Models, Race, and the Law

A Response to
Chen & Stephanopoulos, *The Race-Blind Future of Voting Rights*
130 YALE L.J. 870 (2021)

Moon Duchin[⋆]
Department of Mathematics
Tufts University

Douglas M. Spencer[†]
School of Law
University of Connecticut

January 2021

[⋆] Associate Professor of Mathematics, Director of the Program in Science, Technology, and Society, and Senior Fellow in the Tisch College of Civic Life at Tufts University. E-mail: moon.duchin@tufts.edu.

[†] Professor of Law & Public Policy, University of Connecticut and Distinguished Faculty Fellow (2020-2021) at the Byron R. White Center for the Study of American Constitutional Law at the University of Colorado, Boulder. E-mail: douglas.spencer@uconn.edu. We are grateful to Amariah Becker, Guy-Uriel Charles, Chris Elmendorf, Dara Gold, Sam Hirsch, Tyler Jarvis, Ellen Katz, Vlad Kogan, Justin Levitt, Kevin Quinn, Michael Ratcliffe and Bertrall Ross for their helpful input and feedback. We are also indebted to Anna Dorman, Parker Rule, Gabe Schoenbach, Jackson Skeen, and Matthew Sun for their excellent research assistance.

Contents

Introduction

The Voting Rights Act of 1965 (VRA) guarantees that all American citizens, regardless of race or ethnicity, should have an equal opportunity to participate in the political process and to elect representatives of their choice.[1] The VRA frequently interacts with single-member districts, which serve as the electoral system for congressional and nearly all state legislative races and are the go-to remedy in local VRA enforcement. It has long been known in the redistricting literature that random boundary placement puts minorities at a major structural disadvantage.[2] Single-member districts can secure electoral opportunity for minorities, but only if the minority population is sufficiently concentrated and the boundaries are favorably aligned. The ability of the VRA to remediate historical discrimination and underrepresentation thus depends on proactive redistricting. As a matter of practice, when a set of districts empowers minority communities to elect representatives in rough proportion to their population, courts have held the promise of political equality to have been fulfilled.[3] However, proportionality has functionally operated as a ceiling even when viewed as normatively desirable: White voters will never be represented by less than their share of the population while minority communities nearly invariably will.[4]

In *The Race-Blind Future of Voting Rights*, Jowei Chen and Nicholas Stephanopoulos sketch out a less proactive future of districting, including a mechanism that stands to needlessly sabotage minority political power and undermine the signal remedial goal of the VRA. The authors devote their Article to delineating a new baseline of opportunity provided by a randomized redistricting protocol that operates with no regard to race.[5] Their project is strategic and pragmatic, motivated by the prediction that an increasingly conservative Supreme Court is likely to effect "avulsive change" on the VRA in the near term, quite possibly by dropping any role for rough proportionality and elevating race-blind mapping as a new ideal.[6] Their Article thus seeks to provide a roadmap for voting rights advocates to navigate a new nominally race-blind landscape.

To present their approach as a manageable standard, Chen and Stephanopoulos go big—modeling voter preferences in 1,903 districts and evaluating 38,000 districting plans spanning 19 states—and describe their outputs authoritatively as *the* race-blind baseline, full stop. Their particular setup is said to be capable of capturing the full dynamics of non-racial redistricting.

> We find that most—though not all—enacted state-house plans overrepresent minority voters relative to the race-blind baseline. For example, numerous plans in the deep South include substantially more African American opportunity districts than would typically emerge from a nonracial redistricting process, while a few plans in the border

[1] 52 U.S.C. §10301(b).

[2] *See*, *e.g.*, Bernard N. Grofman, *For Single-Member Districts, Random is Not Equal*, *in* Representation and Redistricting Issues 55-58 (Bernard Grofman, Arend Lijphart, Robert McKay & Howard Scarrow eds. 1982). Jowei Chen also co-authored a ground-breaking study of the interplay of geography and this well known majority seat bonus. *See* Jowei Chen and Jonathan Rodden, *Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures*, 8 Q.J. Pol. Sci. 239 (2013).

[3] Johnson v. DeGrandy, 512 U.S. 997, 1000 (1994). *See also* Ellen D. Katz, et al, Documenting Discrimination in Voting: Judicial Findings under Section 2 of the Voting Rights Act since 1982, 39 U. Mich. J.L. Reform 643 (2006).

[4] *See*, *e.g.*, Nicholas O. Stephanopoulos, *The Relegation of Polarization*, 83 U. Chi. L. Rev. Online 160, 168 (2017) (explaining that a "more accurate statement of the [dominant theory of vote dilution] is that minority voters should be able to elect their preferred candidates to the extent permitted by their geographic distribution up to a ceiling of proportionality.").

[5] Jowei Chen & Nicholas O. Stephanopoulos, *The Race-Blind Future of Voting Rights*, 130 Yale L.J. 870 (2021).

[6] *Id.* at 949.

> South include fewer such districts. Similarly, several western states feature extra Hispanic opportunity districts compared to the race-blind baseline, while only one western state underrepresents Hispanic voters.[7]

As we show below, the authors' methodology does not warrant these kinds of conclusive statements, much less the slippage into the unmistakably normative language of over- and underrepresentation. We certainly share the authors' enthusiasm about the burgeoning ensemble method. The central counterfactual problem in vote dilution law for many decades has been that of conceptualizing the undiluted baseline: *What is the weight of a vote, absent manipulation?* In recent years, algorithms that generate large samples or "ensembles" of plausible districting plans have been increasingly used to approach that question. Using ensembles made to conform to legal rules, but without regard to race or partisan data, can provide a non-gerrymandered baseline. Unfortunately, the approach taken by Chen and Stephanopoulos does not conform with best practices in mathematical modeling.[8]

First, the authors' ambitious scope leads them to take many shortcuts in methodology as they build their ensembles and their label of opportunity. They borrow tools from mathematical and statistical modeling (notably the randomized districting algorithm developed in the research group that one of us runs[9]) but do not provide a detailed description of their design choices; do not report any convergence metrics to confirm that their ensembles of districting plans are representative of any particular weighting of plans; and do not provide any control of errors that propagate through their workflow, especially through their idiosyncratic use of ecological inference.

