of black voters into District 6 is highly unusual in comparison to the simulated plans. Is that right?

A. Uh-huh, that's correct.

Q. And it seems like a further explanation

concerns the placement of the City of North Charleston; is that correct?

A. Yeah, that's one.

Q. So if a simulation plan split Charleston County or North Charleston in a way that affected where black voters were placed that plan would also crack voters, correct?

A. Right, but my finding is that simulated plan tends to keep them together in one district.

Q. So let's go to Figure 6 on page 17.

A. Right.

Q. And will you explain to me what Figure 6 is showing?

A. So Figure 6 shows that a number of, you know, the black voting age population who lives in Charleston County were assigned to District 1. Under the simulated plan which is the bar and the enacted plan which is the vertical line.

Q. And so the difference between the enacted plan and any of the simulated plans is the difference between the line for the enacted plan and the bar or whatever for the simulated plan, correct?

A. That's right, that's correct.

Q. And so in the simulation where are those extra black voters coming from?

A. Well, these are all black voters who live in Charleston County.

Q. Okay. And so they otherwise are being placed in District 6?

A. Right, the ones -- that's probably right although I haven't, you know, examined every simulated plan.

Q. And so these black voters are being placed in District 1 and not in District 6 or some other district, correct?

A. Right, so this shows the black voters who are assigned to District 1 on the simulated plan.

Q. Okay. And did you generate a similar histogram for what happens with black voters in District 6?

A. Black voters in Charleston County or black voters in general?

Q. How about Charleston County since that's what you have here in Figure 6?

A. Yeah, I didn't create that because in most simulated plans the rest of the black voters who live in Charleston County would be placed in District 6, I think.

Q. Okay. So at least with respect to most of the simulation plans the enacted plans places more

black voters from Charleston County in District 6 than the simulation plan, correct?

A. I think that's right. Although, you know, again, it's possible that some other districts come in play and I have not checked every -- every district to make sure that's, you know, that's -- how often that happens.

Q. But each district in your simulation plan is constrained to some extent by where the incumbent lives, correct?

A. Oh, yes, because I tried to avoid the incumbent pairing so that's right.

Q. So for another district to move into Charleston your simulation plan would have to somehow still be contiguous ultimately?

A. That's right.

Q. To the precinct where in incumbent resides, right?

A. Yeah, I think for any simulated plan for District 6 it would be definitely in play and maybe additional districts like District 7 may also play on the eastern side but most likely it would be District 6.

Q. And given that it's most likely that the enacted plan places more black voters in District 6

than the simulation plans, right, in Charleston County?

A. Oh, in Charleston County, yes, because you know, that's just -- there is a fixed number of black voters in Charleston County and if there are more in District 1 than there is in the rest of the other districts.

Q. And so paragraph 37 --

A. Uh-huh.

Q. You say there is a sentence that begins, in fact: In fact, a large spike around 74,600 implies a vast majority of simulated plans, 76.3 percent assigned the entire county to District 1.

Did I read that correctly?

A. Yeah, that's correct.

Q. Do you know whether Charleston County was split in the benchmark plan?

A. Oh, in the benchmark plan. In the enacted plan it was but I don't know the benchmark plan.

Q. And if a map drawer preferred to keep Charleston County split he would have rejected plans that made it whole in District 1, correct?

A. Again, I don't have any opinion on how or why map drawer made certain decisions.

Q. And again, you didn't review any public testimony, comment or legislative testimony about splitting or repairing the split in Charleston County, correct?

A. No.

Q. And did you analyze the political effect of placing all of Charleston County in District 1 with Nancy Mace?

A. I did not use any partisan data in my analysis.

Q. And did you analyze what changes to the map would have been required in other parts of the state if all the Charleston was placed in District 1?

A. Can you repeat the question again? Sorry.

Q. Sure. So if you -- Charleston County, if you place Charleston County in District 1?

A. Uh-huh.

Q. In the enacted plan, you would have to make changes to other districts in order to equalize population, correct?

A. That's correct.

Q. All right. And did you do any analysis of that other than to recognize if that's true?

A. Yeah, that's true but I didn't do any

specific analysis with -- regarding that particular point.

Q. Now, your statewide simulation analysis with respect to District 1 and District 6 you have this discussion of Charleston County, right?

A. Uh-huh.

Q. Did you also examine in your statewide simulation analysis Dorchester County?

A. No.

Q. How about Colleton County?

A. No.

Q. Or Jasper County?

A. No.

Q. Okay. Why did you not examine Dorchester County?

A. Well, so my -- you know, analysis, the entire analysis for this report started with examining, you know, District 1 and 6, that's the first part of the analysis District 1 and 6 boundaries and what I found there is that it's Charleston County where I see the largest difference in terms of the role that the race played between the simulated plan and enacted plan.

And so what I wanted to then do is, you know, that is really the start of the analysis.

And the other two analyses, the second race analysis and statewide analysis, is to make sure that the results are robust to other factors.

So for the second race run analysis I wanted to make sure that it was actually the Charleston, the way Charleston County was split, not necessarily, you know, other parts of Districts 1 and 6 so that's why I froze the other parts of the Districts 1 and 6.

And for this analysis I wanted to make sure that it wasn't due to this particular constraint about keeping the BVAP proportion for District 6 at a certain level so each of these analyses sort of tried to look at the robustness of the initial analysis that I conducted where I show that there are just differences coming from District 1 and District 6 boundaries.

Q. And do you know one way or the other whether Dorchester, Colleton and Jasper are split between Districts 1 and District 6 in the enacted plan?

A. I don't recall which other counties are being split.

Q. Let's turn to page 18 of your report.

A. Okay.

Q. And we'll discuss the line here in Richland County.

A. Uh-huh.

Q. Between District 2 and District 6.

A. Sure.

Q. First of all, do you know whether Richland County was split in the benchmark plan?

A. Yes, you can see it's -- oh, in the benchmark plan I don't know. In the enacted plan they are split.

Q. And you see the -- if we look at Figure 7.

A. Right.

Q. District 2 and District 6 are shown here with a line in what appears to be Richland County; is that correct?

A. Yeah, Richland County is demarked by the gray, thick line there.

Q. And are you aware of any explanations for the shape of the boundary between Districts 2 and 6 in Richland County?

A. No.

Q. And again, you have not reviewed the public record of the legislative record on that, have you?

A. No.

Q. So let's go to paragraph 40.
A. Okay.
Q. I believe it's actually on page 19.
A. Okay.
Q. You start: In fact, 39.4 percent of the simulated plans did not split Richland County at all and all the simulated plans assign the entire county to District 6.

Did I read that correctly?
A. That's correct.
Q. And if a map drawer would have preferred to keep Richland County split, he would have rejected plans that make it whole in District 2, correct?
A. That's right. Although I don't know what, you know, the map drawer considered.
Q. And did you analyze the political effect of placing all of Richland County in District 6?
A. No, I didn't use partisan information if that's what you mean by political.
Q. Thank you. Or any election result information?
A. No.
Q. All right. So page 19 contains Figure 8?
A. Uh-huh.

Q. Would you explain to me what Figure 8 is about?

A. Right, so this is similar to the previous figure where I look at, you know, among the black voting age population who live in Richland County how many of them are assigned to District 2.

Q. And is this a total number of black voting age population or percentage?

A. Yes, that's right. And this is among the plans that split into Districts 2 and 6 so it does not include the one that assigned the entire county to the district.

Q. Okay.

A. So it's focused on the subset.

Q. And so according to this, this shows that district -- that the enacted District 2 places more individuals of black voting age -- black individuals of voting age from Richland County in District 2 then the average statewide simulation plan that also splits Richland between Districts 2 and 6, correct?

A. That's right, that's right.

Q. So what would a histogram of District 6, the corresponding histogram of District 6 look like to Figure 8?

A. Oh, I see. Well, as I said, you know,

40 percent, about 40 percent of the simulated plans assigned the entire Richland County to District 6 so, you know, it -- in that case, you know, all the BVAP population would be in District 6 and even in the cases where the simulated plans split Richland County, they assigned, you know, much fewer BVAP to District 2, excuse me, than the enacted plan. So yeah, I think a lot of Richland County BVAP would be in Richland County, in District 6.

Q. District 6?

A. Yeah.

Q. So if I can just summarize that.

A. Yeah.

Q. The simulation plans assigned lower BVAP to District 2 but higher BVAP to District 6 compared to the enacted districts in those simulation plans that split Richland County between Districts 2 and 6; is that correct?

A. That's correct, that's exactly right.

Q. And if we were to show that as a histogram to District 6 it would be the mirror image of what we see here in Figure 8 accounting for small population variations in your model?

A. That's probably right. Except I don't know how often the other districts are split but my

sense is that most of the simulated plans involve Districts 2 and 6 for this county so that's probably correct although I don't have, you know, exact numbers with me.

Q. Well, I'm glad you raise that because I think paragraph 40 says that 39.4 percent of the simulated plans don't split Richland at all?

A. Right, right. So in that case all --

Q. So in that case -- right, and then paragraph 41 says: 23.9 percent of simulated plans divide Richland into two and six?

A. Right, so there are some.

Q. Yeah, and so I understood Figure 8 to relate only to that 23.9 percent referenced in paragraph 41, right?

A. That's exactly right. That's why it's 2388 plans.

Q. I've got it.

A. Yeah. So there are some districts -- sorry, there are some simulated plans that would involve some other districts.

Q. So there is a third category of simulated plans that split Richland between two districts but not two and six?

A. Or three and six. Possibly three and six.

Yeah, I don't know, yeah.

Q. Yeah, you just don't know?

A. Right.

Q. But that's a category?

A. Yeah, yeah.

Q. But at least the 23.9 percent split between two and six?

A. That's right.

Q. We have Figure -- so we have Figure 8 for the District 2 BVAP and the District 6 BVAP would be the mirror image of that simply control for population variation that your model tolerates, right?

A. That's right. For this subset, yes.

Q. For this 23.9 percent subset?

A. That's right. That's right.

Q. Okay. Now, you mind flipping back to page 15, Figure 4?

A. Okay.

Q. So when you are discussing in your statewide simulation the line between District 1 and 6 in Charleston you provided us this BVAP proportion table, the histogram?

A. Uh-huh.

Q. But I didn't see a similar BVAP proportion

histogram for any district in your discussion of Richland County, so back on pages 18 or 19 --

A. Uh-huh.

Q. There wasn't the same kind of histogram.

A. Of this district you are talking about?

Q. For District 2 or District 6.

A. Oh, I see. Well, District 6 BVAP proportion is constrained to between 45 and 50 so that's not, you know, that's just by design constrained. District 2, yeah, I don't show BVAP proportion.

Q. Is there a reason for that?

A. Because here I'm just focusing on how unusual it is -- just like for this case how unusual it is to split this particular county in a certain way.

Q. And have you examined District 2's BVAP in your statewide simulation plans?

A. No, because that's not the goal of this particular analysis here.

Q. And have you examined whether your statewide simulation plans or any particular statewide simulation plan cracks black voters in Richland County?

A. Can you repeat that question again, sorry.

Q. Sure. Using your definition of cracking?
A. Uh-huh.
Q. Have you examined whether any of your statewide simulation plans crack black voters in Richland County?
MR. CEPEDA: Object to form.
THE WITNESS: I mean, compared to the enacted plan, the simulated plan tends to keep them together. Like for example, like 40 percent of them keep all the counties together in the same district so that's -- there is no split at all. And even in other 26.9 percent of the cases, you know, they tend to -- most of the BVAP population tend to be on District 6 so that's what this is showing with respect to the cracking of black voters in Richland County.
BY MR. GORE:
Q. And is it possible as we discussed before that the simulation plan cracks black voters in Richland County with how it draws the district line through that county and through a black community?
MR. CEPEDA: Object to form.
THE WITNESS: Are you talking about the remainder of the simulated plans beyond the 40 percent and 24 percent of them or are you --

BY MR. GORE:

Q. Or even the 24 percent. I mean, I take it that there is a 39.4 percent that doesn't crack in Richland at all.

A. Right, that's right.

Q. But I'm just asking and I think you've answered this a few times now and I don't mean to belabor the point, but more broadly speaking under your definition of cracking, cracking is simply splitting a community of black voters into more than one district, correct?

A. Right, but there is a different degree of cracking, right, you know, what's the percentage or what's the, you know -- like it's one thing to split a community into two districts by taking one person out of that community but it's another thing to split the community into say two halves.

Q. And where in your report do you discuss the degree of cracking?

A. So this Richland County BVAP basically shows that, like this analysis shows, that say 40 percent of the cases they are in, you know, basically 100 percent is in one district and even in the 33.9 percent of the case that there are a lot fewer BVAP being assigned to District 2. A lot of

them very close to zero so that means that most of them, most of the black voters would be in District 6 and those -- that would spike that's almost accounting for 50 percent out of that 24 percent.

Q. And at the same time in Figure 8 there is some simulation plans that place more black voters from Richland in District 2 than the enacted plan, correct?

A. Right. So this is statistics knowledge, this you really want to look at the distribution of, you know, simulated plans as opposed to focusing on one and two specific simulated plans.

Q. And --

A. And that's the goal of the analysis, the statistics analysis.

Q. Let's step back from Richland County for a moment. I want to ask just a more global question --

A. Sure.

Q. About the simulation. Is it possible that plans in your statewide simulation crack black voters by splitting their communities into more than one district?

A. It depends on the community. So I don't

know unless you specify specific communities I won't know whether simulated plan would split that community.

Q. But you didn't program the simulation, for example, with a constraint against splitting black communities, did you?

A. No but to the extent that, you know, keeping the BVAP proportion of District 6 to be in certain race that definitely encourages certain number of black voters to be placed in District 6 to the extent that I try to, you know, reduce the number of county splits, municipalities splits and depending on how the black voters are, you know, there are geographical distribution that may have an impact on whether those communities are being split and if they are how.

Q. So in the enacted plan are there any other counties that are split between Districts 2 and 6 other than Richland?

A. Oh, yeah, I'm sure there are. Yeah, I don't recall exactly what those counties are.

Q. Okay. Did you analyze the district boundary between Districts 2 and 6 in Orangeburg County?

A. I don't think so.

Q. And why not?

A. I don't recall the reason. I mean, the draft report may have included. Again, I don't recall, you know, whether I did it or whether that was included in the draft reports or, you know, if I didn't do it why I didn't do it. I don't -- I don't recall.

Q. Let's move to your analysis of Sumter County.

A. Okay.

Q. Which is split between Districts 5 and 6 of the enacted plan; is that correct?

A. Yes, that's correct.

Q. And do you know why Sumter -- well, first of all, do you know whether Sumter was split in the benchmark plan?

A. No, I don't know.

Q. And do you know why Sumter was kept split in the enacted plan?

A. No, I don't know.

Q. Paragraph 43 --

A. Uh-huh.

Q. And Table One shows that 90.3 percent of simulated plans make Sumter whole and place it in District 6; is that correct?

A. Yes, that's correct.
Q. And if a map drawer preferred to keep a portion of Sumter County in District 5 he would have rejected all of those plans, correct?
A. Right. Although I didn't consider the map drawers in my analysis.
Q. And you also didn't consider any public testimony or comment or the legislative record on this split, did you?
A. No, I didn't.
Q. And did you analyze the political effect of placing all of Sumter County in District 6?
A. No, I didn't use partisan information or election results.
Q. And here when you are discussing Sumter County you didn't include any histogram showing BVAP numbers or BVAP proportions?
A. Uh-huh.
Q. What is the number of black individuals of voting age implicated by the split in Sumter County, do you know?
A. Oh, I don't know off the top of my head, but I didn't include them because, you know, 90 percent of the whole county is in District 6 so it's obvious that simulated plan keep those people

together most of the time.

Q. And did you analyze the effect of keeping, placing Sumter County whole in District 6 on any district's BVAP proportion or total BVAP?

A. Oh, how that affects the BVAP proportion of other districts?

Q. Or any district.

A. Okay. I didn't do that. I didn't look at that.

Q. All right.

A. Although by construction, again, the District 6 will maintain that specified level of BVAP proportion so I know that it does not affect the BVAP proportion of District 6.

Q. And within your model District 6 could have a range of BVAP between 45 and 50 percent, right?

A. That's correct, yeah.

Q. And so if District 6's BVAP in the simulation plan is 50 percent that means accordingly that there is 3 percent less BVAP for other districts, correct?

MR. CEPEDA: Object to form.

THE WITNESS: 3 percent of the District 6 population, right.

BY MR. GORE:

Q. So in the --

A. Come from other districts, is that what you mean?

Q. Yeah, so in the enacted plan District 6 BVAP is I think 46.9 percent; does that sound about right?

A. I trust you.

Q. Okay. Let's just for -- and this is just a conceptual point so the numbers are not as important, but a simulated plan with a 49.9 percent BVAP means that that's 3 percent higher than the enacted plan, correct?

A. Uh-huh.

Q. And that 3 percent is being drawn from other districts other than District 6, correct?

A. That's correct but we should also remember that ranges go from 45 to 50 and there are more districts, there are more simulated plans that are on the lower end so I had to remove some of the plans that don't reach, excuse me, don't reach the 45 percent and so there may be a district that has BVAP of, you know, close to 50 percent for District 6 and in a simulated plan, there is also a lot of plans that are closer to the BVAP proportion

for District 6 that is closer to 45 percent and less than enacted plan number.

Q. And do you know the average BVAP proportion for your 10,000 plans in the statewide VRA simulation?

A. I don't recall.

Q. All right. I'm going to mark a new exhibit.

A. Okay.

Q. It's tab eight in your zip file.

A. Yeah.

Q. I'm going to mark it as Exhibit Eight.

(Defendant's Exhibit No. 8, REBUTTAL REPORT OF SEAN P. TRENDE, was marked for identification.)

THE WITNESS: This is the rebuttal.

BY MR. GORE:

Q. Yes. It's the rebuttal report by Mr. Trende?

A. Yes.

Q. Dr. Imai, do you recognize this document?

A. Yes, I was given this document by counsel.

Q. And when were you given it by counsel?

A. Oh, I don't recall the exact date. I think it's very recent. Maybe two weeks ago or something along those lines, but yeah, I don't

recall exact date.

Q. Okay. But it was -- I believe the date of this report is, I'll scroll to the last page, May 4th?

A. May 4th, okay.

Q. So you don't recall receiving it in May 2022 it sounds like?

A. Yeah, I just -- like I think it was two or three weeks ago when counsel sent me as a part of the exhibits. I didn't remember and so that doesn't necessarily mean I didn't have it, but I just don't pay attention, I guess, if I had it.

Q. Okay. And you've not offered an opinion on any of Mr. Trende's analysis in this case, correct?

A. No, I didn't. I did not.

Q. And you have not offered an opinion on any of his analysis in his rebuttal report?

A. No, I did not and I also wasn't asked so...

Q. Okay. And did you attempt to recreate the numbers that Mr. Trende or the statistics that Mr. Trende provides in his rebuttal report?

A. No, I did not attempt.

Q. And at this point do you plan to do so?

A. I don't think so.

Q. Okay. And as you sit here today, do you dispute any of the calculations in Mr. Trende's report?

A. I haven't formed an opinion on this so I can't -- I can't make any judgment about accuracy or validity of his analysis.

Q. And likewise, you haven't formed any opinion about his rebuttal report, correct?

A. Wait, are you talking about the previous report or the rebuttal report? I'm a little confused.

Q. I was talking about the rebuttal, but I wanted to make the record clear.

A. Oh.

Q. I think I called it his report rather than his rebuttal report.

A. Okay.

Q. So I'm glad that you and I were equally confused by my poor question. So let me ask the question a little bit more cleanly for the record if that's okay.

A. Sure, sure.

Q. Have you formed any opinions about Mr. Trende's rebuttal report?

A. No.

Q. Or any of the statistics or numbers he provides in his rebuttal report?

A. No, I have not formed opinions on those.

Q. So on page 2 of this rebuttal report through page 5 Mr. Trende discusses core retention and offers a variety of statistics about core retention.

A. Uh-huh.

Q. And so you have not formed -- you have not analyzed or formed any opinions about Mr. Trende's discussion of core retention in this rebuttal report, correct?

A. I have not analyzed nor formed opinion on that.

Q. And Mr. Trende, starting on page 5 through page 8 discusses politics and partisanship?

A. Uh-huh.

Q. And you have not formed any opinions or done any analysis on Mr. Trende's statistics here either?

A. No. Yeah, I have not done that and I haven't formed opinion on that.

Q. And similarly, on page 8 Mr. Trende refers to movement of citizens or voters to repair precinct

splits and you haven't analyzed or formed an opinion about that either, have you?

A. No, yeah.

MR. GORE: Let's go off the record here for a moment.

(A recess was taken.)

MR. GORE: Dr. Imai, did you discuss your deposition with anyone during the break?

THE WITNESS: No.

MR. GORE: Dr. Imai, thank you for your attentiveness today. I have no further questions. I'm now passing the witness.

Andrew or Ms. Crum, anyone have questions for Dr. Imai?

MS. CRUM: This is Liz Crum. No, we have no questions on behalf of the Election Commission defendants.

MR. MATHIAS: This is Andrew Mathias, on behalf of the House Defendants. I have no questions.

MR. CEPEDA: Thank you. I guess, I have a few questions.

EXAMINATION

BY MR. CEPEDA:

Q. Dr. Imai, just a second. Dr. Imai, you

testified that you drafted a single report on the Congressional map in this case, right?

A. There is only one final report, yes.

Q. So from the time that you started working on simulations having to do with the Congressional map you meant those to form part of the same report, right?

A. Right, right.

Q. And did you mention at some point an early draft included simulations that had to do with the benchmark map, right?

A. Right, I think there may have some simulation results.

Q. Yeah, and it sounded like you mentioned that you may have discussed that draft with your lawyers, right?

A. That's correct.

Q. With me, for example?

A. Yeah, I think you and I know Patricia was there, Ms. Yan was there, and there may have been some others.

Q. I've got it. And the lawyers or can I recall we gave you our impressions and shared our thoughts of what we were seeing?

A. Right.

MR. GORE: Object to the form.

BY MR. CEPEDA:

Q. And at some point I believe you testified those simulations involving the benchmark plan got cut from later drafts, right?

A. That's correct.

Q. But you didn't share those drafts with anyone besides plaintiffs counsel in this lawsuit or their staff, right?

A. That's correct.

Q. Dr. Imai, earlier Mr. Gore asked you whether you had looked at any maps proposed by the public and you noted that you don't even know what those maps are; am I right?

A. Right. I did not -- yeah, I don't know what they are.

Q. Is it possible that you could have looked at other maps that were not enacted at some point even if you don't recognize them now by name or label?

A. Oh, hum. Yeah, I don't -- you know, I don't recall looking at them. I guess that's -- but, you know, that's based on my recollection.

Q. That's fine. Thank you. But were any un-enacted maps discussed in your final report?

A. No, I look at the enacted plan.

Q. Thank you. You answered my next question. So Dr. Imai, you base your simulation constraints on the published South Carolina guidelines for the House and Senate, right?

A. Yeah, I don't know whether they are published but those two guidelines that were given to me.

Q. And in those two guidelines was there any indication, for example, that core preservation should be prioritized over other criteria?

A. No. I believe that it was listed as additional constraint in Senate guideline I think and may not be even directly mentioned in the House guideline or at least it was not priority, listed as a priority.

Q. Thank you. And you testified -- well, why did you choose not to incorporate core preservation, if you can explain again?

A. Right. So the goal of my analysis, the entire report, the goal of the entire report was to examine whether race played a significant role in drawing district boundaries of the enacted plan and, if so, how that happened. And to do that I need to isolate the impact of race, like the role that race

played from other traditional redistricting criteria and some of the rules in the -- mentioned in the guideline.

If I incorporate any product does not have to be benchmark plan, but if I incorporate any plan in my simulation analysis, it will basically carry all the factors that went into that particular plan. So in order to isolate the race as a factor I did not use this through my analysis that I did not use any plan including the previous plan.

Q. Thank you. Now, you recall Mr. Gore asked you some questions about the use of partisanship data in your simulation, right?

A. Yes.

Q. And you explained that you didn't do any -- you didn't use partisanship information; is that right?

A. Right.

Q. And we just covered this, but you read the guidelines, right?

A. Uh-huh, yes, I did.

Q. Did anything in the guidelines suggest to you that your simulation should have accounted for Nancy Mace's election chances, for example?

A. I didn't see any mention of that. Yeah, I

did not see any specific instruction about use of election outcomes.

Q. Did anything suggest to you that it was important for the map makers to enact a map that favored Republicans?

A. I don't analyze intent of map drawer so I can't, you know, say what they have thought about but the guideline didn't specify, you know, specific use of electoral outcome or electoral chance of politicians and that wasn't, you know, even -- a political consideration wasn't an additional consideration and so I took other more traditional redistricting criteria as priority.

MR. CEPEDA: Thank you, Dr. Imai. I have no more questions.

EXAMINATION

BY MR. GORE:

Q. I have just a couple of questions of redirect, Dr. Imai.

A. Okay.

Q. Now, you said you haven't attempted to analyze the intent or motives of the map drawer or legislators, correct?

A. That's correct.

Q. And so you don't have an opinion one way

or the other as to whether the map drawer or the legislators considered politics even if politics is not in the guidelines, correct?

A. That's right. I don't have any opinion on that.

Q. Do you have a view or opinion on whether the map drawer or the legislators considered Nancy Mace's reelection prospect whether or not that's listed in the guidelines?

A. No, I don't have any opinion on that.

Q. And do you have any opinion or view on whether the map drawer or legislators wanted a plan that would elect six Republicans regardless of whether that's in the guidelines?

A. I don't have any opinion on that.

Q. And Dr. Imai, is keeping Charleston in a single district anywhere in the guidelines?

A. I don't think so, there is no specific counties being mentioned.

Q. How about keeping Richland in a single district?

A. I don't think so.

Q. And how about keeping District 6's BVAP between 45 percent and 50 percent?

A. Those numbers are not specifically

mentioned but the VRA consideration is listed as one of the principles that should be followed.

Q. And you don't have a view, do you, as to whether a District 6 between 45 percent and 50 percent BVAP complies with the VRA, right?

A. I'm not a lawyer so I don't make legal judgment.

Q. And you haven't conducted any racially polarized voting analysis in this case, correct?

A. No, that's right. I did not conduct racial polarize analysis.

MR. GORE: Thank you, Dr. Imai. I have no further questions.

MR. CEPEDA: I think that might be it. That's it unless Liz or Andrew -- I've spoken out of turn but I hope not.

MR. MATHIAS: I'm not going to disappoint you. No questions.

THE COURT REPORTER: Dr. Imai, will you read and sign? If so, where can we email that?

MR. CEPEDA: He will and you can send it to me.

THE COURT REPORTER: To you, okay.

All right. Mr. Gore, regular turnaround which is the 22nd; is that correct?

MR. GORE: What can you do as an expedited turnaround?

THE COURT REPORTER: So I can get it to you by Monday or do you need it before that? I can also provide you with a rough draft.

MR. GORE: Yeah, Monday would be fine and we are happy to take a rough too, I guess.

THE COURT REPORTER: Okay.

Mr. Cepeda, would you also like it expedited and a rough draft?

MR. CEPEDA: Monday would be good for me and we can work the details.

THE COURT REPORTER: I don't know what you mean. Do you want a rough tomorrow or no?

MR. CEPEDA: I would like the rough too, sure.

THE COURT REPORTER: Okay. Do you want your copy Monday too?

MR. CEPEDA: I don't need it by Monday. Does that have any additional cost?

THE COURT REPORTER: Yes, it does.

MR. CEPEDA: So when would I get the regular?

THE COURT REPORTER: That would be the 22nd.

MR. CEPEDA: No, we want it before that so I'll take it Monday as well.

THE COURT REPORTER: Okay.

Ms. Crum?

MS. CRUM: We do not need a transcript.

THE COURT REPORTER: Okay.

Mr. Mathias?

MR. MATHIAS: I guess it is peer pressure, but we will take it Monday.

THE COURT REPORTER: And do you want a rough?

MR. MATHIAS: Yes.

THE COURT REPORTER: Okay. Thank you.

(The deposition concluded at 4:32 p.m.)

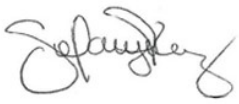
(The witness, after having been advised of his right to read and sign this transcript, does not waive that right.)

CERTIFICATE OF REPORTER

I, Solange Ruiz-Uribe, Notary Public for the State of South Carolina at Large, do hereby certify that the foregoing transcript is a true, accurate, and complete record.

I further certify that I am neither related to nor counsel for any party to the cause pending or interested in the events thereof.

Witness my hand, I have hereunto affixed my official seal this 8th day of August, 2022 at Fort Mill, York County, South Carolina.



_____

Solange Ruiz-Uribe

My Commission expires

February 2, 2027

I N D E X

WITNESS EXAMINATION	Page/Line	
KOSUKE IMAI, PhD	4	1
EXAMINATION		
BY MR. GORE	4	3
EXAMINATION		
BY MR. CEPEDA	177	23
EXAMINATION		
BY MR. GORE	182	16
SIGNATURE OF DEPONENT		
CERTIFICATE OF REPORTER	187	1

REQUESTED INFORMATION INDEX

(No Information Index Requested)

E X H I B I T S

		Page/Line	
DEFENDANT'S EXHIBITS			
EXHIBIT 1	SENATE DEFENDANTS' AMENDED NOTICE OF TAKING VIDEO CONFERENCE DEPOSITION OF KOSUKE IMAI, PHD	9	20
EXHIBIT 2	EXPERT REPORT OF KOSUKE IMAI, PH.D. DATED APRIL 4, 2022	16	9
EXHIBIT 3	JUNE 2022 CURRICULUM VITAE OF KOSUKE IMAI	19	3
EXHIBIT 4	SEQUENTIAL MONTE CARLO FOR SAMPLING BALANCED AND COMPACT REDISTRICTING PLANS PAPER	34	3
EXHIBIT 5	RECOMBINATION: A FAMILY OF MARKOV CHAINS FOR REDISTRICTING ARTICLE	65	9
EXHIBIT 6	SOUTH CAROLINA HOUSE OF REPRESENTATIVES JUDICIARY COMMITTEE REDISTRICTING AD HOC COMMITTEE 2021 GUIDELINES AND CRITERIA FOR CONGRESSIONAL AND LEGISLATIVE REDISTRICTING	68	16
EXHIBIT 7	2021 REDISTRICTING GUIDELINES SOUTH CAROLINA SENATE JUDICIARY COMMITTEE REDISTRICTING SUBCOMMITTEE	69	14
EXHIBIT 8	REBUTTAL REPORT OF SEAN P. TRENDE	173	13

Adriel Cepeda
acepedaderieux@aclu.org

August 15, 2022

RE: South Carolina State Conference Of The NAACP And Scott, Taiwan v. Mcmaster, Henry, Et Al.

8/8/2022, Kosuke Imai, PhD (#5350506)

The above-referenced transcript is available for review.

Within the applicable timeframe, the witness should read the testimony to verify its accuracy. If there are any changes, the witness should note those with the reason, on the attached Errata Sheet.

The witness should sign the Acknowledgment of Deponent and Errata and return to the deposing attorney. Copies should be sent to all counsel, and to Veritext at cs-southeast@veritext.com

Return completed errata within 30 days from receipt of testimony.

If the witness fails to do so within the time allotted, the transcript may be used as if signed.

Yours,

Veritext Legal Solutions

South Carolina State Conference Of The NAACP And Scott, Taiwan v. Mcmaster, Henry, Et Al.

Kosuke Imai, PhD (#5350506)

E R R A T A S H E E T

PAGE_____ LINE_____ CHANGE_____

_____

REASON_____

PAGE_____ LINE_____ CHANGE_____

_____

REASON_____

PAGE_____ LINE_____ CHANGE_____

_____

REASON_____

PAGE_____ LINE_____ CHANGE_____

_____

REASON_____

PAGE_____ LINE_____ CHANGE_____

_____

REASON_____

PAGE_____ LINE_____ CHANGE_____

_____

REASON_____

PAGE_____ LINE_____ CHANGE_____

_____

REASON_____

_____ _____

Kosuke Imai, PhD Date

South Carolina State Conference Of The NAACP And Scott, Taiwan v. Mcmaster, Henry, Et Al.

Kosuke Imai, PhD (#5350506)

ACKNOWLEDGEMENT OF DEPONENT

I, Kosuke Imai, PhD, do hereby declare that I have read the foregoing transcript, I have made any corrections, additions, or changes I deemed necessary as noted above to be appended hereto, and that the same is a true, correct and complete transcript of the testimony given by me.

_____ _____

Kosuke Imai, PhD Date

*If notary is required

SUBSCRIBED AND SWORN TO BEFORE ME THIS

_____ DAY OF _____, 20___.

_____

NOTARY PUBLIC

&

& 3:9,14

0

0 191:10,20 192:10,20
0.1 74:8,15
0.4 89:8
0.45 138:15
0.5 138:16
03302 1:6

1

1 9:20 38:21 40:25 64:11,15 71:20 84:23 88:19,24 89:6 94:8 95:25 103:10 107:7 110:23,25 113:7 113:24 114:5,12 116:11,13,17,22 117:5,12 118:13 118:25 119:5,19 120:14,19 122:4 122:7 123:21 124:5 127:6,11 128:1,10 129:2 129:18 130:2 133:4,6,21,25 134:8 136:2,4,16 141:19 146:3,10 146:16 147:5,19 147:19 148:8 149:1,16 151:16 152:9,12 154:6 154:14,23 155:7 155:14,17 156:4 156:18,19 157:7 157:9,17,20 163:21 188:4,12 189:4 190:1 191:1,11,21 192:1,11,21
1's 123:10 129:9
1.07 81:15,15 97:10
1.2 81:16
1.29 129:24
1.3 129:15,22 130:3,4,8
1.35 81:16
10 70:15 88:8 97:18 130:14,15 130:16,24 133:2 190:10
10,000 92:23,25 93:25 94:4 95:4 98:12 139:13,19 140:10,14,18 142:19 143:2,10 173:4
100 12:18 27:15 45:21 115:12 116:1 117:6 166:23
10006 2:16
104 3:5
11 97:18 114:2 132:14 190:11
11:04 1:13
12 89:21 90:3,5 94:7 114:12 120:18 121:7,11 190:12
1221 3:11,16
13 25:19,23,25 26:1,4 94:22,23 120:10,12,19 127:3 189:23 190:13
14 26:22 27:11 28:16 106:23 107:4 127:14 189:20 190:14
15 127:14 145:18 145:21 163:18 190:3,15
16 188:10 189:7 189:16 190:16
17 32:21,23 151:10 190:17
17.4 121:2
173 189:23
1730 2:5,10
177 188:8
18 157:24 164:2 190:18
1800 2:5,10 3:11 3:16
182 188:10
187 188:12
19 159:3,24 164:2 189:9 190:19
1998 19:10

2

2 16:9 38:20 101:9 114:8,8 120:11,12,13 122:17,25 127:23,25 158:4 158:13,19 159:13 160:6,10 160:16,18,20 161:7,15,17 162:2 163:10 164:6,10 166:25 167:8 168:18,23 176:5 187:25 189:7 190:2 191:2,12,22 192:2,12,22
2's 164:17
20 70:10,17 74:7 77:16,21 89:24 106:20 109:13 109:18 189:4 190:20 192:15
20001 2:21
2002 19:12
2003 19:13
20036 2:6,11
2010 19:20 51:16 52:10 53:8,19,25
202 2:6,11,22
2020 50:24 51:15 53:1,11 54:2 110:24
2021 68:18 69:14 189:17,20
2022 1:12 16:10 18:17 19:3 174:7 187:11 189:8,9 190:3
2027 187:25
21 190:21
212 2:16
21903 187:22
22 70:14,17 74:7 77:16,21 106:20

109:13,19 190:22
22nd 184:25 185:25
23 88:6 188:8 190:23
23.9 162:10,14 163:6,15
2388 162:17
24 165:25 166:2 167:5 190:24
24,900 129:3
25 78:18,21 84:22 89:2 138:6,7,8 190:25
26 59:14 62:9 130:11 138:2
26.9 165:12
26b 48:7
27 94:23 106:24 108:2 114:11
28 120:15
282-1195 3:6
29 119:3 121:1
29.4 121:7
29201 3:11,16
29601 3:5

3

3 19:3 43:4 114:9 127:18 128:12,13,14,15 171:21,24 172:12,15 188:6 189:9,9,11 190:3 191:3,13,23 192:3,13,23
30 120:15,17,18 120:18 190:17
31 17:13,15
32 129:1
33.9 166:24
34 189:11
35 145:22 147:4
36 150:19
37 154:8
39.4 159:5 162:6 166:3
3:21 1:6

4

4 16:10 29:12,14 30:18 34:3 114:9 128:12 163:18 188:4,6 189:8,11 190:4 191:4,14,24 192:4,14,24
4.5 147:6
40 2:15 159:1 161:1,1 162:6 165:9,25 166:22
41 65:21 66:3 162:10,15
41.1 120:4
43 169:21
45 41:20 135:15 136:20,23 137:12 139:18 140:1 164:8 171:16 172:18 172:22 173:1 183:24 184:4
450 14:20
457-0800 2:6,11
46 137:5
46.9 172:6
49.9 172:11
4:32 186:14
4th 174:4,5

5

5 30:18 65:9 114:8,9 123:2,4 146:6,8 169:11 170:3 176:6,16 189:14 190:5 191:5,15,25 192:5,15,25
5.8 119:6,10,18 120:2 121:24
50 41:21 115:16 116:3 135:16 136:20,24 137:12 139:21 140:2 164:8 167:4 171:16,20 172:18,23 183:24 184:5
500 115:7,12
51 2:21
5350506 190:5 191:2 192:2
57 78:20 80:7,22 84:22 89:2 97:7 106:17 138:3,8,9
5th 2:15

6

6 25:20 26:3 41:19,22,25 42:4 43:2 64:11,15 68:16 84:23 89:7 94:8 95:25 103:10 113:7,24 114:5 118:25 119:24 120:8 121:19,22 122:4 122:8,11,17,25 123:2,4,15,21 124:5 127:6,11 130:3 133:4,6,21 133:25 134:2,8 135:15,20 136:2 136:5,8,13,17,24 137:9 138:25 139:9,15 140:1 141:14,21,24 142:9 146:3 147:23 148:2,3,4 148:6 149:1,16 150:22 151:10 151:12,14 152:4 152:9,15,19,23 153:1,20,23,25 156:4,18,19 157:8,9,13,17,20 158:4,13,19 159:8,18 160:10 160:20,22,23 161:2,4,9,10,15 161:18,21 162:2 163:10,22 164:6 164:7 165:14 167:3 168:8,10 168:18,23 169:11,25 170:12,24 171:3 171:12,14,15,24 172:5,16,24 173:1 184:4

189:16 190:6 191:6,16 192:6 192:16
6's 120:3 121:10 121:14 123:9 136:19 138:15 141:8,17 171:19 183:23
6.5 147:6,13
61 108:2
65 189:14
68 189:16
69 189:20
6s 121:20

7

7 32:21 69:14 114:9 147:24 148:4 153:21 158:11 189:20 190:7 191:7,17 192:7,17
730 71:22
731,204 72:25
74,600 154:11
76.3 154:13
799-9800 3:12 3:17

8

8 1:12 70:12 109:17 159:24 160:1,24 161:22 162:13 163:9 167:6 173:13 176:17,24 189:23 190:8 191:8,18 192:8 192:18
8/8/2022 190:5
803 3:12,17
864 3:6
879-3939 2:22
8th 187:11

9

9 70:15 189:4,7 189:14 190:9 191:9,19 192:9 192:19
90 117:13 170:24
90.3 169:23
900 3:5
9500 129:5,8
965-2200 2:16

a

a.m. 1:13
ability 122:12
able 47:9 124:8
abstract 34:22 36:7 38:9,22 43:7 44:22
academic 16:1 22:22 132:1
accepted 22:23 23:2,5
access 18:3 47:5 64:2
account 100:13 101:25 102:13 102:15,18 122:1 136:21
accounted 181:23
accounting 161:22 167:4
accounts 131:22
accuracy 39:8 175:6 190:9
accurate 187:5
accurately 36:13
acepedaderieux 2:7 190:2
achieve 73:1
acknowledged 40:9
acknowledge... 192:3
acknowledgm... 190:12
aclu 3:21 14:3,6 14:9 45:6,21
aclu.org 2:7,12 190:2
act 137:13
actual 43:18 51:10 137:8
ad 68:18 189:17
adams 28:21
add 49:6 107:15 135:11,23
added 110:13,18
addition 12:11 33:21
additional 54:15 73:25 101:11 103:1 135:11 153:21 180:13 182:11 185:20
additions 192:6
address 22:25 26:15 112:12 113:17
addresses 30:12
adjacent 23:23 24:3 79:10
adjusting 76:20
administrative 31:9 102:15 105:3
admission 143:1 143:5
adriel 2:4 47:3 190:1
advised 186:15
affect 58:14,17 76:16 133:11 171:13
affixed 187:10
age 114:18,19,21 114:23 115:8 151:15 160:5,8 160:17,18 170:20
aggregate 20:11
ago 27:3 45:17 123:8 173:24 174:9
agree 36:20,22 66:11,16 67:1 72:1 84:3 115:6 129:20 139:25
agreement 5:22 14:23
ahead 5:16 49:13,21 56:1 62:20 65:7 113:11
al 1:4,7 2:3 190:4 191:1 192:1

alabama 14:14 28:11
alexander 1:7 4:13
algorithm 13:18 21:9,24 22:2 23:10,11,20,22 24:5,18 33:5,6,7 33:8,11,16,25 37:5,7,24 38:6 38:14,15 39:19 41:6,13,13,22,23 42:4,6,8,12,24 45:10 49:15 50:14 51:1 55:8 56:8 58:7 70:22 71:2 75:5,8 78:10,13,16 79:9 80:14,19 81:20 82:7 83:23 93:1 93:6,14 95:2,18 95:20 96:11 98:3,6 101:25 102:8,19 103:4,7 106:15 110:13 110:14,18,23 132:13,16
algorithms 20:1 20:7 23:13,15,16 24:16 33:10 37:18 38:2,16,18 38:25 39:2,5,9 39:17 40:9,16 43:15,17,22 44:4 44:8,13,17
allotted 190:20
allows 88:12 136:13
altered 114:5
amathias 3:6
amend 17:1
amended 9:21 189:4
american 2:4,9
analyses 22:9,14 22:17 48:18,19 64:10 104:13,13 157:1,14
analysis 11:20 12:4,6,13,17,24 17:9 21:16 22:5 22:8,18 26:17 30:11,12 31:25 36:20 38:13,18 40:5 41:15 42:6 47:23 48:3,9,16 48:23 49:2,3,8 49:17,25 51:13 52:19 53:24 54:11,23,25 55:11 56:4 58:17,17,22 59:16 60:4 61:6 61:20 62:1,6,11 62:14,18 63:16 67:10,23 68:1 76:17 81:1 86:8 86:16,18,19 91:24 96:5,8 98:1 102:13 103:17,20 104:12,15 112:12 113:3,19 113:21,21 118:11,12,15,19 122:14 123:11 123:12,15,22 125:14 127:20 130:5,8,10 133:23 134:13 135:5 136:6,10 136:15,22 138:23 139:3 140:21 141:7,15 141:16,18 142:23 145:1 146:6,25 155:10 155:23 156:1,3,8 156:16,17,19,25 157:1,2,4,10,15 164:20 166:21 167:15,16 169:8 170:6 174:14,18 175:7 176:20 180:20 181:6,9 184:9,11
analyze 32:4,7 41:16 67:21 76:10 122:10,11 155:6,11 159:17 168:22 170:11 171:2 182:6,22
analyzed 112:7 119:24 133:22 176:11,14 177:1
analyzes 29:15
analyzing 85:6 145:24
andrew 3:4 177:13,18 184:15
announcement 10:8
answer 5:15,16 5:24 6:6 13:6 47:6 49:13,21 51:20 52:3,15 53:22 54:1,9,21 55:21 56:1 57:4 59:18 62:25 124:8
answered 166:7 180:2
answers 5:4
apart 7:18
appearances 2:1 2:25 3:1
appeared 27:12
appearing 4:18
appears 24:21 107:10 131:6 158:14
appended 192:7
applicable 190:8
application 20:7 35:13 36:10 40:6,11,13 43:24
applications 21:11 22:2 40:4 40:11
applies 38:17
apply 39:21
applying 43:17
apportionment 28:8 73:15
appreciate 93:22 126:7
approach 40:16
appropriate 75:1

approximate 83:10
approximately 45:11 71:23 129:3 137:8
april 16:10 17:21 27:5 189:8
area 31:1,16 88:3,15,23,23 123:25 126:1 148:9 149:10 150:9
areas 36:16,25 104:21 124:4
article 24:6 43:4 65:10,15,22 66:10 73:15 189:15
articles 10:25 20:20 21:15,19 22:3,10 23:14
arts 19:9,10
asked 11:2,5 22:24 48:15 49:22 53:7 54:22 174:19 179:11 181:11
asking 22:17 48:12 59:13,15 59:16 62:10,11 90:25 91:1 142:5 166:6
asks 62:8
aspect 112:7
aspects 132:12
assembly 83:11 83:18,23,24 84:8 84:13 136:19,22
assess 112:24
assign 95:2 105:15,19,21 106:4 159:7
assigned 83:18 122:8 128:10 148:7 151:16 152:12 154:13 160:6,11 161:2,6 161:14 166:25
assigns 80:7
assist 15:1,14,18
assistants 15:5,6
assume 5:25
assuming 73:25
assumptions 12:21,23 13:2,7 13:10,13,15 46:22
attached 190:11
attempt 54:16 174:21,24
attempted 182:21
attention 174:12
attentiveness 177:11
attorney 55:19 190:13
attorneys 2:2,18 3:2,8
august 1:12 187:11 190:3
automated 76:7
automatically 40:12 75:13 136:16 141:20
available 104:6 105:7 106:10 190:6
avenue 2:21
average 74:23 77:23,24 78:2 92:16,18,19,20 92:24 93:8,15,19 94:1,21 95:10 97:13,17 98:13 98:15,18 99:1,2 99:4 119:4,8,11 121:20,23 129:2 129:3 130:3,4 131:7,16 132:7 133:3,5 147:5,13 160:19 173:3
averages 95:5 98:11 133:4
avoid 73:24 75:5 103:7 153:11
avoidance 79:24 81:9 84:24 85:21 89:7 106:8
avoiding 73:23
aware 19:17 117:21 133:20 134:7 136:18 158:18

b

b 10:17 11:3 72:21 103:2 189:1
bachelor 19:9
back 17:21 19:16 24:25 25:15,16 27:5 34:24 44:22 45:16 47:15 70:5 72:13 74:5 77:15 88:5 89:1 95:4,16 99:8 101:4 109:12,13 138:2,10 145:17 163:17 164:2 167:17
background 19:8 91:1
bagley 9:11
balanced 34:4 189:11
baodong 9:7
bar 151:17,22
bars 90:10
base 180:3
based 39:22 57:6 58:3,23 63:4 71:11,12 74:17 91:24 96:16 99:24,25 100:17 100:20,25 107:16 179:23
baseline 111:10 111:15 113:3,4
basic 35:14 131:21
basically 20:10 23:21,23 24:7,12 35:17 40:22 42:23 55:1 61:6 61:19 78:12 79:5,13 80:1,17 82:12 93:5 108:14 123:21

139:7 141:10 142:14 166:20 166:23 181:6
basis 29:25
bayesian 26:18
bear 65:1
begins 154:10
behalf 27:25 177:16,19
behavior 20:11
belabor 166:8
believe 10:13 16:7 17:22 18:16 24:20 29:15 32:19,23 52:21 57:17 59:18 65:15,25 66:10 68:6 92:15 106:19 123:23 129:1 137:11 140:25 159:3 174:2 179:3 180:12
ben 45:2
benchmark 47:23 48:24 49:2,9,14,15,16 49:18,23,24 50:4 50:6 51:17,22,25 52:4,6 53:24 54:6,13,18 55:12 55:15 56:9,24 57:5,8,18,21,22 57:24 58:2,5,6,9 58:13,24 59:9,25 60:3,22,23 61:5 61:9,13,15,17,22 61:25 62:2,5 63:3,17 103:19 106:3 110:25 118:16,17,18 122:23 128:19 128:20,23,25 147:2,3 154:18 154:19,20 158:7 158:9 169:16 178:11 179:4 181:5
benefit 10:4
berkeley 117:2 146:20
best 33:24 94:8
better 37:25 38:5 41:23 89:13 107:7,11 131:7 132:24
beyond 141:7 142:16 165:24
big 18:15 117:23
bigger 104:14
binder 16:7 34:7 68:10 69:10
bit 8:11 13:21 32:18 39:7 42:3 70:3 90:17 97:6 132:23 144:9 175:21
black 31:11 49:10 114:17,18 114:21,23 115:8 118:20,21 119:11 122:3,6 122:12 123:15 123:17,18,25 124:3,7,22 125:5 125:17,21 127:21 128:8,23 129:3 148:4,13 148:14 149:1,16 149:23 150:1,6,7 150:9,22 151:6 151:15,25 152:1 152:8,11,14,16 152:16,21 153:1 153:25 154:5 160:4,7,17,17 164:23 165:4,15 165:19,21 166:10 167:2,7 167:22 168:5,10 168:13 170:19
blank 24:13 35:4 35:15 36:4 41:7 41:10 42:2,21
blind 40:19,21 40:24 41:1 42:8 61:6,10 103:11 118:22 119:5
block 75:14,23 76:3 77:1
blocks 74:21 75:24 76:1
books 10:24 20:20
border 116:24 116:25,25
boundaries 24:7 30:14 31:9 40:22 55:2 73:22 75:11 76:11 102:15 105:3 106:12 113:1,6 114:8 127:6 131:3 142:3 156:20 157:17 180:23
boundary 39:10 61:11,13 79:13 114:5 122:21 123:6 127:10 136:1 145:25 146:2 149:9 158:19 168:23
break 6:2,4,7 70:9 77:13,16 135:2 177:8
breaks 6:4
brief 8:21
briefly 8:22 19:7 88:6
bring 47:1
broad 23:16
broadly 23:21 166:8
browser 18:8
bucket 98:12,13
buffering 24:21
build 42:21
building 35:16 36:4
bullets 77:20
burchstead 3:10
burden 73:25
burr 3:9,14
burr.com 3:12 3:17
bvap 41:20 42:4 42:25 114:14 115:3,12,16,19 116:1,3,6 119:4 119:19,21,25 120:3,7,14

121:10,14,14,19
122:11 123:9,10
127:17,21 128:2
128:15,16 129:9
130:2,9 135:15
136:3,4,8,13,15
136:17,19,24
137:8,12 138:15
138:25 139:8,15
140:1 141:8,14
141:17,19,20,24
142:9 146:15
147:5 157:12
161:4,6,8,14,15
163:10,10,22,25
164:7,10,17
165:13 166:20
166:25 168:8
170:16,17 171:4
171:4,5,13,14,16
171:19,21 172:6
172:12,23,25
173:3 183:23
184:5

c

c 1:7 10:17 11:3
calculated 94:21
calculating 92:16 97:16
calculation 100:1,25
calculations 175:3
call 31:6,13
called 24:6,11 37:11 65:15 175:16
calling 55:19
calls 48:5
cambridge 1:14 6:17
candidates 122:13
cap 14:21,23,24
career 20:18
careful 60:15
carlo 23:15,19 23:22 24:12,17 33:7 34:4 35:1 35:10,11,15,19 35:24 37:3,4,12 38:1,6,10 50:3,8 189:11
carolina 1:1,3 2:2 4:16 6:14 8:1 14:16 20:13 28:13,13 29:16 31:4 46:10 68:7 68:16 69:3,15 96:16 108:15 112:6 180:4 187:4,12 189:16 189:21 190:4 191:1 192:1
carries 30:18 70:15
carry 181:7
case 1:6 4:16,18 6:12,15 8:6,24 9:3 11:2,15 13:21,23 14:7,14 14:15,16,19,22 17:5 22:6,18 24:17 26:8 27:4 27:6,7,9,10,13 27:14,15,16 28:1 28:3,5,5,11,12 28:14,14,20,21 28:22 29:8 33:9 39:20,23 40:7,14 40:14 41:17,18 44:1 45:23,24 46:5,22 53:17 63:6 72:25 76:9 79:6 80:24,24 84:14 102:12 103:10 108:15 111:8 113:6 119:23 141:23 148:2 161:3 162:8,9 164:14 166:24 174:14 178:2 184:9
cases 14:8,12 26:25 27:11,18 27:19,20,21,23 28:9,10,15,16 38:12 41:9 43:14 44:13,16 44:17 82:2,3 84:6,7 93:17 98:22 111:5 161:5 165:12 166:22
category 162:22 163:4
cause 187:8
caveat 107:15
cc'd 11:7
census 12:9 26:11 50:24 51:16 52:10 53:1,8,9,10,19 54:1,3 74:21 75:14,23,24 76:1 76:3 77:1
cepeda 2:4 5:11 5:12 6:23,24 7:1 7:18 13:4,25 25:8,12 46:7 47:5,11 48:4,12 48:20 49:11,19 51:18 52:1,13 53:20 54:7,19 55:18,22 57:2,20 59:12 60:6 62:7 62:13,21 63:10 137:17 141:4 143:6 148:17 149:3,6 165:6,22 171:23 177:21 177:24 179:2 182:14 184:14 184:21 185:9,11 185:15,19,22 186:1 188:8 190:1
certain 12:18 24:8 31:16 37:9 38:14,15 42:25 44:13 60:17 78:14 79:21 80:8,8,15 81:2 84:9 97:1 123:25 124:4,7 126:1 136:9 138:21 149:24 154:25 157:13 164:15 168:9,9
certainly 9:1 12:20 44:21

52:18 74:11 84:14 88:4
certificate 187:1 188:12
certify 187:4,7
cetera 15:17
chain 23:22 33:7 35:6,10,19,24 37:3,4 38:1,6 50:3,8
chains 65:10,16 189:14
chairman 4:14
chance 182:9
chances 118:3 181:24
change 76:21 82:16,19,20 83:2 83:4 84:14 104:12 122:21 123:5,9,10 130:8 136:11 191:4,7 191:10,13,16,19
changed 27:17 82:23 83:24
changes 155:11 155:20 190:10 192:6
changing 82:15 82:18 148:25
characteristics 59:3 76:8
characterization 132:25
characterize 89:16 131:10,20 132:20
charge 14:20
charging 14:18
charleston 41:1 61:11 64:15 81:13 96:3,4 97:9 103:12 107:11,17 117:11,11,11,17 117:17,23,24 118:6,8 119:14 119:17 123:19 123:19 124:10 124:13,15,19 125:8,18,19 126:12,17 127:1 127:7,11,17 128:3,6,9,15 130:23 131:3,8 133:2,19 134:6 145:25 146:21 146:22 147:17 147:17 148:8 149:19,19 150:20 151:2,4,5 151:16 152:2,16 152:18,22 153:1 153:14 154:1,3,5 154:17,22 155:3 155:7,13,16,17 156:5,21 157:6,6 163:22 183:16
chart 116:12 146:9
check 11:11 106:16,21
checked 11:9 153:5
choice 80:24 122:13
choose 23:23 80:22 81:1,5,15 83:1 93:5,16 95:8 97:24 180:18
choosing 138:24 138:25
chose 137:7
circuit 26:6
citation 39:2 43:16
citizens 176:25
city 27:8 117:11 117:17,23 118:5 119:14,16 123:19 124:13 124:15,18 125:18 126:12 146:21 147:16 151:1
civil 2:4,9
clarification 46:2 50:23 86:9 111:17
clarify 5:22 32:9 50:5 78:8 92:7 131:24 140:4
class 19:16 23:16
classes 19:15
clause 73:15
cleanly 175:21
clear 13:19 22:19 111:25 126:3 133:1 175:14
close 120:17,18 167:1 172:23
closely 98:2,7 132:15
closer 172:25 173:1
clyburn's 135:19
collaborated 45:12
collaborator 45:8
collect 10:22
colleton 134:18 156:10 157:19
colloquialism 143:8
color 116:19
colored 116:12 146:9
columbia 1:2 3:11,16
come 54:2 84:10 153:4 172:3
comes 33:19 50:21
comfortable 47:9
coming 119:11 119:12,13 147:13,23,24 151:25 157:16
comment 150:13 155:2 170:8
commission 3:8 28:8 177:16 187:24
committee 4:14 68:17,18 69:4,16

69:24 189:17,17 189:21
commonly 97:25
communicate 101:21 102:4
communication 55:19
communities 101:18 102:1,14 104:22 105:1,7 105:10,14,16,18 148:14 149:1,24 167:23 168:1,6 168:15
community 102:7 125:17,21 150:6,7,10 165:21 166:10 166:15,16,17 167:25 168:3
compact 34:4 73:21 77:23 81:22 82:1,3,5,8 97:14 98:13 130:21 189:12
compactness 80:9 81:14 82:13 97:4,7,19 97:22 98:3,9,17 98:23,24 99:2,5 99:17,22,24 100:5,10,24 101:14,17 102:23 130:17 132:10 139:4
comparable 88:13 100:11
comparator 66:3 66:4
compare 50:19 51:6,24 52:4,11 52:20 59:10 63:20,25 64:3 85:25 88:12 100:15 111:10 120:8 144:20
compared 53:4 82:6 95:10 120:1 122:23 129:5 161:15 165:7
compares 121:14
comparing 63:8 86:7 100:17 113:4
comparison 63:12 98:20,24 136:14 150:23
complaints 21:9 21:9,10
complete 26:25 187:6 192:8
completed 190:17
completely 58:18
complex 39:7
compliance 84:3
complies 137:12 184:5
comply 67:2
components 20:6
compromise 140:22,23
computational 21:21
conceptual 172:10
conceptually 125:16
concerns 151:1
conclude 30:7
concluded 186:14
conclusion 30:15 66:6 141:5 142:14
conclusions 30:2 58:10 67:24 76:16
condition 91:14
conduct 31:25 51:15 113:2 134:12 144:25 184:10
conducted 15:23 40:20 62:14 112:16 157:15 184:8
conference 1:3 2:2 9:21 189:5 190:4 191:1 192:1
configuration 99:23 100:5
confirm 109:21
confirmed 11:11
confirming 65:21
confused 128:11 175:12,20
congressional 4:15 14:14,15 16:20 27:15 28:12 29:16 46:10 63:23 68:19 72:22,24 73:14,17 74:1 90:21 178:2,5 189:18
congressman 135:19
connection 4:18 10:21
consider 32:11 32:15 40:6 51:1 56:19 71:3 89:10 103:4 109:25 110:3,9 110:11,13 117:25 118:2 144:4 170:5,7
consideration 73:3 100:13 101:17 102:18 113:19 136:9 182:11,12 184:1
considerations 101:12 103:1
considered 136:6 144:9 159:16 183:2,7
considering 136:12
consistency 42:7 103:2

consistent 117:19
constituency 104:4
constituent 103:2
constituents 101:21 102:4 104:1 106:6,9
constitute 31:4
constitution 73:16
constrain 95:18
constrained 153:9 164:8,10
constraint 41:19 43:1 49:6 54:15 54:16 57:25 58:7 60:1,4,5 61:18 62:5 63:2 63:18 74:8,16 78:15 79:1,2,3,4 79:16,19,20,24 80:3,5,12,13,14 80:23 81:5,6,14 81:24 82:11,16 83:1 84:24 85:11,12,17,21 85:22,22 86:22 86:23 89:8,11,17 93:5 95:2 97:8 97:20 103:21,25 104:3,25 105:9 105:19 106:14 106:22 110:14 110:18 111:4,20 112:20,23 113:9 113:15,22 127:5 135:5,12,24 136:24 138:11 138:14,16,21 139:1,2,4,10,11 139:12 140:24 142:1,11,13,15 142:16,18,19,22 143:1,15,23 157:12 168:5 180:13
constraints 13:18 33:10,11 36:12,21 37:10 39:4,13 42:15,23 50:18,25 53:3 55:9,12 56:8,11 56:14,24 57:25 59:10 60:23 67:10,14,17,18 78:6,11,24 79:23 80:8,17 82:24 83:5,7,9,10,13 83:19,25 84:1,18 86:14 89:3 91:13 93:17 96:15 103:14 132:8 143:17,19 180:4
construct 55:8,8
construction 136:16 171:11
constructive 37:12
contacted 13:22 14:1
contain 16:19 59:25
contained 16:24 25:11
contains 159:24
contaminate 54:25
contest 73:23
context 86:7
contiguity 56:15 57:13 71:3,6,15 77:18
contiguous 70:19,23 71:9,14 153:15
continued 2:25 3:1
control 54:17 55:14 60:22 118:15,20 128:19,22 147:1 163:11
controlled 56:18
controls 61:2 80:20 98:4,7 132:13
convenient 81:19
conversation 7:7 46:3,6
conversations 6:23,25 7:4,9,14 12:24 13:9 46:8
cool 116:12 146:9
copied 46:15,18
copies 190:14
copy 7:23,23 8:2 8:3 17:16 185:18
core 104:1,4 106:6,9 176:6,7 176:12 180:10 180:18
cores 54:17 103:3,5,14,21
correct 18:17,18 26:20 28:24 29:17,18 30:4 33:2 35:2,7,8 36:19 44:14,15 50:11 51:2,4,5,7 52:22 56:12,13 56:16,17 59:22 60:2 63:18,19 64:11,16,20,21 68:4,9 70:20 71:16,21 73:5 74:3,4,8,13,14 75:20 82:18 83:6,8 86:11,12 87:21,22 88:20 89:9 92:4 94:3 94:22 99:3,6 100:18,21 101:1 102:23 103:22 104:1,11,19,20 105:10,11,16,23 105:24 106:6 107:9,14,24 109:23 110:10 111:21 112:21 113:9,10,12 114:7,16,20 115:3 116:9,20 117:9 119:1,2,22 121:15,21 122:2 122:5,17 123:12

123:16 124:5,23 127:12,13 128:1 128:20,24 129:7 130:6,19,21,22 134:16 135:9,10 135:12,13,16 137:10 140:11 140:19,20,22,24 142:10 143:11 143:12 145:25 146:3,12,18 147:2 148:6,16 149:20,24 150:4 150:7,24 151:2,7 151:22,23 152:10 153:2,10 154:16,23 155:4 155:21,22 158:15 159:10 159:14 160:20 161:18,19 162:3 166:11 167:9 169:12,13,25 170:1,4 171:18 171:22 172:13 172:16,17 174:15 175:9 176:13 178:17 179:6,10 182:23 182:24 183:3 184:9,25 192:8
corrections 192:6
correctly 7:7 36:18 38:3 39:14 43:19 44:11 66:8 73:4 88:16 100:2 101:2,23 108:9 116:10 117:1 138:18 147:10 154:15 159:9
correlated 112:5
correspond 67:14
corresponding 77:25 78:3 98:14 147:8 160:23
corresponds 75:2
cost 185:20
council 27:8
counsel 1:15 2:1 5:7 8:1,4 10:4 11:3,5,6,11,15 11:22,25 12:8,14 12:21 13:2 18:20 46:4,14,18 49:22 54:22 55:13 56:3 69:8 70:1 173:21,22 174:9 179:8 187:8 190:14
count 20:25 21:1
counties 77:24 79:4,14 89:5 91:20 92:6,11 95:19,21,23 96:1 96:11,18 102:16 105:4 133:20,24 134:7,13 157:22 165:10 168:18 168:21 183:19
counting 92:17
county 31:3,4,6 41:1 61:11 79:5 79:12 81:13 89:7,19,20,23 90:1,14,20 91:12 91:17 92:2,14,16 92:24 93:2,7,8 93:12,13,15,17 93:18 94:1 95:5 95:17 96:3,4,7 96:21 97:9 103:12 107:11 107:17 117:2,12 117:18,24 118:9 127:1,7,11,17 128:3,7,9,15 130:23 131:4,8 133:2,19 134:2,6 134:16,18,20 139:4 145:25 146:22 148:9 150:20 151:5,16 152:2,16,18,22 153:1 154:2,3,5 154:13,17,22 155:4,7,16,17 156:5,8,10,12,15 156:21 157:6 158:2,7,14,16,20 159:6,7,12,18 160:5,11,18 161:2,6,8,9,17 162:2 164:2,15 164:24 165:5,16 165:20,21 166:20 167:17 168:12,24 169:9 170:3,12,16,20 170:24 171:3 187:12
couple 7:8 34:20 37:21 44:2 74:23 78:23 95:13 127:15 182:18
course 6:2 12:25 73:22 81:9 105:22 117:4 124:20
court 1:1,16 4:20 5:2 10:4 28:4 29:9 184:19,23 185:3,8,13,17,21 185:24 186:3,6 186:10,13
cover 88:2,14 104:9
covered 88:22 181:19
crack 124:22 125:5 148:13 149:23 150:5 151:7 165:4 166:3 167:22
cracked 125:24 149:2 150:2
cracking 30:21 30:23 31:5,6,7 31:10,13,14,19 31:24 49:10 123:24 125:9 148:23 149:13 165:1,15 166:9,9 166:13,19
cracks 30:20 123:15,20

124:12 125:7,21 149:16 150:10 164:23 165:19
create 57:7 113:2 152:20
created 58:3 83:16
creating 24:13
criteria 30:8 32:25 68:8,12,19 70:18 72:14,16 81:2 83:15 91:13 97:3 140:11 180:11 181:1 182:13 189:18
criterion 135:11 135:14,23 140:14 141:2 143:10
criticism 21:10 21:11,11,23 22:25 39:16,18 39:21 40:3
criticisms 21:18 22:1 40:15
criticizing 40:5
cross 11:11
crum 3:15 177:13,15,15 186:4,5
cs 190:15
curious 92:19 131:25
current 17:19,19 18:23 75:8
currently 54:4
curriculum 19:4 189:9
cusick 2:14
custom 138:14 138:20,22
cut 85:15 108:19 179:5
cv 1:6 17:16,19 17:19,23 18:17 18:22,23,24 24:20 25:8,11,14 27:4

d

d 188:1
d.c. 2:21
darker 116:16
data 11:14,19,20 11:23,24 12:1,6 12:7,8,9,9,12,15 12:18,19 20:11 26:11 32:4 46:12,13,17,18 50:24 51:16,16 52:10 53:1,8,9 53:10,13,19 54:1 54:2,3 74:18 75:10 85:6 86:2 91:25 102:10,13 109:25 155:9 181:13
date 1:12 173:23 174:1,2 191:24 192:12
dated 16:10 189:7
day 2:19 4:12 6:5 187:11 192:15
days 190:17
dc 2:6,11
debates 43:13
decade 60:19
decided 17:11
decision 26:6
decisions 13:11 154:25
declare 192:4
decrease 120:7 136:4
decreases 136:17
decreasing 122:11
deemed 192:6
default 80:9 81:17,19,25 97:7
defendant 3:2,8 28:1,6
defendant's 9:20 16:9 19:3 34:3 65:9 68:16 69:14 173:13 189:3
defendants 1:9 1:15 2:18,18 3:2 3:8 4:13 177:17 177:19
defendants' 9:21 189:4
defense 2:14 14:3
define 30:23
defined 125:18
definitely 111:1 153:20 168:9
definition 30:25 31:2,5 71:11,11 71:12 104:4 105:6,18 106:10 123:14,24 124:6 124:9 125:9,22 148:23 149:13 165:1 166:9
definitions 105:13
degree 19:11 88:25 166:12,19
demarked 158:16
democrat 32:13 110:5,20
demography 101:18 102:1,7
depending 38:12 39:20 67:9 74:24 104:11 116:8 168:13
depends 85:6 96:8 102:6 123:17 167:25
deponent 188:11 190:13 192:3
deposed 6:10
deposing 190:13
deposition 1:11 6:21 7:19 8:13 9:22 10:9 25:10 77:12 135:1 177:8 186:14 189:5
describe 19:7 125:3

described 33:24 142:8
description 149:8
design 164:9
designed 83:15
detail 32:20 139:23
details 185:12
determinant 140:7
determination 100:1 101:1
determine 68:3 102:10 144:10 144:15,21
determining 30:14 101:16 106:1 142:3
developed 23:13 24:5 45:15
developing 26:12
development 20:7,9 45:15 75:8
deviation 56:20 71:19 72:2,19 73:2,16,19 74:12 75:2,6 76:14,15 76:21 90:22 91:9
deviations 74:2 147:7
diagnostics 33:22
difference 23:10 35:14 42:9 76:20 119:4,8,9 120:6 129:2,4 131:19,23 132:4 132:24 133:11 141:11 151:19 151:21 156:21
differences 22:14,16 35:9,12 41:17 131:12 133:17 157:16
different 23:3,12 24:15,23 36:1,2 37:17 38:12 42:14 44:6,19 49:5,6,6 57:25 66:7 67:6,10 76:19 78:23 82:13 83:19,20 83:25 84:10,16 86:2,3,4,8 88:22 88:23 95:20,23 96:1,24 97:19 98:5 104:10,17 104:18,20,21 111:11 114:23 115:2,19 116:2,7 116:8 131:12 132:11,12 142:13,23 166:12
difficult 39:7,11 42:3 75:7
difficulties 43:16
direct 105:9 108:6
directly 13:8 49:16 50:14,16 51:11 55:8 56:7 58:6 61:18 75:22 93:13 98:3,7,7 100:7 100:14,14 103:18,23 104:3 105:19 180:14
disagree 62:21 67:4
disappoint 184:17
disclosed 13:13
discourage 78:13 80:15
discouragement 80:18
discourages 79:21 80:5
discovery 62:17
discriminates 29:24
discuss 22:4 46:4 77:12 135:1 158:1 166:18 177:7
discussed 7:19 12:23 13:8 14:24 45:23 46:1 52:25 104:7 136:1,3 148:15 165:18 178:15 179:25
discusses 176:6 176:17
discussing 77:16 85:24 126:10 163:20 170:15
discussion 156:5 164:1 176:12
display 10:1 65:14
displaying 150:21
disproportiona... 150:21
dispute 129:21 175:3
disqualified 29:5
distribution 33:8 33:13 36:14 39:3 44:8 82:19 82:20 125:11,14 147:7 167:11 168:14
distributions 44:3
district 1:1,1 24:9,14 30:14 31:12 32:16 35:16 36:5 37:13,15 40:25 41:8,19,22,25 42:4,21,24 43:2 54:18 58:11 61:4,10 64:11 75:11 76:11 79:8,11 81:4 84:23 86:1 87:19,20,24 88:2 88:3,12,13,19,19 88:22,24 89:6 90:21 91:19 95:25 98:14,18 98:21,24 99:4 100:21 101:14 101:17,20 102:2 102:11 107:7

110:4,23,25
113:24,24 114:5
115:2,24 116:2,2
116:13,17,22,25
117:5,12,16,23
118:16 119:5,19
119:22,24 120:3
120:8,14,19
121:10,14,19,20
121:22 122:4,4
122:11 123:6,9
123:10,15
124:16 128:1,7
128:10,19,23
129:2,9 130:2,3
131:2 134:2,8,8
135:15,18,19,20
136:1,4,5,8,13
136:15,16,17,19
136:24 137:9,11
138:15,24,25
139:9,15 140:1
141:8,14,17,19
141:21,24 142:3
142:9 145:25
146:2,10,16
147:1,5,19,19,23
147:24 148:2,3,4
148:4,6,8 149:1
149:1,8 150:22
151:9,16 152:4,9
152:9,10,12,15
152:23 153:1,6,8
153:13,20,21,23
153:25 154:6,14
154:23 155:7,14
155:17 156:4,4
156:18,19
157:13,17,17,20
158:4,4,13,13
159:8,13,18
160:6,12,16,16
160:18,22,23
161:2,4,7,9,10
161:15,15,21
163:10,10,21
164:1,5,6,6,7,10
164:17 165:10
165:14,20
166:11,23,25
167:3,8,24 168:8
168:10,22
169:25 170:3,12
170:24 171:3,7
171:12,14,15,19
171:24 172:5,16
172:22,24 173:1
180:23 183:17
183:21,23 184:4
district's 115:3
171:4
districting 30:8
57:18
districts 23:23
23:24 24:3 31:2
32:12 36:12
37:14 39:10
40:23 41:2,6,10
41:16 42:1,10
55:3,16 56:19,19
57:16 61:7,8
64:15 70:18,22
71:8,18 72:22,24
73:14,17,20,22
76:11 77:22
78:17 79:6,8,10
79:11 80:6 86:3
88:14 91:4,18
92:6,10 94:8
96:6 97:13,22
98:13,19 103:3,5
103:9,10,15,21
104:1,7,9,17
105:22 106:1,6
110:19 113:7
114:8,22 115:18
115:19 118:25
119:22,24
121:18 122:7,9
122:17,25 123:2
123:4,21 124:1,4
126:2 127:6,11
130:9 132:12
133:4,6,10,21,25
135:8 136:2,5,12
146:3 147:21
149:16 153:4,21
154:7 155:20
157:7,9,20
158:19 160:10
160:20 161:16
161:17,25 162:2
162:19,21,23
166:15 168:18
168:23 169:11
171:6,22 172:3
172:16,19
diverse 76:20
divide 129:11
162:11
divided 31:11
123:20
dividing 96:20
division 1:2
divisions 106:12
document 9:16
10:5,7,10,13
16:15,17 18:19
34:11,13 38:20
68:24 69:1,5,7
69:20,22,25
173:20,21
documents 7:21
8:8,12 10:21,22
70:3
doing 74:22
122:7
dominates
143:23
dorchester
134:15 156:8,14
157:19
double 106:16
106:21
dr 4:5,11 6:10,16
9:7,11,13,15
10:5 11:14
13:20 16:4,14
17:13 18:10
19:7 25:18
28:25 34:1,11
36:17 40:15
43:3 47:22
48:17,22 56:1
59:18 62:12,13
62:20,25 67:20
68:6,23 77:12
99:7 101:3
112:4 113:23
135:1 173:20
177:7,10,14,25

177:25 179:11 180:3 182:14,19 183:16 184:12 184:19
draft 12:25 13:9 21:15 22:3,10 39:25,25 43:4 44:25 46:6,9 47:24,25 48:6,13 48:15 51:21 54:10 55:10 59:14,16,19 62:8 62:11,14,16 63:11 169:3,5 178:10,15 185:5 185:10
drafted 178:1
drafts 179:5,7
draw 59:2 66:5 67:24 76:16 95:20
drawer 28:25 58:12,15,16,19 58:23,25 59:1,6 76:7 117:22 154:21,25 159:11,16 170:2 182:6,22 183:1,7 183:12
drawer's 117:25
drawers 83:16 170:6
drawing 29:3 58:10,11,13,15 76:11 91:18 112:25 113:6 180:23
drawn 137:11 142:6,10 172:15
draws 30:2 37:23 165:20
drew 58:16,23 59:6
drop 80:1
dropped 86:18 120:3
dropping 86:13
duchin 9:9
due 108:4 113:19 157:11
duly 4:2

e

e 15:12 106:11 188:1 189:1 191:3,3,3
earlier 29:16 35:3 45:16 148:22 179:11
early 178:9
easier 42:25
eastern 153:22
ecological 20:9 26:16
edges 97:23 98:2 98:6,21 132:14 132:23
education 19:14
educational 2:14 19:8
effect 32:7,10 115:3 116:2 118:2 122:10,12 130:7 155:6 159:17 170:11 171:2
effectively 79:13 101:21 102:4
effectiveness 31:25
effects 115:19
efficiently 36:13 37:24
effort 55:14
efforts 72:25
eight 85:21 86:5 109:16 138:17 139:6 173:10,12
either 27:17 64:4 87:14 91:15 94:15 104:6 106:20 112:3 115:1,23 118:23 126:17 133:10 176:21 177:2
elaborate 74:19 75:15
elect 32:12 110:4 110:19 122:12 183:13
election 3:8 12:14,18 32:4 109:25 159:21 170:14 177:16 181:24 182:2
electoral 182:9,9
eliminate 81:8
eliminated 87:9
elizabeth 3:15
email 11:7 46:14 184:20
emphasize 39:19
empirical 22:1 83:1
enact 110:4 144:11 182:4
enacted 24:2 29:16,20,24 30:8 30:14 32:1 41:24 49:3,25 50:9,9,14,20,21 51:6,12 52:8,11 52:21 53:5 55:3 55:17 57:1 58:6 58:11,13,16,20 58:23 59:2,3,7 59:10 61:14,14 63:9 66:18 74:22 76:6,8,12 77:23 78:1,3 82:4,6 83:11 87:20 88:3,12,24 89:19,23 90:1 91:4,8 92:18 93:3,9,10,16,20 94:2,11,18 95:11 95:19,21 96:6,8 96:11,25 97:14 98:14 100:11,12 104:9 106:1 107:6,10,21 111:10 112:11 113:1,5,7 114:10 119:5 120:1,9,20 121:14,19 122:22 123:20 124:12 127:9 129:5 130:18 131:3,6 132:22 133:21 134:8

135:20 136:9 137:5,9 141:1,11 141:25 144:12 144:18 147:6 149:15 150:2,10 150:20 151:18 151:19,21 152:25 153:25 154:19 155:19 156:23 157:20 158:9 160:16 161:7,16 165:8 167:8 168:17 169:12,19 172:5 172:13 173:2 179:18,25 180:1 180:23
encompass 104:17,18
encourage 78:14 80:15
encouraged 78:15 82:7
encouragement 80:18
encourages 79:21 168:9
ended 13:10
ends 117:5
energy 39:4
enforces 95:5
engagement 14:2
ensemble 44:6 44:17
ensure 33:3,22 79:22 81:3 139:7
entire 154:13 156:17 159:7 160:11 161:2 180:21,21
entitled 5:14
enumerate 37:8
enumeration 37:7
equal 72:6,10,18 72:22 77:19 79:14 89:23 95:9
equality 73:1,13
equalize 75:11 75:19 77:3,7 155:20
equalizing 76:13
equally 82:5 175:19
errata 190:11,13 190:17
especially 38:13 39:12
essentially 41:3 41:11 131:14,21 132:21,22
establish 111:14
et 1:4,7 2:3 15:17 190:4 191:1 192:1
evaluate 19:24 20:8 49:14,16,24 52:6 57:21 58:4 58:5,20,25 61:24 76:7 144:18,23
evaluated 39:22 40:13 52:8
evaluation 30:12 49:25 59:2
events 187:9
eventually 21:20 23:1,5
evolved 27:13
exact 58:18 72:8 72:11 120:5,21 120:25 147:16 162:3 173:23 174:1
exactly 12:5 56:3 57:13 94:6 96:7 134:11 161:19 162:16 168:21
examination 4:3 177:23 182:16 188:3,5,7,9
examine 29:23 92:5,8 143:25 144:4 156:7,14 180:22
examined 43:24 152:6 164:17,21 165:3
examines 40:25
examining 156:18
example 12:3 21:20 32:11 37:15 50:24 63:21 64:1 67:7 78:14 79:23 80:9 81:3,13 88:17 95:18 100:21 103:9 104:15,16 110:17 124:10 125:8 165:9 168:5 178:18 180:10 181:24
exceed 71:18
exclude 56:22
excluded 140:14 142:19 143:15
excuse 70:17 161:7 172:21
exhibit 7:25 8:5 8:5,16,19,21 9:16,20 10:2,15 10:23,24 11:3 16:9,13,14 18:1 18:3 19:1,3 25:3 25:14,16,16 34:2 34:3,8,10 46:24 47:4,18 65:7,9 65:13 68:15,16 69:12,14,19 72:13 73:7 99:14 101:5 173:8,12,13 189:4,7,9,11,14 189:16,20,23
exhibits 6:22 174:10 189:3
exist 124:25
existing 35:7,25 36:15,24 37:2,19 38:1,10 39:5,10 41:24 43:2,17,22 103:3,5,15 105:22
expect 91:11
expedited 185:1 185:10

expert 7:22 8:24 9:2,6 11:2,21,24 12:22 13:3,14,16 14:13 15:2,7,19 15:24 16:5,9,19 22:5 23:12 24:25 26:8 27:1 27:2,12,12,20,23 27:25 28:7 29:2 29:6,9 43:15 45:25 47:15,24 47:25 48:16 54:10 55:10,19 59:19,24 63:12 69:6 70:5 74:5 77:15 189:7
expertise 17:10 19:25
experts 9:4
expires 187:24
explain 22:11,15 22:16 65:20 67:5 75:25 80:11 108:11 132:4 151:12 160:1 180:19
explained 23:12 42:20 63:3 105:25 141:13 141:17 181:15
explains 112:11
explanation 70:24 92:9 150:25
explanations 158:18
explicit 44:5
explicitly 48:13
explore 37:24 66:1
extensively 43:15
extent 48:5 53:25 56:17 61:4,7,12 87:8 100:9,12,16 101:19 102:2,16 103:6,8 105:2,5 105:11 153:9 168:7,11
extra 151:25

f

f 101:14 107:25 108:2
facebook 45:21
fact 23:6 33:18 44:16 60:16 62:16 75:20 108:4 121:23 141:17 154:11 154:11 159:5
factor 71:17 82:23,25 99:20 140:2,5,7,22 141:1,3,7 142:10 144:1,5,11,16 181:8
factors 40:7 55:5 56:18 57:6,7,8,9 57:11 58:3,8 61:19,22 62:1 63:6 78:5,6,11 78:12 80:21 82:23 84:9,18 86:4 91:7 141:12 157:3 181:7
fails 190:19
fair 5:25 6:1 67:20 84:2 87:6 100:23
falling 94:14
familiar 93:23
family 23:19 24:16 65:10,16 189:14
far 87:2
favored 182:5
february 187:25
federal 73:13,25
feedback 13:1
fees 14:21
fewer 76:15 89:22 92:2 93:8 93:13,15,19 94:1 94:5 95:9 161:6 166:25
field 20:3 33:23 40:10
fifield 45:2,5,12 45:18 46:9,12,13 46:19,21
figure 46:25 68:22 89:21 90:3,5 94:7,22 94:23 106:23 107:4 114:12 116:11 118:13 120:11,12,13 127:18,23,25 128:5,6,12,12,13 128:14,15 130:14,15,16,24 132:14 133:2 146:6,8 151:10 151:12,14 152:19 158:11 159:24 160:1,4 160:24 161:22 162:13 163:9,9 163:18 167:6
figures 15:17 97:18
file 17:24 26:13 173:10
filed 27:2
files 12:10 26:11 26:14 102:9
final 13:10 52:23 56:9 59:3,5,22 62:4 86:16,17,19 103:17 139:13 139:19 178:3 179:25
finding 117:14 117:20 123:22 141:10 149:25 151:8
fine 38:14 129:19 179:24 185:6
firm 4:12
first 4:2 9:16 13:22 14:1 36:17 40:18,21 44:3,23 45:2 64:12,13 66:2 76:4 78:25 95:24 104:13 112:19 113:2,23

127:4 140:9 156:19 158:6 169:14
five 7:3 65:8 68:10,13 72:14 72:15 90:11 92:6,11 99:8,9 138:9
fixed 114:10 154:4
flip 72:13 145:17 146:6
flipping 39:9 163:17
floor 2:15
florida 27:4
focus 20:4 26:12 29:11 96:3,4 123:7
focused 160:14
focuses 49:25
focusing 164:13 167:12
folder 8:5 9:17
follow 14:11 22:17 47:9
followed 83:10 137:22 184:2
following 33:23 47:11 61:12
follows 4:2
footer 38:21
footnote 44:23 44:24 71:23
foran 3:21
foregoing 187:5 192:5
forget 66:23
form 5:14 11:17 13:18 57:2,20,22 59:12 62:7 66:13,21 137:17 141:4 143:6 149:3,6 165:6,22 171:23 178:6 179:1
formal 14:2
forman 3:9,14
formed 175:5,8 175:24 176:4,10 176:11,14,19,23 177:1
former 45:7
formula 99:25
formulate 11:1
formulated 110:16
formulation 132:17
fort 187:11
forty 65:23
forward 93:25
found 29:9 113:18 141:16 156:20
four 7:2,2 27:3 34:2,7,8,9 72:18 90:11 92:6,10 128:5,6
fourth 38:24 43:4
fraction 98:6,21 132:13,22
fractions 97:23 98:1
freeze 61:10 135:8
freezes 103:9,11
frep 24:7
frequently 116:17 117:12 117:16 124:21 126:11 146:21
friday 7:5
front 73:10
froze 131:2 157:8
frozen 127:9 133:11
full 4:8 21:24 66:2
function 39:5
fund 2:14 14:3
further 127:5 150:25 177:11 184:13 187:7

g

g 15:13
general 38:7 39:18 40:2,8 55:7 83:11,18,22 83:24 84:8,13 136:18,22 152:17
generally 29:19 38:17 67:3
generate 32:24 33:23 35:17,17 35:20 38:2 41:8 41:25 44:17 50:16,17,18 52:19 75:22 76:5 86:24 89:18 92:2 97:21 152:13
generated 44:6 52:25 59:8 75:9 75:19 87:15
generates 33:12 40:22 51:3 82:17 95:3
generating 35:21 36:1,2 75:13 103:5
generation 51:11
generic 139:3
geocoding 26:19
geographic 73:3 88:2,14,23,23
geographical 31:1,16 60:13,20 105:18 126:1 132:12 168:14
geographically 70:19
geography 31:8 63:3 101:18 102:1,7,9 104:10 104:18
gerrymandering 28:10,12,14,16 28:20,22 111:5,8
give 87:19 124:6
given 75:7,24 80:25 101:18 132:7 138:16 153:24 173:21 173:22 180:7 192:9

gives 33:21
glad 90:2 162:5 175:19
global 167:18
go 5:15 24:25 25:18 43:3 44:22 49:13,21 56:1 62:20 64:23 65:2,7,21 68:10 77:8,15 89:1 95:3 99:7 101:9 106:23,23 109:12 113:11 113:23 126:20 126:20 130:11 134:22 138:2 151:10 159:1 172:18 177:4
goal 52:5,5 56:4 58:5,19 59:5,6 61:25 67:10 76:7,9 83:12 94:16 136:19 164:19 167:15 180:20,21
goals 81:2
goes 23:4
going 5:24 9:15 10:1 16:4 17:13 17:22 24:19 34:1 36:9 37:21 38:19 46:24 48:4,22 64:23 65:14,19 66:24 68:15 69:9,9,12 76:15,21 91:20 96:22 139:9 173:7,12 184:17
good 4:5,6 185:11
gore 2:20 4:4,11 9:25 13:6,12 16:12 19:6 25:9 25:13,17 34:6 47:3,8,14 48:8 48:14,21 49:12 49:20 51:19 52:2,14 53:21 54:8,20 55:20,24 55:25 57:3,23 59:15,17 60:9 62:10,16,19,23 62:24 63:14 64:22,25 65:12 68:21 69:18 77:8,11 126:20 126:24 134:22 134:25 137:18 143:7 148:21 149:4,11 165:17 166:1 172:1 173:16 177:4,7 177:10 179:1,11 181:11 182:17 184:12,24 185:1 185:6 188:6,10
graham 28:21
graphical 21:21
gray 158:17
great 6:9 29:23 69:12 114:11
greater 20:1 72:2 74:2,10,13 74:20 75:6 77:25 78:2
greenville 3:5
group 31:7,10,15 124:7 125:25 149:9
growing 37:15
guarantee 33:16 33:19,21 70:22
guess 70:25 109:15 146:14 147:15 174:12 177:21 179:22 185:7 186:8
guidance 15:16
guided 96:15
guideline 8:2,3 69:23 76:23 96:19 99:9 112:1 180:13,15 181:3 182:8
guidelines 8:17 8:18 68:18 69:2 69:15 72:2,3,6,6 72:10 73:8 74:11,12 83:14 96:17 99:11,15 101:6,7 102:21 103:1 136:25 180:5,7,9 181:20 181:22 183:3,9 183:14,17 189:18,20

h

h 15:10 189:1 191:3
half 64:23
halves 166:17
hand 90:5 116:11 187:10
handles 41:17,18
handwritten 8:8
happen 67:8
happened 122:22 123:5 180:24
happens 95:1 152:14 153:7
happy 185:7
hard 21:1 39:8 79:3,16 89:16
harpootlian 64:8 145:15
harvard 19:12 19:13
head 5:5 170:22
heading 108:2
heavily 84:9
heavy 89:11
help 86:21 89:12 128:4 144:8
helped 15:16
henry 190:4 191:1 192:1
hereto 192:7
hereunto 187:10
hierarchical 79:1,2 139:10
high 148:8
higher 85:10 110:23 119:19 120:19,23 121:7 121:19,24 130:3 130:20 141:18 161:15 172:12

highlight 43:16
highly 39:6 112:5 150:22
hired 13:21
histogram 90:6 91:2 107:16 120:10 121:13 130:25 133:2 152:14 160:22 160:23 161:20 163:23 164:1,4 170:16
histograms 107:2 127:15
hoc 68:18 189:17
hold 29:14 32:22 65:23 72:15 87:13 88:9 89:4 109:14 145:21
hope 34:18 68:22 184:16
hopefully 65:14
hour 7:11,11 64:23
hourly 14:20
hours 7:11,12
house 3:2 6:14 8:1,17 14:15,16 28:5,13,13,14 68:7,12,13,17 69:3 72:2,5,14 72:16 99:9,10 100:24 136:25 177:19 180:5,14 189:16
huge 132:24
huh 20:19 30:19 37:1 70:16 72:17 74:6 81:17 85:1,19 94:25 99:16,21 101:13 104:24 106:13 107:3 108:8 111:18 114:13 115:10 116:15 118:14 124:11 125:20 127:19 131:5 132:2 135:13 138:12 142:7 146:4 147:9 148:12,24 149:17 150:24 154:9 155:18 156:6 158:3 159:25 163:24 164:3 165:2 169:22 170:18 172:14 176:9,18 181:21
hum 179:21
hundred 76:13 76:19
hypothetical 91:22

i

ideal 72:23
identification 9:23 16:11 19:5 34:5 65:11 68:20 69:17 173:14
identify 28:9
image 121:17 123:9 161:21 163:11
imagine 62:23
imai 1:11 4:1,5 4:10,11 6:10,16 9:15,22 10:5 11:14 13:20 16:4,10,14 17:13 18:10,17 19:4,7 25:18 28:25 34:1,11 36:17 40:15 43:3 47:22 48:22 56:1 59:18 62:12,13,20,25 67:20 68:6,23 77:12 99:7 101:3 112:4 113:23 135:1 173:20 177:7,10 177:14,25,25 179:11 180:3 182:14,19 183:16 184:12 184:19 188:4 189:6,7,9 190:5 191:2,24 192:2,4 192:12
impact 67:7 168:15 180:25
implementation 80:19
implicated 170:20
implications 61:5,16 103:14
implies 154:12
implying 108:5
important 5:20 20:3 41:17 76:5 172:11 182:4
importantly 45:6
impose 41:19 42:22 57:14 58:6 63:2 142:21
imposed 33:10 33:11 39:13 58:1 61:1 67:11 106:22 140:24
imposes 54:12
imposing 13:10
impossible 39:11 42:19
impressions 178:23
improved 26:18
improves 38:10
impute 26:15
inadmissible 29:9
inbox 11:9
include 37:19 55:3,4 56:11 57:11 62:3 63:17 103:20,24 104:6,25 106:2 107:1 111:13,20 112:20,22 113:9 117:22 126:15 160:11 170:16 170:23
included 12:1,6 12:19 17:23 27:4 37:4,6