There are quite a few junctures where their modeling decisions should be flagged. For example, the nineteen states under consideration all have different statutory and constitutional rules for redistricting and a one-size-fits-all modeling approach can't come close to the mark of capturing legal nuance. This is not simply a question of *whether* to take each rule or principle into account, but *how* to operationalize that priority. For example, the legal language around county preservation is markedly different across these states: Texas mentions county preservation,[10] North Carolina[11] and

[7] *Id.* at 875.

[8] A different attempt to model VRA compliance in a Markov chain can be found in a collaborative effort by data scientists and a voting rights attorney, *see* Amariah Becker, Moon Duchin, Dara Gold, and Sam Hirsch, *Computational Redistricting and the Voting Rights Act*, Preprint (2020). *Soon to be available at* mggg.org/VRA.

[9] This Markov chain algorithm called *recombination* or `ReCom` is discussed in more length *infra* Part III.2 and Appendix A.1. For a detailed discussion of `ReCom`, *see* Daryl DeFord, Moon Duchin, & Justin Solomon, *Recombination: A Family of Markov Chains for Redistricting*, preprint, May 27, 2020. *Available at* mggg.org/ReCom

[10] Tx. Const. art. 3, § 26 (requiring that state house districts be apportioned among the counties, and that counties not be split to the extent possible).

[11] Stephenson v. Bartlett, 582 S.E.2d 247, 250 (N.C. 2003) (interpreting Article 2 of the state constitution that "no county shall be divided" to permit county splits for VRA compliance and/or when necessary to comply with the one-person-one-vote standard so long as county groupings are minimized and resulting districts fall within five percent of population equality.)

Ohio[12] have extremely specific language about how to measure it, and Delaware[13] and Illinois[14] don't have any county preservation rules at all. Nevertheless the same kind of (very strong) county filter is applied by Chen and Stephanopoulos in generating districts in all states – the details, impacts, and alternatives are left completely undiscussed. Perhaps more fundamentally, the authors rely on a single presidential election to infer voter preferences – Obama vs. Romney 2012 – immediately decoupling their findings from VRA practice where attorneys would never claim to identify minority opportunity based on Obama's re-election numbers alone. Beyond this, the authors consider only a single plausible definition of opportunity district; they do not compare their "opportunity" label against the ground truth of recent district performance; and they provide no significant robustness checks at any step in their modeling. Because the authors package their series of complex and computationally intensive functions into a single statistic (the median number of opportunity districts) with very little discussion about their modeling choices, readers may not appreciate the extent to which many of the ingredients are arbitrary, approximate, or numerically unstable. We unpack some of the workflow complexity in Table 2. Do these many choices have effects that cancel out in the end somehow, leaving the finding of over- or underrepresentation intact even if the numbers shift? Do their design choices systematically bias estimates upwards or downwards relative to what would be possible if more elections were taken into account or state laws were handled differently? These key questions are handled glibly when they are addressed.[15]

Second, the authors misuse the ensembles that they do generate. Ensembles are not suited to identifying a single ideal value of a score, as Chen and Stephanopoulos implicitly do by assigning a designation of under- or overrepresentation based on the *median value* alone.[16] Rather, ensembles are a powerful tool for understanding baseline *ranges* for valid districting plans and are useful for clarifying decisionmaking tradeoffs. As the Supreme Court held in 1994, "no single statistic provides courts with a shortcut to determine whether a set of single-member districts unlawfully dilutes minority strength."[17] The single statistic presented by Chen and Stephanopoulos is no exception.

[12] Ohio Const. art. 19, § 2(B)(5) ("Of the eighty-eight counties in this state, sixty-five counties shall be contained entirely within a district, eighteen counties may be split not more than once, and five counties may be split not more than twice."); § 2(B)(7) ("No two congressional districts shall share portions of the territory of more than one county, except for a county whose population exceeds four hundred thousand."); and § 2(B)(8) ("The authority drawing the districts shall attempt to include at least one whole county in each congressional district. This division does not apply to a congressional district that is contained entirely within one county or that cannot be drawn in that manner while complying with federal law.")

[13] Del. Code tit. 29, § 804 ("In determining the boundaries of the several representative and senatorial districts within the State, the General Assembly shall use the following criteria. Each district shall, insofar as is possible: (1) Be formed of contiguous territory; (2) Be nearly equal in population; (3) Be bounded by major roads, streams or other natural boundaries; and (4) Not be created so as to unduly favor any person or political party.").

[14] Ill. Const. art. 4, § 3(a) ("Legislative Districts shall be compact, contiguous and substantially equal in population. Representative Districts shall be compact, contiguous, and substantially equal in population.").

[15] It is of course insufficient to assert that design choices are applied for measuring opportunity in both the enacted plan and its comparator maps, as the authors do. Chen & Stephanopoulos, *supra* note 5, at 903, fn. 174 ("Any idiosyncracies in our particular ecological inference run are reflected in the numbers of opportunity districts we report for *both* the enacted plans and the simulated [sic] maps.") (emphasis in the original). This inadequacy is demonstrated below in Figure 5, where it is shown that instability may affect the measurement of the enacted plan while leaving the ensemble unchanged.

[16] For a discussion of their reliance on the median, *see infra* Part II.2. It was sleight of hand of just this kind – treating a single number based on piles of political modeling choices as an authoritative indicator – that earned the memorable label of *sociological gobbledygook*. Transcript of Oral Argument at 40, Gill v. Whitford, 138 S. Ct. 1916 (No. 16-1161) ("CHIEF JUSTICE ROBERTS:. . .the whole point is you're taking these issues away from democracy and you're throwing them into the courts pursuant to, and it may be simply my educational background, but I can only describe as sociological gobbledygook.")

[17] Johnson v. DeGrandy, 512 U.S. 997, 1020-21 (1994).

One of the challenges of introducing novel technical methods in a law review is that the blueprints that are especially important for validation – the details of algorithm design, the magnitude of uncertainty, convergence metrics, alternative specifications, and other robustness checks – are not likely to draw needed scrutiny from law review editors or indeed to hold the attention of most readers. The temptation is thus to gloss over or omit these technical details altogether, even in an 86-page article and its 53-page appendix. But transparency is all the more important for a project that has not been subject to rigorous peer review. This worry about law review publication is not new. Nearly twenty years ago, Lee Epstein and Gary King wrote an important piece in which they reviewed the legal literature and sounded the alarm that "*the current state of empirical legal scholarship is deeply flawed*" (emphasis original).[18] The lack of attention to sound methodology, they warned, would lead readers to "learn considerably less accurate information about the empirical world than the studies' stridently stated, but overly confident, conclusions suggest."[19]

This is exactly what generates our grave concerns about the current Article and its placement in a flagship law review. Chen and Stephanopoulos's style of leveraging technical tools while ignoring the scientific standards surrounding their development and deployment risks creating an unnecessarily muddy legal terrain. And the stakes are high: they have provided a recipe that may well devastate electoral opportunity for minority groups just as public opinion and voting behavior are pushing the other way.

In sum, we find that *The Race-Blind Future of Voting Rights* is a provocative proof of concept that stands on a shaky empirical foundation. The Article uses the promising *ensemble method* of random district generation to deliver a baseline for minority electoral opportunity; this Response both flags technical issues and questions the conceptual alignment of the methods with their application to voting rights law.

Overview

In Part 1 we discuss the non-linear effects of winner-take-all districting, explaining that the mathematics of districts induces a major representational disadvantage for any group in the numerical minority. Minority groups are therefore systematically disfavored by single-member districts just as they are by at-large plurality voting, and the law must take up the challenge of counteracting these effects for groups that are protected from disparate treatment. We argue that the difficult task of remedial district design becomes excruciating if we gauge success by standards that ignore, or at least proclaim to ignore, the very feature that triggers the obligation to protect.

In Part 2 we trace the intuition that algorithmic methods can generate a baseline for voting rights opportunity through the law and policy literature, culminating in the proposal by Judge Easterbrook of the Seventh Circuit that is cited as motivation by Chen and Stephanopoulos. We outline the promise of ensembles to address foundational questions about vote dilution and we illustrate why the median should not be elevated to an ideal, as is strongly implied by the authors' labels of under- and overrepresentation.

[18] Lee Epstein & Gary King, *The Rules of Inference*, 69 U. Chi. L. Rev. 1, 6 (2002).

[19] *Id.* at 6-7.

In Part 3 we describe the logic of building ensembles using Markov chain Monte Carlo, or MCMC. The main benefit of using MCMC to find a baseline is that it is built to draw *representative samples* of all valid districting plans according to a desired weighting, or "target distribution." It is this promise of representative sampling that holds the relevance of ensemble methods for finding a baseline. A close read of the Article coupled with a close inspection of the authors' replication materials reveals their conflation of various methodologies and their inattention to bottlenecks that block their ability to sample in a representative manner.

In Part 4 we provide several concrete data demonstrations that test the soundness of the Article's findings, using Texas as an illustrative example. We find that a significant driver of instability is the manner of employing ecological inference, or EI, to estimate candidate preference by race. Though EI is a valid family of estimation methods, it should be used with caution because of well-documented limitations in precision and untestable questions of model selection.[20] The authors do not defend their EI modeling choices or include any uncertainty estimates, generating instability that propagates through their workflow and implicates their analysis. For example, we count 51 seats (of 150) in the Texas state House that have demonstrably provided electoral opportunity for minority candidates of choice in the cycle following the 2010 Census. Chen and Stephanopoulos report that 46 seats currently meet their definition of minority opportunity district (MOD for short). But merely by toggling four settings between the authors' EI setup and alternative settings we commonly find in expert reports—while maintaining their precise definition of MOD and using the same R package they used to run EI—we were able to make the measured number of opportunity districts in the enacted plan itself vary from 34 to 51 seats, as shown in Figure 5. This does not mean that EI should be discarded, but its role in the Article's definition of MOD is far too central and too hard-edged. A definition that uses richer electoral history would be more robust and ultimately more meaningful than one built by pushing a single election through a black box of statistical inference.

We conclude with a look to the future, in which algorithms are fast becoming intertwined with governance. This brings cutting-edge scientific computation more and more into the legal mainstream, which both provides collaborative opportunities and an increasing onus to handle legal questions with scientific best practices.

Research Acknowledgment

A wide-ranging and fast-paced empirical research effort, such as was necessary to compile this Response, is not possible without a team. From our capable team of staff and affiliates of the MGGG Redistricting Lab at Tufts University who were thanked above, we want to particularly acknowledge the extraordinarily talented Parker Rule and Gabe Schoenbach. They did a deep-dive in the replication materials, curated data, wrote our parallel test code in R and Python and Julia, and collaborated on the design of experiments to make this Response possible.

[20] *See*, *e.g.*, Christopher S. Elmendorf, Kevin M. Quinn and Marisa J. Abrajano, *Racially Polarized Voting*, 83 U. Chi. L. Rev. 587 (2016); D. James Greiner, *Re-Solidifying Racial Bloc Voting: Empirics and Legal Doctrine in the Melting Pot*, 96 Ind. L. J. 447 (2011).

1 The Race-Blind Future?

1.1 The Scope of Proactive Protection

Race plays a singular role in American election law. Despite a constitutional prohibition against race discrimination in voting as early as 1870, discrimination has stubbornly persisted and so race has remained a key fault line in the development, implementation, and interpretation of election laws. From overtly racist literacy tests[21] and felon disenfranchisement[22] to more subtle forms of vote dilution,[23] voting laws have long limited the political participation and political power of communities of color and other minority groups across the country.