47:25 51:21 54:11 55:11 59:9,20,21 60:5 63:12 113:21 124:16 140:18 143:10 169:3,5 178:10
including 54:24 56:8,23,23 106:3 181:10
incorporate 36:12,21 39:4 49:23 56:7 180:18 181:4,5
incorporated 112:2 136:10 142:24
increase 81:7 136:3
increases 141:19
increasing 136:15
incumbency 56:15 60:25 61:3 79:23 81:4 81:8 84:21,23 85:20 86:22 87:16,23 139:16
incumbent 12:12 73:23 77:21,22 78:15 86:20 105:5 106:8 116:22 153:9,12 153:17
incumbent's 87:20
incumbents 78:17 79:25 80:6 81:7 86:11 86:14,16,25 103:7,25 104:2,5 106:5
independent 35:25 58:18
index 3:23 188:16,18
indication 180:10
indirectly 100:8 100:17
individual 20:11 149:2,7
individuals 26:5 26:11 115:8,8 118:21 160:17 160:17 170:19
inference 20:10 26:17
influence 57:12 58:9
inform 84:13
information 12:11 26:13,14 55:4 59:1 104:5 109:19,21,22 111:9,14,23 112:1 118:10,23 159:19,22 170:13 181:16 188:16,18
informed 67:15 83:13
inheriting 58:8
inherits 61:18,22
initial 157:15
input 44:24 93:6
inputs 102:17
inquire 20:10
instances 7:16
instruct 78:13
instructed 80:15 96:10 103:7 110:12,22 112:3
instructing 55:20
instruction 182:1
instructive 67:2
instructs 5:15
insulate 62:17
intend 16:23 17:1
intent 30:3,5 118:1 136:22 182:6,22
intention 58:18
intentional 30:1
intentionally 29:24
interchangeably 109:10,11
interest 101:19 102:1,7,14 104:23 105:1,5,7 105:10,16,19 132:1
interested 19:21 19:23 187:9
intern 3:21
interpretation 22:1
introduce 9:16 68:23 69:19
introducing 16:13
intuitive 85:9
invalid 40:12
involve 51:12 162:1,21
involved 15:6 27:14 42:11 45:15
involving 179:4
isolate 57:10 105:25 180:25 181:8
isolating 55:1 56:6,16,25
item 11:10
items 11:6,10
iterative 22:21

j

jacksonville 27:3 27:6,10 28:18,19
jasper 134:20 156:12 157:19
jcusick 2:17
jmgore 2:22
john 2:14,20 4:11 48:4 62:22
joined 101:20 102:2
jones 2:19 4:12
jonesday.com 2:22
journal 21:21 23:3
journals 22:22
judge 99:24 100:5

judged 99:22 100:24
judgment 67:13 67:16 175:6 184:7
judiciary 4:14 68:17 69:16,24 189:16,21
july 7:8 27:2
june 18:17,24,25 19:3 189:9
justified 73:18

k

k 15:11,11,12,12
keep 76:24 109:2 109:4 135:14 136:13,19 142:1 150:1,2 151:9 154:21 159:12 165:8,10 170:2 170:25
keeping 103:25 106:5 139:25 141:8,13 157:12 168:8 171:2 183:16,20,23
kentucky 28:22
kept 96:6 97:23 98:2,6,21 132:14 132:23 141:24 169:18
kevin 15:12,12 15:21
kind 65:2 92:20 95:5 101:3 111:4,20 126:16 145:1 164:4
kinds 78:24
know 6:3 7:25 8:5,20 11:1,6 12:2,5,16 15:13 17:10 18:24 19:23 20:25 21:8,8,12,23 23:5 28:13 29:7 29:10 31:8,14,15 35:13,17,22 37:8 38:8,13,14,18 39:21 40:2,8,10 41:13 42:14 43:24,25 45:16 45:18 46:5 48:9 51:13 53:10 54:10,12,23,25 56:3,3,9 57:7,10 57:11,13 58:4 59:1 60:18 61:2 61:9,12,16,19 62:1 66:20 67:8 75:12,17,25 80:2 80:5,24 81:2,5 82:5,9 83:13 85:8,10 86:12 87:24 89:15,21 89:22,25 91:24 92:12 93:7,14 95:9 96:5,15,20 97:20 98:4 100:13 102:12 102:14,19 103:6 103:9,13,18 104:3,11,12 105:3,5,13,18 110:15 111:2,6 111:11 112:8,24 113:2,18 116:21 117:3 120:7,24 121:16 122:8,22 123:19 124:24 125:1,2 129:8 131:11,19 132:3 132:10,13,19,24 133:17 134:10 136:3,8,14 137:15 139:3,8 140:6 141:8,19 143:23 145:8 147:16 148:1,7 149:9 150:2 151:15 152:6 153:3,6 154:4,17 154:20 156:16 156:18,25 157:7 157:18 158:6,9 159:15,16 160:4 160:25 161:3,3,6 161:25 162:3 163:1,2 164:9 165:12 166:13 166:14,22 167:12 168:1,2,7 168:11,13 169:4 169:5,14,15,17 169:18,20 170:21,22,23 171:13 172:23 173:3 178:19 179:13,15,21,23 180:6 182:7,8,10 185:13
knowledge 167:10
known 44:8
konw 141:12
kosuke 1:11 4:1 4:10 9:22 16:10 18:17 19:4 188:4 189:5,7,9 190:5 191:2,24 192:2,4,12
kuriwaki 15:8 15:10

l

l 81:3
label 179:20
language 72:8 72:11
large 36:11 56:20 117:10,17 118:8 146:22 150:21 154:11 187:4
largest 122:21 123:5 156:21
launch 23:24
law 4:12 73:13 73:25
lawsuit 179:8
lawyer 67:12,23 137:14 184:6
lawyers 46:7 49:22 178:16,22
lcrum 3:17
league 64:6 145:12
leaving 44:5
left 45:20 90:17 107:17 131:16

legal 2:14 14:3 30:15 31:16 36:12,21 67:2,14 67:15,17,22,24 68:4 125:24 137:15,20,21,25 184:6 190:23
legally 66:4,19
legislative 19:15 20:4 27:14 28:7 68:19 150:13 155:2 158:23 170:8 189:19
legislators 150:17 182:23 183:2,7,12
legitimate 73:19
letter 10:17 14:2 72:21 103:2 106:11
level 74:18 75:10 75:14,23 76:25 77:1 87:24 91:25 136:9,14 157:13 171:12
liberal 19:10
liberties 2:4,9
lieu 48:17
light 40:15
lighter 116:19
likelihood 146:9 146:16
likewise 103:24 175:8
limitations 73:3
limited 64:15 91:8 98:20
limits 79:4,13
line 94:11 107:17 120:24 131:16 148:25 149:18 151:18 151:21 158:1,14 158:17 163:21 165:20 188:2 189:2 191:4,7,10 191:13,16,19
lines 149:23 173:25
list 11:10 14:10 26:25 27:11 66:11 70:17 77:17
listed 10:22,24 11:21,23 14:13 66:10 106:17,19 180:12,15 183:9 184:1
lists 103:2
literature 21:12
litigation 16:21
little 8:11 13:21 32:18 42:3 70:3 89:12 90:9,17 97:6 132:23 144:8 175:11,21
liu 9:7
live 31:15 118:21 124:7 128:3,9 148:14 149:10 152:1,22 160:5
lived 128:23
lives 116:23 117:2,5 151:15 153:10
living 31:8 60:19 123:25 124:4 125:25
liz 177:15 184:15
llp 3:9,14
loaded 65:13
loading 16:14
local 64:10,14 122:16,24 134:12 135:8 141:16
localized 81:13 84:23 89:6 97:8 97:9 112:19 113:24 141:18 148:15
location 1:14 12:12 55:15 60:17 87:23 124:8 125:18 146:23
lodging 55:22
long 7:9 23:6 47:12 111:1
longer 17:18
look 12:20 26:21 60:12,14 61:7 76:18 92:12 93:7 97:21 106:11 111:11 123:18 125:11 125:14 130:14 130:23 144:14 157:14 158:11 160:4,23 167:11 171:8 180:1
looked 179:12 179:17
looking 47:13 74:25 93:18 98:16,18,23 123:18 127:10 179:22
looks 84:22 107:6 117:10 131:18
lot 35:12 76:14 161:8 166:24,25 172:25
loud 36:9
louisiana 2:21
lower 85:11 119:21,25 121:11 130:4 139:24 147:7 161:14 172:20
lowest 73:1
luke 4:14

m

m 2:5,10,20 3:15 4:10
ma 1:14
mace 88:19 116:22 117:16 117:22 118:5 155:8
mace's 118:3 146:20 181:24 183:8
main 3:5,11,16
maintain 171:12
major 98:17
majority 107:12 107:22 131:7,15 154:12

makers 182:4
making 39:7 73:20 81:10
map 28:25 29:3 29:16,19 58:12 58:15,16,19,23 58:25 59:1,6 76:7 83:16 117:21,25 118:21 154:21 154:25 155:12 159:11,16 170:2 170:5 178:2,6,11 182:4,4,6,22 183:1,7,12
maps 36:11 46:9 83:17 179:12,14 179:18,25
march 27:5
mark 65:7 68:15 69:12 173:7,12
marked 9:22 16:10 19:1,4 25:3,13 34:5 47:18 65:2,11 68:20 69:17 73:7 99:14 101:5 173:14
markov 23:22 33:7 35:6,10,19 35:24 37:3,4 38:1,6 50:3,8 65:10,16 189:14
massachusetts 6:17
master's 19:11
materials 7:24 7:25 8:19,20 11:12,12
math 129:14
mathematical 23:16 33:12,19 33:20,21 67:18 81:18 98:9 99:25 100:18,19 100:22,25 102:22 105:17 111:2 132:16
mathematically 110:16
mathias 3:4 177:18,18 184:17 186:7,8 186:12
matter 53:13 58:22 73:17
matters 54:3 133:17
maximum 76:14
mburchstead 3:12
mc 38:25
mcm 39:16
mcmaster 190:4 191:1 192:1
mcmc 37:18 38:1 39:2,19 40:9,16 41:3,18,23 44:8 44:13,17
mean 16:1 22:20 23:7 32:9 35:12 40:10 49:5 50:5 53:23 59:23 65:20 66:20 78:8,25 83:24 84:12 85:15 91:16 92:7,19 96:13 102:6 104:16 108:11 108:19 110:15 115:21 120:2 121:10 130:1 132:7,10 138:20 139:6 140:4,6 141:14 143:4 148:19 159:20 165:7 166:2,7 169:2 172:4 174:11 185:14
meaning 31:17 85:9 115:21 125:24 140:23
meaningful 131:20
means 4:21 85:17 101:21 102:3 108:23 125:25 130:21 138:22 139:1 167:1 171:20 172:12
meant 85:16 178:6
measure 98:1,22 102:22 132:11
measures 97:19 97:22,24 98:4,9 100:18 132:11
media 101:20 102:3
members 63:22 145:2
memory 72:13 111:7
mention 78:23 102:22 178:9 181:25
mentioned 86:10 91:17 102:24 119:3 178:14 180:14 181:2 183:19 184:1
mentions 26:4 68:7 76:23
merge 23:11,21 24:2 33:6 41:5,6 42:6,8,10 79:10
merged 79:8,9
merges 79:9
merging 41:11
met 81:3
method 20:9 35:1,4,6,10,11 75:19 93:23 97:16
methodological 43:13
methodology 19:23,25 26:5,10 26:13 32:19 33:3,16
methods 20:1 23:15 24:17 36:11,15,24 37:2 37:3,6,12,19 38:11
metric 130:20
mgl 1:6
michael 3:10
mill 187:12
mimic 83:16

mind 163:17
minimizing 106:12 108:22
minimum 90:20 91:11
minus 71:19 74:8 79:6,15 90:22,23 91:8,19
minute 77:9 123:8
mirror 121:17 123:9 161:21 163:11
mischaracterizes 13:4 59:13 148:17
missing 11:10
mistaken 117:3
misunderstood 134:3
mix 39:6 44:9
model 60:24,25 82:11,15 84:9,11 84:12,15 85:9 100:4 132:4 146:20 161:23 163:12 171:15
moderate 89:10
modify 75:10
modifying 24:4 24:7 35:22 36:2
moment 47:7 66:2 126:21 167:18 177:5
monday 185:4,6 185:11,18,19 186:2,9
monitor 18:13
monte 23:15,19 23:22 24:12,16 33:7 34:3 35:1 35:10,11,15,19 35:24 37:3,4,12 38:1,6,10 50:3,8 189:11
moon 9:9
morning 4:5,6
motives 182:22
move 97:3 135:4 153:13 169:8
moved 60:15 122:4 147:18
movement 176:25
moving 115:1,18 115:21,22,23
msmc 43:1
multiple 31:2 35:18,20,22
municipalities 78:2 94:20,21 95:6 96:25 97:2 102:16 105:4 108:5,12,16,17 108:22 109:3,5 168:12
municipality 96:23 108:7,12 108:18,19

n

n 15:12,13 188:1
n.w. 2:21
naacp 1:3 2:3,14 14:3 64:1,3 145:4 190:4 191:1 192:1
naacpldf.org 2:17
name 4:8,9,11 26:14 27:16 179:19
nancy 88:19 116:22 117:15 117:22 118:3,5 146:20 155:8 181:24 183:7
natural 123:7
nature 35:14
near 116:24
nearly 72:22 85:8
necessarily 22:20 23:7 60:18 75:17 108:16 141:23 147:20 157:7 174:11
necessary 192:6
need 5:3 6:2 8:7 85:5 96:18 125:11 180:24 185:4,19 186:5
needs 14:24 40:12,13 86:6 104:4
neither 187:7
nested 108:15,16
net 115:2
neutral 112:21 113:3,4,18
never 22:20 23:8 61:25
new 2:16 25:13 41:25 79:11 173:7
nexsen 3:3
nexsenpruet.c... 3:6
night 129:14
nods 5:4
non 27:23
noncontiguous 56:19 57:16
nonlegal 137:16 137:22
nonpartisan 111:10,15
normally 54:23
north 119:16 125:8,19 126:16 147:17 151:1,5
notary 187:3 192:13,19
note 190:10
noted 179:13 192:7
notes 8:9
notice 9:21 10:8 25:9 85:20 189:4
number 20:25 21:2 58:3,8 60:17 72:21 77:24,25 78:1,3 79:4,5,14,15 80:1 85:2,3 86:3 86:5,8,13,15,18 87:12,12,14,19 87:19 89:12,23 89:24 90:11,14

90:16,20 91:12 92:13,24 93:1,2 93:7,15,18 95:10 96:20 98:19 108:18 114:18 116:6,8 120:25 127:21 130:21 133:18 137:1,3 138:4 139:18 146:13 147:8 150:21 151:14 154:4 160:7 168:10,12 170:19 173:2
numbered 34:8 43:8
numbers 80:10 90:10 104:8 116:16 147:16 162:4 170:17 172:10 174:22 176:2 183:25
numeral 72:18 73:12 99:17 101:11
nw 2:5,10
ny 2:16

o

o 15:10,12
oath 4:21
object 5:12 48:4 57:2,20 59:12 62:7,7 137:17 141:4 143:6 149:3,6 165:6,22 171:23 179:1
objection 5:13 13:4 49:11,19 51:18 52:1,13 53:20 54:7,19 55:18,23 60:6 63:10 148:17
obligation 4:22
observe 142:1
observed 141:10
obvious 170:25
obviously 5:14 98:5 102:8
occur 31:10
occurs 31:19 124:21
offer 16:23 17:11
offered 174:13 174:17
offers 176:7
official 187:11
oh 14:10 15:12 17:17,25 19:16 20:23,25 21:4,17 25:21 26:1 28:19 32:14 38:23 45:13 46:2 49:5 79:2 90:4 91:3 93:4 109:8 112:24 113:16 114:25 115:25 119:13 120:12 122:14 124:20 128:8 129:11 132:6 133:8 134:1,14 135:25 137:4 138:7,22 144:13 144:23 153:11 154:3,19 158:8 160:25 164:7 168:20 170:22 171:5 173:23 175:15 179:21
ohio 14:15 27:13
okay 5:6,17 6:9 6:12,16 9:24 10:3 13:17 14:13 15:10 16:6 17:15,17,25 18:9 22:19 25:2 25:7,23 26:3,23 28:11 29:14,14 32:22 34:9 36:8 38:23,23 40:18 41:1 43:6,9 47:2 47:15 48:8 50:7 50:12,15,22 51:9 52:24 53:7,12,14 54:16 55:14 60:20 61:21 63:20 65:2,4,23 65:23 66:15,25 67:25 68:11,13 68:14,23 69:11 70:2,4,7,11 71:1 71:17 72:12,15 72:15 73:6 78:19,23 80:7 84:17,19 86:9 87:4,18 88:5,7,9 89:1,4,4 90:4,18 91:23 92:22 93:21,23 94:19 96:10 97:5 99:7 99:9 101:8,10,15 105:12 106:11 109:1,14,17 111:16 112:16 113:13,25 114:3 116:5 117:8,15 118:12 119:15 121:3,4 123:14 124:10 126:6,6 126:22 127:2 128:11,14,14 129:13,19 130:13,15,15 131:15 133:1,7,9 133:12 135:6 137:2 138:4,9,9 138:10 143:13 143:19 144:17 144:25 145:10 145:19,21,21,23 146:8,25 149:14 152:3,13,24 156:14 157:25 159:2,4 160:13 163:17,19 168:22 169:10 171:8 172:9 173:9 174:2,5,13 174:21 175:2,18 175:22 182:20 184:23 185:8,17 186:3,6,13
old 25:11
once 6:11 33:11 35:16 49:5 50:18 79:13 122:1
ones 22:7 125:12 152:5

open 44:5 69:11
opened 9:24 16:8
operate 78:10
operates 102:9
operational 111:2
opinion 30:4 57:22 66:13,13 66:21 112:8,10 112:13 137:15 137:16,19,20,23 138:1 154:24 174:13,17 175:5 175:9 176:14,23 177:1 182:25 183:4,6,10,11,15
opinions 11:1 16:20,23 17:11 29:9,12 175:24 176:4,11,19
opposed 75:21 76:3,25 81:16 86:7 125:14 133:18 139:3 167:12
orangeburg 168:23
order 41:21 105:25 112:24 142:23 155:20 181:8
orders 31:7
outcome 48:16 182:9
outcomes 182:2
outlets 101:20 102:3
outlier 49:9
outperforms 107:21
output 82:20,21
outside 61:11 138:15 139:9,15 140:17 147:21
overall 71:19 73:2 98:16,23 100:10 135:14 140:1
overlap 60:13,14 60:20,21 61:2,5 61:17 63:2 104:14 105:4

p

p 173:14 189:23
p.m. 186:14
page 10:13,14,15 17:13,15 24:21 25:20,23 26:3 29:12,14 30:18 32:21 38:19,20 38:21,21 43:3,4 43:5,8 44:23 65:21 66:3 70:12,15,15,17 70:17 78:18,21 84:22 88:8 89:2 89:4 94:23 99:15 101:9 102:25 104:22 106:24 108:2 109:13,16,17 114:2,12 120:10 120:12 127:3 130:11 138:2,4,6 138:8 145:18,22 151:10 157:24 159:3,24 163:18 174:3 176:5,6,16 176:17,24 188:2 189:2 191:4,7,10 191:13,16,19
pages 127:14 145:17 164:2
pair 80:6 81:7 86:11,14,15,24 95:4
paired 61:3 77:21 78:16 81:4 104:2 105:6
pairing 42:23 78:15 79:24 84:21,24 85:20 86:20,22 103:7 106:8 139:17 153:12
pairings 56:15 87:16,23
pairs 79:25
paper 7:23 21:20 21:23 22:20,22 23:1,2,5,7 34:5 34:14,16,21,25 36:7 189:12
papers 45:9,10 45:11,13
paragraph 25:19 25:19,24,25 26:4 26:22 27:11 28:10,16 29:13 30:17 32:21,23 38:24 39:24 43:11,12 70:10 70:14 74:7,7 78:20,24 80:7,22 81:15 84:22 88:6,11 89:2 97:7 99:20 106:17 107:25 108:1,2 109:18 109:19 114:11 119:3 121:1 129:1 138:3,8 145:22 147:4 150:19 154:8 159:1 162:6,10 162:15 169:21
paragraphs 77:16,21 106:20 109:13
parallel 35:18,20 37:24
parameter 79:20 80:19 81:17 93:5
parameters 81:1 82:19 84:15 85:7 93:16 95:8
part 8:4,16,19 8:21 18:6 19:14 19:18 20:9 23:15,18 24:16 30:1 35:23 46:3 54:10 61:12 62:14 63:5 69:5 96:16 99:23 108:4 143:2 148:7 156:19 174:9 178:6

partially 61:1
particular 17:12 21:24 22:2 23:18 24:1,8 31:1,7,19,23 39:21 40:5 42:24 56:4 66:14,21 67:9 75:3 79:23 80:25 81:19 86:6 112:7,9 125:6,12 138:23 139:3 142:12,16 142:22 156:1 157:11 164:15 164:20,22 181:7
partisan 28:16 28:22 32:15 109:19,21,22 110:16 111:5,7,9 111:11,12,13,23 112:1 122:14 130:10 155:9 159:19 170:13
partisanship 176:17 181:12 181:16
parts 49:15 101:19 102:2 117:10 155:12 157:7,8
party 110:10 187:8
passing 177:12
patience 93:22
patricia 2:9 178:19
pay 174:12
pdf 38:20
peer 15:23 16:1 20:12,21 186:8
penalizes 138:14 138:21
pending 6:6 187:8
pennsylvania 28:4,5
people 14:1 44:24 60:15,17 60:18 71:22 76:13,19 104:20 115:12,15 170:25
percent 12:18 27:15 41:21 45:21 71:20 74:8,15 89:24 115:12,16 116:1 116:3 117:6,13 120:4,15,15,18 120:19 121:2,7,7 121:11,24 129:15,18,22 130:3,4,8 135:15 135:16 136:20 136:20,23,24 137:5,12 139:18 139:21 140:2 154:13 159:5 161:1,1 162:6,10 162:14 163:6,15 165:9,12,25,25 166:2,3,22,23,24 167:4,5 169:23 170:24 171:16 171:20,21,24 172:6,11,12,15 172:22,23 173:1 183:24,24 184:4 184:5
percentage 41:20,20,23 42:5 87:15 110:24 114:19,22 115:3 119:6,10,19 120:2 127:21,24 127:25 128:16 129:10 137:8 147:6,13 160:8 166:13
perform 54:22 107:10
performance 21:24 32:15
performs 107:7 131:7
periodic 6:4
permissible 71:4
permit 72:3
permitted 74:11
person 45:2 74:2 74:13,16 75:6 91:9,20 166:15
persons 72:21
ph.d. 16:10 189:7
phd 1:11 4:1 9:22 19:12 188:4 189:6 190:5 191:2,24 192:2,4,12
phrase 30:16
pick 37:13,14
place 7:4,5 95:15 96:7 139:14 142:1,14 155:17 167:7 169:24
placed 4:20 116:13,17 117:16 118:5 146:10,16,21 147:19,20 151:6 152:4,8,22 155:13 168:10
placement 151:1
places 152:25 153:25 160:16
placing 155:7 159:18 170:12 171:3
plaintiffs 1:5 2:2 8:24 9:2 11:15 11:25 12:8 179:8
plan 16:20 23:17 24:1,2,4 29:20 29:24 30:8,15 32:1,8 35:7,22 36:1 41:24 43:2 47:23 48:24 49:2,4,9,14,15 49:16,18,23,24 49:25 50:4,6,9 50:10,14,17,20 50:21 51:7,12,17 51:23,25 52:5,6 52:8,12,21 53:5 53:24 54:6,13,24 54:24 55:3,4,7 55:12,17 56:7,7

56:9,24 57:1,5,8 57:18,22,24 58:2 58:5,6,7,9,11,13 58:14,16,20,23 58:24 59:2,3,7,9 59:11 60:1,4,22 60:23 61:5,9,13 61:14,14,15,18 61:22,25 62:2,5 63:3,9,18 64:1,3 64:6,8 66:6,18 74:1,22 75:9,19 76:6,8,12 77:24 78:1,3 79:17,25 82:4,6,8 83:11 84:10,14 87:20 87:22,25 88:18 88:18 89:19,23 89:25 90:1,15,21 91:4,8,12 92:5 92:10,13,13,19 93:3,9,9,19,20 94:2,2,11,18 95:11,19,21 96:2 96:6,8,12,25 97:14 98:14,20 98:22 99:2 100:11,12 103:8 103:19 104:9 106:2,2,3 107:6 107:10,21,22 111:10 112:11 113:1,5,7 114:10 118:17,18 120:1 120:1,8,9,20 121:4,15 122:22 122:23 123:20 124:12,17 125:6 125:7,10,17 127:9 128:20,24 128:25 129:5 130:18 131:3,6 132:21,22 133:5 133:21 134:9 135:21 136:9 137:5 139:8 140:10,13,17 141:1,11,12,25 142:18 143:1 144:10,12,13,17 144:18,21 145:4 145:7,13,15 147:2,3,6 149:2 149:15 150:1,2 150:10,10,14,20 151:4,6,9,17,18 151:20,21,22 152:7,12 153:2,8 153:14,19,25 154:18,19,20,20 155:19 156:23 156:23 157:21 158:7,9,9 160:19 161:7 164:23 165:8,8,19 167:8 168:2,17 169:12 169:16,19 170:25 171:20 172:5,11,13,24 173:2 174:25 179:4 180:1,23 181:5,6,8,10,10 183:12

plans 31:24 32:2 32:25 33:17 34:5 35:18,20,22 36:2 37:9,23,25 38:2 44:7,18,19 46:9 50:19,20,23 51:4,11 52:20 55:7 57:22 58:14 63:21 66:3,4,17,18 67:1,3,21 68:4 71:9 75:22 76:14 77:4 78:14,16 79:21 79:22 80:2,6,16 81:7,11,22 82:3 82:5,14,17 83:3 83:21 84:16 86:11,13,15,17 86:17,18,19,24 87:9,14,19 88:18 89:18,22 92:1,21 92:23,25 93:8,10 93:11,12,25 94:14,17 95:3,4 95:9,20 96:24 97:21 98:12 99:23 100:6,11 100:15 103:5 104:8 107:8,12 117:15 119:6,12 119:15,18 120:14 124:3,14 124:18,22,25 125:2,4,12,15 126:11,12 129:4 130:2 131:8,16 132:23 138:14 138:21 139:14 139:19 140:3,15 140:19 142:6,10 142:20 143:2,9 143:14 144:2,6 144:14,15,16,21 144:22,24 145:1 145:8 147:8,12 147:20,24 148:13,20,20 149:22 150:5,23 151:20 152:21 152:25,25 154:1 154:12,22 159:6 159:7,13 160:10 161:1,5,14,16 162:1,7,10,17,20 162:23 164:18 164:22 165:4,24 167:7,12,13,22 169:24 170:4 172:19,21,25 173:4 189:12

plausible 66:5 66:19

play 55:2 153:5 153:20,21

played 30:13 56:25 58:11 59:4 76:10 106:1 112:25 113:5 142:15 156:22 180:22 181:1

playing 142:2

plays 56:6

please 4:9 5:21 6:3 15:9 67:5

plus 53:3 71:19 74:8 90:21,23 91:8,19

point 39:24 40:2 40:19 53:16 56:5 64:22 71:6 71:6,10,15,15 119:10 125:13 137:5 156:2 166:8 172:10 174:25 178:9 179:3,18
points 34:21 39:12 119:6,19 120:2 147:6,13
polarize 184:11
polarized 184:9
policies 73:20
political 19:12 19:25 20:2 32:7 32:10 73:21 118:2,10 155:6 159:17,20 170:11 182:11
politicians 182:10
politics 110:13 110:13,15 112:5 112:11 176:17 183:2,2
polsby 97:23,25 98:17 130:17,20 131:9 132:18,20 133:3
poor 175:20
popper 97:23,25 98:17 130:17,20 131:9 132:18,20 133:3
population 12:10 54:2 56:20 57:13 60:14,21 63:4 71:19 72:7,10,18 72:23,24 73:13 73:16 74:2,23 75:2,5,12,20 76:21 77:3,19 90:22 91:9 102:10,11 114:18,19,21,23 115:9,20 116:6,9 120:6 122:1 128:3 129:12,13 151:15 155:21 160:5,8 161:4,23 163:12 165:13 171:25
populations 23:18 104:18
portion 117:24 118:8 146:20 170:3
portions 117:17
possibility 44:5 44:21 75:21 113:17 141:13
possible 32:24 37:8 42:18 60:10,12,13,24 61:21,24 63:1 73:2 75:4 83:2 87:7 88:4 96:13 111:1 147:25 148:1 150:8 153:4 165:18 167:21 179:17
possibly 150:4,5 162:25
potential 83:21
power 66:5
practicable 72:23
practically 43:17
practice 33:24 39:5
pragmatically 66:5,19
precinct 74:20 74:23 75:10 76:25 106:12,15 108:7,13,20 109:7 115:13,15 115:16 116:1,9 116:25 117:5 146:10 153:17 176:25
precincts 24:9 37:14,16 39:9 74:25 75:24 76:1,3,24 108:14 108:21 114:14 115:1,6,7,11
precise 82:9 146:23
precisely 62:2
predicting 26:5 26:10
predominant 30:16
predominated 30:7
preferred 154:21 159:11 170:2
prepare 6:20 11:5 17:4
prepared 17:20
preparing 8:12 12:25 15:1,7 69:6
present 3:20 7:13,15 46:6 74:18 91:25
presented 51:13 103:17
preservation 54:17 76:23 103:15 180:10 180:18
preserving 73:22 103:3,4,21 105:22
president 4:13
pressure 186:8
pretty 81:8 117:4
prevent 56:5,15 56:24 58:9
prevents 55:1
previous 54:24 55:7 56:8 103:18 160:3 175:10 181:10
previously 25:3 47:18 99:14 101:4
price 143:1,4
primary 41:24
princeton 45:7
principles 57:19 77:17 84:3,5 184:2

printed 7:23 8:6
prior 73:22 99:23 100:5,15
prioritized 180:11
priority 180:15 180:16 182:13
probability 148:8
probably 97:1 104:14 146:24 148:1,3 152:5 161:24 162:2
problem 72:12
problems 43:18
procedure 88:11
proceed 24:10
process 21:22 22:21 23:4,4,6 58:21 63:23 79:7 80:4
produce 10:20 73:1
product 59:5 181:4
program 70:21 71:18 75:5 78:5 78:9 93:1 106:14 168:4
programmed 51:1 92:17 132:8 143:18
progresses 41:14
prohibitively 44:9
properties 66:7 81:10
property 33:12 33:20 79:18
proportion 42:4 42:25 116:7 119:4,25 120:7 120:14 135:15 136:4,13,17 140:1 147:5 157:12 163:22 163:25 164:8,11 168:8 171:4,5,13 171:14 172:25 173:4
proportions 170:17
proposals 39:9
proposed 63:22 66:6 179:12
prospect 183:8
protect 48:13
protected 48:6 59:14 62:9
prove 39:8
provide 12:8,14 12:21 13:2 46:21 121:13 185:5
provided 13:8 46:16,17 163:22
provides 174:23 176:3
provision 96:19
pruet 3:3
public 63:22 145:2 150:12 155:1 158:23 170:7 179:13 187:3 192:19
publication 10:25 20:16,22 22:10
publications 20:13 21:1,6
published 20:16 20:22 21:21 22:21 23:8 34:16 45:14 180:4,7
pull 16:4 17:22 18:1 25:15 34:1
purpose 38:13 76:12 113:20 141:8,15
purposes 67:7
put 55:11 93:25 94:16
pyan 2:12

q

qualified 29:2
quantify 87:14
question 5:21,24 6:5,6 8:25 20:17 22:12 36:17 41:3 47:7 48:5 50:2 52:16 53:15 62:4,8,25 81:12 83:2 84:2 90:3,24 91:1 112:9,12 115:4 124:8,21 131:24 134:3 137:20,21 137:25 141:25 144:3 155:15 164:25 167:19 175:20,21 180:2
questions 5:13 5:16 14:11 47:21 59:13 63:15 70:8 95:13 96:23 112:15 177:11 177:13,16,20,22 181:12 182:15 182:18 184:13 184:18
quite 50:6,11

r

r 3:10 15:10,11 191:3,3
race 26:5,11,16 29:25 30:7,13 40:19,21,24 41:1 42:8 55:2 56:6 56:16,25 57:9,10 58:10 59:4 61:6 61:10,23 76:10 95:24 103:8,11 106:1 109:18 112:5,10,21,25 113:3,4,5,18 118:22 119:5 140:25 141:7 142:2,15 143:25 144:5,11,15 156:22 157:1,4 168:9 180:22,25 180:25 181:8
racial 12:10 28:10,11,14,19 102:13 118:23 141:12 184:11

racially 184:8
ragusa 9:13
raise 162:5
raised 81:14,24 82:6 97:9
ran 48:8,17,24 50:8 52:9 53:7 54:5 60:5 62:6 62:11 64:10
randomly 23:23 24:2,3
range 42:25 73:2 90:12 94:17,18 122:2 131:14,21 132:21 137:4,7 138:15,25 139:9 139:15 140:17 141:9,14,24 142:9 171:16
ranges 21:8,12 172:18
rankin 4:14
rare 22:22
reach 14:24 96:3 172:21,21
reactive 81:23
read 36:9,18 38:3 39:14 43:10,19 44:2,11 50:7 66:8 73:4 88:10,16 100:2 101:2,23 108:9 138:18 147:10 154:15 159:9 181:19 184:20 186:16 190:9 192:5
reading 90:9 116:10
reads 101:16
real 56:4 91:16
realistic 36:12 36:21
really 36:3 52:19 57:10 59:2,5 66:21 67:13 76:10 85:6,25 89:14 91:22 139:6 156:25 167:11
rear 3:23
reason 4:24 6:3 42:17 48:10 55:6 57:5,17 62:2 74:17 75:15,16 76:22 81:23 85:5 96:14 105:24 113:13,16 117:21 129:21 164:12 169:2 190:11 191:6,9 191:12,15,18,21
reasons 21:7 40:18 81:18 143:16
rebuttal 17:4 173:13,15,17 174:18,23 175:9 175:11,13,17,25 176:3,5,12 189:23
recall 7:7 9:4 10:10 12:5 13:22 14:25 21:5,7,13,18 40:1 41:5 48:25 55:10 59:20 60:3,8 63:7,8,13 72:5,8,9,11 79:9 87:17 96:17 120:21 133:8 134:11 139:22 149:5 157:22 168:21 169:2,4,7 173:6,23 174:1,6 178:23 179:22 181:11
receipt 190:18
receive 11:14,19 11:24 46:12,13 46:18
received 8:6 11:21 13:1 18:25 19:9,11,12 69:7 70:1
receiving 10:10 174:6
recess 64:24 77:10 126:23 134:24 177:6
recognize 10:5 16:15 18:19 34:11 68:23 69:20 155:24 173:20 179:19
recollection 179:23
recombination 65:9,15 189:14
record 4:8 5:4 13:5 19:2 53:15 64:23 77:8 126:21 134:22 150:13 158:23 158:23 170:8 175:14,21 177:4 187:6
recreate 174:21
rector 2:15
redirect 182:19
redistricting 4:16 8:2,17,18 17:10 19:15,18 19:22,24 20:5,8 21:3 27:7,8,20 27:21,23 28:1,5 34:4,15 37:9,25 38:18 43:18 45:10 46:10 54:24 56:6 63:23 65:10,16 67:3 68:8,12,18 68:20 69:2,4,15 69:16,23 72:14 73:20 74:1 83:15 96:16 99:10 136:25 181:1 182:13 189:12,15,17,19 189:20,21
redrawing 55:2
reduce 108:18 141:20 168:11
reduced 86:15
reductive 85:4
reelection 118:3 183:8
refer 29:19,21 70:2

reference 8:7
referenced 62:13 162:14 190:6
referred 37:2
referring 40:21
refers 30:25 176:24
reflect 60:18
reflects 132:11
regarding 156:1
regardless 66:15 183:13
regenerate 23:17
region 146:24
regions 61:1
registered 26:15
registration 110:9
regular 184:24 185:23
rejected 20:22 21:6 22:10,24 23:7 154:22 159:13 170:4
rejection 22:19
rejections 21:2,8
relate 22:4 162:14
related 16:20 22:7 23:18 31:24 56:24 57:24 59:9 60:1 60:4,23 61:22 62:5 63:18 98:2 98:2,5,7 100:14 102:17 132:15 132:18 187:7
relation 50:9
relative 81:22 85:9 89:15 101:16 133:14 139:6
relevant 11:12 37:25 40:7 43:14 71:18 77:22 97:13
reliance 66:11
relied 13:3
rely 12:22
remainder 165:24
remember 6:12 116:25 172:17 174:10
remove 172:20
removed 139:19
renaming 88:11
rendering 39:10
repair 176:25
repairing 155:3
repeat 8:25 20:17 22:12 52:16 90:24 144:3 155:15 164:25
rephrase 99:1 134:4
report 6:22 7:22 8:14 11:2,17,21 11:24 12:22,25 13:3,11,14,16 14:14 15:2,7,15 15:19,24 16:5,9 16:18,19,24 17:2 17:4,8,12,14,21 22:5,15 23:12 24:25 25:10 27:2,12,23 29:15 29:23 30:2,6,6 30:24 31:14,18 32:5,21 40:17 45:25 47:1,15 48:9,10,13,15,18 50:7 51:14,21 52:22,23,23,24 53:18 54:11 56:10 58:1 59:19,22,24 62:14 63:12,16 63:17 68:6 69:6 70:6 74:5 77:3 77:15 78:18 80:7 88:6 92:15 94:1,24 97:12 103:18 106:24 111:19,22 112:18 122:15 126:10 130:12 138:3 140:25 141:6 156:17 157:24 166:18 169:3 173:13,17 174:3,18,23 175:4,9,11,11,16 175:17,25 176:3 176:5,13 178:1,3 178:6 179:25 180:21,21 189:7 189:23
reported 1:16
reporter 1:16 4:20 5:2 10:5 184:19,23 185:3 185:8,13,17,21 185:24 186:3,6 186:10,13 187:1 188:12
reports 8:3,15 8:15,24 9:2,5,6 13:9 27:17 46:7 47:25 48:1 55:11 169:5
represent 4:12
representative 8:2 23:17 32:24 33:4,13,18 44:10 44:14 101:22 135:20
representatives 68:17 69:3 73:24 189:16
represented 5:7
republican 32:12 110:4,19
republicans 182:5 183:13
request 6:5 11:4 18:25 55:13
requested 188:16,18
require 84:4 110:23
required 73:19 110:18 155:12 192:13
requirement 67:9 72:9
requirements 67:3,15,16 73:12
requires 84:4

research 15:5,6
residence 12:12 105:6
residences 103:25 106:5
resides 88:20 153:18
respect 22:3 49:9 55:12 63:2 106:8 145:1 152:24 156:4 165:15
respecting 73:21
resplit 23:25
rest 28:15 42:9 61:8 96:5 152:21 154:6
restraint 61:1
resubmit 22:25
result 48:2,23 49:1,7 84:15 89:17 119:22 159:21
resulted 58:20
results 48:19,25 51:22,25 57:12 59:19,21,23 63:8 63:13 76:18 89:20 132:20 136:11 142:13 142:22 157:3 170:14 178:13
retained 14:8
retention 176:6 176:8,12
retreat 23:25
return 11:8 190:13,17
review 8:12 15:23 20:21 21:13,22 23:3 69:5,25 150:12 150:16 155:1 190:7
reviewed 6:22 8:14,15,17,20,23 9:1,6 17:8 20:12 21:12 158:22
reviewing 9:4
reviews 16:1
revise 22:24
richland 133:22 158:2,6,14,16,20 159:6,12,18 160:5,18,20 161:2,5,8,9,17 162:7,11,23 164:2,24 165:5 165:15,20 166:4 166:20 167:8,17 168:19 183:20
right 17:3 18:9 25:5,5 26:7,21 28:17,23 29:11 30:3,4 33:1,5 34:1,18 35:1,5 44:20 47:20 48:20 50:16 51:1,8,8 53:2,6 56:18 58:24 60:1 64:4,17 65:17 68:2,8,13 70:19 71:15,20 71:24 72:3,16,20 73:9 74:9 75:21 76:4 80:13 82:10 83:21 84:11 86:22 87:2,10,11,11 88:3,15,17,21,24 88:25 89:8 90:5 90:7,8,15 91:5,6 91:9,10 92:3 93:3 94:2,3,5,6 94:10,15,16,24 94:25 95:7 96:13 97:10,11 97:11,18 104:10 105:17 106:7 107:8,13,18,23 108:24,25 109:12 110:1,5,8 110:17,21 111:5 111:22 112:14 112:20 114:6,10 114:11,15,16,19 114:24 115:20 116:11,14,15,18 116:19,21,24 117:4,7,10,13,18 119:7,8,20,23 120:4,16,20,22 121:8,9,11,12,16 121:25 123:13 124:1,2 125:5,16 125:22,23 126:8 126:17,19,25 127:7,8,16,18,22 127:23,25 128:2 128:4,16,17,17 129:6,15,18,23 130:4,5,11,18,19 131:9,17 133:12 133:15,16 135:4 135:17,21,22 138:2,9 140:12 140:12,15,16,16 140:23 141:5 142:20,21 143:3 144:7,20 145:17 146:1,4,11,15,17 146:22,24 147:3 147:21,22 148:5 148:6 149:12,21 149:21,25 150:23 151:8,11 151:23 152:5,5 152:11 153:3,12 153:16,18 154:1 155:23 156:5 158:12 159:15 159:24 160:3,9 160:21,21 161:19,24 162:8 162:8,9,12,15,16 163:3,8,13,14,16 163:16 166:5,5 166:12,13 167:10 170:5 171:10,17,25 172:7 173:7 178:2,7,8,8,11 178:12,16,25 179:5,9,14,15 180:5,20 181:13 181:17,18,20 183:4 184:5,10 184:24 186:16 186:17
rights 137:13
rmg 1:6

roads 101:20 102:3
robust 136:11 157:3
robustness 157:14
role 30:13 55:2 56:6,16,25 57:10 58:10 59:4 76:10 105:25 112:25 113:5 142:2,15 156:22 180:22,25
roman 72:18 73:12 99:17 101:11
room 6:18
rough 185:5,7,10 185:14,15 186:11
rounds 129:24
ruiz 1:16 187:3 187:24
rule 5:20 48:7 59:14 62:9
rules 48:13 181:2
run 15:16 23:24 47:22,24 48:18 50:3,13 53:17,18 53:23 81:24 82:12 86:23 95:24 113:14 149:19,23 157:4
running 146:3
runs 149:9,18

s

s 2:14 3:5 15:10 15:12 189:1 191:3
sake 42:7
samantha 3:21
sample 23:17 32:24 33:4,13,18 36:13 39:3 44:4 44:9,10 78:16 79:25 82:8,16 95:9
samples 33:22 44:14
sampling 34:4 36:10 39:7 43:17,22 75:22 80:4 81:20 189:11
satisfied 81:11 143:9
satisfies 140:10
satisfy 32:25 79:22 80:2 140:13 142:18 143:14
satisfying 142:25
saturday 7:6
saying 47:12 93:24
says 18:17 32:23 36:10,23 44:3 48:16 66:16 72:21 73:12,14 88:11 92:15 97:12 99:22 100:24 109:18 109:20 121:1 138:14 140:25 141:6 142:14 146:5 150:19 162:6,10
sc 3:5,11,16
scale 36:11 39:7 85:5 89:15,15
scholarship 19:19,22 20:5
science 19:13 20:2
scope 67:25
score 133:3
scores 130:17
scott 190:4 191:1 192:1
scratch 24:12
screen 18:2,11
scroll 24:20 38:19 99:13 102:25 174:3
se 40:4
seal 187:11
sean 8:3,15 173:14 189:23
second 20:9 26:6 40:24 43:11 44:7 61:9 64:14 64:17 65:1 89:6 95:24 96:2 99:15,20 103:10 107:25 108:1 157:1,4 177:25
section 73:15 126:10
see 10:17,19 14:15 18:2,14,15 20:23 30:21 31:23 45:3 47:3 67:7 82:13 89:21 93:4 94:8 99:18 102:24 108:3 111:8 113:5,16 115:14 115:17,17,25 120:24 126:9 127:14 132:6 138:22 141:16 156:21 158:8,11 160:25 161:22 163:25 164:7 181:25 182:1
seed 37:13,16
seeing 107:4 146:19 178:24
seen 53:10
select 137:2
selected 36:14 82:10 83:10
senate 1:15 2:18 4:12,13,14 8:3 8:18 9:20 68:7 69:15,24 72:3,6 73:8 74:11,12 91:7 101:6,7 102:21 103:1 136:25 180:5,13 189:4,21
send 184:21
sense 5:5,6,19 6:7,8 29:20 42:5 67:17 76:24 83:12 100:22 103:13 142:12 162:1

sent 6:23 7:22,24 8:1,4,16,18,20 9:17 11:10 12:19 43:7 90:2 174:9 190:14
sentence 30:17 36:7,23 37:3,19 37:22 43:11 66:1,2,14,16,22 66:24 88:10 108:1 109:7 154:10
sentences 37:22 40:1 44:3
separate 124:1
separated 126:1
sequence 35:21 42:1,22
sequential 24:11 34:3,25 35:4,10 35:14,23,24 38:10 189:11
sequentially 35:23 38:2
set 32:23,25 33:9 33:17 37:9 44:7 44:19 50:17 51:3 52:10,11,19 52:20 53:5,8,17 53:18 56:12 57:6,11 58:14 59:8,10,25 67:10 81:1 87:10 93:24 94:1,4 109:20 139:13 140:10,14,18 142:19 143:2,10 144:21
sets 52:25 89:18
setup 88:18
seven 29:13 30:17 69:13 73:7 90:21 91:4 91:18 92:5,10 101:5 109:13 110:19
seventh 10:13,14
shakes 5:5
shape 55:15 100:20,22 102:9 132:12 158:19
share 10:2 16:14 18:1,2,3,10 25:16 34:10 47:4 65:13 69:19 110:24 179:7
shared 12:2,6,10 18:20 178:23
sharepoint 46:15
she'll 5:3
sheet 190:11
shifted 90:17
shiro 15:7,10,14
show 49:8 116:16 127:17 146:14,14 157:15 161:20 164:10
showing 73:18 121:13 128:6 133:4,5 151:13 165:14 170:16
shown 118:12 158:13
shows 94:7 114:14 116:12 120:13 130:16 146:9 151:14 152:11 160:15 166:21,21 169:23
side 46:15 90:5 94:15 116:11 153:22
sign 184:20 186:16 190:12
signature 187:22 188:11
signed 190:20
significant 30:13 140:2,5,7 141:1 141:3,7 142:2,9 142:11,15,17 144:1,5,11,16 180:22
significantly 131:12
similar 41:4,12 61:14 122:16,24 127:4 134:12 139:16 152:13 160:3 163:25
similarity 54:12
similarly 106:4 125:7 126:15 139:20 176:24
simple 30:25 75:17
simply 11:3 31:20 37:8 163:11 166:9
simulated 31:24 40:22 50:17,19 50:20 51:11 75:9 76:14 79:17 81:10 82:3,5,14,21 83:3 84:16 86:13,15,16,17 87:22,24 88:1,13 88:18 89:22,25 92:13,21,23,25 93:8,9,11,12,19 94:17 96:24 97:21 98:12 100:10 117:15 119:6,12,15,18 120:1,8,14 121:15,20 124:17,18,25 125:1,4,7,10,12 125:15,17 126:11,11 131:15 132:21 132:23 136:2 139:8 141:11 144:13,14,17,20 144:21,23 147:7 147:8,12 148:20 150:1,10,23 151:8,17,20,22 152:7,12,21 153:19 154:12 156:23 159:6,7 161:1,5 162:1,7 162:10,20,22 165:8,24 167:12 167:13 168:2 169:24 170:25

172:11,19,24
simulation 17:9 17:10 21:16 22:4,9,14 32:1 33:17 34:15 36:20 40:21,25 42:8 43:14 47:23 48:2,23 49:2,3,23 50:3,8 50:13,23 51:3,15 51:22,24 52:10 52:11,20,25 53:4 53:8,18,23 54:11 56:12 59:8,25 60:24,25 61:10 62:6 63:9,16 64:11,14,19 66:17,18 67:1,21 71:9 74:17 76:17,25 77:4,18 81:14 82:7,12,17 84:23 85:18,23 86:1,1,7,8,11,23 86:24 87:10,18 89:18 90:7 91:14,15 92:2 94:9,14 95:22,24 95:25 96:2 97:9 103:11 104:8 107:8,11,21 111:9,12,13 113:3,8,14,17,19 113:20,25 114:4 114:9 118:22 121:5 122:3,7,17 122:25 124:3,14 127:1,5 128:18 129:4 130:2,5,24 131:8 132:5 133:3 134:13 135:5,7,18,25 136:10 138:13 139:14,19 140:3 140:15,18 141:3 141:6,22 142:6 143:2 144:2,6,10 145:20 147:20 148:10,11 149:2 149:22 150:5 151:4,24 152:25 153:2,8,14 154:1 156:3,8 160:19 161:14,16 163:21 164:18 164:22,23 165:4 165:19 167:7,21 167:22 168:4 171:20 173:5 178:13 180:3 181:6,13,23
simulations 15:17 40:20 41:2 42:13 44:18 54:5 63:21 67:6 78:7 103:8 107:23 109:23 112:15 112:19,19 130:18 135:8 148:15 178:5,10 179:4
single 38:17 90:11 178:1 183:17,20
sinko 15:11,11
sir 25:22
sit 116:21 175:2
six 29:13 37:14 37:14 40:23 61:11 68:15 69:10 72:13 73:6 79:6,14 90:11 91:17,20 92:2 99:14,17 101:4 107:7 110:19 122:18 122:20 123:6 130:13 131:4 134:1 142:3 162:11,24,25,25 163:7 183:13
size 75:3 102:10 102:11
skip 37:21
slash 72:19
slate 24:13 35:4 35:15 36:4 41:7 41:10 42:3,21
slightly 89:25 90:17
slow 44:9
small 18:13 73:17 80:1 86:13,18 87:12 139:18 161:22
smaller 89:24
smc 36:3 37:18 37:24 38:5 40:17 41:3,7,17 42:9,12,20
soft 79:19,20,24 80:5,14 81:5 86:23 93:17 139:11,12
software 18:4 45:14 63:5,6
solange 1:16 187:3,24
solutions 190:23
sorry 8:25 25:25 47:6 63:7 85:15 108:1,12 109:15 113:11 134:1 138:6 155:15 162:20 164:25
sort 125:13 132:15 139:2 157:14
sorts 61:19
sound 129:15 142:8 172:6
sounded 178:14
sounds 129:18 140:9 174:7
south 1:1,3 2:2 4:16 6:14 8:1 14:16 20:13 28:12,13 29:16 31:4 46:9 68:7 68:16 69:3,15 96:16 108:15 112:5 180:4 187:4,12 189:16 189:20 190:4 191:1 192:1
southeast 190:15
space 37:25 39:7 39:12
speak 91:22
speaking 23:21 104:22 125:16