Some of the reasons for systematic underrepresentation are structural and function independently of gerrymandering. A chief example is at-large plurality voting, which is still used to elect many city councils, county commissions, and other local bodies across the country. In this system, any group with a majority can capture every seat at the expense of all other groups. Indeed, one reason why Congress mandated that members of the House of Representatives be chosen from single-member districts in 1842 was to provide for a system of representation that would produce outcomes more in line with voter preference between the political parties; that is, to produce more proportional outcomes.[24]

But winner-take-all districting itself tends to deal out representation far short of proportionality to virtually *all* minorities, from environmentalists in Alaska to Republicans in Massachusetts, as a matter of mathematics.[25] In fact, if district lines are drawn at random, a minority constituting one quarter of the population will frequently be entirely deprived of the control of even a single district.[26] Minority representation in a districted system thus depends on proactive measures. These proactive measures cannot simultaneously save every conceivable minority from underrepresentation or outright exclusion, which raises two crucial questions: (1) which minorities, if any, deserve proactive protection? and (2) how much action is necessary to offset the barriers to representation faced by these groups, especially when structural and intentional causes are intertwined?

The short answer to the first question is that racial minorities have long been singled out for particular attention. The Fifteenth Amendment to the U.S. Constitution explicitly prohibits the government from denying or abridging the right to vote "on account of race, color, or previous condition

[21] *See*, e.g., Davis v. Schnell, 81 F. Supp. 872, 880, *aff'd*, 336 U.S. 933 (1949) (holding Alabama's literacy test unconstitutional because "its main object was to restrict voting on a basis of race or color.").

[22] Hunter v. Underwood, 471 U.S. 222, 229 (1985) (striking down a felon disenfranchisement provision in Alabama's state constitution because it "was enacted with the intent of disenfranchising blacks.").

[23] Some examples of more subtle forms of racial discrimination in voting include moving from single-member districts to at-large voting or vice versa, changing elected positions to appointed positions, prohibiting "bullet voting," and vote dilution via cracking and packing when redistricting. *See, e.g.,* Allen v. State Bd. of Elections, 393 U.S. 544 (1969); Presley v. Etowah County Cmm'n, 502 U.S. 491 (1992).

[24] 5 Stat. 491 (1842).

[25] The political science literature on this topic, where this effect goes by the name of a "winner's bonus" or "seat bonus" for the majority, is too large to survey here. For just one important example, *see* Pippa Norris, *Choosing Electoral Systems: Proportional, Majoritarian and Mixed Systems*, 18 Int'l Pol. Sci. Rev. 297 (1997). For a few other key themes in the literature, *see* Moon Duchin, Taissa Gladkova, Eugene Henninger-Voss, Ben Klingensmith, Heather Newman & Hannah Wheelen, *Locating the Representational Baseline: Republicans in Massachusetts*, 18 Election L.J. 388 (2019).

[26] Duchin et al., *supra* note 25 (showing that for many statewide elections, Republicans in Massachusetts received over 30% of the vote but no district is possible with a Republican majority). This effect is explored *infra* Figures 2-3.

of servitude."[27] More generally, federal courts have identified race as a protected class. Owing to the long and often violent history of discrimination against racial minorities, their general political underrepresentation, and the legal determination that race is an immutable trait, the Supreme Court set out a mandate in the mid-1900s to attend to disparate treatment of racial groups in a wide range of contexts.[28] This "strict scrutiny" standard immediately places courts in a skeptical posture with respect to any government policy that intentionally creates racial differences. Against this backdrop, Congress has also mandated specific race-based protections for voting, first in the Civil Rights Acts of 1870[29] and 1871[30] that created a right of action in cases of bribery, intimidation and/or violence aimed at deterring individuals from voting based on their race, and provided severe fines and jail time for violations. Civil Rights Acts in 1957,[31] 1960,[32] and 1964[33] also protected against state and local voting laws that would discriminate along racial lines. Congress has yet to extend the same promise or protections to women, environmentalists, the poor, left-handed citizens, or other groups that are minorities or minoritized.[34]

1.2 Proportionality and its Discontents

Recognizing the special legal status of racial minorities leaves open the question of how much proactive protection is needed to offset the systematic sub-proportional effects of single-member districting. The Voting Rights Act of 1965, arguably the most important proactive voting measure ever enacted in the United States, dictates that racial minorities should have an equal opportunity to participate in the political process and to elect candidates of their choice. The ultimate goal of the VRA is to shield elections from racial discrimination and to ensure effective minority representation at all levels of government.

Chen and Stephanopoulos provide a detailed and accessible account of how courts have adopted a comparator of "rough proportionality" to evaluate whether minority political opportunity is equal to that of Whites.[35] In theory, a standard of rough proportionality might push legislators and other districting bodies to draw lines in a way that puts a near-proportional share of seats in reach for minority-preferred candidates to the greatest extent possible. As the Article notes, however, the rough proportionality standard has operated instead as a ceiling on minority opportunity.[36] In other words, the status quo of VRA practice has ensured that White voters will *never* be represented by less

[27] U.S. Const. amend. XV.

[28] *See, e.g.,* Korematsu v. United States, 323 U.S. 214, 216 (1944) ("It should be noted, to begin with, that all legal restrictions which curtail the civil rights of a single racial group are immediately suspect. That is not to say that all such restrictions are unconstitutional. It is to say that courts must subject them to the most rigid scrutiny."); United States v. Carolene Products, 304 U.S. 144, n.4 (1938) (calling for a "more searching judicial inquiry" in cases where the ordinary political process fails to address prejudice against "discrete and insular minorities.").

[29] 16 Stat. 140.

[30] 17 Stat. 13.

[31] Pub. L. No. 85-315, 71 Stat. 634.

[32] Pub. L. No. 86-449, 74 Stat. 89.

[33] Pub. L. No. 88–352, tit. I, 78 Stat. 241.

[34] *See*, e.g., Helen Hacker, Women as a Minority Group (1951).