166:8
specific 23:25 24:10 31:8 39:3 39:20,22 40:3,7 43:25 48:25 63:13 73:19 80:25 92:13 96:18 98:21,24 99:4 125:2,12,15 137:1,2 138:23 138:24,25 139:2 149:9 156:1 167:13 168:1 182:1,8 183:18
specifically 59:4 183:25
specifics 14:25 21:7,14 59:20
specified 32:25 33:9,14 96:17 97:20 111:2 171:12
specify 41:21 43:1 97:1 168:1 182:8
specifying 33:9 85:7
spell 4:8 15:8
spelled 4:10
spike 154:11 167:3
spin 115:5
split 23:11,21 24:3,18 33:6 41:5 42:6,8,10 77:24 78:1 79:9 79:12,14 89:5,7 89:7,22 91:20 92:6 93:17 94:20,21 95:18 95:20,23 96:1,4 96:7,11 97:2 108:5,17,23 123:24,25 124:4 124:18 125:2,5 126:12 133:20 133:24 134:2,7 148:14 151:4 154:18,22 155:3 157:6,19,23 158:7,10 159:6 159:12 160:10 161:5,17,25 162:7,23 163:6 164:15 165:11 166:14,17 168:2 168:15,18 169:11,15,18 170:9,20
splits 79:5,10 89:19,20,23 90:1 90:14,20 91:12 91:17 92:2,10,14 92:16,25 93:2,7 93:9,12,13,15,18 94:1,5 95:6,10 95:12,17 96:21 96:23 106:15,15 107:2,8,13 108:7 108:13,18,23 125:8,17 139:4 150:20 160:20 168:12,12 177:1
splitting 31:1,2,3 31:3,15,20,21,23 41:11 79:8 108:20 109:3,5 155:3 166:10 167:23 168:5
spoken 184:15
spread 41:6 96:18
staff 45:6 179:9
stand 43:21,23
standard 33:22 147:6
star 44:24
start 24:1,4,7 25:19 35:15,21 36:1,4,6 37:12 37:15 41:24 42:2 43:2 70:9 114:2 145:18 156:25 159:5
started 58:12 122:20 156:17 178:4
starting 30:17 36:4 41:10 145:22 176:16
starts 24:12,13 29:12 35:4,6,15 35:25 37:22 38:25 41:7 42:20 43:12 66:3 70:15 127:3
state 1:3 2:2 14:15,16 27:14 28:5,14 39:12 73:24 75:8,24 102:8 107:20 155:13 187:4 190:4 191:1 192:1
statement 38:7 38:17 43:21 66:12 126:16
statements 150:16
states 1:1 41:24
statewide 41:15 42:5,13 64:19 85:18,25 90:6 92:1 95:22 102:12 104:14 107:22 113:8,14 113:16,20 133:23 135:4,7 135:25 138:13 139:14 140:2,15 141:3,6,15,22 144:1,6 145:20 148:11 156:3,7 157:2 160:19 163:21 164:18 164:22,23 165:4 167:22 173:4
statistic 126:16
statistical 20:1 26:12 30:11,12 31:23 40:4 99:25 124:20 125:13 131:11
statistically 131:19
statistics 19:11 21:22 167:10,16 174:22 176:2,7 176:20
step 50:19 113:2 167:17

steps 52:19
straddles 137:8
street 2:5,10,15 3:5,11,16
strength 80:9,11 80:16,20,22 81:23,25 82:9,10 82:16,23 83:4,19 83:25,25 84:24 85:10,21 89:11 89:11 95:2 105:15,21 106:5 138:10,16 139:5
strengths 80:8 80:13 83:9
strict 73:1
strong 80:17 84:7
stronger 85:11 85:22
struggle 36:15 36:24 39:6
struggling 11:4
student 45:7
studies 19:19
study 20:1
studying 21:25
subcommittee 69:16 189:22
subdivision 73:21
subject 78:10
submit 11:3 22:23
submitted 11:13 17:20 20:15,20 23:2 27:5,12,17 27:22 145:2
subpoena 10:20
subscribed 192:14
subset 160:14 163:14,15
substance 48:5 48:15
substantive 76:16
successful 36:10
suggest 181:22 182:3
suggested 30:15
suggestion 56:2
suitable 38:15
suite 2:5,10 3:5 3:11,16
summarize 38:9 161:12
summary 29:12
sumter 133:23 169:8,14,15,18 169:24 170:3,12 170:15,20 171:3
supplement 17:1
suppose 129:17
supreme 28:4
sure 11:9 20:18 22:13 23:8 27:16 32:11 34:23 45:22 47:8 61:3 79:12 81:10,21 82:4 83:23 84:20 86:21 90:25 93:14 95:14 100:10 107:19 112:17 114:1 115:4,5 117:3,4 126:2 129:25 136:7 141:23 143:17,20 144:4 153:6 155:16 157:2,5,11 158:5 165:1 167:20 168:20 175:23 175:23 185:16
surname 26:19
suspect 119:13
swapping 24:8
swath 146:22
sworn 4:2 192:14
systematically 44:6,18

t

t 189:1 191:3,3
tab 9:17 16:7 25:8,9,10,15,21 34:7,9 47:16 65:3,4 68:10,13 69:10 72:13,15 73:6 99:8,9 101:4 173:10
table 130:24,25 163:23 169:23
tabs 47:12
taiwan 190:4 191:1 192:1
take 5:2 6:2,4,4 6:6 7:4 19:14,17 46:24 50:23 66:24 67:8 69:9 75:9,18 77:6 95:17 100:12 102:13,15 115:7 124:10 129:20 131:19 136:21 166:2 185:7 186:2,9
taken 1:15 42:14 64:24 77:10 102:18 126:23 134:24 177:6
talk 13:20 50:25 53:3 70:14 89:2 89:5 113:23 126:25
talked 32:18 35:3 77:18 97:6 118:24 148:22
talking 92:20 95:12 109:20 123:8 128:12 164:5 165:23 175:10,13
target 33:8 36:14 39:3 82:19
taught 45:7
technical 42:17 70:24
tell 4:22 14:12 41:21 42:3 66:17 76:2 107:15 150:8,9
tells 80:17 85:5
ten 45:17 60:16
tend 150:1,2 165:12,13
tends 139:23 151:9 165:8

term 30:23 125:23
terms 41:12 102:14 107:20 109:9 118:24 119:25 123:23 129:9 133:14 141:12 142:5 156:22
test 31:19,21,22 31:23 72:13
testified 4:2 27:22 28:6 178:1 179:3 180:17
testify 4:25
testimony 5:3 10:21 59:13 148:18 150:13 155:2,2 170:8 190:9,18 192:8
thank 4:7 9:15 13:20 24:19 26:21 28:20 36:6 44:24 50:22 55:24 62:23 65:20,25 77:2 86:9 88:17 89:1 90:2 98:8 99:7 107:1 111:16 126:4,7 128:14 133:7 159:21 177:10 177:21 179:24 180:2,17 181:11 182:14 184:12 186:13
theorem 33:20
theory 39:2
thereof 187:9
thick 158:17
thing 76:4 109:6 124:14 132:19 166:14,16
things 5:13 76:23 105:3 132:8
think 6:11,14 7:2 7:7,15 12:16 13:24,25 14:6,17 15:25 19:20 25:9 27:13,16 28:4,6,17 31:13 31:14 38:16,20 39:18 40:4 45:13,17 46:11 52:24 53:9,15,25 71:22 75:7,18 79:3,17 84:12 86:10 88:8 89:6 91:16 94:7 96:19 106:17,22 109:15 111:6 115:6 116:24 117:2,19 119:3 120:6 121:1 123:1,3 124:15 126:4,14,14,17 126:18 128:11 129:14 133:22 137:5 139:22,23 143:21 144:9 146:24 147:4,15 148:22 149:15 152:23 153:3,19 161:8 162:6 166:6 168:25 172:6 173:24 174:8 175:1,16 178:12,19 180:13 183:18 183:22 184:14
third 36:7 64:19 162:22
thirteen 25:23 26:3
thomas 1:7 4:13
thought 134:1 182:7
thoughts 178:24
thousand 74:24 115:15
three 15:6 19:1 25:14 27:3 65:3 65:4 71:23 88:7 101:11 104:13 162:25,25 174:9
time 1:13 5:12 5:12 8:23 9:1 11:4 21:25 24:14 31:10 35:16 36:5 41:8 41:16 117:6,13 167:6 171:1 178:4 190:19
timeframe 190:8
times 6:9 7:2,3 7:17 62:15 166:7
titled 99:17
tjh 1:6
today 4:7,15,22 4:25 5:8,20 6:16 7:3,21 175:2 177:11
today's 6:20 7:19 8:13
tokyo 19:10
told 48:19 93:14 129:13
tolerates 163:12
tomorrow 185:14
top 10:15 66:3 170:22
topic 19:22
topics 20:3
total 79:5,15 114:17,22 116:8 116:8 127:20 128:16 146:13 160:7 171:4
touched 8:11
tradeoff 84:7 108:6,11
tradeoffs 84:4,4 84:6
traditional 30:7 57:18 83:14 181:1 182:12
transcript 3:23 186:5,16 187:5 190:6,20 192:5,8
transition 39:11
translates 129:9
travel 11:9
traverse 39:6
treatises 10:25
trende 8:4,15 17:8 173:14,18 174:22,23 176:6

176:16,24 189:23
trende's 174:14 175:3,25 176:11 176:20
trends 19:24 20:8
trial 16:24
tried 18:10 26:15 42:16,20 47:6 100:9 129:14 153:11 157:14
trivial 87:12
trouble 90:9
true 44:7,19 82:22 88:1 114:25 141:2 155:24,25 187:5 192:8
trump 110:24
trust 129:18 172:8
truth 4:22
truthfully 4:25
try 23:16 42:15 68:3 77:3,7 109:4 168:11
trying 20:10 25:15 76:5,6 109:2 149:13
turn 9:18 10:12 17:13 32:20 70:5 73:6 74:5 78:18 101:4 120:10 126:19 157:24 184:16
turnaround 184:24 185:2
turning 94:19
twenty 88:7 130:13 138:9
two 7:11,12,16 8:3,15 14:1 16:7 16:13 20:6 23:23 24:2,6 25:4,5,8,10,16 25:21 27:18 41:1,2,6,10 42:10 45:13,13 47:16,19 52:19 61:7 63:1,5 64:3 73:13,15 75:16 79:8,10,11 83:4 86:1 88:13 97:19,22 104:13 114:22 115:6,7 119:23 121:17 122:9,18 126:1 130:8 133:13 136:5,15 145:7 157:1 162:11,23 162:24 163:7 166:15,17 167:13 173:24 174:8 180:7,9
tyler 15:11,18
type 23:11 24:18 33:6,6,24 37:17 42:6,15 78:14 79:21 80:16 98:1 102:17
types 12:8 24:6 24:15
typically 74:21

u

u 15:11
u.s. 73:16
uh 20:19 30:19 37:1 70:16 72:17 74:6 81:17 85:1,19 94:25 99:16,21 101:13 104:24 106:13 107:3 108:8 111:18 114:13 115:10 116:15 118:14 124:11 125:20 127:19 131:5 132:2 135:13 138:12 142:7 146:4 147:9 148:12,24 149:17 150:24 154:9 155:18 156:6 158:3 159:25 163:24 164:3 165:2 169:22 170:18 172:14 176:9,18 181:21
ultimately 153:15
un 179:25
unable 4:24
unchanged 114:9
unclear 53:16
understand 4:17 4:21 5:18,21 17:18 22:13 32:20 50:6,24 52:18 59:6 67:25 70:25 78:25 85:3 89:12 93:21,24 95:1 114:4 115:4 127:4 128:4 130:16 143:21 144:8
understanding 20:2 45:20 52:17 71:25 83:22 128:5
understood 5:25 23:9 86:6 162:13
unfortunately 17:23 36:14
union 2:4,9
united 1:1
units 24:8,8 74:21
universe 83:20 95:3,4
university 19:10
unknown 57:6 57:11 58:8
untouched 61:8
unusual 31:22 31:24 150:22 164:14,14
updated 17:23
uploaded 34:10
uribe 1:16 187:3 187:24
use 12:4,17 17:9 26:13 30:24

31:7,22 33:6,22 39:4,20,22 40:16 41:23 42:6,7,9 54:23 55:7 61:17 71:14 74:7,15 75:1,25 83:23 87:23 102:22 103:18 105:7 109:7,22 110:2 111:4,8,22 118:18 122:14 125:23 128:25 144:17 147:3 149:13 155:9 159:19 170:13 181:9,10,12,16 182:1,9
uses 30:20 102:10
usually 31:7

v

v 190:4 191:1 192:1
valid 39:16,18 44:7,19
validity 40:13 175:7
value 81:17,19 81:22,25 82:7 85:2,4,10,11
values 82:13,18
variables 12:3
variation 74:24 122:1 163:12
variations 161:23
variety 176:7
various 53:3 84:5 130:17 149:1
varying 88:25
vast 107:12,22 131:7 154:12
verbal 5:3
verify 190:9
veritext 190:14 190:23
veritext.com 190:15
version 25:11,13
versus 28:21
vertical 151:18
viability 67:17
viable 66:5,19
video 9:21 189:5
videoteleconfe... 1:16 2:5,10,15 2:20 3:4,10,15
view 38:5 56:5 71:10 112:4 131:22 183:6,11 184:3
violated 57:18
vitae 19:4 189:9
vote 110:24
voter 26:11,13 26:14 110:9 128:19
voters 26:15 31:11,15,20,21 49:10 64:6 104:19 118:21 119:11 122:3,6,8 122:12 123:15 123:17,18,21,24 123:25 124:3,7 124:12,22 125:5 125:25 127:21 128:8,23 129:3,5 129:8 145:13 148:4,13,14 149:10,16,23 150:1,6,7,9,11 150:22 151:6,7 151:25 152:1,8 152:11,14,16,17 152:21 153:1,25 154:5 164:23 165:4,15,19 166:10 167:2,7 167:23 168:10 168:13 176:25
voting 106:12 114:17,18,21,23 115:8 137:13 151:15 160:5,7 160:17,18 170:20 184:9
vra 41:20 42:22 85:18 90:6 92:1 107:20,22 109:20 112:20 112:22 113:9,14 113:19,21 135:5 135:12 138:13 139:14 142:25 142:25 173:5 184:1,5
vs 1:6
vtc 1:11
vtd 106:15 107:2 107:8,12 108:17 116:13,16 118:16 146:10 147:1
vtds 108:5,20,23 109:4,4,7 114:14

w

w 15:11,13
wait 133:8 138:4 175:10
waive 186:17
walk 34:20 78:4 84:17
wang 15:12,13
want 9:18 13:20 25:18 26:24 32:20 36:6 38:7 38:16 40:6 43:10 44:2 47:8 56:18,22 57:10 57:16 61:17 62:3 63:15 70:24 81:3 106:2 108:17 109:6 111:14 113:1 130:23 140:6 167:11,18 185:14,17 186:1 186:10
wanted 41:19 75:16 81:21 82:4 106:21 117:22 126:2 136:7 143:20 156:24 157:5,10 175:14 183:12
washington 2:6 2:11,21

water 71:3,3
way 23:25 24:4 24:10,23 27:17 38:9 45:16 50:2 60:12,13 66:15 66:23 76:20,20 83:16 89:16 94:8 111:3 131:10 133:10 137:23 143:9 144:18 150:19 151:5 157:6,18 164:16 182:25
ways 60:7,10,21 63:1,4
we've 118:24
weak 81:6
weaker 85:12
weakness 40:8
week 7:6,8 48:17
weeks 7:8 27:3 173:24 174:9
weight 79:20
weighted 84:8
weights 106:7
went 45:21 57:7 62:1 87:9 139:17,20 181:7
who've 60:18
wide 98:22 107:20
witness 9:24 13:7 26:8 27:1 28:7 29:6 57:21 60:7 63:11 126:22 141:5 148:19 149:7 165:7,23 171:24 173:15 177:9,12 186:15 187:10 188:3 190:8,10 190:12,19
witnesses 43:16
women 64:6 145:12
wonderful 107:1
word 30:20,20 129:20
words 65:20 66:20
work 14:18 54:6 185:12
working 83:20 121:17 178:4
works 20:15,21 29:21 45:18
world 91:16
written 13:15 20:12,21 34:14 40:2
wrong 13:25
wrote 16:18

x

x 188:1 189:1

y

yan 2:9 7:15,18 13:24 178:20
yeah 9:5,19 12:25 14:16 18:6,12 24:22,24 25:12 26:2 57:5 64:2,13,13,17 68:5 82:25 83:6 84:2 85:14 87:1 87:5,13 89:14 90:13,16,19,25 92:9 94:13 97:15 104:2,15 105:13 106:16 106:18,21,21 107:19 108:3,14 110:2,6,11,21 113:12 114:25 116:4,6 117:2,9 120:5,6,17,21,24 120:24 121:1,6 121:22 122:2 124:20 126:4,14 126:18 129:7,17 129:18 130:6,22 131:1 132:10 133:13,14,16,16 133:22 134:2,5,5 134:5,6,10 137:10 140:6,23 143:19,19 145:11 146:7 147:15 148:3,22 149:12 151:3 152:20 153:19 154:16 155:25 158:16 161:8,11 161:13 162:13 162:19 163:1,1,2 163:5,5 164:10 168:20,20 171:18 172:5 173:11,25 174:8 176:22 177:3 178:14,19 179:15,21 180:6 181:25 185:6
year 29:17
years 12:10 45:16,17 60:16
yield 44:10,14
york 2:16 187:12

z

zero 85:16 167:1
zip 9:17 17:24 65:5 173:10

South Carolina Rules of Civil Procedure

Part V. Depositions and Discovery

Court Rule 30

(e) Submission to Witness; Changes; Signing. When the testimony is fully transcribed the deposition shall be submitted to the witness for examination and shall be read to or by him unless such examination and reading are waived by the witness and by the parties. Any changes in form or substance which the witness desires to make shall be entered upon the deposition by the officer with a statement of the reasons given by the witness for making them. The deposition shall then be signed by the witness, unless the parties by stipulation waive the signing or the witness is ill or cannot be found or refuses to sign. If the deposition is not signed by the witness within 30 days of its submission to him, the officer shall sign it and state on the record the fact of the waiver or of the illness or absence of the witness or the fact of the refusal to sign together with the reason, if any, given therefor; and the deposition may then be used as fully as though signed unless on a motion to suppress under Rule 32(d)(4) the court holds

that the reasons given for the refusal to sign require rejection of the deposition in whole or in part.

DISCLAIMER: THE FOREGOING CIVIL PROCEDURE RULES ARE PROVIDED FOR INFORMATIONAL PURPOSES ONLY. THE ABOVE RULES ARE CURRENT AS OF APRIL 1, 2019. PLEASE REFER TO THE APPLICABLE STATE RULES OF CIVIL PROCEDURE FOR UP-TO-DATE INFORMATION.

VERITEXT LEGAL SOLUTIONS
COMPANY CERTIFICATE AND DISCLOSURE STATEMENT

Veritext Legal Solutions represents that the foregoing transcript is a true, correct and complete transcript of the colloquies, questions and answers as submitted by the court reporter. Veritext Legal Solutions further represents that the attached exhibits, if any, are true, correct and complete documents as submitted by the court reporter and/or attorneys in relation to this deposition and that the documents were processed in accordance with our litigation support and production standards.

Veritext Legal Solutions is committed to maintaining the confidentiality of client and witness information, in accordance with the regulations promulgated under the Health Insurance Portability and Accountability Act (HIPAA), as amended with respect to protected health information and the Gramm-Leach-Bliley Act, as amended, with respect to Personally Identifiable Information (PII). Physical transcripts and exhibits are managed under strict facility and personnel access controls. Electronic files of documents are stored in encrypted form and are transmitted in an encrypted fashion to authenticated parties who are permitted to access the material. Our data is hosted in a Tier 4 SSAE 16 certified facility.

Veritext Legal Solutions complies with all federal and State regulations with respect to the provision of court reporting services, and maintains its neutrality and independence regardless of relationship or the financial outcome of any litigation. Veritext requires adherence to the foregoing professional and ethical standards from all of its subcontractors in their independent contractor agreements.

Inquiries about Veritext Legal Solutions' confidentiality and security policies and practices should be directed to Veritext's Client Services Associates indicated on the cover of this document or at www.veritext.com.

Exhibit 3

Adopted September 17, 2021

2021 REDISTRICTING GUIDELINES
SOUTH CAROLINA SENATE JUDICIARY COMMITTEE
REDISTRICTING SUBCOMMITTEE

The South Carolina Senate Judiciary Redistricting Subcommittee adopts the 2021 Redistricting Guidelines to aid the Redistricting Subcommittee and interested parties in developing and evaluating redistricting plan proposals. These Guidelines are drawn in part from the guidelines adopted for prior redistricting, the 2002 opinion of the three-judge court in *Colleton County Council v. McConnell*, the 2012 opinion of the three-judge court in *Backus v. South Carolina*, other court decisions, and input received in public hearings across the State.

I. REQUIREMENTS OF FEDERAL LAW

A. Population equality

1. Legislative districts

The Fourteenth Amendment to the U.S. Constitution requires an honest and good faith effort to construct legislative districts as nearly of equal population as is practicable. The good faith effort required by the Fourteenth Amendment does not preclude the pursuit of legitimate redistricting policies such as making districts compact, respecting political subdivision boundaries, preserving the cores of prior districts, and avoiding contests between incumbent legislators. Any redistricting plan with population deviation ranges of greater than ten percent (10%) between the most-populous and least-populous districts is presumptively unlawful unless the good faith effort required by the Constitution is proven. So that the State may avoid assuming this additional burden under federal law, population deviations of individual districts shall be within plus (+) or minus (-) five percent (5%) of the ideal population and within an overall range of less than ten percent (10%). This guideline does not affect the requirement of an honest and good faith effort to construct districts as nearly of equal population as is practicable.

2. Congressional districts

Under the apportionment clause of Article I, Section 2 of the U.S. Constitution, any population deviation among congressional districts, no matter how small, must be justified through a showing that the specific deviation is required by legitimate redistricting policies such as making districts compact, respecting political subdivision boundaries, preserving the cores of prior districts, and avoiding contests between incumbent representatives. So that the State may avoid assuming this additional burden under federal law, a congressional redistricting plan should not have population deviations greater than one (1) person.

B. Voting rights. A redistricting plan for the General Assembly or Congress must not have either the purpose or the effect of diluting minority voting strength and must otherwise comply with Section 2 of the Voting Rights Act, as expressed through *Thornburg v. Gingles* and its progeny, and the Fourteenth and Fifteenth Amendments to the U.S. Constitution.

C. Avoidance of racial gerrymandering. All plans must comply with the Fourteenth Amendment to the U.S. Constitution, as interpreted by the United States Supreme Court in *Shaw v. Reno* and subsequent cases. Under those cases, while consideration of race is permissible, race must not be the predominant factor in that race-neutral considerations are subordinated to racial considerations, unless that subordination is narrowly tailored to serve a compelling state interest.

II. CONTIGUITY. All legislative and congressional districts should be composed of contiguous geography. Contiguity by water is acceptable to link territory within a district provided that there is a

reasonable opportunity to access all parts of the district and the linkage is designed to meet the other criteria stated herein. Point-to-point contiguity is acceptable so long as adjacent districts do not use the same vertex as points of transversal.

III. ADDITIONAL CONSIDERATIONS. Other criteria that should be given consideration, where practical and appropriate, in no particular order of preference, are:

A. Communities of Interest. Communities of interest should be considered. Areas defined by geographic, demographic, historic or other characteristics that cause people to identify with one another, including economic, social, cultural, language, political, and recreational activity interests common to the area's population may constitute communities of interest. Communities of interest may be overlapping and may consist of one or more formally, or informally, defined geographic areas with unifying common interests.

B. Constituent Consistency. Preserving the cores of existing districts, keeping incumbents' residences in districts with their core constituents, and avoiding contests between incumbent legislators should be considered.

C. Minimizing Divisions of County Boundaries.

D. Minimizing Divisions of Cities and Towns.

E. Minimizing Divisions of Voting Precinct Boundaries. Voting precinct boundaries are represented by the Census Bureau's Voting Tabulation District (VTD) lines. Both existing lines and pending precinct boundary realignments should be considered. If precincts are split, every effort should be made to divide precincts along recognizable and demonstrable boundaries.

F. District Compactness. In determining the relative compactness of a district, consideration should be given to geography, demography, communities of interest, and the extent to which parts of the district are joined by roads, media outlets, or other means for constituents to communicate effectively with each other and with their representative.

IV. DATA. The total state population and the population of the defined subunits thereof, as reported by the 2020 Federal Decennial Census, shall be the exclusive permissible population database used for the development, evaluation, and analysis of proposed redistricting plans. Other succinct and importable sources of demographic and political information may be considered in drafting and analyzing proposed redistricting plans.

Exhibit 4

South Carolina House of Representatives
Judiciary Committee
Redistricting Ad Hoc Committee

2021 Guidelines and Criteria for Congressional and Legislative Redistricting

The South Carolina House of Representatives, the House Judiciary Committee, and the Redistricting Ad Hoc Committee have the authority to determine the criteria that the South Carolina House of Representatives will use to create Congressional and legislative districts. Therefore, the Redistricting Ad Hoc Committee of the South Carolina House of Representatives adopts these guidelines and criteria.

I. Constitution of the United States

Redistricting plans shall comply with the United States Constitution, the Equal Protection Clause of the Fourteenth Amendment, and the opinions of the Supreme Court of the United States.

II. Federal Law

Redistricting plans shall also comply with federal law and the Voting Rights Act of 1965, as amended. Pursuant to the Voting Rights Act of 1965, and in accordance with the opinions of the Supreme Court of the United States, race may be a factor considered in the creation of redistricting plans, but it shall not be the predominant factor motivating the legislature's decisions concerning the redistricting plan and shall not unconstitutionally predominate over other criteria set forth in these guidelines. The dilution of racial or ethnic minority voting strength is contrary to the laws of the United States and of the State of South Carolina, and also is against the public policy of this state. Any proposed redistricting plan that is demonstrated to have the intent or effect of dispersing or concentrating minority population in a manner that prevents minorities from electing their candidates of choice will neither be accepted nor approved.

III. State Law

Except as otherwise required by federal law, redistricting plans also shall comply with the South Carolina Constitution, the laws of this state, and the opinions of the South Carolina Supreme Court.

IV. Equal Population/Deviation

a. The population of the Congressional and legislative districts will be determined based solely on the enumeration of the 2020 federal decennial census pursuant to the United States Constitution, Article I, Section 2, and the Equal Protection Clause of the Fourteenth Amendment.

b. The number of persons in Congressional districts shall be as nearly equal in population as is practicable. The ideal population for Congressional districts shall

be 731,204. In every case, efforts shall be made to achieve strict equality or produce the lowest overall range of deviation possible when taking into consideration geographic limitations.

c. Legislative districts shall have substantial equality of population. The ideal population for a South Carolina House of Representatives district shall be 41,278. In every case, efforts should be made to limit the overall range of deviation from the ideal population to less than five percent, or a relative deviation in excess of plus or minus two and one-half percent for each South Carolina House district. Nevertheless, any overall deviation greater than five percent from equality of population among South Carolina House districts shall be justified when it is the result of geographic limitations, the promotion of a constitutionally permissible state policy, or to otherwise comply with the criteria identified in these guidelines.

V. Contiguity

Congressional and legislative districts shall be comprised of contiguous territory. Contiguity by water is sufficient. Areas which meet only at the points of adjoining corners are not considered contiguous.

VI. Compactness

Congressional and legislative districts should be reasonably compact in form and should follow census geography. Bizarrely-shaped districts are to be avoided except when required by one or more of the following factors: (a) census block geography; (b) natural geography including water; (c) efforts to comply with the standards delineated above in Section IV; and/or (d) efforts to comply with the Voting Rights Act of 1965, as amended. Compactness may require the division of population concentrations when to do otherwise would mean dramatically altering the character of a district or would require tortuous configuration of an adjoining district.

Compactness should be judged in part by the configuration of prior plans. Compactness should not be judged based upon any mathematical, statistical, or formula-based calculation or determination.

VII. Communities of Interest

Communities of interest should be considered in the redistricting process. A variety of factors may contribute to a community of interest including, but not limited to the following: (a) economic; (b) social and cultural; (c) historic influences; (d) political beliefs; (e) voting behavior; (f) governmental services; (g) commonality of communications; and (h) geographic location and features. Communities of interest should be considered and balanced by the Redistricting Ad Hoc Committee, the House Judiciary Committee, and the South Carolina House of Representatives. County boundaries, municipality boundaries, and precinct lines (as represented by the Census Bureau's Voting Tabulation District lines) may be considered as evidence of communities of interest to be balanced, but will be given no greater weight, as a matter of state policy, than other identifiable communities of interest.

It is possible that competing communities of interest will be identified during the redistricting process. Although it may not be possible to accommodate all communities of interests, the Redistricting Ad Hoc Committee, the House Judiciary Committee, and the South Carolina House of Representatives will attempt to accommodate diverse communities of interest to the extent possible.

VIII. Incumbency Consideration

Incumbency may be considered in the reapportionment process. Reasonable efforts may be made to ensure that incumbent legislators remain in their current districts. Reasonable efforts may be made to ensure that incumbent legislators are not placed into districts where they will be compelled to run against other incumbent members of the South Carolina House of Representatives. However, incumbency considerations shall not influence the redistricting plan to such an extent as to overtake other redistricting principles.

IX. Priority of Criteria

a. In establishing Congressional and legislative districts, all criteria identified in these guidelines should be considered. However, if there is a conflict among the requirements of these guidelines, the requirements addressed in Sections I, II, III, and IV herein should be given priority.

b. If application of the criteria set forth in these guidelines will cause a violation of applicable constitutional, federal, or state law, and there is no other way to conform to the criteria without such violation of law, deviations from the criteria are permitted. However, any deviation from the criteria shall not be any more than necessary to avoid the violation of law, and the remainder of the redistricting plan shall remain faithful to the criteria.

X. Public Input

The Redistricting Ad Hoc Committee should make reasonable efforts to be transparent and allow public input into the redistricting process.

Exhibit 5

arXiv:2008.06131v4 [stat.AP] 12 Jun 2022

Sequential Monte Carlo for Sampling Balanced and Compact Redistricting Plans*

Cory McCartan† Kosuke Imai‡

First Draft: July 6, 2020
This Draft: June 14, 2022

Abstract

Random sampling of graph partitions under constraints has become a popular tool for evaluating legislative redistricting plans. Analysts detect partisan gerrymandering by comparing a proposed redistricting plan with an ensemble of sampled alternative plans. For successful application, sampling methods must scale to large maps with many districts, incorporate realistic legal constraints, and accurately and efficiently sample from a selected target distribution. Unfortunately, most existing methods struggle in at least one of these areas. We present a new Sequential Monte Carlo (SMC) algorithm that generates a sample of redistricting plans converging to a realistic target distribution. Because it draws many plans in parallel, the SMC algorithm can efficiently explore the relevant space of redistricting plans better than the existing Markov chain Monte Carlo (MCMC) algorithms that generate plans sequentially. Our algorithm can simultaneously incorporate several constraints commonly imposed in real-world redistricting problems, including equal population, compactness, and preservation of administrative boundaries. We validate the accuracy of the proposed algorithm by using a small map where all redistricting plans can be enumerated. We then apply the SMC algorithm to evaluate the partisan implications of several maps submitted by relevant parties in a recent high-profile redistricting case in the state of Pennsylvania. We find that the proposed algorithm converges to the target distribution faster and with fewer samples than a state-of-the-art MCMC algorithm. Open-source software is available for implementing the proposed methodology.

Key Words: gerrymandering, graph partition, importance sampling, spanning trees

*We thank Ben Fifield, Greg Herschlag, Mike Higgins, Chris Kenny, Jonathan Mattingly, Justin Solomon, and Alex Tarr for helpful comments and conversations. Imai thanks Yunkyu Sohn for his contributions at an initial phase of this project. Open-source software is available for implementing the proposed methodology (Kenny et al., 2020).

†Ph.D. candidate, Department of Statistics, Harvard University. 1 Oxford Street, Cambridge 02138. Email: cmccartan@g.harvard.edu

‡Professor, Department of Government and Department of Statistics, Harvard University. 1737 Cambridge Street, Institute for Quantitative Social Science, Cambridge 02138. Email: imai@harvard.edu, URL: https://imai.fas.harvard.edu/

1 Introduction

In first-past-the-post electoral systems, legislative districts serve as the fundamental building block of democratic representation. In the United States, congressional redistricting, which redraws district boundaries in each state following the decennial Census, plays a central role in influencing who is elected and hence what policies are eventually enacted. Because the stakes are so high, redistricting has been subject to intense political battles. Parties often engage in *gerrymandering* by manipulating district boundaries in order to amplify the voting power of some groups while diluting that of others.

In recent years, the availability of granular data about individual voters has led to sophisticated partisan gerrymandering attempts that cannot be easily detected. At the same time, many scholars have focused their efforts on developing methods to uncover gerrymandering by comparing a proposed redistricting plan with a large collection of alternative plans that satisfy the relevant legal requirements. A primary advantage of such an approach over the use of simple summary statistics is its ability to account for the characteristics of each state's physical and political geography and state-specific redistricting rules.

For its successful application, a sampling algorithm for drawing alternative plans must (1) be efficient enough to scale to large maps with thousands of geographic units and many districts, (2) simultaneously incorporate a variety of real-world legal constraints such as population balance (Section 3.1), geographical compactness (Section 3.3), and the preservation of administrative boundaries (Section 4.5), and (3) ensure these samples are representative of a specific target population, against which a redistricting plan of interest can be evaluated. Although some have been used in several recent court challenges to existing redistricting plans, all existing algorithms run into limitations of varying severity with regards to at least one of these three key requirements.

Optimization-based (e.g., Mehrotra et al., 1998; Macmillan, 2001; Bozkaya et al., 2003; Liu et al., 2016) and constructive Monte Carlo (e.g., Cirincione et al., 2000; Chen and Rodden, 2013; Magleby and Mosesson, 2018) methods can be made scalable and incorporate many constraints. But they are not designed to sample from any specific target distribution. As a consequence, the resulting plans tend to differ systematically, for example, from a uniform distribution under certain constraints (Cho and Liu, 2018; Fifield et al., 2020a,b). The absence of an explicit target distribution makes it difficult to interpret the ensembles generated by these methods and use them for statistical outlier analysis to detect gerrymandering.

MCMC algorithms (e.g., Mattingly and Vaughn, 2014; Wu et al., 2015; Chikina et al., 2017; DeFord et al., 2021; Carter et al., 2019; Fifield et al., 2020a) can in theory sample from a specific target distribution, and incorporate constraints through the use of an energy function. In practice, however, existing algorithms struggle to mix and traverse through a highly complex sampling space, making scalability difficult and accuracy hard to prove. Some of these algorithms make proposals by flipping precincts at the boundary of existing districts (e.g., Mattingly and Vaughn, 2014; Fifield et al., 2020a), rendering it difficult or even impossible to transition between points in the state space, especially as more constraints are imposed. More recent algorithms by DeFord et al. (2021) and Carter et al. (2019) use spanning trees to make their proposals, and this has allowed these algorithms to yield more global moves and greatly improve mixing. Yet recent theoretical results suggest that even these larger moves may not be enough to traverse the entire state space, and therefore converge to the correct distribution, if a realistic population balance constraint is imposed (Akitaya et al., 2022).

In Sections 3 and 4, we present a new Sequential Monte Carlo (SMC) algorithm, based on a similar but not identical spanning tree construction to DeFord et al. (2021) and Carter et al. (2019), that addresses the above three key challenges. Like MCMC algorithms, the SMC algorithm generates samples which approximate the target distribution arbitrarily well as the sample size increases. But like constructive Monte Carlo methods, the SMC algorithm draws many separate plans from scratch, rather than tweaking a single plan sequentially. This approach is better suited to the large discrete state space with a multimodal target distribution that characterizes redistricting problems. In cases where existing MCMC proposals render the state space disconnected, the SMC algorithm can still converge to the target distribution. As we demonstrate in Sections 5 and 6, this sampling approach translates to faster convergence and smaller standard errors for a given computational budget. For large and complex redistricting sampling problems, the SMC algorithm can be easily parallelized.

The proposed algorithm proceeds by splitting off one district at a time, building up the redistricting plan piece by piece (see Figure 2 for an illustration). Each split is accomplished by drawing a spanning tree and removing

one edge, which splits the spanning tree in two. We also extend the SMC algorithm so that it preserves administrative boundaries and certain geographical areas as much as possible, which is another common constraint considered in many real-world redistricting cases. An open-source software package, `redist`, is available for implementing the proposed algorithm (Kenny et al., 2020).

Like existing MCMC approaches, the SMC algorithm only guarantees convergence to the target distribution as the sample size approaches infinity. Thus it is important to understand the limitations of the proposed algorithm in finite samples. In Section 4.4.4, we provide a set of diagnostics which can be used in practice to help identify if more samples are needed to reach convergence or if the constraints imposed by an analyst are too strong. In Section 5, we validate the SMC algorithm using a small map for which all potential redistricting plans can be enumerated (Fifield et al., 2020b). We demonstrate that the proposed algorithm samples accurately from a realistic target distribution on these plans and does so more efficiently than a state-of-the-art MCMC algorithm. Section 6 applies the SMC algorithm to the 2011 Pennsylvania congressional redistricting, and compares its performance on this problem with the existing approaches. We find that the proposed SMC algorithm is again more efficient in sampling than the existing MCMC algorithm. Section 7 concludes and discusses directions for future work.

2 The 2011 Pennsylvania Congressional Redistricting

We study the 2011 Pennsylvania congressional redistricting because it illustrates the salient features of the redistricting problem. We begin by briefly summarizing the background of this case and then explain the role of sampling algorithms used in the expert witness reports.

2.1 Background

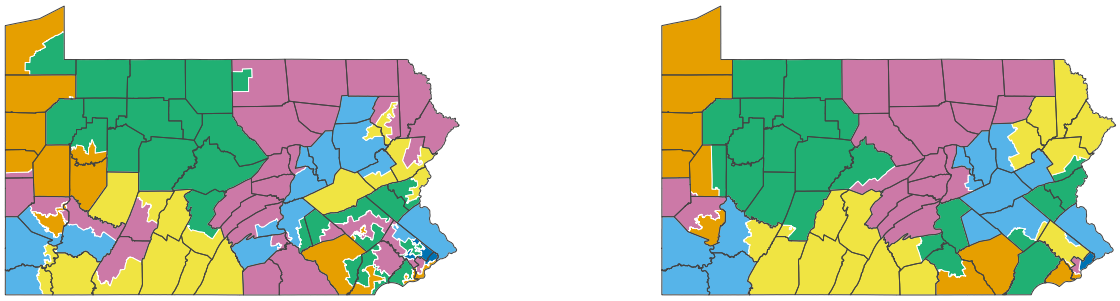
Pennsylvania lost a seat in Congress during the reapportionment of the 435 U.S. House seats following the 2010 Census. In Pennsylvania, the General Assembly, which is the state's legislative body, draws new congressional districts, subject to gubernatorial veto. At the time, the General Assembly was controlled by Republicans, and Tom Corbett, also a Republican, served as governor. In the 2012 election, which took place under the newly adopted 2011 districting map, Democrats won 5 seats while Republicans took the remaining 13. Under the previous plan, the split was 7–12.

In June 2017, the League of Women Voters of Pennsylvania filed a lawsuit alleging that the 2011 plan adopted by the Republican legislature violated the state constitution by diluting the political power of Democratic voters. The case worked its way through the state court system, and on January 22, 2018, the Pennsylvania Supreme Court issued its ruling, writing that the 2011 plan "clearly, plainly and palpably violates the Constitution of the Commonwealth of Pennsylvania, and, on that sole basis, we hereby strike it as unconstitutional." (League of Women Voters v. Commonwealth, 2018).

The court ordered that the General Assembly adopt a remedial plan and submit it to the governor, who would in turn submit it to the court, by February 15, 2018. In its ruling, the court laid out specific requirements that had to be satisfied by all proposed plans:

> composed of compact and contiguous territory; as nearly equal in population as practicable; and which do not divide any county, city, incorporated town, borough, township, or ward, except where necessary to ensure equality of population.

The leaders of the Republican Party in the General Assembly drew a new map, but the Democratic governor, Tom Wolf, refused to submit it to the court, claiming that it, too, was an unconstitutional gerrymander. Instead, the court received remedial plans from seven parties: the petitioners, the League of Women Voters; the respondents, the Republican leaders of the General Assembly; the governor, a Democrat; the lieutenant governor, also a Democrat; the Democratic Pennsylvania House minority leadership; the Democratic Pennsylvania Senate minority leadership; and the intervenors, which included Republican party candidates and officials. Ultimately, the Supreme Court drew its own plan and adopted it on February 19, 2018, arguing that it was "superior or comparable to all plans submitted by the parties." Figure 1 shows the remedial plan created by the Supreme Court as well as the 2011 map adopted by the General Assembly, which were found on the court's case page.



(a) 2011 General Assembly map

(b) 2018 Pennsylvania Supreme Court map

Figure 1: Comparison of the 2011 map drawn by the General Assembly and the final map imposed by the Supreme court in 2018. County lines are shown in dark gray, and district boundaries that do not coincide with county boundaries are in white.

The constraints explicitly laid out by the court, as well as the numerous remedial plans submitted by the parties, make the 2011 Pennsylvania redistricting a useful case study that evaluates redistricting plans.

2.2 The Role of Sampling Algorithms

The original finding that the 2011 General Assembly plan was a partisan gerrymander was in part based on different outlier analyses performed by two academic researchers, Jowei Chen and Wesley Pegden, who served as the petitioner's expert witnesses. Chen randomly generated two sets of 500 redistricting plans according to a constructive Monte Carlo algorithm based on Chen and Rodden (2013). He considered population balance, contiguity, compactness, avoiding county and municipal splits, and, in the second set of 500, avoiding plans that placed more than two incumbents in the same district (at least one pair of incumbents in the same district was necessary, given that Pennsylvania lost a seat from 2000 to 2010). Pegden ran a reversible Markov chain similar to that used in the MCMC algorithm of Mattingly and Vaughn (2014) for one trillion steps, and computed upper bounds of p-values using the method of Chikina et al. (2017). This method was also used in a follow-up analysis by Moon Duchin, who served as an expert for Governor Wolf (Duchin, 2018). Both petitioner experts concluded that the 2011 plan was an extreme outlier according to compactness, county and municipal splits, and the number of Republican and Democratic seats implied by statewide election results.

The respondents also retained an expert academic witness, Wendy Tam Cho, who directly addressed the sampling-based analyses of Chen and Pegden. Cho criticized Chen's analysis for not sampling from a specified target distribution. She also criticized Pedgen's analysis by arguing that his Markov chain only made local explorations of the space of redistricting plans, and could not therefore have generated a representative sample of all valid plans, though the p-values computed using the Chikina et al. (2017) method explicitly do not require mixing of the Markov chain (see also Cho and Rubinstein-Salzedo, 2019, and Chikina et al. (2019)). We do not directly examine the intellectual merits of the specific arguments put forth by the expert witnesses. However, these methodological debates are also relevant for other cases where simulation algorithms have been extensively used by expert witnesses (e.g., Rucho v. Common Cause (2019); Common Cause v. Lewis (2019); Covington v. North Carolina (2017); Harper v. Lewis (2020)), and highlight the difficulties in practically applying existing sampling algorithms to actual redistricting problems.

First, the distributions that some of these algorithms sample from are not made explicit, leaving open the possibility that the generated ensemble is systematically different from the true set of all valid plans. Second, even when the distribution is known, MCMC algorithms used to sample from it may be prohibitively slow to mix and cannot yield a representative sample. These challenges motivate us to design an algorithm that accurately samples from a specific target distribution and incorporates most common redistricting constraints, while being efficient and scalable.

3 Sampling Balanced and Compact Districts

3.1 The Setup

Redistricting plans are ultimately aggregations of geographic units such as counties, voting precincts, or Census blocks. The usual requirement that the districts in a plan be contiguous necessitates consideration of the spatial relationship between these units. The natural mathematical structure for this consideration is a graph $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_m\}$ consists of m nodes representing the geographic units of redistricting and E contains edges connecting units which are legally adjacent.

A labeled redistricting plan on G consists of n districts, where each district is a collection of nodes. A labeled plan is described by a function $\xi : V \to \{1, 2, \ldots, n\}$, where $\xi(v) = i$ implies that node v is in district i. In practice, we will be interested in *unlabeled* plans, since the assignment of sets of precincts to labels is arbitrary and has no impact on the real-world aspects of the district such as its population, demographic composition, or partisan lean. Define an equivalence relation $\cong$ by $\xi_1 \cong \xi_2$ if there exists a permutation σ such that $\xi_1(v) = \sigma(\xi_2(v))$ for all v. Then an unlabeled redistricting plan can be viewed as an equivalence class under $\cong$, denoted $[\xi]$; nothing in what follows will depend on the particular labeled plan representative ξ.

We let $V_i(\xi)$ and $E_i(\xi)$ denote the nodes and edges contained in district i under a given redistricting plan ξ, so $G_i(\xi) = (V_i(\xi), E_i(\xi))$ represents the induced subgraph that corresponds to district i under the plan. We suppress the dependence on ξ when it is clear from context, writing $G_i = (V_i, E_i)$. Since each node belongs to only one district, we have $V = \bigcup_{i=1}^{n} V_i(\xi)$ and $V_i(\xi) \bigcap V_{i'}(\xi) = \varnothing$ for any redistricting plan ξ. In addition, we require that each district be contiguous, i.e., that G_i is a connected graph, for all i.

Beyond connectedness, redistricting plans are always required to have roughly equal population in every district. To formalize this requirement, let $\mathrm{pop}(v)$ denote the population of node v. Then the population of a district i may be written as $\mathrm{pop}(V_i(\xi)) := \sum_{v \in V_i(\xi)} \mathrm{pop}(v)$. We quantify the discrepancy between a given plan and the ideal of equal population in every district by the *maximum population deviation*,

$$\mathrm{dev}(\xi) := \max_{1 \le i \le n} \left| \frac{\mathrm{pop}(V_i)}{\mathrm{pop}(V)/n} - 1 \right|,$$

where $\mathrm{pop}(V)$ is the total population. Some courts and states have imposed hard maximums on this quantity, e.g., $\mathrm{dev}(\xi) \le D = 0.05$ for state legislative redistricting (National Conference of State Legislatures, 2021).

The proposed algorithm samples plans by way of spanning trees on each district, i.e., subgraphs of $G_i(\xi)$ which contain all vertices, no cycles, and are connected. Let T_i represent a spanning tree for district i whose vertices and edges are given by $V_i(\xi)$ and a subset of $E_i(\xi)$, respectively. The collection of spanning trees from all districts together form a spanning forest. Each node belongs to one spanning tree in the forest, and this assignment corresponds to a redistricting plan. However, a single redistricting plan may correspond to multiple spanning forests because each district may admit more than one spanning tree.

For a given redistricting plan, we can compute the exact number of spanning forests in polynomial time using the determinant of a submatrix of the graph Laplacian, according to the Matrix Tree Theorem of Kirchhoff (see Tutte (1984)). Thus, for a graph H, if we let $\tau(H)$ denote the number of spanning trees on the graph, we can represent the number of spanning forests that correspond to a redistricting plan ξ as $\tau(\xi) := \prod_{i=1}^{n} \tau(G_i(\xi))$. This fact will play an important role in the definition of our sampling algorithm and its target distribution.

3.2 The Target Distribution

The algorithm is designed to sample an unlabeled plan $[\xi]$ with probability

$$\pi([\xi]) \propto \exp\{-J(\xi)\}\tau(\xi)^{\rho} \mathbf{1}_{\{\xi \text{ connected}\}} \mathbf{1}_{\{\mathrm{dev}(\xi) \le D\}}, \tag{1}$$

where the indicator functions ensure that the plans meet population balance and connectedness criteria, $\tau(\xi)$ measures the compactness of the districts in ξ (see Section 3.3), and J encodes additional constraints on the types of plans preferred. As done in Section 6, we often use a reasonably strict population constraint such as $D = 0.001$. The parameter $\rho \in \mathbb{R}_0^+$ is chosen to control the compactness of the generated plans.

This target distribution π has both substantive and theoretical justifications. First, it incorporates two constraints which are almost always present in real-world redistricting: the support of π is restricted to contiguous plans and those that meet a population deviation threshold. Second, it represents the unique maximum entropy distribution on the set of redistricting plans satisfying these two universal constraints and the moment conditions implied by the other constraints, i.e., $\mathbb{E}_\pi[\log \tau(\xi)] = \mu_\tau$ and $\mathbb{E}_\pi[J(\xi)] = \mu_J$ for some constants μ_τ and μ_J (see Cover and Thomas, 2006, Theorem. 12.1.1, originally of Boltzmann).

Thus, our target distribution ensures that all plans meet contiguity and population requirements, and *on average* satisfy a compactness standard as well as any other additional constraints (through the function J). It is no surprise, therefore, that this class of target distributions has been used by other work developing redistricting sampling algorithms (Herschlag et al., 2017; Fifield et al., 2020a).

The generality of the additional constraint function J is intentional, as its exact form and number imposed on the redistricting process varies by state and by the type of districts being drawn; any type of constraint may be incorporated by choosing a J which is small for preferred plans and large otherwise. For example, a preference for plans close to an existing plan ξ_{sq} may be encoded as

$$J_{\text{sq}}(\xi) = -\frac{\beta}{\log n}\text{VI}(\xi, \xi_{\text{sq}}) \coloneqq -\frac{\beta}{2\log n}\sum_{i,j=1}^{n}\frac{P_{ij}}{\text{pop}(V)}\left(\log\left(\frac{P_{ij}}{\text{pop}(V_j(\xi_{\text{sq}}))}\right) + \log\left(\frac{P_{ij}}{\text{pop}(V_j(\xi))}\right)\right), \tag{2}$$

where $\beta \in \mathbb{R}^+$ controls the strength of the constraint and $P_{ij} = \text{pop}(V_i(\xi) \cap V_j(\xi_{\text{sq}}))$ is the population shared between district i of ξ and j of ξ_{sq}. The function $\text{VI}(\cdot, \cdot)$ represents the variation of information (also known as the shared information distance), which is the difference between the joint entropy and the mutual information of the distribution of population over the new districts ξ relative to the existing districts ξ_{sq} (Cover and Thomas, 2006). When ξ is any relabeling of ξ_{sq}, then $J_{\text{sq}}(\xi) = 0$. In contrast, when ξ evenly splits the nodes of each district of ξ_{sq} among the districts of ξ, then $J_{\text{sq}}(\xi) = \beta$. This distance measure will prove useful later in quantifying the diversity of a sample of redistricting plans (see also Guth et al., 2020).