[35] Chen & Stephanopoulos, *supra* note 5, at 874-877.

[36] *Id.* at 921 (referring to the proportionality baseline as "an upper limit to how much representation minority groups can legally claim."); *see also* Katz, et al, *supra* note 3; Stephanopoulos, *supra* note 4, at 168 (explaining that a "more accurate statement of the theory [of rough proportionality] is that minority voters should be able to elect their preferred candidates to the extent permitted by their geographic distribution up to a ceiling of proportionality.").

than their share of the population while minority voters almost always will. But even a proportionality target is far from a perfect realization of the loftiest goals of the VRA. The proper goal of the VRA is real political power for minority groups, which is a stubbornly local and particular matter, and is therefore hard to capture in a mere count of districts that pass any quantitative threshold test.[37]

These weaknesses in the VRA status quo are not what drives the authors to explore a race-blind alternative. Instead, they focus on a different set of critiques to motivate their project.[38] Though a proportionality standard is intuitive and easy to measure,[39] the Court has warned that it can lead to conflation of political outcomes with political opportunities,[40] and critics argue that it puts undue stress on race in violation of the Equal Protection Clause of the 14th Amendment.[41] These critics also note that drawing designer districts to approach proportionality can result in non-compact districts that split counties and cities.[42]

The authors do not confront these critiques of proportionality-based standards in any depth, nor do they endorse them.[43] They perceive no obligation to argue that any standard for interpreting the VRA is better than any other, including the novel standard that they articulate at great length: "to be clear, in this Article, we are not advocating for any particular legal interpretation of the VRA."[44] Instead, "we are merely analyzing the empirical consequences of the hypothetical adoption of a race-blind baseline for minority representation under section 2."[45]

Because Chen and Stephanopoulos are so restrained in articulating their normative stance, some will read their Article in line with their stated intent: as purely descriptive of the racial landscape, taking the idea of race-blind districting literally and seriously to its conclusions. Other readers may not find the treatment so neutral, instead reading the Article as an endorsement of the approach that it delineates, at least as a compromise that saves the VRA from a complete dismantlement by the Roberts Court. Few readers are likely to take the authors to be warning of the potential of dire consequences to this particular computer-centric approach.

[37] *See*, e.g., *see* Justin Levitt, *Quick and Dirty: The New Misreading of the Voting Rights Act*, 43 Fla. St. U. L. Rev. 573, 578 (2016) ("Proper focus on local nuance and meaningful political power – as precedent demands – can restore the Voting Rights Act to a vehicle for fighting both racial discrimination and racial essentialism.").

[38] Chen & Stephanopoulos, *supra* note 5, at 871-872.

[39] *See*, e.g., *Holder v. Hall*, 512 U.S. 874, 928 (1994) (Thomas, J., concurring) (referring to proportional representation as "the most logical ratio for assessing a claim of vote dilution" and noting that other standards would have "less intuitive appeal"); *Thornburg v. Gingles*, 478 U.S. 30, 84 (1986) (O'Connor, J., concurring) ("[A]ny theory of vote dilution must necessarily rely to some extent on a measure of minority voting strength that makes some reference to the proportion between the minority group and the electorate at large.")

[40] Johnson v. DeGrandy, 512 U.S. 997, 1020 (1994) ("minority voters are not immune from the obligation to pull, haul, and trade to find common political ground").

[41] *See* Chen & Stephanopoulos, *supra* note 5, at 874 (citing to Justice Thomas's concurring opinion in *Holder v. Hall*, 510 U.S. 874 (1994) and the majority opinion in *Shaw v. Reno*, 509 U.S. 630 (1993) (referring to remedial racial districting as "political apartheid" that may "balkanize us into competing racial factions.").

[42] Chen & Stephanopoulos, *supra* note 5, at 877.

[43] The authors offer a brief summary of potential responses in footnotes 56, 63, and 70 but remain studiously agnostic about the merits of the critiques. *See supra* note 5 at 883 ("To be clear, we do not endorse the conservative objections to the proportionality baseline.").

[44] *See supra* note 5 at 872, fn. 21.

[45] *Ibid.*

1.3 The Limits of Race-Blindness

This neglected question – can the aims of the VRA be served by a race-blind baseline? – should be seen as a pressing matter, since the goal of the VRA is to "hasten the waning of racism in American politics"[46] and the protocol delineated in *The Race-Blind Future of Voting Rights* could very well hasten the waning of political power for people of color at all levels of government instead.[47]

Battling the anti-minoritarian tendencies of districts to generate adequate opportunity for minority groups, all without attention to race, is a challenge indeed.[48] In the current regime, this often leads to elaborate post hoc claims of having "backed in" to a satisfactory demographic arrangement across districts by happy circumstance, in a kind of race-blind theater.[49]

As we explain in the next Part, the power of the ensemble method is to hold the human and political geography of a jurisdiction fixed while varying district lines. It therefore has the unique capacity to measure the extent of the control exercised by the mapmaker. But laying randomized lines over fixed human geography bakes in the effects of residential patterns which may themselves be driven by discriminatory policy and which certainly reflect histories of racism and prejudice. Residential patterns have an enormous impact on the landscape of possible districted outcomes.

Does the human geography interact with the system of election in a way that enables minority groups to be agentic – to have an opportunity to elect? Instead of being satisfied with letting the chips fall where they may with respect to the interactions of residential segregation and compact, contiguous, equipopulous districts, the logic of the VRA requires us to interrogate the system itself. That is because districts may indeed secure adequate opportunity, but only when mindfully drawn. If proactive districting is too race-conscious for the 21st century Court, as Chen and Stephanopoulos predict, then plurality districts themselves must be implicated. We do not share Chen and Stephanopoulos's view that race-blind benchmarks are "the only alternative to proportionality currently on the table."[50]

Finally, the "race-blind" approach outlined by the authors is anything but blind. To check compliance in their framework (confirming that a proposed map is at the 50th percentile of a batch of neutral alternatives in its number of MODs) requires a detailed use of racial data and the same ecological inference machinery that is used in the measurement of racial polarization in the *Gingles* framework that the authors profess to leave behind. So even checking compliance requires statistical modeling of vote preferences by race. This makes it race-conscious in far deeper ways that the mere use of population proportionality, and it trades a simple and manageable barometer for a compli-

[46] Johnson v. DeGrandy, 512 U.S. 997, 1020 (1994).

[47] Chen & Stephanopoulos, *supra* note 5, at 924 (noting the "dramatic implications" of a race-blind baseline, one where "most Section 2 suits seeking the formation of new opportunity districts would fail.")

[48] For a discussion of the tension between requirements that race discrimination be intentional while remedies be blind to race, *see*, e.g., Ian Haney-Lopez, *Intentional Blindness*, 87 N.Y.U. L. Rev. 1779 (2012).

[49] For other examples of VRA theater, in which redistricting actors proclaim one set of data-driven aims while targeting another set of political and racial aims, *see* Levitt *supra* note 37, at 605 (noting that "a state may have incorrectly attempted to comply with section 2 and yet still have drawn lines that provide an equal opportunity for minority voters to elect candidates of choice.") and Shelby County v. Holder, 679 F.3d 848, 885 (D.C. Cir. 2012) (Williams, J., dissenting) (criticizing the reverse engineered coverage formula in § 4(b) of the VRA by noting that "sometimes a skilled dart-thrower can hit a bull's eye throwing a dart backwards over his shoulder . . .")

[50] Chen & Stephanopoulos, *supra* note 5, at 879. *See* Lani Guinier, The Tyranny of the Majority 121 (1994) ("It's districting in general – not race-conscious districting in particular – that is the problem."). Guinier and others have looked to alternative voting systems precisely for their promise in this regard, and ranked choice voting in particular is currently seeing a surge of interest, from Maine to Alaska. *See*, Gerdus Benade, Ruth Buck, Moon Duchin, Dara Gold, and Thomas Weighill, *Ranked Choice Voting and Minority Representation*, Preprint. *Soon to be available at* mggg.org/STV.

cated and contestable alternative. This new alternative relies on more than just the measurement of political preferences by race; the second major ingredient is the comparator ensemble of valid plans. We turn to that methodology now.

2 Ensemble Methods: Arguing from Alternatives

An *ensemble* of plans is a collection, or sample, from among all possible districting plans. If the purpose of an ensemble is to serve as a basis for comparison, then it should be fashioned so as to be representative of the universe of valid plans—as we will discuss in the next Part, this requires that the samples are drawn with a clear weighting, and that the samples are large enough to draw sound statistical conclusions.

Plans are typically assessed by *summary statistics*, like the number of seats with a Democratic advantage, the number of competitive seats, one of a variety of compactness scores, or, here, the number of "minority opportunity districts." These statistics can be integer-valued, like anything denominated in seats or districts, or they can be essentially continuous, like many compactness metrics or the *efficiency gap* partisan metric. If we focus on a single statistic and record the value achieved by each plan, ensembles will often generate a bell-shaped distribution. That familiar bell curve visual can help us think through what is the normal range and what is vanishingly rare in the universe we have specified. The bulk of that distribution can be treated as a *baseline range* for the statistic. The tails of the curve contain the outliers—finding that a plan falls in the vanishing outer reaches of the sample is a strong indicator that some element of the mapmaker's intent was not accounted for in the ensemble design.

In other words, ensembles generate empirical distributions that have descriptive power. Districting ensembles do not answer our normative questions for us, although they can be extraordinarily useful for addressing normative inquiries. For example, some states have enshrined in their rules the norm that political agents should not be excessively or "unduly" partisan when drawing districts. To investigate whether a plan is in line with this norm, we can survey the summary statistics for an ensemble of partisan-neutral plans (i.e., made with zero partisan data). Some proposed plans will have partisan properties that are typical of the ensemble, while others will fall in the long tails of the empirical distribution. The ensemble furnishes evidence for evaluation by the lights of the norm without ever providing a normative ideal by imagining that there is some most partisan-neutral plan.[51]

Many norms for redistricting are framed negatively or proscriptively: race should not predominate over traditional districting principles;[52] voting rights should not be denied or abridged on account of race;[53] the shapes of districts should not be bizarre, eccentric, or irrational.[54] But very few of these thou-shalt-nots come with a corresponding "shalt" that has any clarity or precision. An exception is overall malapportionment, where population equality across districts is the positive norm. Vote dilution on the basis of group membership is a crucial instance of the lack of a prescribed ideal.

[51] Common Cause v. Rucho, 139 S. Ct. 2484 (2019) (Kagan, J., dissenting).

[52] Miller v. Johnson, 515 U.S. 900 (1995).

[53] U.S. Const. amend. XV; Voting Rights Act § 2 (codified at 52 U.S.C. § 10301).

[54] Shaw v. Reno, 509 U.S. 630 (1993); Bush v. Vera, 517 U.S. 952 (1996).

Since at least the 1940s, the court has struggled to discover an undiluted baseline for the weight of a vote: *what is the neutral state of affairs, absent gerrymandering?*[55]

In practice, this means that ensembles are useful for identifying whether a particular districting plan might be disallowed according to statutory or constitutional guidance because it distributes the group members across the districts in a way that is far out of line with the neutral tendencies of geographic partitions.[56] Using ensembles to flag outliers does not commit us to any view on which of two competing options from the bulk of the ensemble is better or closer to ideal. In particular we will argue that the mean (average), the median (50th percentile), or the mode (most frequent) values of any statistic derived from an ensemble have no inherent claim on best quality. In part, this is because only a subset of the rules is amenable to quantification, and therefore can be taken into account by algorithmic methods. We can certainly search for what is most typical or most frequent under blind application of the quantifiable subset of the rules, but we would need significant additional reasons to hold it up as ideal.