There exist other formulations of constraints, and considerations in choosing a set of weights that balance constraints against each other (see e.g., Bangia et al., 2017; Herschlag et al., 2017; Fifield et al., 2020a). Here, we focus on sampling from the broad class of distributions characterized by Equation (1), which have been used in other work; we do not address the important but separate problem of picking a specific instance of this class for a given redistricting problem.

The flexibility of J can be deceptive, however. The algorithm operates efficiently only when the additional constraints imposed by J are not too severe. Even a small number of strong constraints incorporated into J can dramatically limit the number of valid plans and considerably complicate the process of sampling. The Markov chain algorithms developed to date partially avoid this problem by moving toward maps with lower J over a number of steps, but in general including more constraints makes it even more difficult to transition between valid redistricting plans. Approaches such as simulated annealing (Bangia et al., 2017; Herschlag et al., 2017) and parallel tempering (Fifield et al., 2020a) have been proposed to handle multiple constraints, but these can be difficult to calibrate in practice and provide few, if any, theoretical guarantees.

In practice, we usually find that the most stringent constraints are those involving population deviation, compactness, and administrative boundary splits. As shown later, we address this issue by designing our algorithm to directly satisfy these constraints. Weak additional constraints do not generally have a substantial effect on the sampling efficiency, though there are exceptions. Monitoring the distribution of the weights and the overall sampling efficiency is crucial to obtaining a good sample, as we discuss later.

3.3 Spanning Forests and Compactness

One common redistricting requirement is that districts be geographically compact, though nearly every state leaves this term undefined. Dozens of numerical compactness measures have been proposed, with the Polsby–Popper score (Polsby and Popper, 1991) perhaps the most popular. Defined as the ratio of a district's area to that of a circle with the same perimeter as the district, the Polsby–Popper score is constrained to $[0, 1]$, with higher scores indicating more compactness.

Other scholars have proposed a graph-theoretic measure known as *edge-cut compactness* (Dube and Clark, 2016; DeFord et al., 2021). This measure counts the number of edges that must be removed from the original graph to partition it according to a given plan. Formally, it is defined as

$$\mathrm{rem}(\xi) \;:=\; 1 - \frac{\sum_{i=1}^{n} |E_i(\xi)|}{|E(G)|},$$

where we have normalized to the total number of edges.

Plans that involve cutting many edges will necessarily have long internal boundaries, driving up their average district perimeter (and driving down their Polsby–Popper scores), while plans that cut as few edges as possible will have relatively short internal boundaries and much more compact districts. Additionally, given the high density of voting units in urban areas, plans which cut fewer edges will tend to avoid drawing district lines through the heart of these urban areas. This has the welcome side effect of avoiding splitting cities and towns, and in doing so helping to preserve "communities of interest," another common redistricting consideration.

Empirically, this graph-based compactness measure is highly correlated with $\log \tau(G) - \log \tau(\xi)$ (we often observe a correlation in excess of 0.99). It is difficult to precisely characterize this relationship except in special cases because $\tau(\xi)$ is calculated as a matrix determinant (McKay, 1981). However, this quantity is strongly controlled by the product of the degrees of each node in the graph, $\prod_{i=1}^{m} \deg(v_i)$ (Kostochka, 1995). Removing an edge from a graph decreases the degree of the vertices at either end by one, so we would expect $\log \tau(G)$ to change by approximately $2\{\log \bar{d} - \log(\bar{d} - 1)\}$ with this edge removal, where $\bar{d}$ is the average degree of the graph. This implies a linear relation $\log \tau(G) - \log \tau(\xi) \;\approx\; \mathrm{rem}(\xi) \cdot 2\{\log \bar{d} - \log(\bar{d} - 1)\}$, and hence $\tau(\xi)^{\rho} \approx C_1 \exp(-C_2\, \rho\, \mathrm{rem}(\xi))$, where C_1 and C_2 are some constants depending on the details of the map.

As a result, a greater value of ρ in the target distribution corresponds to a preference for fewer edge cuts and therefore a redistricting plan with more compact districts. This and the considerations given in the literature (Dube and Clark, 2016; DeFord et al., 2021) suggest that the target distribution in Equation (1) with $\rho = 1$ (or another positive value) is a good choice for sampling compact districts. The choice of $\rho = 1$ is computationally convenient, as it allows us to avoid calculating $\tau(\xi)$ as part of sampling (an asymptotic bottleneck), and yet usually produces satisfactorily compact districts. Of course, if another compactness metric is desired, one can simply set $\rho = 0$ and incorporate the alternative metric into J. This will preserve the algorithm's efficiency to the extent that the alternative metric correlates with the edge-removal measure of compactness.

4 The Proposed Algorithm

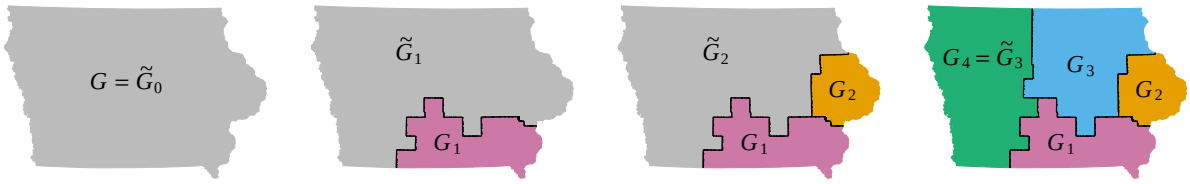
The proposed algorithm samples redistricting plans by sequentially drawing districts over $n - 1$ iterations of a splitting procedure. This is fundamentally different from existing MCMC approaches, which change an existing plan according to some transition kernel. The iterations of the proposed algorithm are from district to district within a single plan whereas the iterations in an MCMC algorithm are from plan to plan.

Our algorithm begins by partitioning the original graph $G = (V, E) = (\widetilde{V}_0, \widetilde{E}_0) = \widetilde{G}_0$ into two induced subgraphs: $G_1 = (V_1, E_1)$, which will constitute a district in the final map, and the remainder of the graph $\widetilde{G}_1 = (\widetilde{V}_1, \widetilde{E}_1)$, where $\widetilde{V}_1 = V \setminus V_1$ and $\widetilde{E}_1$ consists of all the edges between vertices in $\widetilde{V}_1$. Next, the algorithm takes $\widetilde{G}_1$ as an input graph and partitions it into two induced subgraphs, one which will become a district G_2 and the remaining graph $\widetilde{G}_2$. The algorithm repeats the same splitting procedure until the final $(n-1)$-th iteration whose two resulting partitions, G_{n-1} and $\widetilde{G}_{n-1} = G_n$, become the final two districts of the redistricting plan.

Figure 2 provides an illustration of this sequential procedure. To sample a large number of redistricting plans from the target distribution given in Equation (1), at each iteration, the algorithm samples many candidate partitions, discards those which fail to meet the population constraint, and then resamples a certain number of the remainder according to importance weights, using the resampled partitions at the next iteration. The rest of this section explains the details of the proposed algorithm.

4.1 The Splitting Procedure

We first describe the splitting procedure, which is similar to the merge-split Markov chain proposals of DeFord et al. (2021) and Carter et al. (2019). It proceeds by drawing a random spanning tree T, identifying the k_i most



(a) Initial map (b) Iteration 1 (c) Iteration 2 (d) Iteration 3

Figure 2: The sequential splitting procedure applied to the state of Iowa, where four congressional districts are created at the county level.

promising edges to cut within the tree, and selecting one such edge at random to create two induced subgraphs. Spanning trees are an attractive way to split districts, as the removal of a single edge induces a partition with two connected components, and spanning trees can be sampled uniformly (Wilson, 1996).

Algorithm 1 Splitting procedure to generate one district

Input: initial graph $\widetilde{G}_{i-1}$ and a parameter $k_i \in \mathbb{Z}^+$.

(a) Draw a single spanning tree T on $\widetilde{G}_{i-1}$ uniformly from the set of all such trees using Wilson's algorithm.

(b) Each edge $e \in E(T)$ divides T into two components, $T_e^{(1)}$ and $T_e^{(2)}$. For each edge, compute the following population deviation for the two districts that would be induced by cutting T at e,

$$d_e^{(1)} = \left| \frac{\sum_{v \in T_e^{(1)}} \text{pop}(v)}{\text{pop}(V)/n} - 1 \right| \quad \text{and} \quad d_e^{(2)} = \left| \frac{\sum_{v \in T_e^{(2)}} \text{pop}(v)}{\text{pop}(V)/n} - 1 \right|.$$

Let $d_e = \min\{d_e^{(1)}, d_e^{(2)}\}$, and index the edges in ascending order by this quantity, so that we have $d_{e_1} \leq d_{e_2} \leq \cdots \leq d_{e_{m_i-1}}$, where $m_i = |V_i|$.

(c) Select one edge e^* uniformly from the top k_i edges, $\{e_1, e_2, \ldots, e_{k_i}\}$, and remove it from T, creating a spanning forest $(T_{e^*}^{(1)}, T_{e^*}^{(2)})$ which induces a partition $(G_i^{(1)}, G_i^{(2)})$.

(d) If $d_{e^*}^{(1)} \leq d_{e^*}^{(2)}$, i.e., if $T_{e^*}^{(1)}$ induces a district that is closer to the optimal population than $T_{e^*}^{(2)}$ does, set $G_i = G_i^{(1)}$ and $\widetilde{G}_i = G_i^{(2)}$; otherwise, set $G_i = G_i^{(2)}$ and $\widetilde{G}_i = G_i^{(1)}$.

As part of the full sampling procedure (Algorithm 2), after splitting, we check that the population of the new district G_i falls within the bounds

$$P_i^- = \max\left\{ \frac{\text{pop}(V)}{n}(1-D),\ \text{pop}(\widetilde{V}_{i-1}) - \frac{n-i}{n}\text{pop}(V)(1+D) \right\} \quad \text{and}$$
$$P_i^+ = \min\left\{ \frac{\text{pop}(V)}{n}(1+D),\ \text{pop}(\widetilde{V}_{i-1}) - \frac{n-i}{n}\text{pop}(V)(1-D) \right\}.$$

These bounds also ensure that it will be possible for future iterations to generate valid districts out of $\widetilde{G}_i$. If $\text{pop}(V_i) \notin [P_i^-, P_i^+]$, then the entire redistricting plan is rejected and the sampling process begins again. While the rate of rejection varies by map and by iteration, we generally encounter acceptance rates at each iteration between 5% and 30%, which are not so low as to make sampling from large maps intractable. Algorithm 1 details the steps of the splitting procedure, where at the first iteration we take $\widetilde{G}_0 = G$.

4.2 The Sampling Probability

The above sequential splitting procedure does not generate plans from the target distribution π. We denote the sampling measure by q, and write the sampling probability for a given connected plan ξ at iteration i as

$q(G_i \mid \widetilde{G}_{i-1})$, since each new district G_i depends only on the leftover map area $\widetilde{G}_{i-1}$ from the previous iteration. This probability can be written as the probability that we cut an edge along the boundary of the new district, integrated over all spanning trees which could be cut to form the district, i.e.,

$$q(G_i \mid \widetilde{G}_{i-1}) = \sum_{T \in \mathcal{T}(\widetilde{G}_{i-1})} q(G_i \mid T)\, \tau(\widetilde{G}_{i-1})^{-1}, \tag{3}$$

where $\mathcal{T}(\cdot)$ represents the set of all spanning trees of a given graph, and we have relied on the fact that Wilson's algorithm draws spanning trees uniformly.

The key is that for certain choices of k_i (the number of edges considered to be cut at iteration i), the probability that an edge is cut is independent of the trees that are drawn. Let $ok(T)$ represent the number of edges on any spanning tree T that induce balanced partitions with population deviation below D, i.e.,

$$ok(T) \;\coloneqq\; |\{e \in E(T) : d_e \leq D\}|.$$

Then define $K_i \coloneqq \max_{T \in \mathcal{T}(\widetilde{G}_{i-1})} ok(T)$, the maximum number of such edges across all spanning trees. Furthermore, let $\mathcal{C}(G, H)$ represent the set of edges joining nodes in a graph G to nodes in a graph H. We have the following result for the splitting probability for new districts whose populations lie inside the bounds defined above (Appendix A for the proof).

Lemma 1. *The probability of splitting a* valid *new district G_i from an existing area $\widetilde{G}_{i-1}$ using Algorithm 1 with parameter $k_i \geq K_i$ is*

$$q(G_i \mid \widetilde{G}_{i-1}, \mathrm{pop}(V_i) \in [P_i^-, P_i^+]) \;=\; \frac{\tau(G_i)\tau(\widetilde{G}_i)}{\tau(\widetilde{G}_{i-1})k_i} |\mathcal{C}(G_i, \widetilde{G}_i)|. \tag{4}$$

4.3 Sequential Monte Carlo

We follow a sequential Monte Carlo approach (Doucet et al., 2001; Liu et al., 2001) to generate draws from the target distribution, rather than simply performing $n-1$ iterations of Algorithm 1 and resampling or reweighting at the final stage. A sequential approach is also useful in operationalizing the rejection procedure to enforce the population constraint.

The proposed procedure is presented as Algorithm 2. The algorithm is governed by a parameter $\alpha \in (0, 1]$, which has no effect on the target distribution nor the asymptotic accuracy of the algorithm. Rather, α may be adjusted to maximize the efficiency of sampling. To generate S redistricting plans, at each iteration of the splitting procedure $i \in \{1, 2, \ldots, n-1\}$, we resample and split the existing plans one at a time, rejecting those which do not meet the population constraints, until we obtain S new plans for the next iteration. This rejection process can be viewed as a form of partial rejection control (Liu et al., 1998, 2001), or a version of the AliveSMC algorithm (LeGland and Oudjane, 2005; Peters et al., 2012).

The weights at each stage serve three purposes. The first is to account for the extraneous $|\mathcal{C}(G_i, \widetilde{G}_i)|$ term which appears in the splitting probability $q(G_i \mid \widetilde{G}_{i-1}, \mathrm{pop}(V_i) \in [P_i^-, P_i^+])$ but not in the target distribution π. The second is to account for differences in the compactness parameter ρ; the splitting procedure generates plans according to $\rho = 1$, but this may be different from the target value. The third purpose is to adjust for the imbalances between the labeled plans generated by the splitting procedure and the unlabeled plans which are the target of sampling. While there are $n!$ labeled plans corresponding to every unlabeled plan, not every labeled plan may be sampled according to the sequential procedure. In fact, the number of labeled plans which are in the support of the SMC proposal distribution varies from one unlabeled plan to another. It is this imbalance which necessitates an additional correction.

For a labeled plan ξ, let G/ξ denote the district-level quotient graph corresponding to the plan; i.e., the nodes are districts $\{1, \ldots, n\}$ and edges connect adjacent districts. Notice that the splitting procedure ensures that the leftover area after each split, $\widetilde{G}_i$, is a connected graph. This means that for every $1 \leq i \leq n-1$, the subgraph of G/ξ induced by the vertices $\{i+1, i+2, \ldots, n\}$ (the districts drawn after split i) is connected. We call any

Algorithm 2 Sequential Monte Carlo (SMC) Algorithm

Input: graph G to be split into n districts, target distribution parameters $\rho \in \mathbb{R}_0^+$ and constraint function J, and sampling parameters $\alpha \in (0, 1]$ and $k_i \in \mathbb{Z}^+$, with $i \in \{1, 2, \ldots, n-1\}$.

(a) Generate an initial set of S plans $\{\widetilde{G}_0^{(1)}, \widetilde{G}_0^{(2)}, \ldots, \widetilde{G}_0^{(S)}\}$ and corresponding weights $\{w_0^{(1)}, w_0^{(2)}, \ldots, w_0^{(S)}\}$, where each $\widetilde{G}_0^{(j)} := G$ and $w_0^{(j)} = 1$.

(b) For each splitting iteration $i \in \{1, 2, \ldots, n-1\}$:

(1) Until there are S valid plans:

(i) Sample a partial plan $\widetilde{G}_{i-1}$ from $\{\widetilde{G}_{i-1}^{(1)}, \widetilde{G}_{i-1}^{(2)}, \ldots, \widetilde{G}_{i-1}^{(S)}\}$ according to weights $\left(\prod_{l=1}^{i-1} w_l^{(j)}\right)^{\alpha}$.

(ii) Split off a new district from $\widetilde{G}_{i-1}$ through one iteration of the splitting procedure (Algorithm 1), creating a new plan $(G_i, \widetilde{G}_i)$.

(iii) If the newly sampled plan $(G_i, \widetilde{G}_i)$ satisfies $\text{pop}(V_i) \in [P_i^-, P_i^+]$, save it; otherwise, reject it.

(2) Calculate weights for each of the new plans $w_i^{(j)} = \frac{\tau(G_i^{(j)})^{\rho-1}}{|\mathcal{C}(G_i^{(j)}, \widetilde{G}_i^{(j)})|}$.

(c) Calculate final weights

$$w^{(j)} = \exp\{-J(\xi^{(j)})\}\psi([\xi^{(j)}])^{-1}\left(\prod_{i=1}^{n-2} w_i^{(j)}\right)^{(1-\alpha)} w_{n-1}^{(j)}\left(\tau(G_n^{(j)})\right)^{\rho-1}. \tag{5}$$

(d) Output the S final plans $\{\xi^{(j)}\}_{j=1}^S$, where $\xi^{(j)} = (G_1^{(j)}, \ldots, G_{n-1}^{(j)}, G_n^{(j)})$, and the final weights $\{w^{(j)}\}_{j=1}^S$.

labeled plan ξ a *sequentially valid labeling* if this property holds, and denote by $\psi([\xi])$ the number of sequentially valid labelings corresponding to an unlabeled plan $[\xi]$. The SMC weights incorporate $\psi([\xi])$ as a correction term. Calculation of $\psi([\xi])$ is discussed below in Section 4.4.2.

From the output of Algorithm 2, one last resampling of S plans using the final weights can be performed to generate a final sample. Alternatively, the weights can be used directly to estimate the expectation of some statistics of interest under the target distribution, i.e., $H = \mathbb{E}_\pi(h(\xi))$, using the self-normalized importance sampling estimate $\widehat{H} = \sum_{j=1}^S h(\xi^{(j)})w^{(j)} / \sum_{j=1}^S w^{(j)}$.

The sampled plans are not completely independent, because the weights in each step must be normalized before resampling, and because the resampling itself introduces some dependence. Precisely quantifying the amount of dependence is difficult. However, as we demonstrate in Section 5, the dependence is not large enough to cause measurable bias in summary statistics of interest.

It is difficult to precisely characterize the computational complexity of the entire SMC algorithm since the rejection sampling introduces a random component, which depends on the difficulty of sampling a new district within the population bounds. This random complexity is also shared by existing MCMC approaches, which must redraw proposals if they are invalid. Appendix C analyzes the complexity of generating each sample without accounting for the rejection step. In contrast with MCMC approaches, step (b)(1) of Algorithm 2 can be embarrassingly parallelized, since each resample-split-check requires interaction only with the previous set of samples.

The weights in the proposed algorithm are chosen to match existing general SMC algorithms with partial rejection control. These existing algorithms provide guarantees as to the convergence of the samples to the target distribution. One such result, which will suffice for our purposes, is the following central limit theorem.

Proposition 1. *Let* $\pi_S = \sum_{j=1}^S w^{(j)}\delta_{[\xi^{(j)}]}$ *be the weighted particle approximation generated by Algorithm 2.*

Then for all measurable h on unlabeled plans, as $S \to \infty$,

$$\sqrt{S}(\mathbb{E}_{\pi_S}[h([\xi])] - \mathbb{E}_{\pi}[h([\xi])]) \xrightarrow{d} \mathcal{N}(0, V_{SMC}(h)),$$

for some asymptotic variance $V_{SMC}(h)$.

A proof is given in Appendix A. This central limit theorem implies consistency (in S) of any quantity derived from the weighted samples. However, since this convergence is in probability (w.r.t. the algorithm's sampling probability), the proposition does not establish that $\pi_S \xrightarrow{d} \pi$ almost surely. While the almost sure convergence result exists for a standard SMC algorithm (Del Moral et al., 2006), we do not know of an extension to the case of partial rejection control.

In some cases, the constraints incorporated into $J(\xi)$ admit a natural decomposition to the district level as $\prod_{i=1}^{n} J'(G_i)$—for example, a preference for districts which split as few counties as possible, or against districts which would pair off incumbents. In these cases, an extra term of $\exp\{-J'(G_i^{(j)})\}$ can be added to the weights $w_i^{(j)}$ in each stage, and the same term can be dropped from the final weights $w^{(j)}$. This can be particularly useful for more stringent constraints; incorporating J' in each stage allows the importance resampling to "steer" the set of redistricting plans towards those which are preferred by the constraints. This same idea can be used to amortize the contribution of $\psi([\xi])$ to the final weights over the preceding SMC iterations; details can be found in Appendix C.

4.4 Practical Implementation and Use

The SMC algorithm described above allows for an accurate characterization of the target distribution given in Equation (1) as the sample size goes to infinity. In practice, of course, we must use a finite number of samples. Additionally, the algorithm relies on choosing k_i, which is difficult since K_i is typically unknown, and on calculating $\psi([\xi])$, which is difficult since it is not analytically available.

This section discusses these practical implementation challenges, as well as selection of the parameter α and the use of diagnostics that can be applied to help identify when the algorithm is or is not performing well. Further details about implementation may be found in Appendix C.

4.4.1 Choosing k_i

The accuracy of the algorithm is theoretically guaranteed only when the number of edges considered for removal at each stage is at least the maximum number of edges across all graphs which induce districts G_i with $\mathrm{dev}(G_i) \leq D$, i.e., $k_i \geq K_i$. Unfortunately, K_i is generally unknown in practice. We could conservatively set $k_i = m - 1$, the number of edges in each spanning tree but such a choice results in a prohibitively inefficient algorithm—the random edge selected for removal will with high probability induce an invalid partition, leading to a rejection of the entire map. In practice we estimate k_i before each SMC stage by generating random spanning trees, as we detail in Appendix C. Theoretical support for this estimate is given by Proposition 2, which bounds the inaccuracy of the approximation with a user-selectable parameter.

4.4.2 Calculating $\psi([\xi])$

The number of sequentially valid labelings $\psi([\xi])$ corresponding to an unlabeled plan $[\xi]$ can vary significantly across unlabeled plans, and has no closed-form expression permitting easy calculation. Unfortunately, $\psi([\xi])$ grows rapidly in n, but not nearly as fast as $n!$, the total number of labelings per unlabeled plan, which makes both direct and approximate calculation methods challenging. We adopt a hybrid approach in practice. When $n \leq 13$, we compute $\psi([\xi])$ with a recursive divide-and-conquer algorithm. When $n > 13$, we can estimate $\psi([\xi])$ to arbitrary precision using a particular importance sampling scheme. Both of these approaches are detailed in Appendix C, and do not increase the runtime of the SMC algorithm asymptotically.

4.4.3 Choosing α

As noted above, as long as $\alpha \in (0, 1]$ the SMC algorithm is asymptotically valid. Larger values are more aggressive in downweighting unlikely plans (those which are overrepresented in q versus π), which may lead to less diversity in the final sample, while smaller values of α are less aggressive, which can result in more variable

final weights and more wasted samples. Liu et al. (2001) recommend a default choice of $\alpha = 0.5$, which we have found appropriate, though in general we have not found the algorithm's performance to be very sensitive to α within the range $0.5 \leq \alpha \leq 0.9$.

4.4.4 Diagnostics

As with any complex sampler, application of the SMC algorithm in practice can be greatly facilitated by a set of diagnostic measures. Diagnostics can never prove that an algorithm is working correctly, but they can identify situations when the algorithm fails and shed light on why. We discuss several useful diagnostics below, all of which are implemented in our open-source software.

Our primary recommendation is to perform multiple independent runs of the SMC algorithm, which enables both checking for nonconvergence to the target distribution as well as the calculation of standard errors for quantities of interest. For convergence, we adopt the Gelman-Rubin $\hat{R}$ statistic (Gelman and Rubin, 1992), which compares between-run variation to within-run variation. If the latter is an appreciable fraction of the former, then independent runs produce different results, and the algorithm has not converged. For our purposes, we will consider an algorithm not to have converged if $\hat{R}$ values for statistics of interest exceed 1.05, though higher or lower thresholds may be appropriate for given computational budgets and other applied considerations.

We use the rank-normalized and folded $\hat{R}$ of Vehtari et al. (2019), which makes the statistic more robust to heavy tails and more sensitive to discrepancies in scale, not just location. For our purposes, $\hat{R}$ is advantageous, since it is computed on summary statistics and thus provides a measure of convergence tailored to practical quantities of interest. Additionally, it is unitless, and so can be compared across runs and indeed across algorithms, as we will do in Section 5.

Multiple runs can also be easily used to compute standard errors of expectations taken relative to the sampled plans. Recently, Lee and Whiteley (2018) and Olsson and Douc (2019), among others, have developed a method to estimate SMC variance by using information on the ancestry of each sampled particle (here, plan). While we have implemented this method in our open-source software and found it to be correct on average, the resulting estimates tend to be much noisier than those obtained from a simple multiple-run standard error.

Other diagnostics are useful for assessing the overall quality of the sample and pinpointing where issues arise when the algorithm is not working well. One measure of quality is *sample diversity*, or how different the samples are from each other, which we measure using the variation of information metric shown in Equation (2). Similar plans will have a variation of information near zero, while plans which are extremely different will have a variation of information closer to 1. A non-diverse sample will have many copies of similar or identical plans, which tends to increase sampling error and reduces the interpretability of the generated samples.

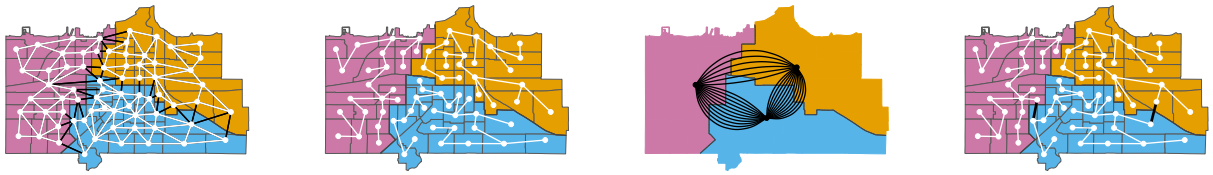
When sampling problems arise, a closer look at each iteration of Algorithm 2 can help identify the cause. A low acceptance rate in step (b)(1)(3), or weights which are highly variable or have a heavy right tail in step (b)(2), can lower the sampling efficiency and diversity. Examining the variance of the log-weights at each iteration can be useful. Variances which increase consistently across iterations can be a sign to use a higher α, which will sample more aggressively earlier on to avoid a build-up of variable weights. We have also found that the number of unique plans from the previous iteration which survive to the subsequent iteration is a highly useful indicator of sampling problems, since it captures inefficiencies introduced by both the rejection and resampling steps. If the number of unique plans is substantially below what would be expected from a uniform sample with replacement from the same population, then some kind of bottleneck or other sampling problem is generally present.

4.5 Incorporating Administrative Boundary Constraints

Another common requirement for redistricting plans is that districts "to the greatest extent possible" follow existing administrative boundaries such as county and municipality lines.[1] In theory, this constraint can be formulated using a J function which penalizes maps for every county line crossed by a district. In practice, however, we can more efficiently generate desired maps by directly incorporating this constraint into our sampling algorithm.

Fortunately, with a small modification to the proposed algorithm, we can sample redistricting plans proportional to a similar target distribution but with the additional constraint that the number of administrative splits

[1] If a redistricting plan must always respect these boundaries, we can simply treat the administrative units as the nodes of the original graph to be partitioned.



(a) The Erie graph. Edges that cross from one administrative unit to another are colored black.

(b) Spanning trees drawn on each administrative unit.

(c) The quotient multigraph. Edges connecting boundary units in the original graphs become edges in the multigraph.

(d) The final spanning tree. A spanning tree on the quotient multigraph connects the trees on each administrative unit.

Figure 3: The two-step spanning tree sampling procedure applied to the city of Erie, Pennsylvania, with three arbitrary administrative units indicated by the colored sections of each map.

not exceed $n-1$. The hard constraint of $n-1$ splits cannot be adjusted upwards or downwards, unfortunately, but a preference for even fewer administrative splits may be incorporated through the J function.

Let A be the set of administrative units, such as counties. We can relate these units to the nodes by way of a labeling function $\eta : V \to A$ that assigns each node to its corresponding unit. This function induces an equivalence relation $\sim_\eta$ on nodes, where $v \sim_\eta u$ for nodes v and u iff $\eta(v) = \eta(u)$. If we quotient G by this relation, we obtain the administrative-level multigraph $G \,/\, \sim_\eta$, where each vertex is an administrative unit and every edge corresponds to an edge in G which connects two nodes in different administrative units. We can write the number of administrative splits as

$$\mathrm{spl}(\xi) = \left(\sum_{a \in A} \sum_{i=1}^{n} C(\eta^{-1}(a) \cap \xi^{-1}(i)) \right) - |A|,$$

where $C(\cdot)$ counts the number of connected components in the subgraph $\eta^{-1}(a) \cap \xi^{-1}(i)$.

To implement this constraint, we draw the spanning trees in step (a) of Algorithm 2 in two substeps such that we sample from a specific subset of all spanning trees. First, we use Wilson's algorithm to draw a spanning tree on each administrative unit $a \in A$, and then we connect these spanning trees to each other by drawing a spanning tree on the quotient multigraph $\widetilde{G}_i \,/\, \sim_\eta$. Figure 3 illustrates this process. This approach is similar to the independently-developed multi-scale merge-split algorithm of Autry et al. (2020).

Drawing the spanning trees in two steps limits the trees used to those which, when restricted to the nodes $\eta^{-1}(a)$ in any administrative unit a, are still spanning trees. The importance of this restriction is that cutting any edge in such a tree will either split the map exactly along administrative boundaries (if the edge is on the quotient multigraph) or split one administrative unit in two and preserve administrative boundaries everywhere else. Since the algorithm has $n-1$ stages, this limits the support of the sampling distribution to maps with no more than $n-1$ administrative splits.

The two-step construction makes clear that the total number of such spanning trees is given by

$$\tau_\eta(\widetilde{G}_i) \;=\; \tau(\widetilde{G}_i \,/\, \sim_\eta) \prod_{a \in A} \tau(\widetilde{G}_i \cap \eta^{-1}(a)), \tag{6}$$

where $\widetilde{G}_i \cap \eta^{-1}(a)$ denotes the subgraph of $\widetilde{G}_i$ which lies in unit a, and we take $\tau(\varnothing) = 1$. Replacing τ with τ_η in the expression for the weights $w_i^{(j)}$ and $w^{(j)}$ then gives the modified algorithm that asymptotically samples from

$$\pi_\eta([\xi]) \;\propto\; \exp\{-J(\xi)\} \tau_\eta(\xi)^\rho \mathbf{1}_{\{\xi \text{ connected}\}} \mathbf{1}_{\{\mathrm{dev}(\xi) \le D\}} \mathbf{1}_{\{\mathrm{spl}(\xi) \le n-1\}}. \tag{7}$$

This idea can in fact be extended to arbitrary levels of nested administrative hierarchy. We can, for example, limit not only the number of split counties but also the number of split cities and Census tracts to $n-1$ each, since tracts are nested within cities, which are nested within counties. To do so, we begin by drawing spanning

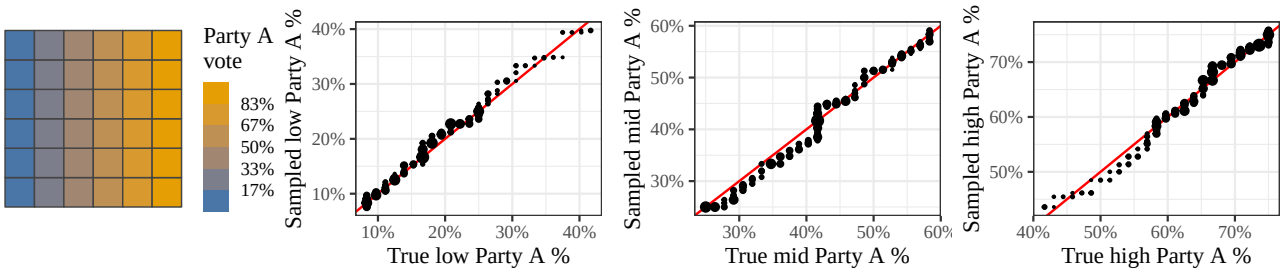


Figure 4: The 6-by-6 map used in validation, colored by the vote share of Party A in each precinct (left). The remaining panels are quantile-quantile plots comparing the sampled distribution to the enumerated distribution weighted to the target measure for three summary statistics: the lowest, middle, and highest vote shares of Party A across the three districts.

trees using Wilson's algorithm on the smallest administrative units. We then connect spanning trees into larger and larger trees by drawing spanning trees on the quotient graphs of each higher administrative level. (In this approach, cities which cross county boundaries must be split in two, and regions outside of cities must be treated as their own pseudo-city.) Even with multiple levels of administrative hierarchy, the calculation of the number of spanning trees is still straightforward, by analogy to Equation (6).

5 An Empirical Validation Study

Although the proposed algorithm has desirable theoretical properties, it is important to empirically assess its performance (Fifield et al., 2020b). We examine whether or not the proposed algorithm can produce a sample of redistricting maps that is actually representative of a target distribution, and compare its efficiency to that of a state-of-the-art MCMC redistricting algorithm (Carter et al., 2019).[2]

We apply the efficient enumeration algorithm of (Fifield et al., 2020b) to obtain all possible redistricting plans with three contiguous districts on the 6-by-6 grid map shown in Figure 4. Each precinct in the grid has a population of 6, and we have imposed a hypothetical partisan composition which changes from primarily "Party A" on the right to primarily "Party B" on the left. There are over 99 million unlabeled plans with three districts, 264,500 of which have districts with exactly balanced populations. We demonstrate that the proposed algorithm can efficiently approximate a realistic target distribution on this set of balanced plans. Appendix B contains an additional validation study with four districts using a 50-precinct map taken from the state of Florida.

We also compare the accuracy and efficiency of the SMC algorithm to that of a merge-split MCMC algorithm (Carter et al., 2019). The merge-split algorithm here uses a spanning tree-based proposal identical to the splitting procedure described in Algorithm 1: it merges adjacent districts, draws a spanning tree on the merged district, and splits it to ensure the population constraint is met. In fact, the implementation (contained in our open-source `redist` package) shares much of the same code base, which allows for more realistic performance comparisons. The stationary distribution of this algorithm is exactly that of Equation (1).

We set the target distribution by choosing $\rho = 1$, which as discussed above will generate compact districts. This is also the value of ρ for which the proposal and sampling distributions are as close as possible; both the SMC and MCMC algorithms will operate at peak efficiency. The validation study in Appendix B explores a wider range of values of ρ.

To start, we generate 400 samples using Algorithm 2 from this target, and compute the vote share of Party A in the three districts of each plan. We compute the same vote shares for all of the 264,500 enumerated plans, and compare the distribution of the lowest, middle, and highest Party A vote shares, weighting the enumerated plans according to $\tau(\xi)$.

[2]Throughout this paper, we do not make a comparison to the recombination ("ReCom") algorithm of DeFord et al. (2021), which pioneered the spanning tree-based proposal used in the merge-split algorithm, as it does not sample from a known target distribution.

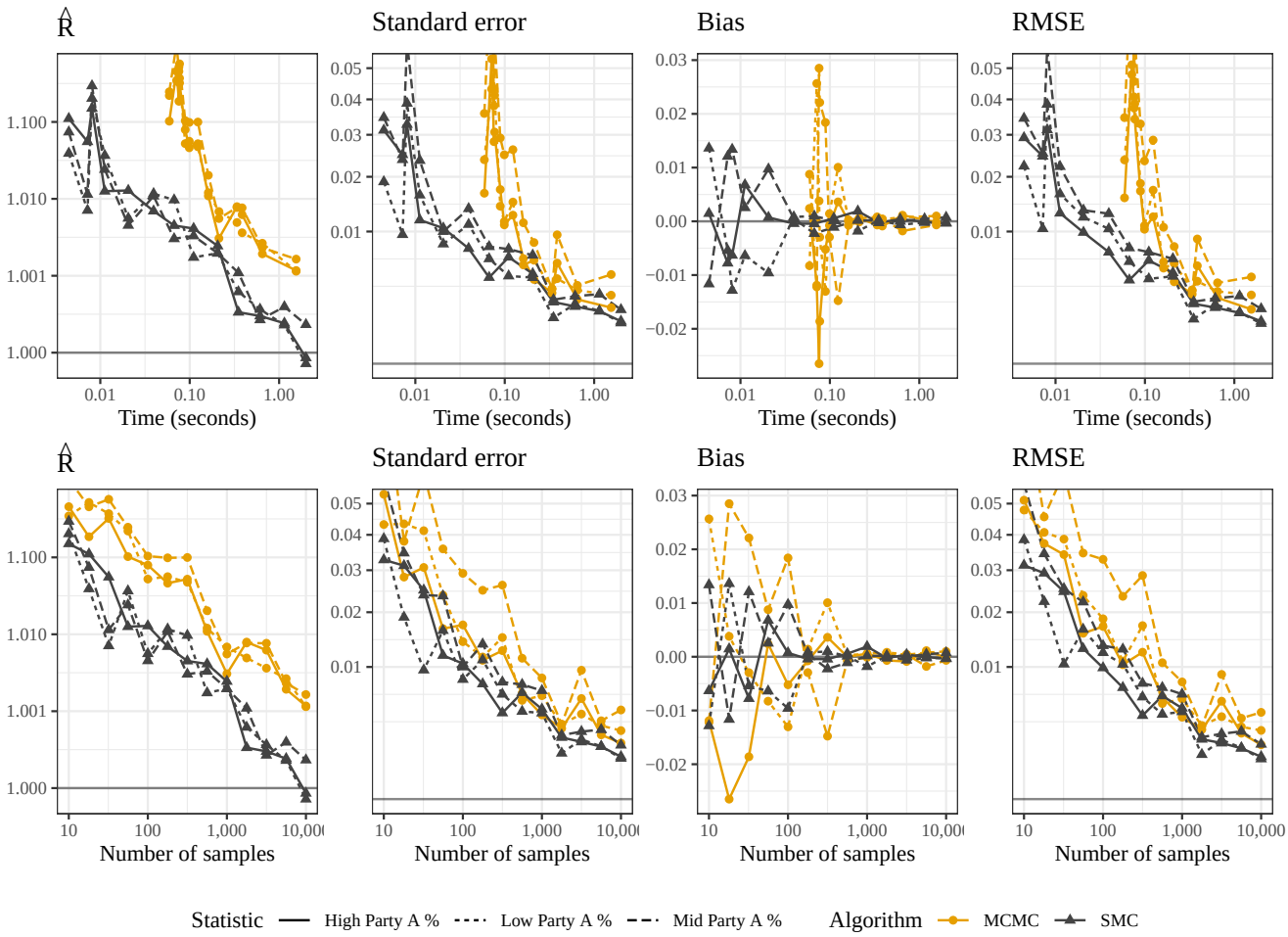


Figure 5: Convergence diagnostic $\hat{R}$ (on a log scale), standard errors, bias, and RMSE for the same three Party A vote share summary statistics, calculated across eight independent runs of the MCMC and SMC algorithms. Values are plotted versus total sampling time (top row) and number of samples per run (bottom row).

This comparison is summarized in the right three panels of Figure 4 with quantile-quantile plots. Despite the relatively small sample size, the agreement is excellent, indicating that for these partisan summary statistics, which are the kind generally of interest in redistricting analysis and litigation, the SMC algorithm is able to accurately represent the target distribution. We can also perform three more identical runs of 400 samples, allowing us to compute the $\hat{R}$ diagnostic for our three summary statistics. These are 1.006, 1.001, and 1.003 for the low, mid, and high Party A vote shares, indicating likely convergence of the algorithm.

We next study the quality of the SMC sample as the sample size increases, and make comparisons to the MCMC algorithm with the same proposal distribution. The MCMC chains are each initialized with a random SMC sample, and run for 500 warm-up iterations before samples are recorded. We let the sample size vary from 100 to 10,000. For each sample size, we perform eight independent runs of each algorithm, which allows for accurate calculation of standard errors and $\hat{R}$ for each of our three summary statistics. We also compute the bias and the root mean-square error (RMSE) of estimates for the average values of the summary statistics, compared to the true average values from the $\tau(\xi)$-weighted enumeration distribution.

Figure 5 summarizes the results of this experiment. As the left two panels show, for any given number of iterations or total runtime, the SMC algorithm generates samples with lower $\hat{R}$. According to our heuristic convergence check that $\hat{R} \leq 1.05$, the SMC algorithm converges after roughly 55 iterations, while the MCMC algorithm requires roughly 560 iterations (plus 500 warm-up iterations), for a total runtime that is 14.5 times

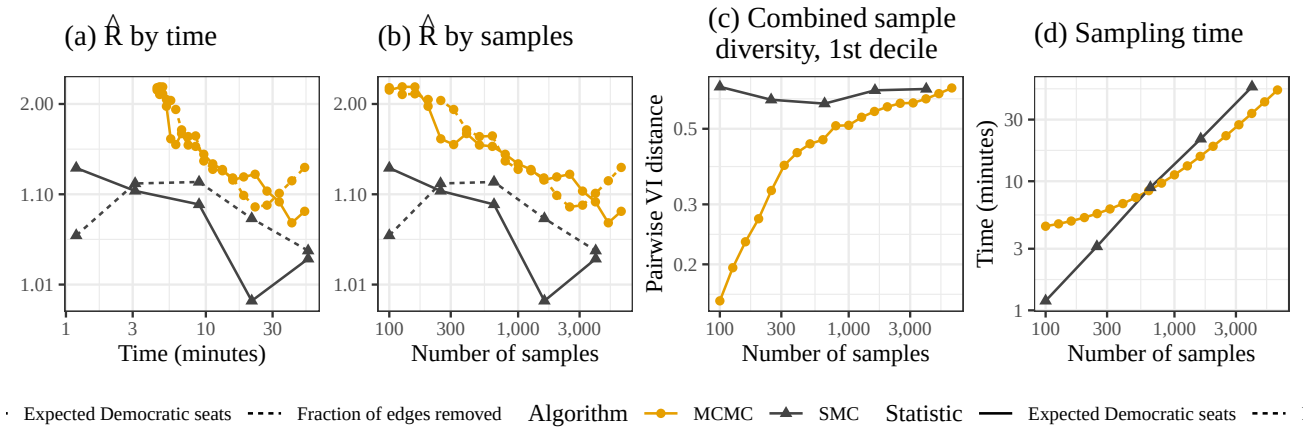


Figure 6: Convergence diagnostic $\hat{R}$ (on a log scale) for two statistics plotted against sampling time (a) and number of samples (b). First decile of the pairwise variation of information distance (a measure of sample diversity) versus number of samples (c). Distance measures were computed with the combined sample of four independent runs. Sampling time versus number of samples (d).

longer. The same pattern is apparent in the standard errors and RMSE, with the SMC algorithm generating less variable estimates for both a fixed sample size and a fixed total runtime for each summary statistic. As the center-right panels show, both algorithms produce roughly unbiased estimates at all sample sizes.

6 Analysis of the 2011 Pennsylvania Redistricting

As discussed in Section 2, in the process of determining a remedial redistricting plan to replace the 2011 General Assembly map, the Pennsylvania Supreme Court received submissions from seven parties. In this section, we compare four of these maps to both the original 2011 plan and the remedial plan ultimately adopted by the court. We study the governor's plan and the House Democrats' plan; the petitioner's plan (specifically, their "Map A"), which was selected from an ensemble of 500 plans used as part of the litigation; and the respondent's plan, which was drawn by Republican officials.

6.1 The Setup

To evaluate these six plans, we drew reference maps from the target distribution given in Equation (7) by using the proposed algorithm along with the modifications presented in Section 4.5 to cap the number of split counties at 17 (out of a total of 67), in line with the court's mandate. We set $\rho = 1$ to put most of the sample's mass on compact districts, and enforced $\mathrm{dev}(\xi) \leq 0.001$ to reflect the "one person, one vote" requirement.

This population constraint translates to a tolerance of around 700 people, in a state where the median precinct has 1,121 people. Like most research on redistricting, we use precincts because they represent the smallest geographical units for which election results are available. To draw from a stricter population constraint we would need to use the 421,545 Census blocks in Pennsylvania rather than the 9,256 precincts, which would significantly increase the computational burden.

6.2 Comparison with a State-of-the-Art MCMC Algorithm

We first compare the computational efficiency of the SMC algorithm with the merge-split MCMC algorithm used in the empirical validation above. As in the validation study, the MCMC chains are each initialized with a random SMC sample, and run for 500 warm-up iterations before samples are recorded. We let the sample size vary from 100 to 4,000 for the SMC algorithm and from 100 to 6,310 for the MCMC algorithm. For each sample size, we perform four independent runs of each algorithm.

Figures 6(a) and (b) show the $\hat{R}$ values of two summary statistics (Democratic seats and compactness) by the total runtime and the number of samples per run. As with the smaller validation study, the SMC algorithm is more efficient, with lower $\hat{R}$ values than the MCMC algorithm for a given runtime or sample size. According

to our heuristic convergence check that $\hat{R} \leq 1.05$, the SMC algorithm converges after 1,600 iterations, while the MCMC algorithm has still not converged after more than 6,300 iterations (plus 500 warm-up iterations), at which point the MCMC algorithm has been running 2.4 times longer than the SMC algorithm.

Figure 6(c) examines the sampling efficiency from the perspective of sample diversity. As discussed in Section 4.4.4, the pairwise variation of information (VI) distance provides a way to quantify how similar the plans in each sample are to each other. If the distribution of these pairwise VI distances has a long lower tail, then there are a significant number of plans which are very similar. Figure 6(c) shows the first decile of the pairwise VI distances in the combined sample (i.e., pooled across all four runs) for each algorithm as the sample size increases. Because of the sequential nature of the MCMC algorithms, the sample diversity doesn't match that of the SMC algorithms until there are at least 3,000 samples. This is reflected in the lag-1 autocorrelation of the MCMC sample, which was 0.982 for Democratic seats and 0.984 for compactness.

Not shown in Figure 6 are the standard errors for each quantity of interest. Like the sample diversity and $\hat{R}$, for a fixed sample size or runtime, the SMC algorithm produced substantially smaller standard errors for both summary statistics than the MCMC algorithm (30–80% smaller across the range of sample sizes), indicating higher efficiency.

Figure 6(d) shows how long each algorithm takes to run for a given number of samples. Because of the fixed warm-up cost of the MCMC algorithm, it is slower than the SMC algorithm for less than 600 samples. Asymptotically, both algorithms have runtime linear in the number of samples. But the per-sample cost of the MCMC algorithm is lower, since only two districts are redrawn, rather than the entire map.

6.3 Compactness and County Splits

While the results of the comparison above suggest that 1,600 SMC samples is enough for Pennsylvania, we perform the rest of our analysis with 4,000 samples from one of the SMC runs.

Figures 7(a) and (b) show distribution of the fraction of edges removed $\mathrm{rem}(\xi)$ and the number of county splits $\mathrm{spl}(\xi)$ across the reference maps generated by our algorithm. The figure also shows these values for each of the six plans. The 2011 General Assembly plan is a clear outlier for both statistics, being far less compact and splitting far more counties than any of the reference plans and all of the remedial plans. Among the remedial plans, the petitioner's is almost an outlier in being *too* compact, although this is perhaps not surprising—the map was generated by an algorithm explicitly designed to optimize over criteria such as population balance and compactness (Chen, 2017). The other plans' compactness falls within the range of the reference sample.

Figure 7(b) shows that all of the submitted remedial plans, as well as the reference sample, followed the court's instructions to limit the number of county splits. Yet around half of the reference maps split fewer than 13 counties, the minimum number of splits found in any of the remedial plans. This may be a result of the strict population constraint (all six plans were within 1 person of equal population across all districts), or a different prioritization between the various constraints imposed.