2.1 Judge Easterbrook's Dream and its Antecedents

The Race-Blind Future of Voting Rights builds its analysis on a proposal of Judge Frank H. Easterbrook. Easterbrook's own formulation in *Gonzalez v. City of Aurora* begins with algorithmic ensembles: "Today, however, computers can use census data to generate many variations on compact districts with equal population." From there, both outlier logic and the primacy of the median get billing:

> Suppose that after 1,000 different maps of Aurora's wards have been generated, 10% have two or three "safe" districts for Latinos and the other 90% look something like the actual map drawn in 2002: one safe district and two "influence districts" where no candidate is likely to win without substantial Latino support. Then we could confidently conclude that Aurora's map did not dilute the effectiveness of the Latino vote. But suppose, instead, that Latinos are sufficiently concentrated that the random, race-blind exercise we have proposed yields three "Latino effective" districts at least 50% of the time. Then a court might sensibly conclude that Aurora had diluted the Latino vote by undermining the normal effects of the choices that Aurora's citizens had made about where to live.[57]

More than thirty years earlier (and in a different spirit), Bernard Grofman, Michael Migalski, and Nicholas Noviello anticipated the same logical move in 1985, complete with the computational turn – only with the mode in place of the median.

[55] *See* Heather K. Gerken, *The Right to an Undiluted Vote*, 114 Harv. L. Rev. 1663, 1723 (2001) ("The right to an undiluted vote does not fit easily into either a group rights or an individual rights category. While it is certainly true that an individual's right is linked to the status of the group, that is because the injury being asserted by an individual is the inability to aggregate her vote. The only way to measure that individual harm is to evaluate the position of other group members with whom she wishes to coalesce.")

[56] This "outlier analysis" has been the focus of recent litigation about partisan gerrymandering in state and federal courts. *See*, e.g., League of Women Voters v. Pennsylvania, 178 A.3d 737, 828 (Pa. 2018) (Baer, J., concurring in part) ("a petitioner may establish that partisan considerations predominated in the drawing of the map by, *inter alia*, introducing expert analysis and testimony that the adopted map is a statistical outlier in contrast with other maps drawn utilizing traditional districting criteria"); Rucho v. Common Cause, 139 S. Ct. 2484, 2517-18 (2019) (Kagan, J., dissenting) ("the plaintiffs demonstrated the districting plan's effects mostly by relying on what might be called the 'extreme outlier approach.'")

[57] 535 F.3d 594, 600 (2008).

> Social scientists have developed computer methods to create hypothetical single-member-district plans satisfying specified constraints. By generating a large number of such plans, we can determine the expected racial representation under the modal single-member-districting scheme and compare a minority group's actual or anticipated ability to elect representation of its choice under the actual plan with the outcomes expected under neutrally drawn smd plans.[58]

And even a few years before that, the landmark 1982 paper of Blacksher and Menefee, from which the Gingles factors were plucked by the Supreme Court in short order, laid out a remarkably similar vision but with a still different spin.

> [T]he relevant question should be whether the minority population is so concentrated that, if districts were drawn pursuant to accepted nonracial criteria, there is a reasonable possibility that at least one district would give the racial minority a voting majority.[59]

For this purpose, the authors tell us, "computer-assisted mathematical models" would be sufficient but are not necessary to answer the question.

It is worth noting that Judge Easterbrook's use of the median leaves room for shades of gray: if the median has three effective districts and the proposed plan has one effective and two mere influence districts, he tells us that signs point to dilution. But what if the median plan has two safe districts and one barely over the effectiveness threshold, while the proposed plan has two safe districts and one barely under the effectiveness threshold. Is this as clear a case? Judge Easterbrook does not tell us what a court might sensibly conclude. But the Chen–Stephanopoulos framework, because it works in integers and yes/no answers, declares this to be a full-fledged case of underrepresentation in the proposed plan. Phrased differently: Judge Easterbrook only comes to conclusions when a proposed plan is sufficiently far from the median. He does not tell us whether to prefer the median to its near neighbors or how far from the median a plan can permissibly be.

In another example, Justice Kagan's dissent in *Rucho* also calls on an ensemble median for judging partisan gerrymanders. "And we can see where the State's actual plan falls on the spectrum—at or near the median or way out on one of the tails? The further out on the tail, the more extreme the partisan distortion and the more significant the vote dilution."[60]

Against this background, Chen and Stephanopoulos have elected to rely heavily on the median values of their ensembles for their top-line conclusions. For example, they report that Alabama's twenty-seven black opportunity districts "exceeds by four the number of black opportunity districts in the *median* simulated map,"[61] that "the enacted plan [in Illinois] has twenty-one black opportunity districts: two more than the *midpoint* of the simulations,"[62] and that "the enacted plan [in Florida],

[58] Bernard Grofman; Michael Migalski; Nicholas Noviello, *The Totality of Circumstances Test in Section 2 of the 1982 Extension of the Voting Rights Act: A Social Science Perspective*, 7 L. & Pol'y 199, 224 (1985).

[59] James U. Blacksher & Larry T. Menefee, *From Reynolds v. Sims to City of Mobile v. Bolden: Have the White Suburbs Commandeered the Fifteenth Amendment*, 34 Hastings L.J. 1 (1982).

[60] Rucho v. Common Cause, 139 S. Ct. 2484, 18 (2019) (Kagan, J., dissenting). *See also* Moon Duchin, *How to Reason from the Universe of Maps (The Normative Logic of Map Sampling)*, Election L. Blog (July 5, 2019), electionlawblog.org/?p=106069.

[61] Chen & Stephanopoulos, *supra* note 5, at 908 (emphasis added).

[62] *Id.* at 910 (emphasis added).

on the other hand, has seven [Hispanic opportunity] districts, or three fewer than the *midpoint* of the simulations."[63] With this choice they go farther than any of these previous authors, including Easterbrook himself. The median stands alone with no notion of a baseline range; it is held up as a standard from which plans that deviate by even one legislative seat will receive a label of over- or underrepresentation.[64] This slippage from a negative to a narrow positive norm for ensemble methods leads to strange conclusions.