6.4 Partisan Analysis

While important, the outlier status of the General Assembly plan as regards compactness and county splits is not sufficient to show that it is a *partisan* gerrymander. To evaluate the partisan implications of the six plans, we take a precinct-level baseline voting pattern and aggregate it by district to explore hypothetical election outcomes under the six plans and the reference maps. The baseline pattern is calculated by averaging the vote totals for the three presidential elections and three gubernatorial elections that were held in Pennsylvania from 2000 to 2010.[3] These election data were also used during litigation. While being far from a perfect way to create counterfactual election outcomes, this simple averaging of statewide results is often used in academic research and courts. The values of $\hat{R}$ for all district-level vote shares were less than 1.05.

Then, within each plan, we number the districts by their baseline Democratic two-party vote share, so District 1 is the least Democratic and District 18 the most. Figure 7(c), analogous to Figure 7 in Herschlag et al. (2017), presents the distribution of the Democratic two-party vote share for each of the districts across the

[3]Data from Ansolabehere, S. and Rodden, J. (2011). Pennsylvania Data Files. Available at https://doi.org/10.7910/DVN/FJHHDS.

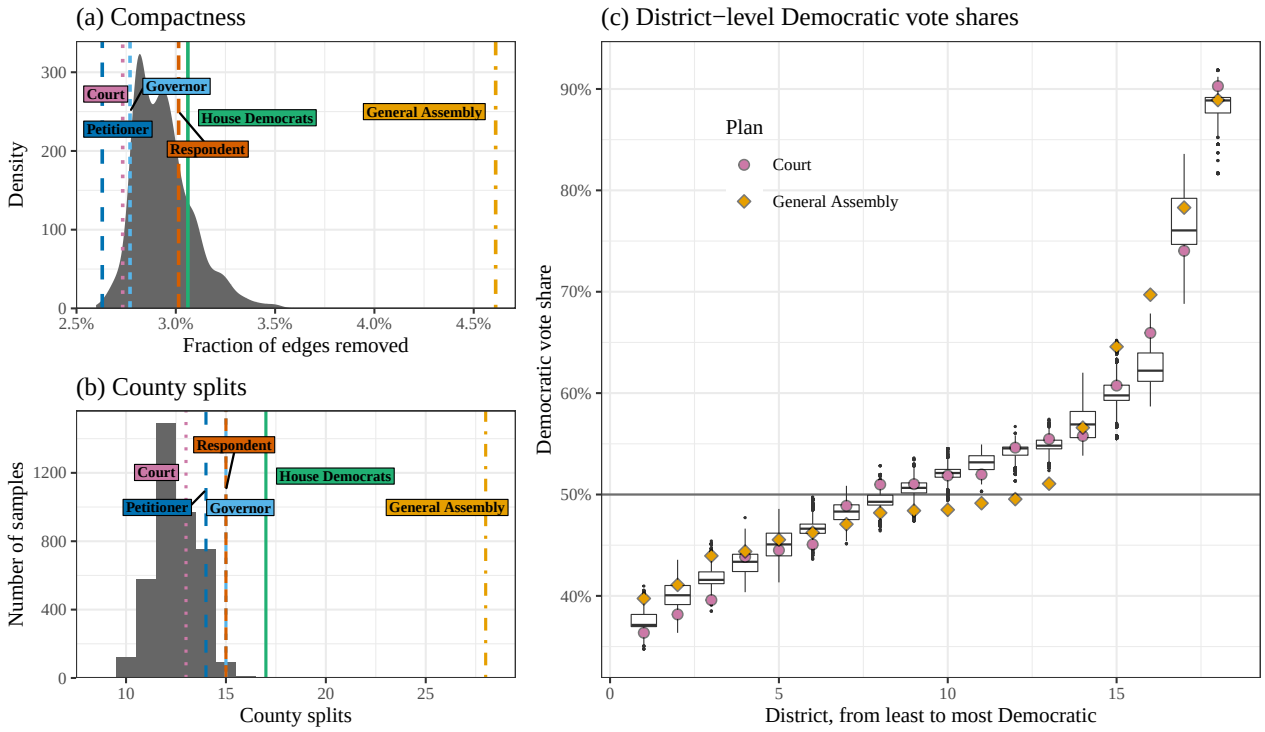


Figure 7: Summary statistics for the sampled and comparison plans: (a) compactness (smaller means more compact) and (b) county splits. Panel (c) shows the Democratic two-party vote share by district, where within each plan districts are ordered by Democratic vote share.

reference maps, and also shows the values for the General Assembly plan (orange triangles) and the court's adopted plan (purple circles). When compared to the reference maps and the court's plan, the General Assembly plan tends to yield smaller Democratic vote share in competitive districts (p-values of 0.0162, 0.0002, 0.0002, and 0.0002 for Districts 9–12 compared to the reference set) while giving larger Democratic vote share in non-competitive districts. The effect of this is to produce 6 Democratic seats, on average, under the General Assembly plan, compared to a range of 8–12 under the reference sample (p-value 0.0002) and 11 under the Court plan. All of these findings provide evidence that the General Assembly plan was gerrymandered in favor of Republicans by packing Democratic voters into non-competitive districts.

7 Concluding Remarks

Redistricting sampling algorithms allow for the empirical evaluation of a redistricting plan by generating alternative plans under a certain set of constraints. Researchers and policymakers can compute various statistics from the redistricting plan of interest and compare them with the corresponding statistics based on these sampled plans. Unfortunately, existing approaches often struggle when applied to real-world problems, owing to the scale of the problems and the number of constraints involved.

The SMC algorithm presented here can provably asymptotically sample from a specific target distribution, the first such guarantee in the literature. Because of the constructive sampling approach, it does not face the same kinds of scalability problems as many existing MCMC algorithms, and even performs more efficiently than a MCMC algorithm that uses an identical proposal distribution. The proposed method also incorporates, by design, the common redistricting constraints of population balance, geographic compactness, and minimizing administrative splits. We expect these advantages of the SMC algorithm to substantially improve the reliability of outlier analysis in real-world redistricting cases.

The proposed approach is not without its limitations. Like other redistricting sampling algorithms, if the target distribution strays too far from the proposal distribution, the quality of the generated sample will suffer

dramatically. Close examination of diagnostic measures like $\hat{R}$, sample diversity, and per-iteration efficiency is required to provide confidence that the algorithm is sampling accurately and has converged for particular quantities of interest. Our open-source software provides these diagnostics so that analysts can use the proposed algorithm successfully.

One drawback particular to the SMC algorithm arises in situations with dozens or hundreds of separate districts. Like any other SMC or particle resampling scheme, as the number of iterations (districts) grows, eventually all of the samples will share a common ancestor. While this is not a problem in many SMC applications (such as Bayesian inference), for redistricting, this means that all of the sampled plans will share one or more districts that are completely identical. Any summary statistics which rely on these districts will have very large standard errors and may not converge even with a large number of samples. This problem can be ameliorated, though not solved, with many independent runs of the SMC algorithm. One interesting avenue for future research would be to examine whether several applications of an MCMC kernel at various points in the SMC algorithm could help refresh the sample and counteract the tendency to collapse to a single ancestor.

Future research should also explore the possibility of improving several design choices in the algorithm to further increase its efficiency. Wilson's algorithm, for instance, can be generalized to sample from edge-weighted graphs. Choosing weights appropriately could lead to trees which induce maps that are more balanced or more compact. And the procedure for choosing edges to cut, while allowing for the sampling probability to be calculated, introduces inefficiencies by leading to many rejected maps. Further improvements in either of these areas should allow us to better sample and investigate redistricting plans over large maps and with even more complex sets of constraints.

References

Akitaya, H. A., Korman, M., Korten, O., Souvaine, D. L., and Tóth, C. D. (2022). Reconfiguration of connected graph partitions via recombination. *Theoretical Computer Science*.

Autry, E., Carter, D., Herschlag, G., Hunter, Z., and Mattingly, J. (2020). Multi-scale merge-split Markov chain Monte Carlo for redistricting. *Working paper*.

Bangia, S., Graves, C. V., Herschlag, G., Kang, H. S., Luo, J., Mattingly, J. C., and Ravier, R. (2017). Redistricting: Drawing the line. *arXiv preprint arXiv:1704.03360*.

Bozkaya, B., Erkut, E., and Laporte, G. (2003). A tabu search heuristic and adaptive memory procedure for political districting. *European journal of operational research*, 144(1):12–26.

Carter, D., Herschlag, G., Hunter, Z., and Mattingly, J. (2019). A merge-split proposal for reversible Monte Carlo Markov chain sampling of redistricting plans. *arXiv preprint arXiv:1911.01503*.

Chen, J. (2017). Expert report of Jowei Chen, Ph.D. Expert witness report in League of Women Voters v. Commonwealth.

Chen, J. and Rodden, J. (2013). Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science*, 8(3):239–269.

Chikina, M., Frieze, A., and Pegden, W. (2017). Assessing significance in a Markov chain without mixing. *Proceedings of the National Academy of Sciences*, 114(11):2860–2864.

Chikina, M., Frieze, A., and Pegden, W. (2019). Understanding our Markov Chain significance test: A reply to Cho and Rubinstein-Salzedo. *Statistics and Public Policy*, 6(1):50–53.

Cho, W. K. T. and Liu, Y. Y. (2018). Sampling from complicated and unknown distributions: Monte Carlo and Markov chain Monte Carlo methods for redistricting. *Physica A: Statistical Mechanics and its Applications*, 506:170–178.

Cho, W. K. T. and Rubinstein-Salzedo, S. (2019). Understanding significance tests from a non-mixing Markov Chain for partisan gerrymandering claims. *Statistics and Public Policy*, 6(1):44–49.

Cirincione, C., Darling, T. A., and O'Rourke, T. G. (2000). Assessing South Carolina's 1990s congressional districting. *Political Geography*, 19(2):189–211.

Cover, T. M. and Thomas, J. A. (2006). *Elements of information theory*. John Wiley & Sons, 2nd edition.

DeFord, D., Duchin, M., and Solomon, J. (2021). Recombination: A family of Markov chains for redistricting. *Harvard Data Science Review*. https://hdsr.mitpress.mit.edu/pub/1ds8ptxu.

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.

Doucet, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. Springer, New York.

Dube, M. P. and Clark, J. T. (2016). Beyond the circle: Measuring district compactness using graph theory. In *Annual Meeting of the Northeastern Political Science Association*.

Duchin, M. (2018). Outlier analysis for Pennsylvania congressional redistricting.

Fifield, B., Higgins, M., Imai, K., and Tarr, A. (2020a). Automated redistricting simulation using Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 29(4):715–728.

Fifield, B., Imai, K., Kawahara, J., and Kenny, C. T. (2020b). The essential role of empirical validation in legislative redistricting simulation. *Statistics and Public Policy*, 7(1):52–68.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.

Guth, L., Nieh, A., and Weighill, T. (2020). Three applications of entropy to gerrymandering. *arXiv preprint arXiv:2010.14972*.

Herschlag, G., Ravier, R., and Mattingly, J. C. (2017). Evaluating partisan gerrymandering in Wisconsin. *arXiv preprint arXiv:1709.01596*.

Ionides, E. L. (2008). Truncated importance sampling. *Journal of Computational and Graphical Statistics*, 17(2):295–311.

Kenny, C. T., McCartan, C., Fifield, B., and Imai, K. (2020). redist: Computational algorithms for redistricting simulation. https://CRAN.R-project.org/package=redist.

Kostochka, A. V. (1995). The number of spanning trees in graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):269–274.

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

Lee, A. and Whiteley, N. (2018). Variance estimation in the particle filter. *Biometrika*, 105(3):609–625.

LeGland, F. and Oudjane, N. (2005). A sequential particle algorithm that keeps the particle system alive. In *2005 13th European Signal Processing Conference*, pages 1–4. IEEE.

Liu, J. S., Chen, R., and Logvinenko, T. (2001). A theoretical framework for sequential importance sampling with resampling. In *Sequential Monte Carlo methods in practice*, pages 225–246. Springer.

Liu, J. S., Chen, R., and Wong, W. H. (1998). Rejection control and sequential importance sampling. *Journal of the American Statistical Association*, 93(443):1022–1031.

Liu, Y. Y., Cho, W. K. T., and Wang, S. (2016). PEAR: a massively parallel evolutionary computation approach for political redistricting optimization and analysis. *Swarm and Evolutionary Computation*, 30:78–92.

Macmillan, W. (2001). Redistricting in a GIS environment: An optimisation algorithm using switching-points. *Journal of Geographical Systems*, 3(2):167–180.

Magleby, D. B. and Mosesson, D. B. (2018). A new approach for developing neutral redistricting plans. *Political Analysis*, 26(2):147–167.

Massey, D. S. and Denton, N. A. (1988). The dimensions of residential segregation. *Social forces*, 67(2):281–315.

Mattingly, J. C. and Vaughn, C. (2014). Redistricting and the will of the people. *arXiv preprint arXiv:1410.8796*.

McKay, B. D. (1981). Spanning trees in random regular graphs. In *Proceedings of the Third Caribbean Conference on Combinatorics and Computing*. Citeseer.

Mehrotra, A., Johnson, E. L., and Nemhauser, G. L. (1998). An optimization based heuristic for political districting. *Management Science*, 44(8):1100–1114.

Najt, L., Deford, D., and Solomon, J. (2019). Complexity and geometry of sampling connected graph partitions. *arXiv preprint arXiv:1908.08881*.

National Conference of State Legislatures (2021). Redistricting criteria. Available at https://www.ncsl.org/research/redistricting/redistricting-criteria.aspx.

Olsson, J. and Douc, R. (2019). Numerically stable online estimation of variance in particle filters. *Bernoulli*, 25(2):1504–1535.

Peters, G. W., Fan, Y., and Sisson, S. A. (2012). On sequential monte carlo, partial rejection control and approximate bayesian computation. *Statistics and Computing*, 22(6):1209–1222.

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

Tutte, W. T. (1984). *Graph Theory*. Addison-Wesley.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2019). Rank-normalization, folding, and localization: An improved r for assessing convergence of mcmc. arxiv. *arXiv preprint arXiv:1903.08008*.

Wilson, D. B. (1996). Generating random spanning trees more quickly than the cover time. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 296–303.

Wu, L. C., Dou, J. X., Sleator, D., Frieze, A., and Miller, D. (2015). Impartial redistricting: A Markov Chain approach. *arXiv preprint arXiv:1510.03247*.

A Proofs of Propositions

Lemma 1. *The probability of splitting a* valid *new district G_i from an existing area $\widetilde{G}_{i-1}$ using Algorithm 1 with parameter $k_i \geq K_i$ is*

$$q(G_i \mid \widetilde{G}_{i-1}, \text{pop}(V_i) \in [P_i^-, P_i^+]) \;=\; \frac{\tau(G_i)\tau(\widetilde{G}_i)}{\tau(\widetilde{G}_{i-1})k_i}|\mathcal{C}(G_i, \widetilde{G}_i)|. \tag{4}$$

Proof. Any spanning tree can be decomposed into two other trees and an edge joining them. Let $T \cup e \cup T'$ denote the spanning tree obtained by joining two other spanning trees, T and T', with an edge e. Then Equation (3) can be written as

$$q(G_i \mid \widetilde{G}_{i-1}) = \sum_{\substack{T^{(1)} \in \mathcal{T}(G_i) \\ T^{(2)} \in \mathcal{T}(\widetilde{G}_i)}} \sum_{e \in \mathcal{C}(T^{(1)}, T^{(2)})} q(G_i \mid T^{(1)} \cup e \cup T^{(2)})\, \tau(\widetilde{G}_{i-1})^{-1}.$$

Now, $q(G_i \mid T^{(1)} \cup e \cup T^{(2)})$ is determined by whether whether $e^* = e$, i.e., if e is the edge selected to be cut. If e has d_e in the top k_i (if it induces one of the best k_i balanced splits), then it has a $1/k_i$ probability of being selected in step (c) and cut. If d_e is not in the top k_i, then this probability is zero.

Everything written to this point holds regardless of whether G_i is a valid district (i.e., satisfies $\text{pop}(V_i) \in [P_i^-, P_i^+]$). From here onwards we will restrict our attention to valid districts only. Notice that the forward-looking bounds P_i^- and P_i^+ are stricter than merely ensuring $\text{dev}(G_i) \leq D$. That is, conditional on $\text{pop}(V_i) \in [P_i^-, P_i^+]$, we must also have $\text{dev}(G_i) \leq D$.

Therefore, if a sorted edge e_j *in any spanning tree* induces such a balanced partition, we must have $j \leq K_i$, where as in the main text K_i counts the maximum number of such edges across all possible spanning trees. Thus, so long as we set $k_i \geq K_i$, we will have $d_e \leq D$.

Furthermore, across all spanning trees $T^{(1)} \in \mathcal{T}(G_i)$ and $T^{(2)} \in \mathcal{T}(\widetilde{G}_i)$, and connecting edges $e \in E(T^{(1)}, T^{(2)})$, the value of d_e is constant, since removing e induces the same districting. Combining these two facts, we have, conditional on satisfying the bounds P_i^- and P_i^+,

$$q(e^* = e \mid T^{(1)} \cup e \cup T^{(2)}, \text{pop}(V_i) \in [P_i^-, P_i^+]) \;=\; k_i^{-1},$$

which does not depend on $T^{(1)}$, $T^{(2)}$, or e. We may therefore write the conditional sampling probability as

$$\begin{aligned} q(G_i \mid \widetilde{G}_{i-1}, \text{pop}(V_i) \in [P_i^-, P_i^+]) \;&=\; \sum_{\substack{T^{(1)} \in \mathcal{T}(G_i) \\ T^{(2)} \in \mathcal{T}(\widetilde{G}_i)}} \sum_{e \in \mathcal{C}(T^{(1)}, T^{(2)})} \frac{1}{k_i \tau(\widetilde{G}_{i-1})} \\ &=\; \frac{\tau(G_i)\tau(\widetilde{G}_i)}{\tau(\widetilde{G}_{i-1})k_i}|\mathcal{C}(G_i, \widetilde{G}_i)|, \end{aligned} \tag{8}$$

where as in the main text we let $\mathcal{C}(G, H)$ represent the set of edges joining nodes in a graph G to nodes in a graph H. □

Proposition 1. *Let $\pi_S = \sum_{j=1}^{S} w^{(j)} \delta_{[\xi^{(j)}]}$ be the weighted particle approximation generated by Algorithm 2. Then for all measurable h on unlabeled plans, as $S \to \infty$,*

$$\sqrt{S}(\mathbb{E}_{\pi_S}[h([\xi])] - \mathbb{E}_{\pi}[h([\xi])]) \xrightarrow{d} \mathcal{N}(0, V_{SMC}(h)),$$

for some asymptotic variance $V_{SMC}(h)$.

The proof proceeds by showing that the weights in Algorithm 2 are of a form derived from an existing SMC algorithm with an established central limit theorem.

Proof. We can associate our target measure $\pi([\xi])$ on unlabeled redistricting plans with a corresponding measure on labeled plans

$$\tilde{\pi}(\xi) \coloneqq \psi([\xi])^{-1}\pi([\xi]),$$

so that the pushfoward measure obtained by mapping $\xi \mapsto [\xi]$ recovers π.

Our SMC algorithm will operate on labeled plans, targeting $\tilde{\pi}$, so that the resulting plans, when considered as representatives of their corresponding unlabeled plans, will be representative of π in the sense given by the theorem statement. Recall that a labeled redistricting plan ξ is just a tuple of graph partitions $(G_1, G_2, \ldots, G_n)$. We begin by extending $\tilde{\pi}(\xi)$ to a series of measures on partial plans,

$$\begin{aligned}\tilde{\pi}_i(G_1, G_2, \ldots, G_i) &:\propto \prod_{j=1}^{i} \frac{\tau(G_j)^\rho \tau(\widetilde{G}_j)}{\tau(\widetilde{G}_{j-1})} \mathbf{1}_{\mathrm{pop}(V_j)\in[P_j^-, P_j^+]} \\ &\propto \tau(\widetilde{G}_j) \prod_{j=1}^{i} \tau(G_j)^\rho \mathbf{1}_{\mathrm{pop}(V_j)\in[P_j^-, P_j^+]},\end{aligned}$$

for $1 \le i \le n-2$, and where we have simplified the telescoping product in the second equality. Recall that the $\widetilde{G}_i$ are determined completely by $G_1, G_2, \ldots, G_i$.

For $i = n-1$, the above definition would yield

$$\tilde{\pi}_{n-1}(G_1, G_2, \ldots, G_i) \propto \tau(\widetilde{G}_{n-1}) \prod_{j=1}^{n-1} \tau(G_j)^\rho \mathbf{1}_{\mathrm{pop}(V_j)\in[P_j^-, P_j^+]}, = \tau(\xi)\tau(G_n)^{1-\rho}\mathbf{1}_{\mathrm{dev}(\xi)\le D},$$

which is close to but not quite the target measure. So we instead define $\pi_{n-1} \coloneqq \tilde{\pi}$; i.e., we add the additional terms $\exp(-J(\xi))$ and $\psi([\xi])$ and adjust for $\tau(G_n)^{1-\rho}$.

With these partial-plan measures defined, notice that the incremental weight $w_i^{(j)}$ for partial plans with $1 \le i \le n-2$ and $\mathrm{pop}(V_j) \in [P_j^-, P_j^+]$ may be written as

$$\begin{aligned} w_i^{(j)} &= \tau(G_i^{(j)})^{\rho-1} \frac{k_i}{|\mathcal{C}(G_i^{(j)}, \widetilde{G}_i^{(j)})|} \\ &= \frac{\tau(G_i^{(j)})^\rho \tau(\widetilde{G}_i^{(j)})}{\tau(\widetilde{G}_{i-1}^{(j)})} \left(\frac{\tau(G_i^{(j)})\tau(\widetilde{G}_i^{(j)})}{\tau(\widetilde{G}_{i-1}^{(j)})} \frac{|\mathcal{C}(G_i^{(j)}, \widetilde{G}_i^{(j)})|}{k_i} \right)^{-1} \\ &= \frac{\tilde{\pi}_i(G_1, \ldots, G_i)}{\tilde{\pi}_{i-1}(G_1, \ldots, G_{i-1}) q(G_i \mid \widetilde{G}_{i-1}, \mathrm{pop}(V_i) \in [P_i^-, P_i^+])}. \end{aligned} \tag{9}$$

For the final weighting at split $i = n-1$, the incremental weight (i.e., not including the residual previous weights $\left(\prod_{i=1}^{n-2} w_i^{(j)}\right)^{1-\alpha}$ given by step (c) of Algorithm 2 is

$$\begin{aligned} &\exp\left(-J(\xi^{(j)})\right) w_{n-1}^{(j)} \psi([\xi])^{-1} \left(\tau(G_{n-1}^{(j)})\right)^{\rho-1} \\ &\qquad = \frac{\tilde{\pi}_{n-1}(G_1, \ldots, G_{n-1})}{\tilde{\pi}_{n-2}(G_1, \ldots, G_{n-2}) q(G_{n-1} \mid \widetilde{G}_{n-2}, \mathrm{pop}(V_{n-1}) \in [P_{n-1}^-, P_{n-1}^+])}, \end{aligned}$$

since this weight includes exactly the same additional terms as $\tilde{\pi}_{n-1}$ mentioned above. So in fact Equation (9) holds for all $1 \le i \le n-1$.

These incremental weights are precisely those of the SMC partial rejection control algorithm of Peters et al. (2012) (see also LeGland and Oudjane (2005)), with the weights set to zero for invalid samples and the partial rejection threshold set to the minimum possible nonzero weight. So after pushing forward to unlabeled plans, we gain immediately the theorem proved in that work (its Equation 6), viz., that for all measurable h on unlabeled plans and as $S \to \infty$, we have

$$\sqrt{S}(\mathbb{E}_{\pi_S}[h([\xi])] - \mathbb{E}_{\pi}[h([\xi])]) \xrightarrow{d} \mathcal{N}(0, V_{\mathrm{SMC}}(h)),$$

for asymptotic variance $V_{\mathrm{SMC}}(h)$ given by Equation 7 of the same work. □

B Additional Validation Example

This section reports the results of another validation study applied to a 50-precinct map taken from the state of Florida. As in Section 5, we use the efficient enumeration algorithm of Fifield et al. (2020b) to obtain all possible redistricting maps with three and four contiguous districts, and use these plans as a baseline to validate the proposed algorithm. The left plot of Figure B.1 below shows the validation map.

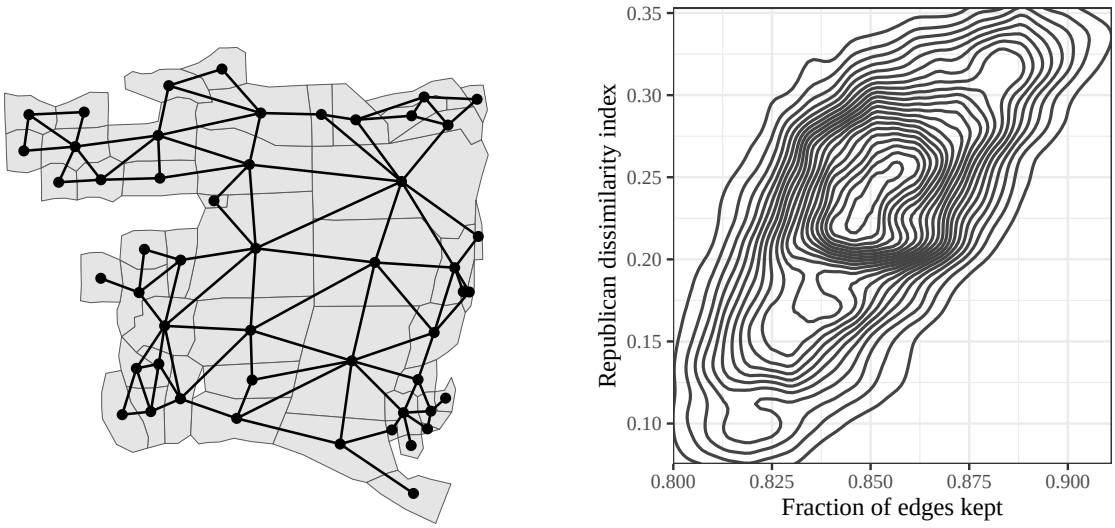


Figure B.1: The 50-precinct Florida map used for validation (left) and the joint distribution of Republican dissimilarity and compactness on the map over all partitions into four districts with $\mathrm{dev}(\xi) \leq 0.10$ (right).

There are 112,515,494 partitions of the map into four districts, 33,635 of which have $\mathrm{dev}(\xi) \leq 0.10$. We evaluate the accuracy of the proposed algorithm, and compare it to the same MCMC algorithm used in Section 5, across a range of target distributions, with ρ running from 0.8 to 1.2 The right plot of Figure B.1 shows the joint distribution of compactness and the summary statistic we use for the validation, the Republican dissimilarity index (Massey and Denton, 1988). With this validation map, Republican dissimilarity is reasonably sensitive to the compactness of districts. This makes dissimilarity a good test statistic for comparing distributions that differ primarily in their average compactness.

For each target distribution, we sample 1,500 redistricting plans from both the SMC and MCMC algorithms, and repeat the sampling four times in order to produce $\hat{R}$ estimates. As in Section 5, the MCMC algorithm is initialized with a random SMC-drawn map and first run for 500 warm-up iterations. To validate and compare the samples, we reweight the enumerated redistricting plans by $\tau(\xi)^\rho$, and then produce quantile-quantile plots of the Republican dissimilarity index, which are shown in Figure B.2.

Across the range of ρ values, the agreement between the SMC sample and target distributions is excellent, even in the lower tail. The $\hat{R}$ values for the SMC algorithm were also all less than 1.003. The MCMC algorithm fares well in general but has noticeable bias for $\rho = 0.9$ and $\rho = 1.2$, and has some additional misses for the other target distributions as well, which manifest as small protuberances in the quantile-quantile plot. The $\hat{R}$ values for the MCMC algorithm were all less than 1.004, indicating that the overall location and scale of the Republican dissimilarity were estimated well, but $\hat{R}$ is not designed to capture these smaller-scale deviations from the target distribution.

Here and in Section 5, we validated and compared the SMC and MCMC algorithms with several summary statistics. This reflects the applied use case for redistricting analysis. But it is also informative to study how well the SMC algorithm can target the actual distribution of plans themselves. Of course, we cannot expect the algorithm to perform well in this regard if there are fewer samples then there are plans, which is the case in almost all real-world problems. So we subset the enumerated plans to those with $\mathrm{dev}(\xi) \leq 0.01$, of which there are just 38. We first generate 10,000 samples from the SMC algorithm, targeting a distribution with $\rho = 1$.

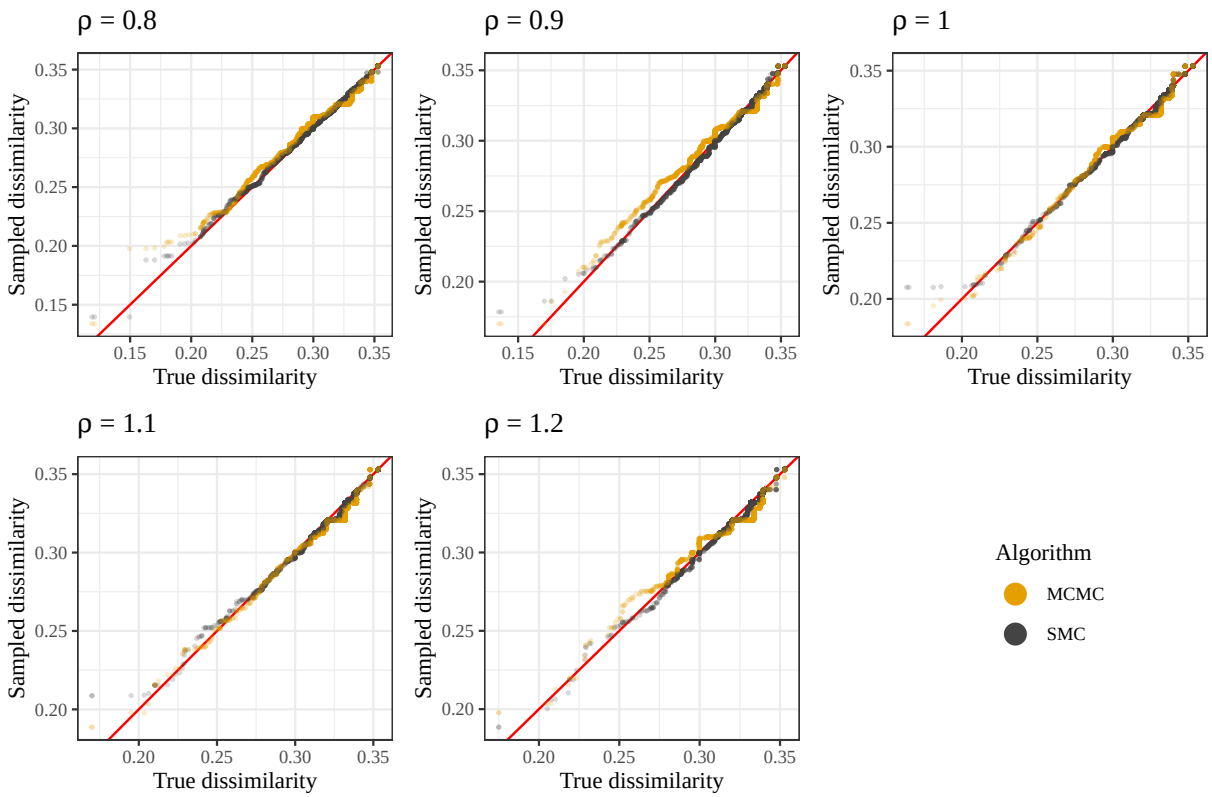


Figure B.2: Quantile-quantile plots for 1,500 MCMC and SMC samples of Republican dissimilarity across a range of target distributions with different compactness parameters ρ.

We measure the discrepancy between the sampled distribution of the 38 plans and the enumerated set weighted to $\tau(\xi)$ with the total variation distance, which is 0.0178 for this sample. Increasing the sample size to 50,000 decreases the total variation distance to 0.0149.

C Algorithm Implementation Details

C.1 Estimating K_i

A natural approach to estimating K_i is to draw a moderate number of spanning trees $\mathcal{T}_i \subseteq \mathcal{T}(\widetilde{G}_i)$ and compute $ok(T)$ for each $T \in \mathcal{T}_i$. The sample maximum, or the sample maximum plus some small buffer amount, would then be an estimate of the true maximum $\widehat{K}_i$ and an appropriate choice of k_i. In practice, we find little noticeable loss in algorithmic accuracy even if $k_i < K_i$. The following proposition theoretically justifies this finding. As above, d_e represents the population deviation of the district induced by removing edge e from a spanning tree.

Proposition 2. *The probability $q(e = e^* \mid \mathcal{F})$, i.e., the probability that an edge e is selected to be cut at iteration i, given that the tree T containing e has been drawn, and that e would induce a valid district, satisfies*

$$\max\left\{0, q(d_e \le d_{e_{k_i}} \mid \mathcal{F})\left(1 + \frac{1}{k_i}\right) - 1\right\} \le q(e = e^* \mid \mathcal{F}) \le \frac{1}{k_i},$$

where $\mathcal{F}$ is the σ-field generated by $\{T, \text{pop}(V_i) \in [P_i^-, P_i^+]\}$.

Proof. We can write

$$q(e = e^* \mid \mathcal{F}) \;=\; q(e = e^*, d_e \le d_{e_{k_i}} \mid \mathcal{F}) \;=\; \frac{1}{k_i} q(d_e \le d_{e_{k_i}} \mid \mathcal{F}),$$

This holds because the edge e will not be cut unless $d_e \leq d_{e_{k_i}}$, i.e., if e is among the top k_i edges. We then have immediately that $q(e = e^* \mid \mathcal{F}) \leq k_i^{-1}$. Additionally, using the lower Fréchet inequality, we find the lower bound

$$\begin{aligned} q(e = e^* \mid \mathcal{F}) &= q(e = e^*, d_e \leq d_{e_{k_i}} \mid \mathcal{F}) \\ &\geq \max\left\{0, q(e = e^* \mid \mathcal{F}) + q(d_e \leq d_{e_{k_i}} \mid \mathcal{F}) - 1\right\} \\ &= \max\left\{0, \frac{1}{k_i} q(d_e \leq d_{e_{k_i}} \mid \mathcal{F}) + q(d_e \leq d_{e_{k_i}} \mid \mathcal{F}) - 1\right\} \\ &= \max\left\{0, q(d_e \leq d_{e_{k_i}} \mid \mathcal{F})\left(1 + \frac{1}{k_i}\right) - 1\right\}. \end{aligned}$$ □

If $k_i \geq K_i$, then $q(e = e^* \mid \mathcal{F})$ is exactly k_i^{-1}, a fact which is used in the proof of Proposition 1. This result, which is proved using a simple Fréchet bound, shows that as long as $q(d_e \leq d_{e_{k_i}} \mid \mathcal{F})$ is close to 1, using k_i^{-1} in Proposition 1 is a good approximation to the true sampling probability.

Having sampled $\mathcal{T}_i$, we can compute for each value of k the sample proportion of trees where a randomly selected edge e among the top k of edges of the tree is also among the top k for the other trees—in effect estimating $q(d_e \leq d_{e_{k_i}} \mid \mathcal{F})$. We may then choose k_i to be the smallest k for which this proportion exceeds a pre-set threshold (e.g., 0.99). We have found that this procedure, repeated at the beginning of each sampling stage, efficiently selects k_i without compromising the ability to sample from the target distribution.

C.2 Calculating $\psi([\xi])$

To calculate $\psi([\xi])$, we first observe that sequentially valid labelings of a particular unlabeled plan are in bijective correspondence with certain increasing sequences of connected subgraphs of the quotient graph G/ξ. Specifically, as noted in the text, for every $1 \leq i \leq n-1$, the subgraph $A_i := \{i+1, i+2, \ldots, n\}$ of G/ξ is connected if and only if ξ is a sequentially valid labeling. Let $j = n - i$; then the sequence A_j is increasing. Going from A_j to A_{j+1} for any j involves adding a vertex in G/ξ which is adjacent to A_j. This fact provides an easy scheme to generate the number of sequentially valid labelings: for each vertex v in G/ξ, let $A_1 = \{v\}$. Then pick a neighbor of A_1 and add it to the set to form A_2. Continue in this fashion until A_{n-1} contains all but one vertex of G/ξ. Undoing our bijection, this remaining vertex is labeled 1; the vertex added between A_{n-2} and A_{n-1} is labeled 2, and so on.

Figure C.3 illustrates this scheme on district-level graph for the Pennsylvania court-imposed plan. We pick an arbitary vertex, denoted by "A" in Figure C.3(a). Then we pick a neighbor, denoted "B," and add it to the subgraph. We continue this way, adding the vertices indicated by the alphabetical ordering, until we have covered the whole district-level graph. Then the last vertex added, "R," is labeled 1 in the district labeling, as shown in Figure C.3(b). Veretx "Q" is labeled 2, and so on, until vertex "A" is labeled 18. One can easily check that this labeling is sequentially valid—at every point, the region of the map corresponding to unlabeled districts is contiguous.

As Section 4.4.2 mentioned, we adopt different strategies for calculating $\psi([\xi])$ when $n \leq 13$ and $n > 13$. When there are no more than 13 districts, we simply recurse down the tree of possibilities for generating all sequentially valid plans. In the example of Figure C.3, we have 18 options for the first vertex in the subgraph. Once we have picked vertex "A", then we could add "B" or "C". If we add "B", then we could add "C", "D", or "G".

Clearly this tree grows quite large very quickly. Fortunately, there are many duplicate nodes—in our example, adding vertex B and then C in sequence produces the same subgraph as if we had added C and then B. Thus by memoizing our counting function we can significantly reduce the number of tree branches we must explore in full.

When $n > 13$, it is no longer computationally feasible to perform the recursion for each sampled plan in what will often be a large number of sampled plans. In these cases (it should be noted that for the 2020 redistricting cycle, only nine states had more than 13 districts), we can only estimate the number of sequentially valid labelings. To do so, we generate a large number of random sequentially valid labelings by following the

(a) Increasing subgraphs (b) Sequentially valid labeling

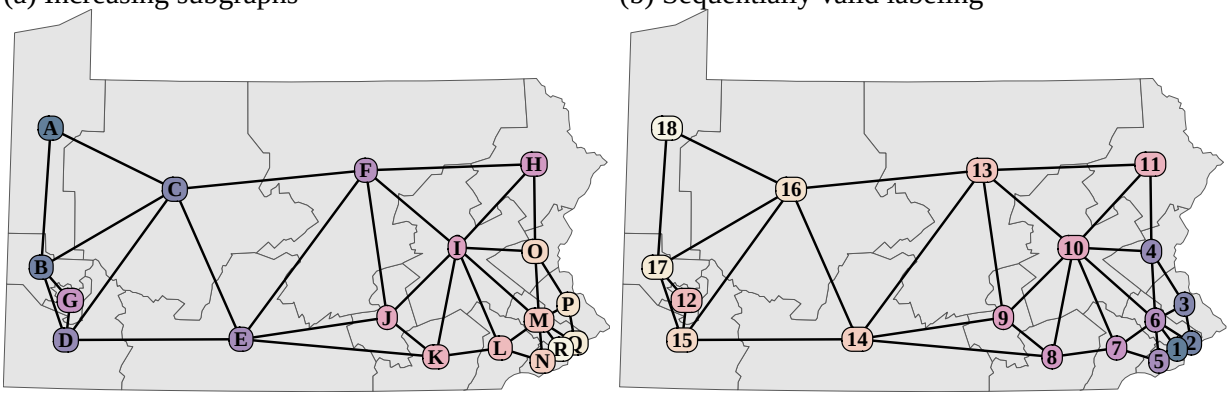


Figure C.3: Schematic of the process to generate a sequentially valid labeling, using the district-level graph for the Pennsylvania court-imposed plan. Subgraphs are built in alphabetical order in panel (a): $\{A\}, \{A, B\}, \ldots, \{A, B, \ldots, Q\}$. This corresponds to a sequentially valid labeling shown in panel (b).

scheme outlined above, picking vertices of G/ξ to add to each A_j uniformly at random. Since we observe the number of candidate vertices to add at each stage, and because each set of choices produces a unique sequentially valid labeling, we can compute the probability of drawing each sequentially valid labeling directly.

Denote the probability distribution created by this sampling scheme by p_{sv}, and the (not-probability) measure which assigns mass 1 to all $n!$ relabelings by μ_{all}. Our goal here is to estimate

$$\int \mathbf{1}\{\sigma \text{ is sequentially valid}\} d\mu_{\text{all}}(\sigma)$$

the total number of sequentially valid labelings. This can be done with our sample from p_{sv}, since it is supported on precisely the same set for which the indicator function takes a value of 1. We need only take the mean of the inverse of the probability of each sampled labeling.

This importance sampling estimate is quite accurate, since the proposal support matches the target support exactly, and there is not too much variation in the proposal probabilities. The estimate can be made arbitrarily good by increasing the number of importance samples. In our software, we use on the order of 1,500 samples for $n = 14$, and more when n is larger (roughly growing with $\sqrt{n}$).

Finally, we note that the procedures used here to calculate $\psi([\xi])$ can also be applied to partial plans. Notating this formally is more difficult, but a partial plan can be viewed as a (unbalanced) plan with fewer districts. Calculating ψ for these partial plans and incorporating them into the SMC weights in Algorithm 2 improves the sampling efficiency without changing the target distribution, as long as each partial ψ's contribution to the weights is canceled in the following iteration. This is the approach taken by our software implementation.

C.3 Stabilizing Importance Weights

When $\rho \neq 1$ or when the constraints imposed by J are severe, there can be substantial variance in the importance sampling weights. For large maps with $\rho = 0$, for instance, the weights will generally span hundreds if not thousands of orders of magnitude. This reflects the general computational difficulty in sampling uniformly from constrained graph partitions. As Najt et al. (2019) show, sampling of node-balanced graph partitions is computationally intractable in the worst case. In such cases, the importance sampling estimates will be highly variable, and resampling based on these weights may lead to degenerate samples with only one unique map.

When the importance weights are variable but not quite so extreme, we find it useful to truncate the normalized final weights (such that their mean is 1) from above at a value $w_{\max}$ at the end of sampling. The theoretical basis for this maneuver is provided by Ionides (2008), who proved that as long as $w_{\max} \to \infty$ and $w_{\max}/S \to 0$ as $S \to \infty$, the resulting estimates are consistent and have bounded variance (since the truncation occurs only after the final SMC step, these conclusions, which were made in the context of importance sampling,

carry over.) One such choice we have found to work well for the weights generated by this sampling process is $w_{\max} = S^{0.4}/100$, though for particular maps other choices of exponent and constant multiplier may be superior.

Truncation is no panacea, however. As with any method that relies on importance sampling, it is critical to examine the distribution of importance weights to ensure that they will yield acceptable resamples.

C.4 Computational Complexity

The two asymptotically slowest steps of the SMC algorithm are computing $\tau(G_i)$ for every district G_i and drawing a spanning tree using Wilson's algorithm for each iteration. All other steps, such as computing d_e and $|\mathcal{C}(G_i^{(j)}, \widetilde{G}_i^{(j)})|$, are linear in the number of vertices, and are repeated at most once per iteration.[4] Computing $\tau(G_i)$ requires computing a determinant, which currently has computational complexity $O(|V_i(\xi)|^{2.373})$ though most implementations are $O(|V_i(\xi)|^3)$. Since this must be done for each district of size roughly m/n, the total complexity for sampling one plan is $O(n \cdot (m/n)^{2.373})$. For the spanning trees, the expected runtime of Wilson's algorithm is the mean hitting time of the graph, which is $O(m^2)$ in the worst case. So the total complexity for each sample is $O(nm^2 + m^{2.373}n^{-1.373})$. Note that when $\rho = 1$, we need not compute $\tau(G_i)$, and the total complexity is $O(nm^2)$.

[4] To compute d_e, we walk depth-first over the tree and store, for each node, the total population of that node and the nodes below it. This allows for $O(1)$ computation of d_e for all edges.

Exhibit 6

STATISTICS AND PUBLIC POLICY
2020, VOL. 7, NO. 1, 52–68
https://doi.org/10.1080/2330443X.2020.1791773

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

The Essential Role of Empirical Validation in Legislative Redistricting Simulation

Benjamin Fifield[a], Kosuke Imai[a,b,c], Jun Kawahara[d], and Christopher T. Kenny[b]

[a]Institute for Quantitative Social Science, Harvard University, Cambridge, MA; [b]Department of Government, Harvard University, Cambridge, MA; [c]Department of Statistics, Harvard University, Cambridge, MA; [d]Graduate School of Infomatics, Kyoto University, Kyoto, Japan

ABSTRACT

As granular data about elections and voters become available, redistricting simulation methods are playing an increasingly important role when legislatures adopt redistricting plans and courts determine their legality. These simulation methods are designed to yield a representative sample of all redistricting plans that satisfy statutory guidelines and requirements such as contiguity, population parity, and compactness. A proposed redistricting plan can be considered gerrymandered if it constitutes an outlier relative to this sample according to partisan fairness metrics. Despite their growing use, an insufficient effort has been made to empirically validate the accuracy of the simulation methods. We apply a recently developed computational method that can efficiently enumerate all possible redistricting plans and yield an independent sample from this population. We show that this algorithm scales to a state with a couple of hundred geographical units. Finally, we empirically examine how existing simulation methods perform on realistic validation datasets.

ARTICLE HISTORY
Received October 2019
Accepted June 2020

KEYWORDS
Enumeration; Gerrymandering; Graph partition; Markov chain Monte Carlo; Redistricting; Zero-suppressed binary decision diagram

1. Introduction

Congressional redistricting, which refers to the practice of redrawing congressional district lines following the constitutionally mandated decennial census, is of major political consequence in the United States. Redistricting reshapes geographic boundaries and those changes can have substantial impacts on representation and governance in the American political system. As a fundamentally political process, redistricting has also been manipulated to fulfill partisan ends, and recent debates have raised possible reforms to lessen the role of politicians and the influence of political motives in determining the boundaries of these political communities.

Starting in the 1960s, scholars began proposing simulation-based approaches to make the redistricting process more transparent, objective, and unbiased (early proposals include Vickrey 1961; Weaver and Hess 1963; Hess et al. 1965; Nagel 1965). While this research agenda lay dormant for some time, recent advances in computing capability and methodologies, along with the increasing availability of granular data about voters and elections, has led to a resurgence in proposals, implementations, and applications of simulation methods to applied redistricting problems (e.g., Cirincione, Darling, and O'Rourke 2000; McCarty, Poole, and Rosenthal 2009; Altman and McDonald 2011; Chen and Rodden 2013; Fifield et al. 2014; Fifield, Higgins, et al. 2020; Mattingly and Vaughn 2014; Liu, Tam Cho, and Wang 2016; Herschlag, Ravier, and Mattingly 2017; Chikina, Frieze, and Pegden 2017; Magleby and Mosesson 2018; Carter et al. 2019; DeFord, Duchin, and Solomon 2019).

Furthermore, simulation methods for redistricting play an increasingly important role in court cases challenging redistricting plans. In 2019, simulation evidence was introduced and accepted in redistricting cases in North Carolina, Ohio, and Michigan.[1] In the few years prior, simulation methods were presented to courts in North Carolina, and Missouri.[2] Given these recent court cases challenging redistricting in state and federal courts, simulation methods are expected to become an even more influential source of evidence for legal challenges to redistricting plans across many states after the upcoming decennial census in 2020.

These simulation methods are designed to yield a representative sample of redistricting plans that satisfy statutory guidelines and requirements such as contiguity, population parity, and

CONTACT Kosuke Imai imai@harvard.edu Department of Government and Department of Statistics, Institute for Quantitative Social Science, Harvard University, Cambridge MA 02138.

[1]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).

[2]Expert Report of Jowei Chen, Raleigh Wake Citizens Assoc v. Wake County Board of Elections (2016); Expert Report of Jowei Chen, City of Greensboro v. Guilford County Board of Elections (2015); Supplemental Report of Jonathan Rodden and Jowei Chen: Assessment of Plaintiffs Redistricting Proposals, Missouri State Conference of the NAACP v. Ferguson-Florissant School District (2017).

© 2020 The Author(s). Published with license by Taylor and Francis Group, LLC
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.

compactness.[3] Then, a proposed redistricting plan can be considered gerrymandered if it constitutes an outlier relative to this sample according to a partisan fairness measure (see Katz, King, and Rosenblatt 2020, for a discussion of various measures). Simulation methods are particularly useful because enumeration of all possible redistricting plans in a state is often computationally infeasible. For example, even partitioning cells of an 8×8 checkerboard into two connected components generates over 1.2×10^{11} unique partitions (see *https://oeis.org/A068416*). Unfortunately, most redistricting problems are of much greater scale.[4] Therefore, to compare an implemented redistricting plan against a set of other candidate plans, researchers and policy makers must resort to simulation methods.

Despite the widespread use of redistricting simulation methods in court cases, insufficient efforts have been made to examine whether or not they actually yield a representative sample of all possible redistricting plans in practice.[5] Instead, some assume that the existing simulation methods work as intended. For example, in his amicus brief to the Supreme Court for *Rucho et al. v. Common Cause*, Eric Lander declares,[6]

> With modern computer technology, it is now straightforward to generate a large collection of redistricting plans that are representative of all possible plans that meet the State's declared goals (e.g., compactness and contiguity).

And yet, if there exists no scientific evidence that these simulation methods can actually yield a representative sample of valid redistricting plans, we cannot rule out the possibility that the comparison of a particular plan against sampled plans yields misleading conclusions about gerrymandering.

We argue that the empirical validation of simulation methods is essential for the credibility of academic scholarship and expert testimony in court. We apply the recently developed computational method of Kawahara et al. (2017), `enumpart`, that efficiently enumerates all possible redistricting plans and obtains an independent sample from this population (Section 2). The algorithm uses a compact data structure, called the zero-suppressed binary decision diagram (ZDD) (Minato 1993). In the aforementioned 8×8 checkerboard problem, explicitly storing every partition would require more than 1 terabyte of storage. In contrast, the ZDD needs only 1.5 megabytes. Our enumeration results are available as Fifield, Imai, et al. (2020) so that other researchers can use them to validate their own simulation methods. In addition, we will also make the code that implements the algorithm publicly available and incorporate it as part of an open-source R software package for redistricting, redist (Fifield, Tarr, and Imai 2015).

We begin by showing that the `enumpart` algorithm scales to a state with a couple of hundred geographical units, yielding realistic validation datasets (Section 3). We then test the empirical performance of existing simulation methods in two ways (Section 4). First, we randomly sample many submaps of various sizes from actual state shapefiles so that we average over the idiosyncratic geographic features and voter distributions of each map. For each sampled small map, we conduct a statistical test of the distributional equality between sampled and enumerated maps under various population parity constraints. If the simulation methods yield a representative sample of valid redistricting plans, then the distribution of the resulting p-values should be uniform. Second, we exploit the fact that even for a medium-sized redistricting problem, the `enumpart` algorithm can independently sample from the population of all valid redistricting plans. We then compare the resulting representative sample with the sample obtained using existing simulation methods. This second approach is applied to the actual redistricting problem in Iowa with 99 counties and a 250-precinct subset map from Florida, both of which are too computationally intensive for enumeration.

The overall conclusion of our empirical validation studies is that Markov chain Monte Carlo (MCMC) methods (e.g., Fifield et al. 2014; Fifield, Higgins, et al. 2020; Mattingly and Vaughn 2014; Carter et al. 2019) substantially outperform so-called random-seed-and-grow (RSG) algorithms (e.g., Cirincione, Darling, and O'Rourke 2000; Chen and Rodden 2013). These are two types of simulation methods that are most widely used in practice. Although the currently available MCMC methods are far from perfect and have much room for improvement, it is clear that the RSG algorithms are unreliable. Of course, showing that MCMC methods work reasonably well on these particular validation datasets does not necessarily imply that they will also perform well on other datasets especially larger scale redistricting problems. Rather, failing these validation tests on small and medium-scale redistricting problems provides evidence that RSG methods are most likely to perform poorly when applied to other larger states.

To the best of our knowledge, the only publicly available validation dataset for redistricting is the 25-precinct map obtained from Florida, for which Fifield et al. (2014) and Fifield, Higgins, et al. (2020) enumerated all possible redistricting maps for two or three contiguous districts. Other researchers have used this validation data or enumeration method to evaluate their own algorithms (e.g., Magleby and Mosesson 2018; Carter et al. 2019). However, this dataset is small and represents only a particular set of precincts representing a specific political geography, and may not be representative of other redistricting problems. For example, as noted by Magleby and Mosesson (2018), this dataset is not particularly balanced—only eight partitions fall within standard levels of population parity ($\pm 1.5\%$), and most fall above 10%. Our new validation datasets are much larger and hence provide unique opportunities to conduct a more realistic empirical evaluation of simulation methods.

2. The Methodology

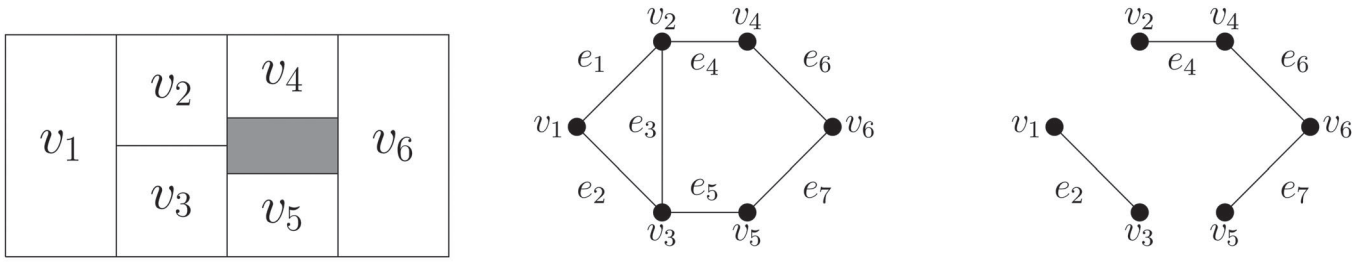
In this section, we describe the enumeration and sampling methods used in our empirical validation studies. Our methods

[3] The outlier detection method proposed by Chikina, Frieze, and Pegden (2017) is a statistical test and its goal is not uniform sampling. However, the proposed enumeration method can still be useful for assessing its empirical performance.

[4] While statutory guidelines and requirements such as district contiguity, population parity, and compactness reduce the number of partitions dramatically, the resulting problem currently remains out-of-reach of full enumeration methods.

[5] For an exception, see, for example, Carter et al. (2019), Jonathan C. Mattingly. "Rebuttal of Defendant's Expert Reports for *Common Cause v. Lewis*." Andrew Chin, Gregory Herschlag, and Jonathan C. Mattingly. "The Signature of Gerrymandering in *Rucho v. Common Cause*," pp. 1261–1262.

[6] Brief for Amicus Curiae Eric S. Lander. In Support of Appellees, p. 4, *Rucho et al. v. Common Cause*, No. 18-422. March 7, 2019, p. 4.



(a) Original map (b) Graph representation (c) Induced subgraph

Figure 1. A running redistricting example. We consider dividing a state with six geographical units into two districts. The original map is shown in the left panel where the shaded area is uninhabited. The middle panel shows its graph representation, whereas the right panel shows an example of redistricting map represented by an induced subgraph, which consists of a subset of edges.

are based on the `enumpart` algorithm originally developed by Kawahara et al. (2017) who showed how to enumerate all possible redistricting plans and store them using a compact data structure, called a ZDD (Minato 1993). We also show how the `enumpart` algorithm can be used to independently sample from the population of contiguous redistricting plans.

2.1. The Setup

Following the literature (see, e.g., Altman 1997; Mehrotra, Johnson, and Nemhauser 1998; Fifield et al. 2014), we formulate redistricting as a graph-partitioning problem. Given a map of a state, each precinct (or any other geographical units used for redistricting) is represented by a vertex, whereas the existence of an edge between two vertices implies that they are geographically contiguous to one another. Formally, let $G = (V, E)$ represent a graph with the vertex set $V = \{v_1, \ldots, v_n\}$ and the edge set $E = \{e_1, \ldots, e_m\}$. We consider redistricting of a state into a total of p districts where all precincts of each district are connected. This is equivalent to partitioning a graph G into p connected components $\{V_1, V_2, \ldots, V_p\}$ such that every vertex in V belongs to exactly one connected component, that is, $V_1 \cup \cdots \cup V_p = V$, $V_k \cap V_{k'} = \emptyset$ for any $k \neq k'$ and all the vertices in V_k are connected.

We use the fact that a p-graph partition can alternatively be represented as an edge set S. That is, by removing certain edges from E, we can partition G into p connected components. Formally, for each connected component V_k, we define an induced subgraph $(V_k, S(V_k))$ as a graph whose edge set consists of all edges whose two endpoints (i.e., the two vertices directly connected by the edge) belong to V_k. Then, the p-graph partition can be defined as the union of these induced subgraphs, that is, $\mathcal{P} = \bigcup_{k=1}^{p} S(V_k)$ where $S(V_k) \cap S(V_{k'}) = \emptyset$ for any $k \neq k'$. Our initial task is to enumerate *all* possible p-graph partitions of G.

Figure 1(a) presents the running example used throughout this section to illustrate our methodology. In this hypothetical state, we have a total of six precincts, represented as vertices $\{v_1, v_2, \ldots, v_6\}$, which we hope to divide into two districts, $\{V_1, V_2\}$. A gray area is uninhabited (e.g., lake). This map can be represented as a graph of Figure 1(b) where two contiguous vertices share an edge. Consider a redistricting map with $V_1 = \{v_1, v_3\}$ and $V_2 = \{v_2, v_4, v_5, v_6\}$. As shown in Figure 1(c), this redistricting map can be represented by an induced subgraph after removing three edges, that is, $\{e_1, e_3, e_5\}$. Thus, we can represent each district as an induced subgraph, which is a set of edges, that is, $S(V_1) = \{e_2\}$ or $S(V_2) = \{e_4, e_6, e_7\}$.

2.2. Graph Partitions and Zero-Suppressed Binary Decision Diagram (ZDD)

A major challenge for enumerating redistricting maps is memory management because the total number of possible maps increases exponentially. We use the ZDD, which uses a compact data structure to efficiently represent a family of sets (Minato 1993). We first discuss how the ZDD can represent a family of graph partitions before explaining how we construct the ZDD from a given graph.

The ZDD that corresponds to the running example of Figure 1 is given in Figure 2. A ZDD is a directed acyclic graph. As is clear from the figure, each edge of the original graph corresponds to possibly multiple nodes of a ZDD. To avoid confusing terminology, we use a "node" rather than a "vertex" to refer to a unit of ZDD, which represents an edge of the original graph. Similarly, we call an edge of the ZDD an "arc" to distinguish it from an edge of the original graph. There are three special nodes in a ZDD. The *root node*, labeled as e_1 in our example, has no incoming arc but, like other nodes, represents one of the edges in the original graph. We will discuss later how we label nodes. ZDD also has two types of *terminal nodes* without an outgoing arc, called 0-terminal and 1-terminal nodes and represented by $\boxed{0}$ and $\boxed{1}$, respectively. Unlike other nodes, these terminal nodes do not correspond to any edge in the original graph. Finally, each nonterminal node, including the root node, has exactly two outgoing arcs, *0-arc* (dashed arrow) and *1-arc* (solid arrow).

Given a ZDD, we can represent a graph partition as the set of edges that belong to a directed path from the root node to 1-terminal node and have an outgoing 1-arc. For example, the path highlighted by blue, $e_1 \dashrightarrow e_2 \longrightarrow e_3 \dashrightarrow e_4 \longrightarrow e_5 \dashrightarrow e_6 \longrightarrow e_7 \longrightarrow \boxed{1}$, represents the edge set $\{e_2, e_4, e_6, e_7\}$, which corresponds to the 2-graph partition shown in Figure 1(c). Indeed, there is a one-to-one correspondence between a graph partition and a path of a ZDD.

2.3. Construction of the ZDD

How should we construct a ZDD for a p-graph partition from a given graph? We use the frontier-based search algorithm

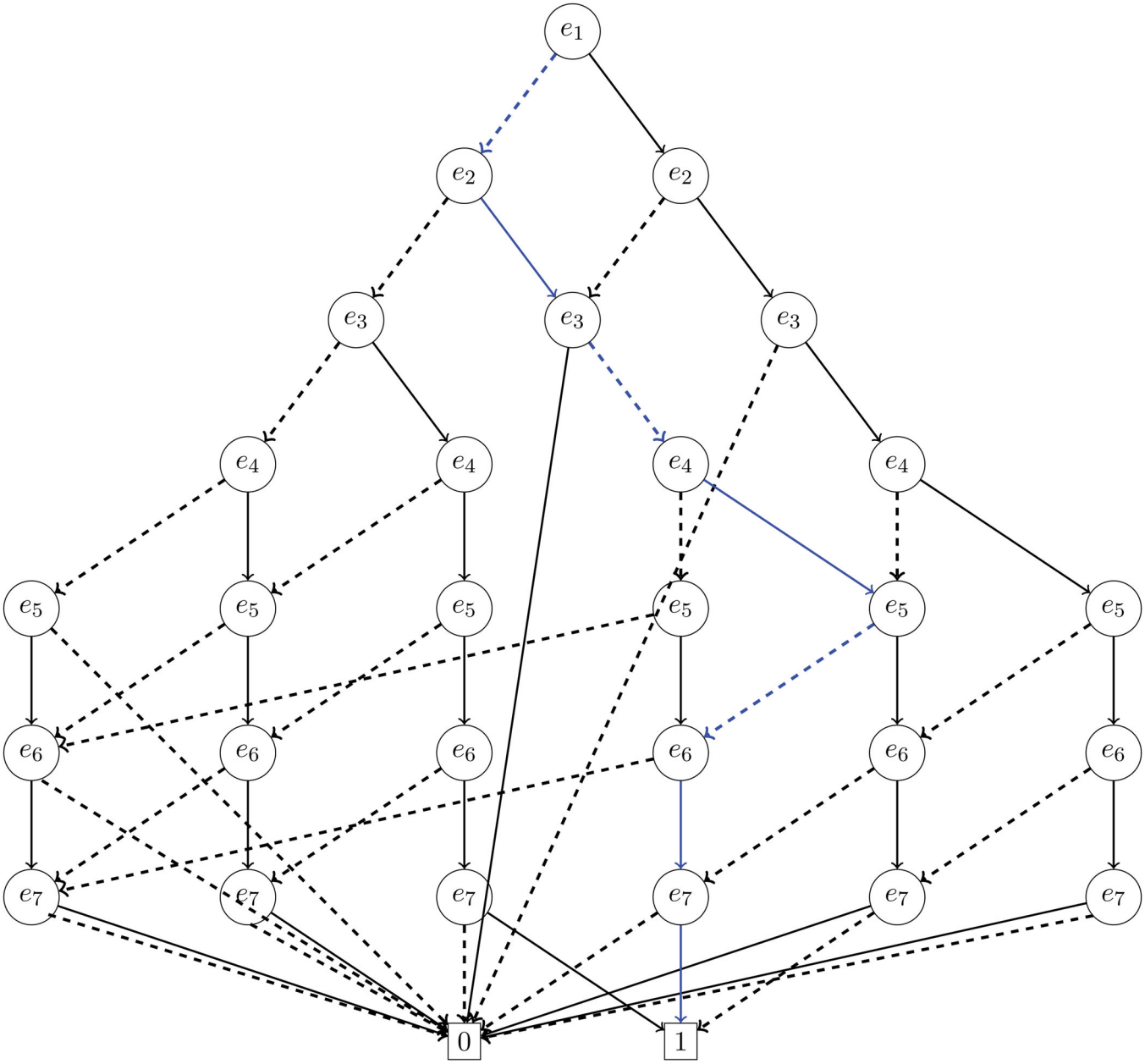


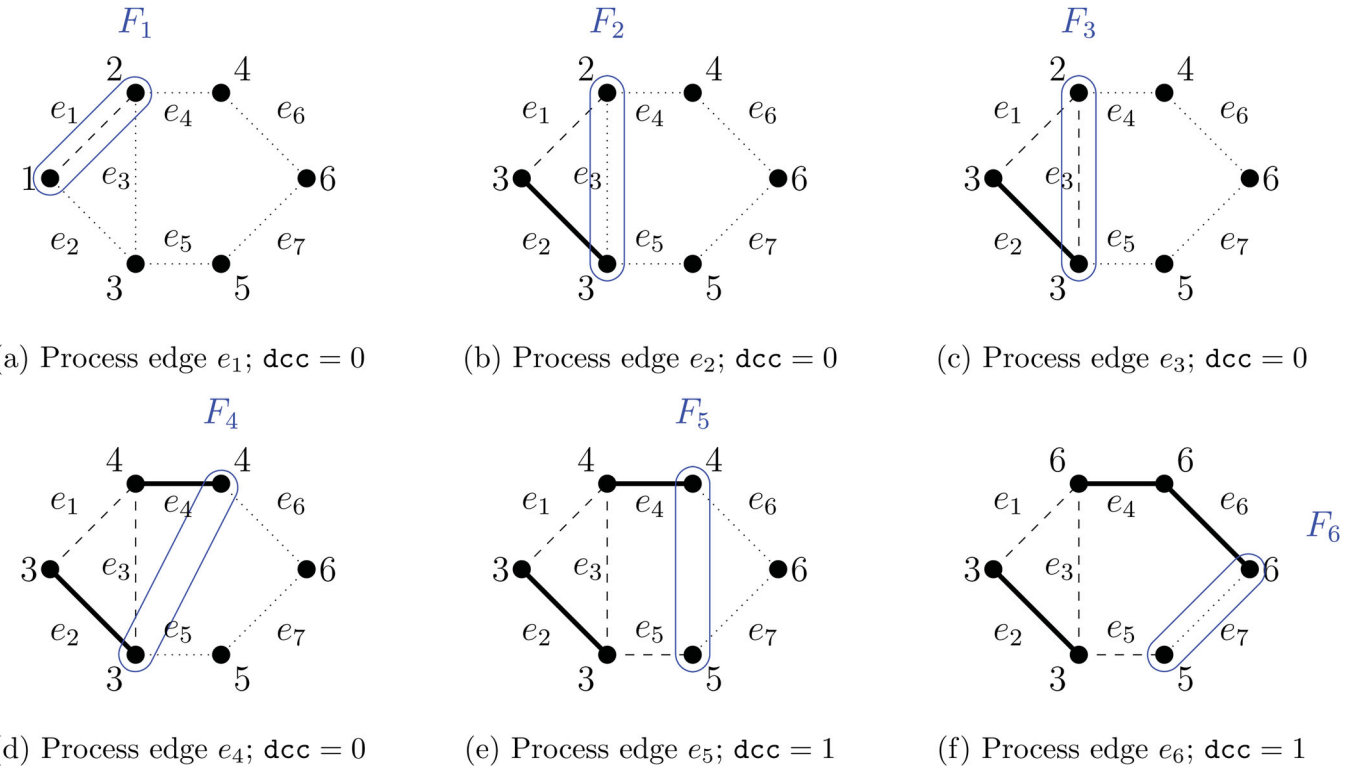
Figure 2. Zero-suppressed binary decision diagram (ZDD) for the running example of Figure 1(b). The blue path corresponds to the redistricting map represented by the induced subgraph in Figure 1(c).

proposed by Kawahara et al. (2017). The algorithm grows a tree starting with the root node in a specific manner. We first discuss how to construct a ZDD given m labeled edges, $\{e_1, \ldots, e_m\}$, where e_1 represents the root node. We then explain how we merge nodes to reduce the size of the resulting ZDD and how we label edges given a graph to be partitioned so that the computation is efficient.

2.3.1. The Preliminaries

Starting with the root node $i = 1$, we first create one outgoing 0-arc and one outgoing 1-arc from the corresponding node e_i to the next node e_{i+1}. To ensure that each enumerated partitioning has exactly p connected components, we store the number of *determined connected components* as the `dcc` variable for each ZDD node. Consider a directed path $e_1 \longrightarrow e_2 \dashrightarrow e_3 \dashrightarrow e_4 \dashrightarrow e_5$. In this example, e_1 is retained whereas edges $\{e_2, e_3, e_4\}$ are not. We know that the two vertices, $\{v_1, v_2\}$, together form one district, regardless of whether or not e_5 is retained. Then, we say that a connected component is determined and set `dcc` to 1 for e_5. If `dcc` exceeds p, then we create an arc into the 0-terminal node rather than create an arc into the next node since there is no longer a prospect of constructing a valid partition. Similarly, when creating an arc out of the final node, e_m, we point the arc into the 0-terminal node if `dcc` is less than p, which represents the total number of partitions. Finally, if the number of remaining edges exceeds $p - \texttt{dcc}$, we stop growing the path by creating an outgoing arc into the 0-terminal node.

How do we find out when another connected component is determined so that `dcc` needs to be increased? To do this, we need two new variables. First, for each vertex v_i, we store the *connected component number*, denoted by $\texttt{comp}[v_i]$, indicating



(a) Process edge e_1; dcc $= 0$ (b) Process edge e_2; dcc $= 0$ (c) Process edge e_3; dcc $= 0$

(d) Process edge e_4; dcc $= 0$ (e) Process edge e_5; dcc $= 1$ (f) Process edge e_6; dcc $= 1$

Figure 3. Calculation of the frontier, the connected component number, and the determined connected components. This illustrative example is based on the redistricting problem shown in Figure 1. A positive integer placed next to each vertex represents the connected component number, whereas the vertices grouped by the solid blue line represent a frontier. A connected component is determined when processing edge e_5.

the connected component to which v_i belongs. Thus, two vertices, v_i and $v_{i'}$, share an identical connected component number if and only if they belong to the same connected component, that is, $\texttt{comp}[v_i] = \texttt{comp}[v_{i'}]$.

We initialize the connected component number as $\texttt{comp}[v_i] \leftarrow i$ for $i = 1, 2, \ldots, n$ where n is the number of vertices in the original graph. Suppose that we process and retain an edge incident to two vertices v_i and $v_{i'}$ for $i \neq i'$ by creating an outgoing 1-arc. Then, we set $\texttt{comp}[v_j] \leftarrow \max\{\texttt{comp}[v_i], \texttt{comp}[v_{i'}]\}$ for any vertex v_j whose current connected component number is given by $\texttt{comp}[v_j] = \min\{\texttt{comp}[v_i], \texttt{comp}[v_{i'}]\}$. This operation ensures that all vertices that are connected to v_i or $v_{i'}$ have the same connected component number (larger of the two original numbers).

2.3.2. The Frontier-Based Search

Next, we define the *frontier* of a graph, which changes as we process each edge and grow a tree. Suppose that we have created a directed path by processing the nodes from e_1 up to e_ℓ where $\ell = 2, 3, \ldots, m-1$. For each $\ell = 1, 2, \ldots, m-1$, the ℓth frontier F_ℓ represents the set of vertices of the original graph that are incident to both a processed edge (i.e., at least one of $e_1, e_2, \ldots, e_\ell$) and an unprocessed edge (i.e., at least one of $e_{\ell+1}, e_{\ell+2}, \ldots, e_m$). Note that we define $F_0 = F_m = \emptyset$ and that the set of processed edges includes the one currently being processed. Thus, for a given graph, the frontier only depends on which edge is being processed but does not hinge on how edges have been or will be processed. That is, the same frontier results for each node regardless of paths.

The frontier can be used to check whether a connected component is determined. Specifically, suppose there exists a vertex v that belongs to the previous frontier but is not part of the current one, that is, $v \in F_{\ell-1}$ and $v \notin F_\ell$. Then, if there is no other vertex in F_ℓ that has the same connected component number as v (i.e., no vertex in F_ℓ is connected to v), there will not be another vertex in subsequent frontiers, that is, $F_{\ell+1}, \ldots, F_m$, that are connected to v. Thus, under this condition, the connected component $\texttt{comp}[v]$ is determined, and we increment dcc by one.

Figure 3 gives an example of computing the connected component number, constructing the frontier, and updating the determined connected components, based on the redistricting problem shown in Figure 1. In each graph, a positive integer placed next to a vertex represents its connected component number, whereas the vertices grouped by the solid line represent a frontier. For example, when processing edge e_5 (see Figure 3(e)), we have $F_4 = \{v_3, v_4\}$ and $F_5 = \{v_4, v_5\}$. Since there is no vertex in F_5 that shares the same connected component number as v_3 (which is 1), we can determine the first connected component and increment dcc by one.

Finally, when processing the last edge e_m represented by node n^*, if two vertices incident to the edge belong to the same connected component number, then the 0-arc from node n^* points to the 0-terminal node whereas the destination of the 1-arc is the 1-terminal node unless $\texttt{dcc} \neq p$. If they have different connected component numbers, the 0-arc of node n^* goes to the 1-terminal node whereas the destination of its 1-arc is the 1-terminal node so long as $\texttt{dcc} = p$ and the induced subgraph condition described in the next paragraph is satisfied (otherwise, it is the 0-terminal node).

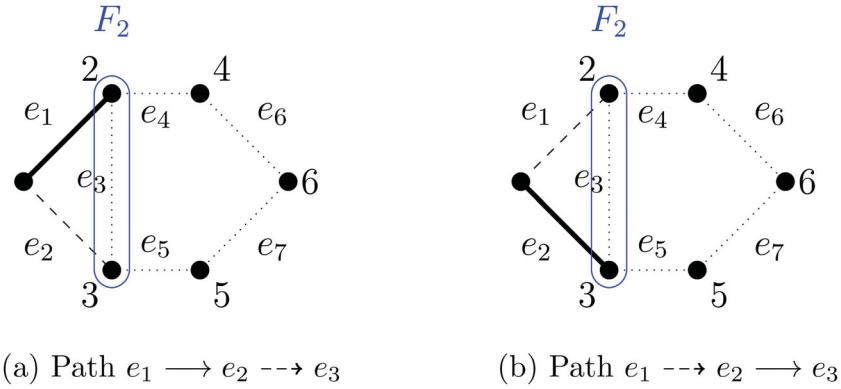


Figure 4. An example of node merging. as shown in Figure 2, these two paths merge at e_3 because the connected component numbers in F_2 are identical and the number of determined connected components is zero.

Throughout the process of building a ZDD, we must make sure that every path actually corresponds to an induced subgraph, which is defined as a subset of nodes and all arcs connecting pairs of such nodes. We call this the induced subgraph condition. Consider a path, $e_1 \longrightarrow e_2 \longrightarrow e_3$. Since three vertices, $\{v_1, v_2, v_3\}$, are connected, we must retain edge e_3 because its two incident vertices, v_2 and v_3, are connected. Thus, we have $e_1 \longrightarrow e_2 \longrightarrow e_3 \dashrightarrow \boxed{0}$. Similarly, consider a path, $e_1 \longrightarrow e_2 \dashrightarrow e_3$. We cannot retain e_3 because e_2, which is incident to v_1 and v_3, is not retained. This yields $e_1 \longrightarrow e_2 \dashrightarrow e_3 \longrightarrow \boxed{0}$.

To impose the induced subgraph condition, we introduce the *forbidden pair set* for each node. Once we decide not to use an edge that connects two distinct components, the two components must not be connected any more. Otherwise, the new component generated by connecting the two components has an unused edge, violating the induced subgraph condition. Therefore, if we determine that an edge $\{v, v'\}$ is not used, the addition of $\{\texttt{comp}[v], \texttt{comp}[v']\}$ to the forbidden pair set reminds us that the components $\texttt{comp}[v]$ and $\texttt{comp}[v']$ must not be connected. That is, if we use an edge $\{u, u'\}$ and the forbidden pair set contains $\{\texttt{comp}[u], \texttt{comp}[u']\}$, the path will be sent to the 0-terminal. In the above example, if we pass through $e_1 \longrightarrow e_2 \dashrightarrow e_3$, $\{2, 3\}$ is added to the forbidden pair set, where 2 is the component number of $\{v_1, v_2\}$ and 3 is that of $\{v_3\}$. Then, since retaining e_3 violates the induced subgraph condition, we have $e_1 \longrightarrow e_2 \dashrightarrow e_3 \longrightarrow \boxed{0}$.

2.3.3. Node Merge

The above operation implies that when processing e_ℓ, the only required information is the connectivity of vertices in $F_{\ell-1}$. We can reduce the size of the ZDD by exploiting this fact. First, we can avoid repeating the same computation by merging multiple nodes if the connected component numbers of all vertices in $F_{\ell-1}$ and the number of determined connected components $\texttt{dcc}$ are identical. This is a key property of the ZDD, which allows us to efficiently enumerate all possible redistricting plans by merging many different paths. Second, we only need to examine the connectivity of vertices within a frontier to decide whether or not any connected component is determined. Thus, we adjust the connected component number so that it equals the maximum vertex number in the frontier. That is, if some vertices in the frontier share the same connected component number, we change it to the maximum vertex index among those vertices. For example, in Figure 3(b), we set $\texttt{comp}[v_2] = 2$ and $\texttt{comp}[v_3] = 3$. We need not worry about how the renumbering of $\texttt{comp}[v_3]$ affects the value of $\texttt{comp}[v_1]$ because $v_1 \notin F_2$. This operation results in merging of additional nodes, reducing the overall size of the ZDD.

Figure 4 gives an example of such a merge. Figure 4(a) corresponds to the path, $e_1 \longrightarrow e_2 \dashrightarrow e_3$ whereas Figure 4(b) represents the path, $e_1 \dashrightarrow e_2 \longrightarrow e_3$. As shown in Figure 2, these two paths are merged at e_3 because the connected component numbers in their frontier F_2 are identical and both have the same number of determined connected component, that is, $\texttt{dcc} = 0$. Note that in Figure 4(b) the connected component number is normalized within the frontier F_2 such that the connected component number of v_2 is the maximum vertex index, that is, 2, and that of v_3 is 3.

Node merging plays a key role in scaling up the enumeration algorithm. Although we can construct the ZDD that only enumerates graph partitions by storing the sum of population values into each node (see Kawahara et al. 2017, sec. 4), this prevents nodes from being merged, dramatically reducing the scalability of the enumeration algorithm. Therefore, we do not take this approach here.

2.3.4. Edge Ordering

How should we label the edges of the original graph? The amount of computation depends on the number of nodes in the ZDD. Recall that two nodes are merged if the stored values such as $\texttt{comp}$ and $\texttt{dcc}$ are identical. Since a ZDD node stores the $\texttt{comp}$ value for each vertex in the frontier, the number of unique stored values grows exponentially as the frontier size increases (see Section 3.1 in Kawahara et al. (2017) for the detailed analysis). Therefore, we wish to label the edges of a graph such that the maximum size of the frontier is minimized. We take a heuristic approach here. Specifically, we first choose two vertices s, t such that the shortest distance between s and t is as large as possible across all vertex pairs. We use the Floyd–Warshall algorithm, which can find the shortest paths between all vertex pairs in $\mathcal{O}(|V|^3)$ where $|V|$ is the number of vertices of a graph. Next, we compute the minimum s–t vertex graph cut, which is the minimum set of vertices whose removal generates two or more connected components. To do this, we use a max-flow based algorithm, and arbitrarily order

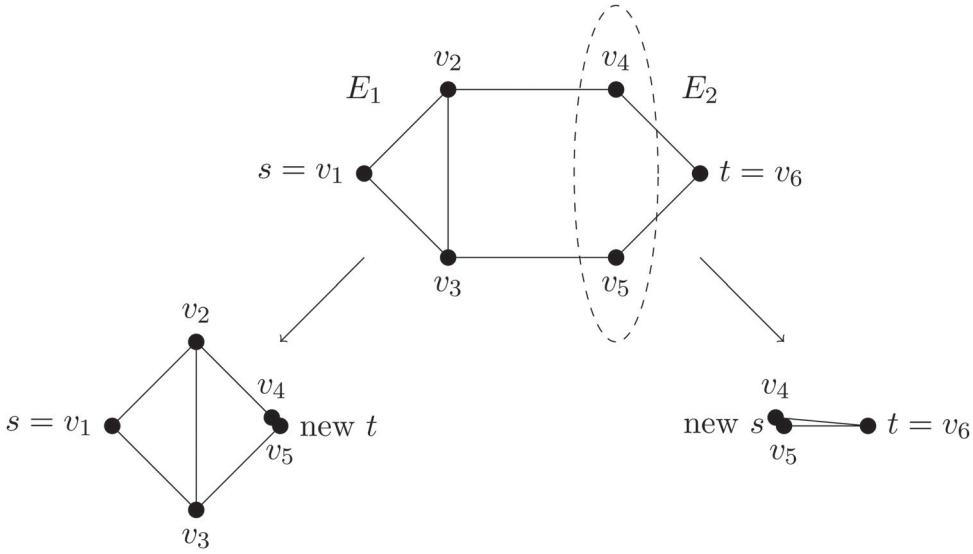


Figure 5. An example of edge ordering by vertex cuts. To order edges, we choose two vertices with the maximum shortest distance and call them *s* and *t*. We then use the minimum vertex cut, indicated by the dashed oval, to create two or more connected components, which are arbitrarily ordered. The same procedure is then applied to each connected component until the resulting connected components are sufficiently small.

the resulting connected components. Finally, we recursively apply this procedure to each connected component until the resulting connected components are sufficiently small (e.g., 5 edges), at which point they are ordered in an arbitrary fashion.

Figure 5 illustrates this process. In this example, a pair, $s = v_1$ and $t = v_6$, gives the maximum shortest distance. Given this choice, there are four minimum s–t graph cuts whose size is 2, that is, $\{v_2, v_3\}, \{v_2, v_5\}, \{v_3, v_4\}, \{v_4, v_5\}$. We arbitrarily select one of them and call it S. Suppose we set $S = \{v_4, v_5\}$. Then, this yields two connected components, that is, $C_1 = \{v_1, v_2, v_3\}$ and $C_2 = \{v_6\}$. For each connected component C_i, let E_i represent the set of edges in C_i and between C_i and S. In the current example, $E_1 = \{\{v_1, v_2\}, \{v_1, v_3\}, \{v_2, v_3\}, \{v_2, v_4\}, \{v_3, v_5\}\}$ and $E_2 = \{\{v_4, v_6\}, \{v_5, v_7\}\}$. We order these edge sets so that all the edges E_1 will be placed before those of E_2. To continue this process recursively, we combine all the vertices in S into a single vertex and let this new vertex be t in E_1 and s in E_2. Now, we can apply the same procedure separately to E_1 and E_2: computing the minimum s–t vertex cut and splitting the graph into two (or more) components.

The reason why we expect the above edge ordering procedure to produce a small frontier is that each vertex cut in the process equals one of the frontiers of the corresponding ZDD. In our example, the first vertex cut S is equal to $F_5 = \{v_4, v_5\}$. Since we choose minimum vertex cuts in each step, we expect the input graph with the edge order obtained through this procedure to have small frontiers.

2.4. Enumeration and Independent Sampling

It can be shown that every path from the root node to the 1-terminal node in the resulting ZDD has a one-to-one correspondence to a p-graph partition. This is because each p-graph partition can be uniquely represented by the union of induced subgraphs, which in turn corresponds to a unique path from the root node to the 1-terminal node. The complexity of the `enumpart` algorithm is generally difficult to characterize, but Kawahara et al. (2017) analyzed it in the case of planar graphs. Thus, once we obtain the ZDD as described above, we can quickly enumerate all the paths from the root node to the 1-terminal node. Specifically, we start with the 1-terminal node and then proceed upward to the root node, yielding a unique graph partition.

In addition to enumeration, we can also independently sample p-graph partitions (Knuth 2011). First, for each node ν of the ZDD, we compute the number of paths to the 1-terminal node. Let $c(\nu)$ be the number of such paths, and ν_0 and ν_1 be the nodes pointed by the 0-arc and 1-arc of ν, respectively. Clearly, we have $c(\nu) = c(\nu_0) + c(\nu_1)$. The values of c for the 0-terminal and 1-terminal nodes are 0 and 1, respectively. As done for enumeration, we compute and store the value of c for each node by moving upward from the terminal node to the root node. Finally, we conduct random sampling by starting with the root node and choosing node ν_1 with probability $c(\nu_1)/\{c(\nu_0) + c(\nu_1)\}$ until we reach the 1-terminal node. Since the probability of reaching the 0-terminal node is zero, we will always arrive at the 1-terminal node, implying that we obtain a path corresponding to a p-graph partition. Repeating this procedure will yield the desired number of independently and uniformly sampled p-graph partitions.

The reason why this procedure samples redistricting maps uniformly is that a path from the root node to the 1-terminal node corresponds to a unique p-graph partition. Since each node ν stores the number of paths from the node to the 1-terminal node as $c(\nu)$, the sampling procedure uniformly and randomly selects one path among $c(\nu_1)$ paths where ν_1 is the root node. If nonuniform sampling is desired, one could apply the sampling-importance resampling algorithm to obtain a representative sample from a target distribution (Rubin 1987).

3. Empirical Scalability Analysis

This section analyzes the scalability of the `enumpart` algorithm described above, and shows that the algorithm scales to

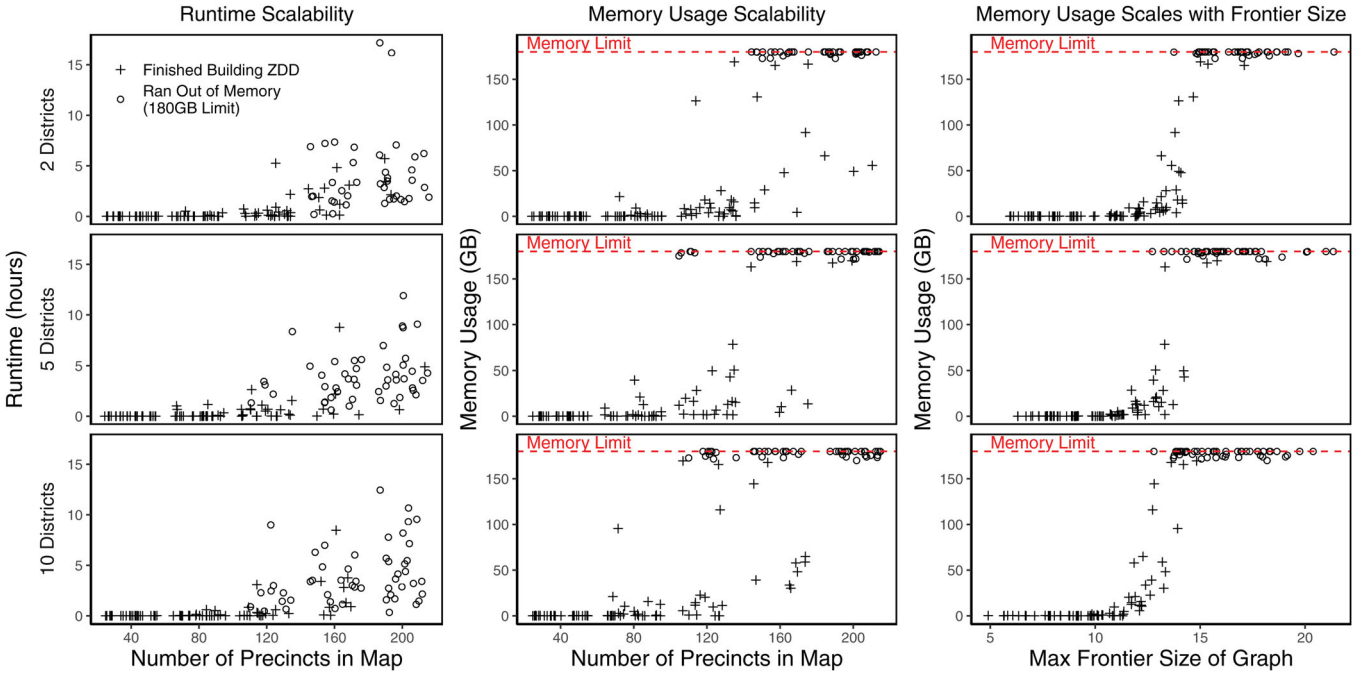


Figure 6. The scalability of the `enumpart` algorithm on subsets of the New Hampshire precinct map. This figure shows the runtime scalability of the `enumpart` algorithm for building the ZDD on random contiguous subsets of the New Hampshire precinct map. Crosses indicate maps where the ZDD was successfully built within the RAM limit of 180GB. In contrast, open circles represent maps where the algorithm ran out of memory. For the left and middle columns, the results are jittered horizontally with a width of 20 for the clarity of presentation. (The actual evaluation points on the horizontal axis are 40, 80, 120, 160, and 200.) The left column shows how total runtime increases with the number of units in the underlying map, while the center column shows how the total RAM usage increases with the number of units in the underlying map. Lastly, the right-hand column shows that memory usage is primarily a function of the maximum frontier size of the ZDD. We show results for 2-district partitions (top row), five-district partitions (middle row), and 10-district partitions (bottom row).

enumerate partitions of maps many times larger than existing enumeration procedures. We analyze the algorithm's scalability in terms of runtime and memory usage, and show how the memory usage of `enumpart` is closely tied to the frontier of the corresponding ZDD as explained in the previous section.

To make this empirical analysis realistic, we use independently constructed and contiguous subsets of the 2008 New Hampshire precinct map for maps ranging between 40 precincts and 200 precincts, increasing by 40, that is, $\{40, 80, 120, 160, 200\}$. The original New Hampshire map consists of 327 precincts, which are divided into two congressional districts. To generate an independent contiguous subset of the map, we first randomly sample a precinct, and add its adjacent precincts to a queue. We then repeatedly sample additional precincts from the queue to be added to the subset map, and add the neighbors of the sampled precincts to the queue, until the map reaches the specified size. We repeat this process until the subsetted map reaches a prespecified size.

We consider partitioning each of these maps into two, five, or ten districts and apply `enumpart` to each case. We then compute the time and memory usage of generating a ZDD for each application. For each precinct size and number of districts, we repeat the above sampling procedure 25 times, producing 25 independent and contiguous subsets of the New Hampshire map. All trials were run on a Linux computing cluster with 530 nodes and 48 Intel Cascade Lake cores per node, where each node has 180GB of RAM. Note that we do not save the results of enumeration to disk as doing so for every trial is computationally too expensive. This means that we cannot conduct an in-depth analysis of the characteristics of all enumerated maps.

Figure 6 shows the results of our scalability analysis. The top row shows scalability results for generating a ZDD for partitions of the map into two districts, while the middle row shows the results with five districts and the bottom row shows results with ten districts. Each dot represents a run of the `enumpart` algorithm on a subset of the New Hampshire map. Crosses represent trials where the ZDD successfully built using under the 180GB RAM limit. In contrast, open circles show trials that were unable to build the ZDD with the same RAM limit. Note that for the left and middle columns, the results are jittered horizontally with a width of 20 for the clarity of presentation.

The left-hand and center columns show how the `enumpart` algorithm scales in terms of runtime and memory usage, respectively, as the number of precincts in the underlying map increases. For small maps ranging from 25 precincts to 80 precincts, runtime and memory usage are for the most part negligible. The ZDD for two-district, five-district, and ten-district partitions for these small maps can be constructed in nearly all cases in under two minutes, and using less than one gigabyte of RAM. As the number of precincts in the map starts to increase, so do the runtime and memory usage requirements. For maps of 200 precincts, over 90% of the tested maps hit the 180GB memory limit before building the complete ZDD. For all map sizes, we also note that the runtime and memory usage requirements for building the ZDD do not appear to depend much on the number of districts that the map is being partitioned into.

What drives these patterns in scalability? While the number of units in the underlying map predicts both runtime and memory usage, there is still a great deal of variability even

conditional on the number of precincts in the map. In the right-hand column, we show that the memory usage requirements for building the ZDD are closely tied to the maximum frontier size of the underlying map, as defined in Section 2.3.2. While the memory usage is minimal so long as the maximum frontier size of the graph is under 11, memory usage increases quickly once the maximum frontier size grows beyond that. This suggests that improved routines for reducing the size of a map's frontier can allow for the enumeration of increasingly large maps.

4. Empirical Validation Studies

In this section, we introduce a set of new validation tests and datasets that can be used to evaluate the performance of redistricting simulation methods. We focus on the two most popular types of simulation methods that are implemented as part of the open-source software package redist (Fifield, Tarr, and Imai 2015): one based on the MCMC algorithm (Fifield et al. 2014; Fifield, Higgins, et al. 2020; Mattingly and Vaughn 2014) and the other based on the RSG algorithm (Cirincione, Darling, and O'Rourke 2000; Chen and Rodden 2013). Below, we conduct empirical validation studies both through full enumeration and independent sampling. For the sake of simplicity, we use the uniform distribution over all valid redistricting maps under various constraints. However, one could also use a nonuniform target distribution by appropriately weighting each map.

4.1. Validation Through Enumeration

We conduct two types of validation tests using enumeration. We first use the `enumpart` algorithm to enumerate all possible redistricting plans using a map with 70 precincts, which is much larger than the existing validation map with 25 precincts analyzed in Fifield et al. (2014) and Fifield, Higgins, et al. (2020). We then compare the sampled redistricting plans obtained from simulation methods against the ground truth based on the enumerated plans. The second approach is based on many smaller maps with 25 precincts. We then assess the overall performance of simulation methods across these many maps rather than focusing on a specific map.

4.1.1. A New 70-Precinct Validation Map

The top left plot of Figure 7 introduces a new validation map with 70 precincts and their population, which is a subset of the 2008 Florida precinct map consisting of 6,688 precincts with 25 districts. We use the `enumpart` algorithm to enumerate every partition of this map into two districts, which took approximately 8 hr on a MacBook Pro laptop with 16GB RAM and 2.8 GHz Intel i7 processors. Nearly all of this time was spent writing the partitions to disk—building the ZDD for this map took under half a second.

The histograms of the figure shows the number of redistricting plans that satisfy the deviation from population parity up to 20 percentage points (by one percentage point increments, i.e., $[0, 0.01), [0.01, 0.02), \ldots, [0.19, 0.2)$). The deviation from population parity is defined as,

$$\max_{1\le k\le p} \frac{|P_k - \overline{P}|}{\overline{P}}, \tag{1}$$

where P_k represents the population of the kth district, $\overline{P} = \sum_{k=1}^{p} P_k/p$, and p is the total number of districts. When partitioning this map into two districts, there exist a total of 44,082,156 possible redistricting plans if we only impose the contiguity requirement.

As shown in the upper right plot, out of these, over 700,000 plans are within a 1% population parity constraint. As we relax the population parity constraint, the cumulative number of valid redistricting plans gradually increases, reaching over 3 and 7 million plans for the 5% and 10% population parity constraints, respectively. Thus, this validation map represents a more realistic redistricting problem than the validation map analyzed in Fifield et al. (2014) and Fifield, Higgins, et al. (2020). That dataset, which enumerates all 117,688 partitions of a 25-precinct subset of the Florida map into three districts, includes only 8 plans within 1% of population parity, and 927 plans within 10% of population parity.

In addition to the population parity, we also consider compactness constraints. Although there exist a large number of different compactness measures, for the sake of illustration, we use the relative proximity index (RPI) proposed by Fryer and Holden (2011). The RPI for a given plan π_s in the valid set of redistricting plans $\boldsymbol{\pi}$ is defined as,

$$\text{RPI}(\pi_s) = \frac{\sum_{k=1}^{p} \sum_{i\in V_k} \sum_{j\in V_k} P_i P_j D_{ij}^2}{\arg\min_{\pi_s\in\boldsymbol{\pi}} \sum_{k=1}^{p} \sum_{i\in V_k} \sum_{j\in V_k} P_i P_j D_{ij}^2}, \tag{2}$$

where P_i corresponds to the population for precinct i assigned to district k, and D_{ij} corresponds to the distance between precincts i and j assigned to district k. Thus, a plan with a lower RPI is more compact.

We consider two compactness thresholds based on the RPI values: 25th and 75th percentiles, which equal 1.76 and 1.44, respectively. As shown in the bottom left histogram of Figure 7, the 25th percentile constraint does little beyond the population constraint. The number of valid plans that satisfy a 5% population parity threshold remains identical even after imposing this compactness constraint. However, the 75th percentile compactness constraint dramatically reduces the number of valid plans as seen in the bottom right histogram. For example, it reduces the total number of plans that meet the 1% population parity threshold by more than 70%.

Figure 8 shows the performance of the MCMC and RSG simulation methods using the new 70-precinct validation map (see Algorithms 1 and 3 of Fifield et al. (2014) and Fifield, Higgins, et al. (2020), respectively, for the details of implementation). The solid gray density shows the true distribution of the Republican dissimilarity index (Massey and Denton 1988) on the validation map, which is defined as

$$D = \frac{1}{2}\sum_{k=1}^{p} \frac{P_k}{P} \cdot \frac{|R_k - R|}{R(1-R)}, \tag{3}$$

where k indexes districts in a state, P_k is the population of district k, R_k is the share of district k that voted for the Republican presidential candidate in 2008, P is the total population in the state, and R is the voteshare for the Republican presidential candidate across all districts.

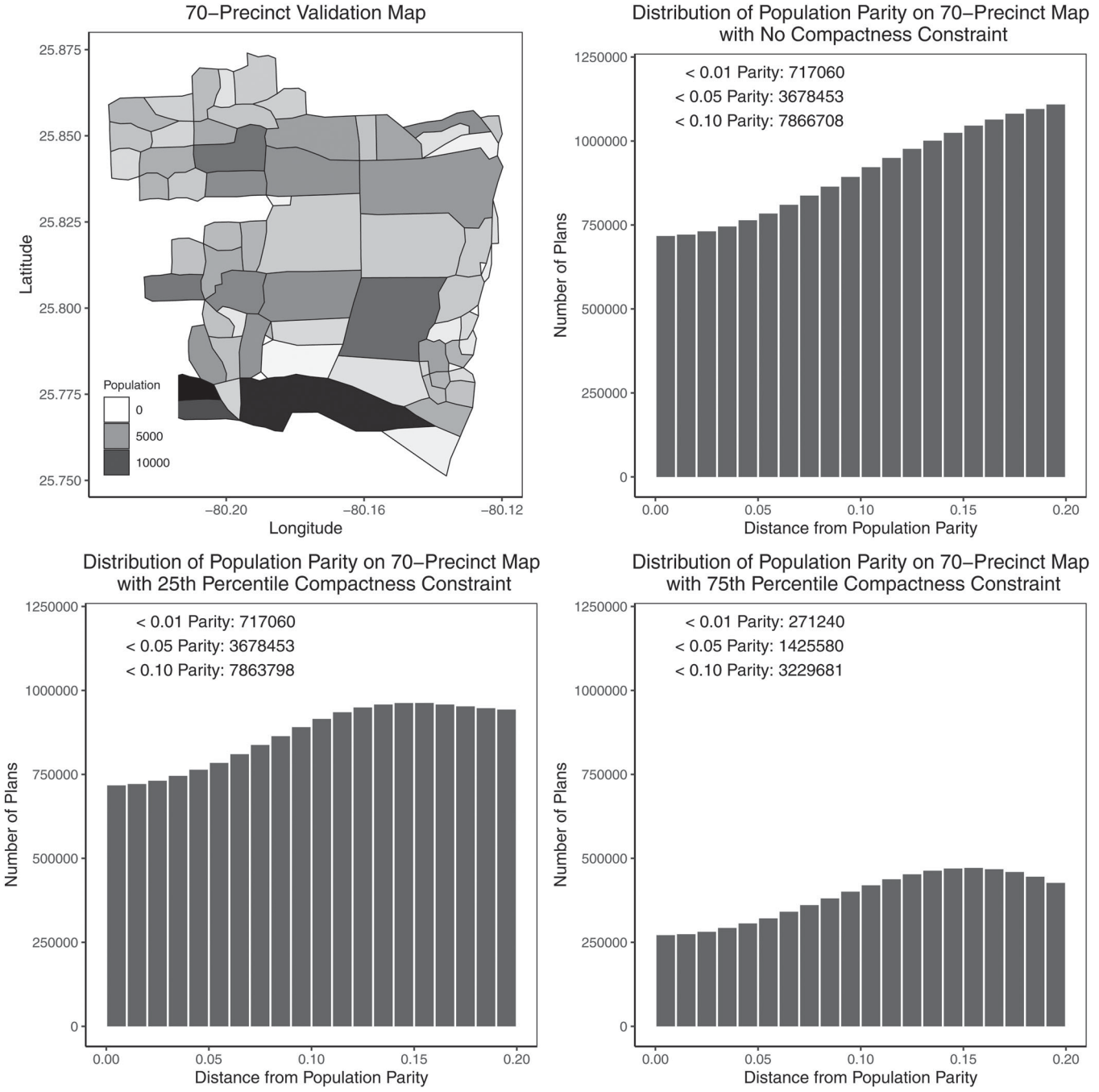


Figure 7. A new 70-precinct validation map and the histogram of redistricting plans under various population parity and compactness constraints. The underlying data is a 70-precinct contiguous subset of the Florida precinct map, for which the `enumpart` algorithm enumerated every 44,082,156 partitions of the map into two contiguous districts. In the histograms, each bar represents the number of redistricting plans that fall within a 1 percentage point range of a certain population parity, that is, $[0, 0.01), [0.01, 0.02), \ldots, [0.19, 0.20)$. The 25th (75th) percentile compactness constraint is defined as the set of plans that are more compact than the 25th (75th) percentile of maps within the full enumeration of all plans for the 70-precinct map, using the Relative Proximity Index to measure compactness. The annotations reflect the exact number of plans which meet the constraints. For example, when no compactness constraint is applied, there are 3,678,453 valid plans when applying a 5% population parity constraint, and 717,060 valid plans when applying a 1% population parity constraint. Under the strictest constraints, the 1% population parity constraint and 75th percentile compactness constraint, there are 271,240 valid plans.

The red dashed lines show the distribution of the Republican dissimilarity index of the RSG algorithm. Solid black lines show the distribution of the Republican dissimilarity index on plans drawn by the MCMC algorithm. In cases where we impose a population parity target, we specify a target distribution of plans using the Gibbs distribution where plans closer to population parity are more likely to be sampled by the algorithm (see Fifield et al. 2014; Fifield, Higgins, et al. 2020, for details).

Similarly, when a compactness constraint is imposed, we specify a target Gibbs distribution such that more compact plans are sampled. Note that in typical redistricting applications, we would not know the denominator of Equation (2). Fryer and Holden (2011) derived a power-diagram approach to finding a plan that approximately minimizes the denominator. Since we have enumerated all possible plans in the current setting, we simply use the true minimum value. However, this has no

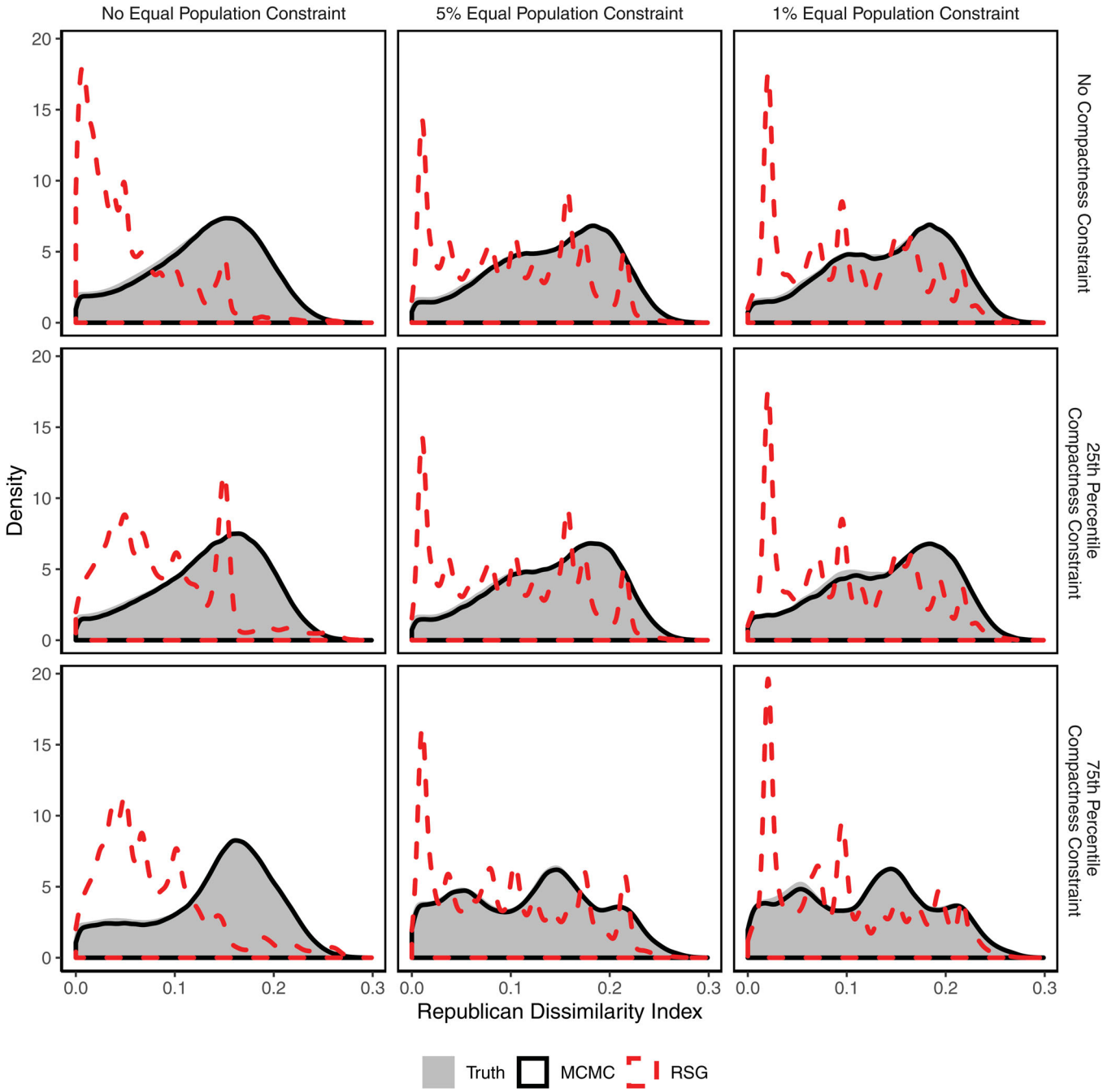


Figure 8. A validation study enumerating all partitions of a 70-precinct map into two districts. The underlying data are the 70-precinct contiguous subset introduced in the left plot of Figure 7. Unlike the random-seed-and-grow (RSG) method (red dashed lines), the Markov chain Monte Carlo (MCMC) method (solid black line) is able to approximate the target distribution. The 25th percentile (75th percentile) compactness constraint is defined as the set of plans that are more compact than the 25th (75th) percentile of maps within the full enumeration of all plans for the 70-precinct map, using the Relative Proximity Index to measure compactness.

impact on the performance of the algorithm, since it is absorbed into the normalizing constant of the target distribution.

Specifically, we use the following target Gibbs distribution,

$$f_\beta(\pi_s) = \frac{1}{z(\beta)} \exp\left\{-\sum_{k=1}^{p}(\beta_p \psi_k^p + \beta_c \psi_k^c)\right\}, \qquad (4)$$

where

$$\psi_k^p = \frac{|P_k - \overline{P}|}{\overline{P}} \quad \text{and}$$

$$\psi_k^c = \frac{\sum_{i\in V_k}\sum_{j\in V_k} P_i P_j D_{ij}^2}{\arg\min_{\pi_s\in\boldsymbol{\pi}} \sum_{k=1}^{p}\sum_{i\in V_k}\sum_{j\in V_k} P_i P_j D_{ij}^2}.$$

In this formulation, the strength of each constraint is governed by separate temperature parameters β_p (for population parity) and β_c (for compactness), where higher temperatures increase the likelihood that plans closer to the population parity or compactness target will be sampled. Once the algorithm is run, we discard sampled plans that fail to meet the target population and compactness constraints, and then reweight and resample the remaining plans so that they approximate a uniform sample from the population of all plans satisfying the constraints. After some initial tuning, we selected $\beta_p = 10$ for the 5% equal population constraint, and $\beta_p = 50$ for the 1% equal population constraint. We selected $\beta_c = 0.001$ for the 25th percentile

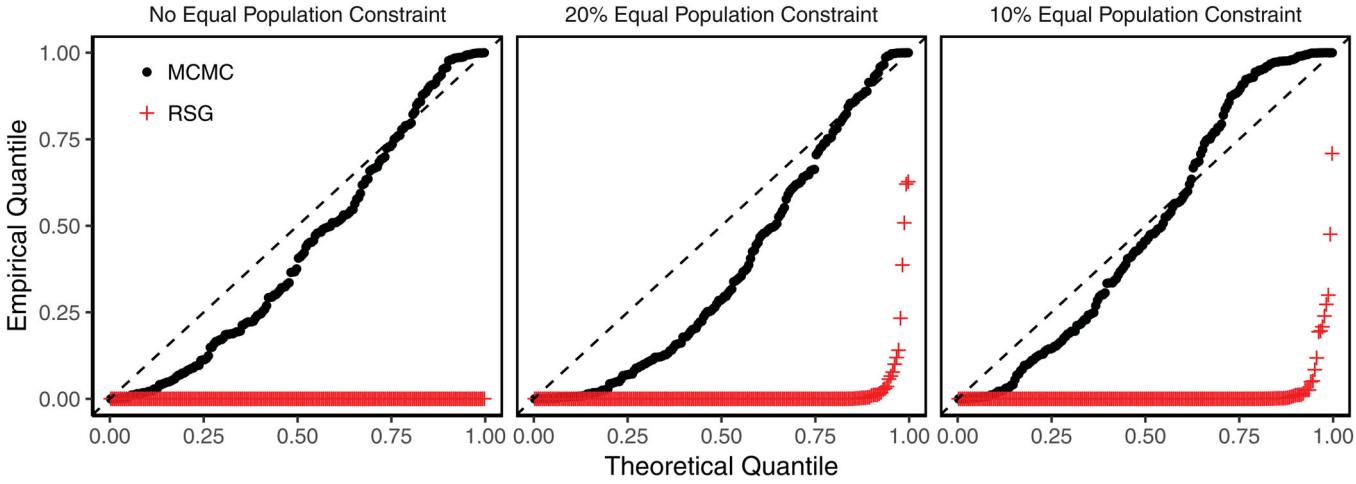


Figure 9. Quantile-quantile plot of *p*-values based on the Kolmogorov–Smirnov (KS) tests of distributional equality between the enumerated and simulated plans across 200 validation maps and under different population parity constraints. Each dot represents the *p*-value from a KS test comparing the empirical distribution of the Republican dissimilarity index from the simulated and enumerated redistricting plans. Under independent and uniform sampling, we expect the dots to fall on the 45 degree line. The MCMC algorithm (black dots), although imperfect, significantly outperforms the RSG algorithm (red crosses). See Figure A.1 in the appendix for discussion of thinning values.

compactness constraint and $\beta_c = 0.01$ for the 75th percentile compactness constraint. When the population or compactness constraints are not applied, we set their corresponding temperature parameter to 0.

The RSG algorithm was run for 1,000,000 independent draws for each population constraint, while the MCMC algorithms were run for 250,000 iterations using 8 chains for each pair of constraints. Starting plans for each MCMC chain were independently selected using the RSG algorithm. The Gelman–Rubin diagnostic (Gelman and Rubin 1992), a standard diagnostic tool for MCMC methods based on multiple chains, suggests that all MCMC chains had converged after at most 30,000 iterations. Unfortunately, the RSG algorithm does not come with such a diagnostic and hence we simply run it until it yields the same number of draws as the MCMC algorithms for the sake of comparison.

It is clear that on this test map, the RSG algorithm is unable to obtain a representative sample of the target distribution, at any level of population parity or compactness. This finding is consistent with the fact that the RSG algorithm is a heuristic algorithm with no theoretical guarantees and no specified target distribution. In contrast, the MCMC algorithm is able to approximate the target distribution, across all levels of population parity and compactness tested.

4.1.2. *Many Small Validation Maps*

A potential criticism of the previous validation study is that it is based on a single map. This means that even though it is of much larger size than the previously available validation map, the results may depend on the idiosyncratic geographical and other features of this particular validation map. To address this, we conduct another study based on many small validation maps. Specifically, we use our algorithm to enumerate all possible redistricting plans for each of 200 independent 25-precinct subsets of the 2008 Florida map. We then evaluate the performance of simulation methods for each validation map. Since we do not tune the temperature parameter of the MCMC algorithm for each simulated map unlike what one would do in practice, this yields a simulation setting that poses a significant challenge for the MCMC algorithm.

To assess the overall performance across these validation maps, we use the Kolmogorov–Smirnov (KS) statistic to test the distributional equality of the Republican dissimilarity index between the enumerated plans and the simulated plans. To increase the independence across simulated plans, we run the MCMC and RSG algorithms for 5 million iterations each on every map and then thinning by 500 (i.e., taking every 500th posterior draw). Without thinning, there is a significant amount of autocorrelation across draws, with the autocorrelation typically ranging between 0.75 and 0.85 between adjacent draws and between 0.30 and 0.60 for draws separated by 5 iterations. In contrast, when thinning the Markov chain by 500, the autocorrelation between adjacent draws falls to under 0.05. When thinning the Markov chain by 1000, the results are approximately the same, as seen in Figure A.1.

Although this does not make simulated draws completely independent of one another, we compute the *p*-value under the assumption of two independently and identically distributed samples. If the simulation methods are successful and the independence assumption holds, then we should find that the distribution of *p*-values across 200 small validation maps should be approximately uniform. After some initial tuning, we set the temperature parameter of the MCMC algorithm such that $\beta_p = 1$ for the 20% equal population constraint, and $\beta_p = 5$ for the 10% equal population constraint. These values are used throughout the simulations. After running the simulations, we again discard plans falling outside of the specified parity threshold and then reweight and resample the remaining plans to approximate a uniform draw from the target distribution of plans satisfying the specified parity constraint. We then calculate the KS test *p*-value by comparing the reweighted and resampled set of plans against the true distribution.

Figure 9 shows the results of this validation study. The left plot shows how the MCMC (black dots) and RSG (red crosses)

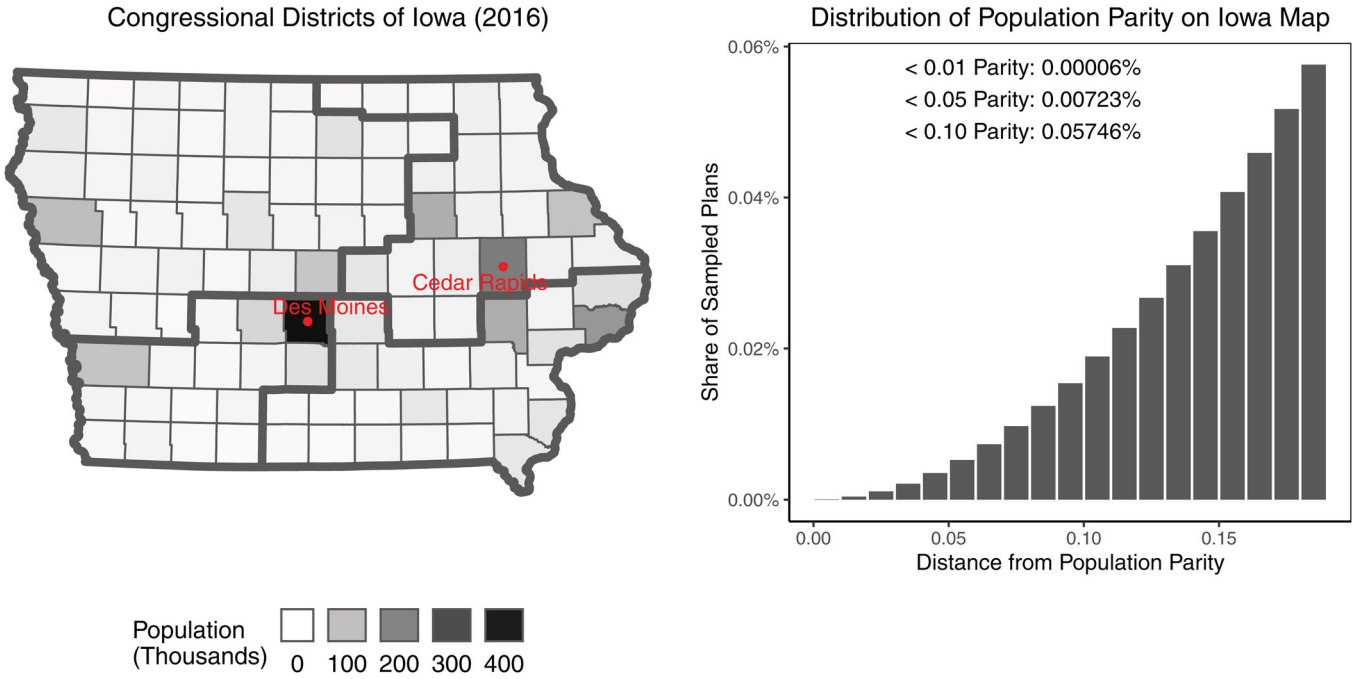


Figure 10. Iowa's 2016 congressional districts and the histogram of a random sample of redistricting plans under various population parity constraints. The underlying data is the Iowa county map, for which the `enumpart` algorithm generated an independent and uniform random sample of 500 million partitions of the map into four contiguous districts. In the histogram, each bar represents the number of redistricting plans that fall within the 1 percentage point range of a certain population parity, that is, [0, 0.01), [0.01, 0.02), . . . , [0.19, 0.20). There are 36,131 valid plans when applying a 5% population parity constraint, and only 300 valid plans when applying a 1% population parity constraint.

algorithms perform when not applying any population parity constraint. Each dot corresponds to the p-value of the KS test for a separate 25-precinct map. Under the assumption of independent sampled plans, if a simulation algorithm is successfully approximating the target distribution, these dots should fall roughly on the 45 degree line.

It is clear from this validation test that the RSG algorithm consistently fails to obtain a representative sample of the target distribution. That the red dots are concentrated near the bottom of the graph indicates that the KS p-value for the RSG algorithm is near zero for nearly every map tested. When population parity constraints of 20% and 10% are applied, the RSG algorithm continues to perform poorly compared to the MCMC algorithm. By using a soft constraint based on the Gibbs distribution, we allow the Markov chain to traverse from one valid plan to another through intermediate plans that may not satisfy the desired parity constraint. We find that although imperfect, the MCMC algorithm works much better than RSG algorithm.

4.2. Validation Through Independent Sampling

Next, we conduct larger-scale validation studies by leveraging the fact that the `enumpart` algorithm can independently sample from the population of all possible redistricting plans. This feature allows us to scale up our validation studies further by avoiding for larger maps the computationally intensive task of writing to the hard disk all possible redistricting plans, which exponentially increases as the map size gets larger or as we try to partition a map into more districts. We independently sample a large number of redistricting plans and compare them against the samples obtained from simulation methods. Below, we present two validation studies. The first study uses the actual Congressional district maps from Iowa, where by law redistricting is done using 99 counties. The second study is based on a new 250-precinct validation map obtained from the Florida map.

4.2.1. The Iowa Congressional District Map

We first analyze a new validation dataset constructed on the redistricting map from the state of Iowa. In Iowa, redistricting is conducted using a total of 99 counties instead of census blocks to piece together districts, to avoid splitting county boundaries in line with the Iowa State Constitution.[7] As a result, the Iowa redistricting problem is more manageable than other states.

The left plot of Figure 10 shows the Iowa map, where the shading indicates the population of each county. In 2016, Republicans won three districts while Democrats won one district, while in 2018, Democrats won three districts and the Republicans held only one. We use the `enumpart` algorithm to independently and uniformly sample 500 million contiguous partitions of this map into four districts. This number is minuscule relative to the total number of valid partitions of the map into four districts, of which there are approximately 10^{24}, but still is more than enough to use it as the target distribution. We note that while it took around 36 hr to sample 500 million partitions on the aforementioned computer cluster using significant parallelization, building the ZDD for this map took less than

[7] Article 3, Section 37 of the Iowa State Constitution states "When a congressional, senatorial or representative district shall be composed of two or more counties, it shall not be entirely separated by any county belonging to another district; and no county shall be divided in forming a congressional, senatorial, or representative district."

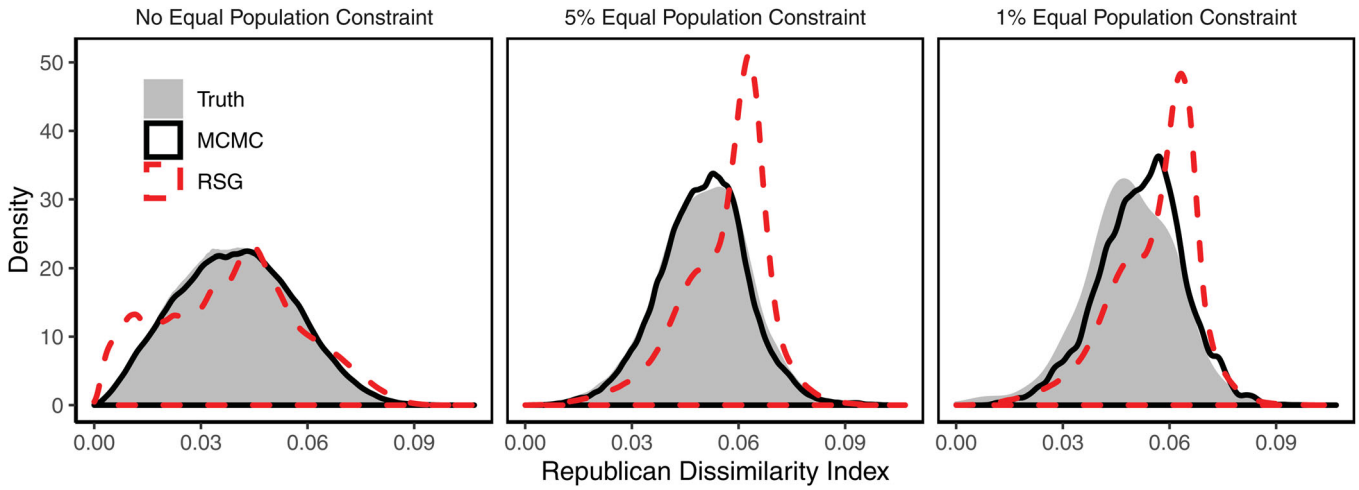


Figure 11. A validation study, uniformly sampling from the population of all partitions of the Iowa map into four districts. The underlying data are Iowa's county map in the left plot of Figure 10, which is partitioned into four congressional districts. As in the previous validation exercises, the Markov chain Monte Carlo (MCMC) method (solid black line) is able to approximate the independently and uniformly sampled target distribution, while the random-seed-and-grow (RSG) method (red dashed line) performs poorly.

half a second on our MacBook Pro laptop mentioned earlier. Nearly all of the runtime of the enumeration was spent writing the solutions to harddisk.

The histogram in Figure 10 shows the share of the sampled redistricting plans that satisfy the deviation from population parity up to 20 percentage points. Of the 500 million plans we have randomly sampled, only 300, or less than 0.00006%, satisfy a 1% population parity constraint, illustrating the sheer scale of the redistricting problem and how much the population equality constraint alone shrinks the total solution space of valid redistricting plans. There are 36,131 plans, or less than 0.001%, satisfying a 5% population parity constraint, which is still a minuscule share compared to the total number of enumerated plans.

Figure 11 shows the performance of the MCMC and RSG simulation methods on the state-sized redistricting problem for Iowa. The solid gray density shows the distribution of the Republican dissimilarity index based on the independently and uniformly sampled set of 500 million redistricting plans. The red dashed lines show the distribution of the Republican dissimilarity index on plans sampled by the RSG algorithm, while the solid black lines shows the distribution for plans sampled by the MCMC algorithm. As in the previous validation exercise, where we impose 5% and 1% population parity constraints, we specify a target distribution of plans using the Gibbs distribution. Here, we set the temperature parameter $\beta_p = 25$ for the 5% parity constraint, and $\beta_p = 50$ for the 1% parity constraint, which we selected after initial tuning. After discarding plans not satisfying the constraint and then reweighting, we ended up with 629,729 plans for the 5% parity constraint, and 93,046 plans for the 1% parity constraint. The RSG algorithm was run for 2 million independent draws, while the MCMC algorithms were run for 250,000 iterations and initializing 8 chains for each algorithm. The chains were run without a burn-in period, and the Gelman–Rubin diagnostic suggested that the Markov chains had converged after at most 30,000 iterations.

As with the previous validation test, the MCMC algorithm outperforms the RSG algorithm across all levels of the population parity constraint. When no equal population constraint is applied or a 5% population parity constraint is applied, the MCMC algorithm samples from the target distribution nearly perfectly. Even with the 1% parity map, where there are only 300 valid plans in the target distribution, the MCMC algorithm approximates the target distribution reasonably well, missing by only slightly in portions of the distribution. In contrast, at all levels of population parity, the RSG algorithm is unable to draw a representative sample of plans from the target distribution.

4.2.2. A New 250-Precinct Validation Map

Next, we present the results of validation tests that use a new, 250-precinct validation map, which is constructed from a contiguous subset of the 2008 Florida precinct map. As with the previous validation exercise, we use the `enumpart` algorithm to independently sample 100 million partitions of this map into two districts. This is still a minuscule number of plans relative to about 5×10^{39} possible partitions of this map into two districts. However, given the ability of the `enumpart` algorithm to independently sample plans using the ZDD, we are able to approximate an accurate target distribution arbitrarily well.

The left plot of Figure 12 shows the validation map, where the shading indicates the population of each of the precincts. Unlike the Iowa map, this map has geographical units of various sizes. This validation map also has a slightly larger frontier size (maximum frontier of 14) than that of the Iowa map (maximum frontier of 11), making it more likely to run out of memory due to the size of ZDD and thereby also increasing computational time. The histogram on the right gives the distribution of population parity distance among the sampled plans, through 20% parity. Of the sampled plans, 1.95% (1.95 million plans) satisfy the 1% population parity constraint, while 21.8% of the sampled plans (21.8 million plans) satisfy the 10% population parity constraint.

We sample 4 million plans for each population parity level using the MCMC and RSG algorithms. For the MCMC algorithm, we initialized 8 chains running for 500,000 iterations each, and where a population parity constraint is imposed, we specify the target distribution of plans using the Gibbs

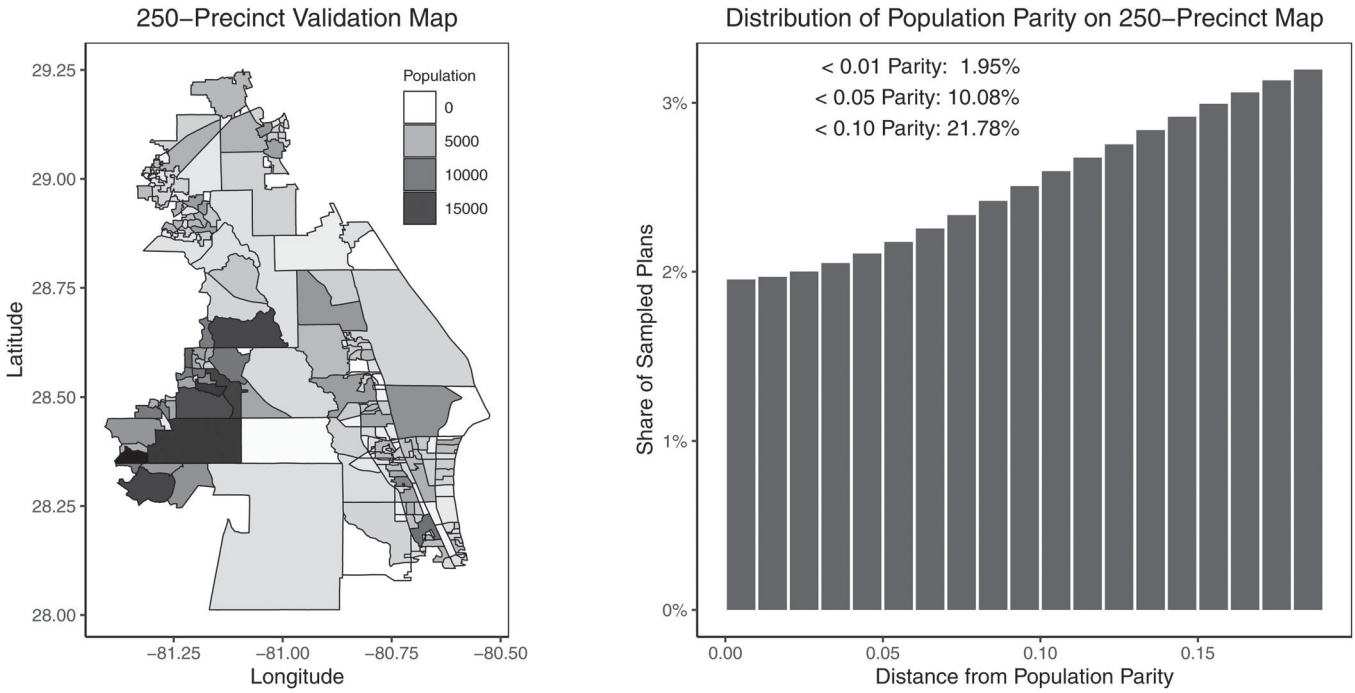


Figure 12. A new 250-precinct validation map and the histogram of redistricting plans under various population parity constraints. The underlying data are a 250-precinct contiguous subset of the Florida precinct map, for which the `enumpart` algorithm generated an independent and uniform random sample of 100 million partitions of the map into two contiguous districts. In the histogram, each bar represents the number of redistricting plans that fall within the 1 percentage point range of a certain population parity, that is, $[0, 0.01), [0.01, 0.02), \ldots, [0.19, 0.20)$. There are 10,082,542 valid plans when applying a 5% population parity constraint, and 1,953,736 valid plans when applying a 1% population parity constraint.

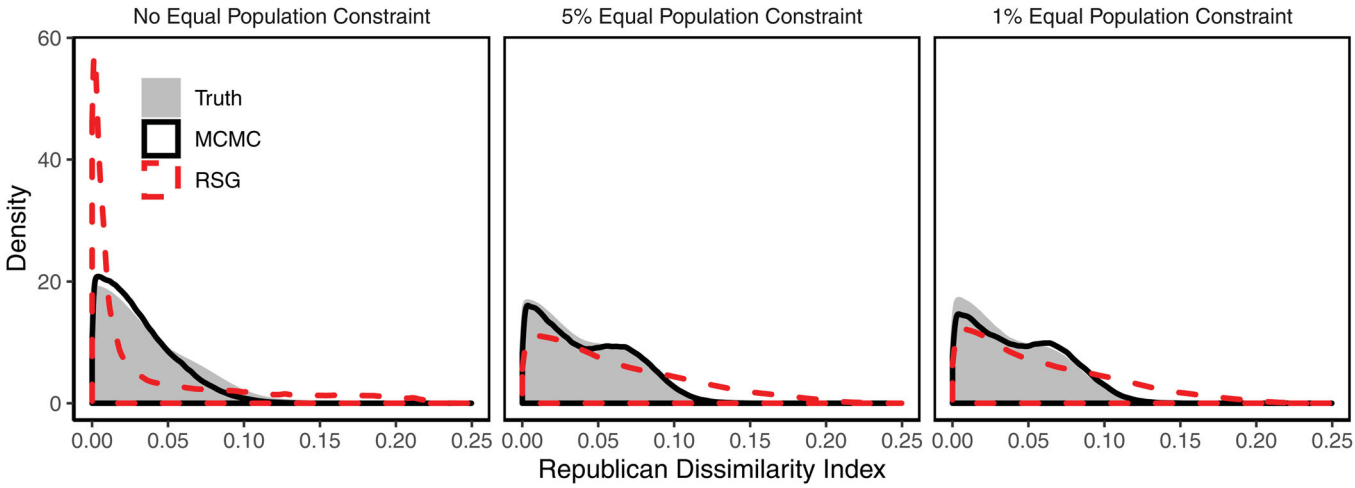


Figure 13. A validation study enumerating all partitions of a 250-precinct map into two districts. The underlying data are the 250-precinct contiguous subset introduced in the left plot of Figure 12. As in the previous validation exercises, the Markov chain Monte Carlo (MCMC) method (solid black line) is able to approximate the target distribution based on the independent and uniform sampling, while the random-seed-and-grow (RSG) method (red dashed line) performs poorly.

distribution. We set the temperature parameter $\beta_p = 25$ for sampling plans within 5% of parity, and $\beta_p = 50$ when sampling plans within 1% of parity. After discarding invalid plans and reweighting, these parameter settings yielded 3,088,086 plans satisfying the 5% parity constraint, and 1,881,043 plans satisfying the 1% parity constraint. All 8 chains were run without burn-in, and the Gelman–Rubin convergence diagnostic suggested the chains had converged after approximately 75,000 iterations.

Results for the validation test using the 250-precinct validation map are shown in Figure 13. As with the previous validation exercise, the solid gray density shows the target distribution of the Republican dissimilarity index on the 100 million plans sampled by the `enumpart` algorithm, while the red dashed lines show the distribution of the Republican dissimilarity index for the RSG algorithm. Finally, the solid black lines show the distribution of the Republican dissimilarity index for the MCMC algorithm. Across all levels of population parity, including the 1% constraint, the MCMC algorithm is able to successfully sample from the target distribution and return a representative sample of redistricting plans. In contrast, where no population parity constraint is applied or where a 5% parity constraint is applied, the RSG algorithm is not able to sample from the target distribution with any accuracy. While it performs somewhat better on the 1% constraint, it is still biased toward plans with

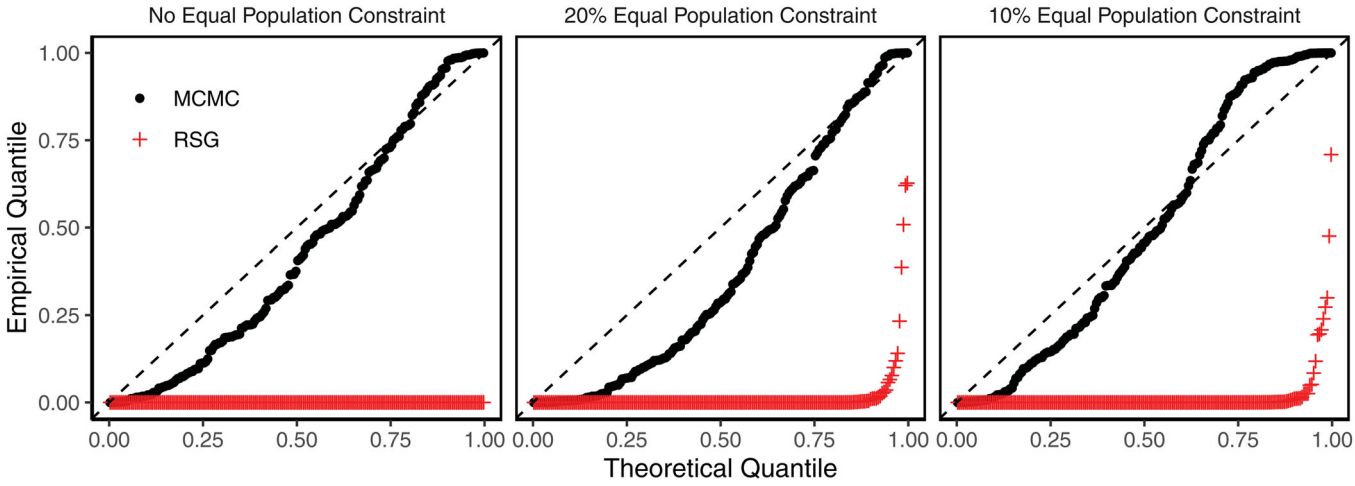


Figure A.1. Quantile-quantile plot of *p*-values based on the Kolmogorov–Smirnov (KS) tests of distributional equality between the enumerated and simulated plans across 200 validation maps and under different population parity constraints. Each dot represents the *p*-value from a KS test comparing the empirical distribution of the Republican dissimilarity index from the simulated and enumerated redistricting plans. Under independent and uniform sampling, we expect the dots to fall on the 45 degree line. The MCMC algorithm (black dots), although imperfect, significantly outperforms the RSG algorithm (red crosses).

higher values on the dissimilarity index, and fails to capture the bimodality of the target distribution.

5. Concluding Remarks

More than a half century after scholars began to consider automated redistricting, legislatures and courts are increasingly relying on computational methods to generate redistricting plans and determine their legality. Unfortunately, despite the growing popularity of simulation methods in the recent redistricting cases, there exists little empirical evidence that these methods can in practice generate a representative sample of all possible redistricting maps under the statutory guidelines and requirements.

We believe that the scientific community has an obligation to empirically validate the accuracy of these methods. In this article, we show how to conduct empirical validation studies by utilizing a recently developed computational method that enables the enumeration and independent sampling of all possible redistricting plans for maps with a couple of hundred geographical units. We make these validation maps publicly available, and implement our methodology as part of an open source software package, redist. These resources should facilitate researchers' efforts to evaluate the performance of existing and new methods in realistic settings.

Indeed, much work remains to be done to understand the conditions under which a specific simulation method do and does not perform well. A real-world redistricting process is complex. Distinct geographical features and diverse legal requirements play important roles in each state. It is far from clear how these factors interact with different simulation methods. Future work should address these issues using the data from various states.

It is also important to further improve the capabilities of the `enumpart` algorithm and of the MCMC algorithm. The maximum frontier size of our largest validation maps, which predicts the computational difficulty for the `enumpart` algorithm, is 14, which is far less than that of other states. For example, the maximum frontier size for New Hampshire (2 districts, 327 precincts) and Wisconsin (8 districts, 6895 precincts) are 21 and 84, respectively. These are much more challenging redistricting problems than the validation studies presented in this article.

As the 2020 census passes, lawsuits challenging proposed redistricting plans will inevitably be brought to court, and simulation evidence will be used to challenge and defend those plans. Thus, it is necessary that the empirical performance of these methods be rigorously evaluated. This article introduces what we hope will be the first of many future complementary validation tests used to ensure that this evidence is of the highest possible quality according to scientific standards.

Appendix

Figure A.1 provides comparison to Figure 9. In Figure 9, thinning for the MCMC runs was set to 500. For this run, thinning for the MCMC runs was set to 1000. We find no significant difference between the two values. Thinning at 500 should then be sufficient and more efficient for this case.

Acknowledgments

We thank Steve Schecter, participants of the Quantitative Gerrymandering and Redistricting Conference at Duke University, and two anonymous reviewers for helpful comments.

Funding

Kawahara acknowledges financial support by JSPS KAKENHI grant number JP18K04610. The computations in this article were run on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University.

Supplementary Materials

Replication materials and enumeration results are available as Fifield, Imai, et al. (2020).

ORCID

Benjamin Fifield http://orcid.org/0000-0002-2247-0201
Kosuke Imai http://orcid.org/0000-0002-2748-1022
Christopher T. Kenny http://orcid.org/0000-0002-9386-6860

References

Altman, M. (1997), "The Computational Complexity of Automated Redistricting: Is Automation the Answer," *Rutgers Computer & Technology Law Journal*, 23, 81–142. [54]

Altman, M., and McDonald, M. P. (2011), "BARD: Better Automated Redistricting," *Journal of Statistical Software*, 42, 1–28. [52]

Carter, D., Herschlag, G., Hunter, Z., and Mattingly, J. (2019), "A Merge-Split Proposal for Reversible Monte Carlo Markov Chain Sampling of Redistricting Plans," Tech. Rep., arXiv no. 1911.01503. [52,53]

Chen, J., and Rodden, J. (2013), "Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures," *Quarterly Journal of Political Science*, 8, 239–269. [52,53,60]

Chikina, M., Frieze, A., and Pegden, W. (2017), "Assessing Significance in a Markov Chain Without Mixing," *Proceedings of the National Academy of Sciences of the United States of America*, 114, 2860–2864. [52,53]

Cirincione, C., Darling, T. A., and O'Rourke, T. G. (2000), "Assessing South Carolina's 1990s Congressional Districting," *Political Geography*, 19, 189–211. [52,53,60]

DeFord, D., Duchin, M., and Solomon, J. (2019), "Recombination: A Family of Markov Chains for Redistricting," Tech. Rep., arXiv no. 1911.05725. [52]

Fifield, B., Higgins, M., Imai, K., and Tarr, A. (2014), "A New Automated Redistricting Simulator Using Markov Chain Monte Carlo," Tech. Rep., Presented at the *31th Annual Meeting of the Society for Political Methodology*, University of Georgia. [52,53,54,60,61]

——— (2020), "Automated Redistricting Simulator Using Markov Chain Monte Carlo," *Journal of Computational and Graphical Statistics* (forthcoming). [52,53,60,61]

Fifield, B., Imai, K., Kawahara, J., and Kenny, C. T. (2020), "Replication Data for: The Essential Role of Empirical Validation in Legislative Redistricting Simulation," available at *https://doi.org/10.7910/DVN/NH4CRS*. [53,67]

Fifield, B., Tarr, A., and Imai, K. (2015), "redist: Markov Chain Monte Carlo Methods for Redistricting Simulation," Comprehensive R Archive Network (CRAN), available at *https://CRAN.R-project.org/package=redist*. [53,60]

Fryer, R. G., and Holden, R. (2011), "Measuring the Compactness of Political Districting Plans," *The Journal of Law and Economics*, 54, 493–535. [60,61]

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulations Using Multiple Sequences" (with discussion), *Statistical Science*, 7, 457–472. [63]

Herschlag, G., Ravier, R., and Mattingly, J. C. (2017), "Evaluating Partisan Gerrymandering in Wisconsin," Tech. Rep., Department of Mathematics, Duke University. [52]

Hess, S. W., Weaver, J. B., Siegfeldt, H. J., Whelan, J. N., and Zitlau, P. A. (1965), "Nonpartisan Political Redistricting by Computer," *Operations Research*, 13, 998–1006. [52]

Katz, J., King, G., and Rosenblatt, E. (2020), "Theoretical Foundations and Empirical Evaluations of Partisan Fairness in District-Based Democracies," *American Political Science Review*, 114, 164–178. [53]

Kawahara, J., Horiyama, T., Hotta, K., and Minato, S. (2017), "Generating All Patterns of Graph Partitions Within a Disparity Bound," in *Proceedings of the 11th International Conference and Workshops on Algorithms and Computation*, Vol. 10167, pp. 119–131. [53,54,55,57,58]

Knuth, D. E. (2011), *The Art of Computer Programming: Combinatorial Algorithms, Part I* (Vol. 4A), New York: Addison-Wesley. [58]

Liu, Y. Y., Tam Cho, W. K., and Wang, S. (2016), "PEAR: A Massively Parallel Evolutionary Computation Approach for Political Redistricting Optimization and Analysis," *Swarm and Evolutionary Computation*, 30, 78–92. [52]

Magleby, D., and Mosesson, D. (2018), "A New Approach for Developing Neutral Redistricting Plans," *Political Analysis*, 26, 147–167. [52,53]

Massey, D. S., and Denton, N. A. (1988), "The Dimensions of Residential Segregation," *Social Forces*, 67, 281–315. [60]

Mattingly, J. C., and Vaughn, C. (2014), "Redistricting and the Will of the People," Tech. Rep., Department of Mathematics, Duke University. [52,53,60]

McCarty, N., Poole, K. T., and Rosenthal, H. (2009), "Does Gerrymandering Cause Polarization?," *American Journal of Political Science*, 53, 666–680. [52]

Mehrotra, A., Johnson, E., and Nemhauser, G. L. (1998), "An Optimization Based Heuristic for Political Districting," *Management Science*, 44, 1100–1114. [54]

Minato, S. (1993), "Zero-Suppressed BDDs for Set Manipulation in Combinatorial Problems," in *Proceedings of the 30th ACM/IEEE Design Automation Conference*, pp. 272–277. [53,54]

Nagel, S. S. (1965), "Simplified Bipartisan Computer Redistricting," *Stanford Law Journal*, 17, 863–899. [52]

Rubin, D. B. (1987), "Comment: A Noniterative Sampling/Importance Resampling Alternative to the Data Augmentation Algorithm for Creating a Few Imputation When Fractions of Missing Information Are Modest: The SIR Algorithm," *Journal of the American Statistical Association*, 82, 543–546. [58]

Vickrey, W. (1961), "On the Prevention of Gerrymandering," *Political Science Quarterly*, 76, 105–110. [52]

Weaver, J. B., and Hess, S. W. (1963), "A Procedure for Nonpartisan Districting: Development of Computer Techniques," *Yale Law Journal*, 73, 288–308. [52]