2.2 The Tyranny of the Median

To see why a strong focus on the median value is problematic, suppose we have a coin and we want to determine if it is a "fair coin" – that is, whether it is equally weighted between heads and tails or exhibits a structural bias toward one or the other outcome. There is a basic test for this: we flip the coin repeatedly and record the results. To fix terminology, let's say a *trial* is made by conducting 1000 coin flips and recording the number of heads, so that the possible outcomes range from 0 to 1000. The evidence provided by one trial about whether the coin is fair is similar in some ways to the evidence provided by superimposing one set of election results on a districting plan and recording the plan's summary statistics.

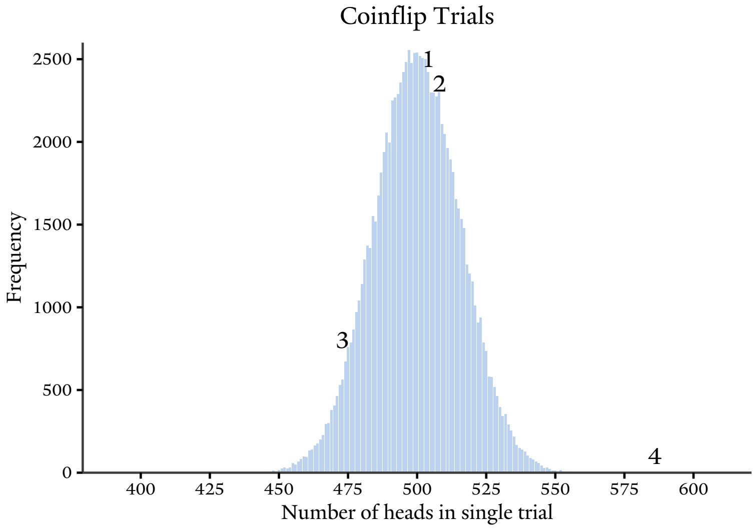


Figure 1: This histogram shows the outcome of 100,000 simulation trials with a true fair coin, approximating a familiar bell curve. If we want to test four coins for fairness, suppose we flip each one 1000 times. Coin 1 gives 504 heads; Coin 2 gives 508 heads; Coin 3 gives 473 heads; and Coin 4 gives 586 heads. What can we conclude?

Of course, in this case it is clear from basic probabilistic reasoning that the expected number of heads in 1000 flips is 500 if the coin is fair – and that would also be the median and the mode outcome in a very large experiment with many trials. The mean will in all likelihood be a non-integer

[63] *Id.* at 913 (emphasis added).

[64] *Id.* at 942, fig. 16 and Table 1 in Appendix C.

value just to one side or the other of 500. We can see how this plays out for progressively larger samples in Table 1, where we include two samples at each size as a reminder of variability.

# Trials of 1000 Flips	10	10	1000	1000	100,000	100,000	10,000,000	10,000,000
Median	501	495.5	500	500	500	500	500	500
Mode	482	488	495	496	496	500	500	499
Mean	503	497.3	500.246	499.1	499.555	499.564	499.993	500.003
Max	529	515	546	550	564	570	592	583
Min	482	482	456	445	432	428	415	412

Table 1: Each column is a sample of outcomes from repeated trials with an actually fair coin (up to the limits of a computer's ability to randomize). The more trials in our sample, the more predictable the results. (Note that if there is a tie for the most frequently observed value, the smallest of these values is reported as the mode.)

We would be justified in concluding that a coin that flipped heads 586 times out of 1000 is unlikely to be fair. But if my coin came up heads 508 times and your coin came up heads 504 times, we would not be reasonably able to conclude that your coin is fairer than mine. This would be an error: rather, both coins have behavior that is consistent with fairness, since the outcomes are well within the reasonable range for a fair coin. Even stranger would be to require that any legally permissible fair coin should pass the test of having exactly 500 heads in its official trial—after all, this occurs only about 2.5% of the time even for a perfectly fair coin.

This fuzziness is of course inconvenient in the search for a manageable legal standard: clear goals and clear thresholds are preferable when possible. But elevating the median number, and suppressing talk of a *reasonable range* of outcomes, leads to fundamental problems.

2.3 Example: Distribution of BVAP

To see ensembles in action and their power to illustrate the interplay between human geography and the mathematics of districts, we turn to our first data demonstration.[65] The Article sets out to study 20 states (but ultimately excludes New Jersey due to unexplained "unreliable ecological-inference estimates").[66] For each of those twenty states and each level of districting, we have created two million districting plans that are compact and contiguous, with each district always within 2% of ideal size, using the method described in the next Part of this Response.[67] Figure 2 shows the counts of majority-Black districts observed in those plans, vividly illustrating the war between proportionality and plurality districts.[68] Not once in 114,000,000 attempts across the states and levels did a plan

[65] All ensembles that we generate in this Response use the implementation of ReCom in the high-performance programming language Julia, which is publicly available at github.com/mggg/GerryChainJulia.

[66] Chen & Stephanopoulos, *supra* note 5, at 892, fn. 145. The pressures of the authors' one-size-fits-all modeling begin to show in these kinds of exceptions. The authors also hard-coded various exceptional cases in their programs, for instance by manually loosening the intact-county threshold and the compactness threshold in some states.

[67] For these runs, we use ensembles built from Census block groups, since we do not need electoral data. We have provided confirmation data from selected states showing that using blocks, block groups, or precincts gives similar results.

[68] The shaded range shows the seats outcomes *ever* observed in the ensemble, regardless of its frequency, and whole numbers of districts are shown as small dots. Thus, for example, of the two million maps made for Louisiana's congressional delegation, just *six* districting plans included a majority-Black district. The remaining 1,999,994 plans had zero majority-minority districts. Despite its extremely low probability, Figure 2 includes this one seat. This is a good reminder that subsampling, or skipping over many plans to thin the ensemble, may not be the best practice for these ensemble applications,

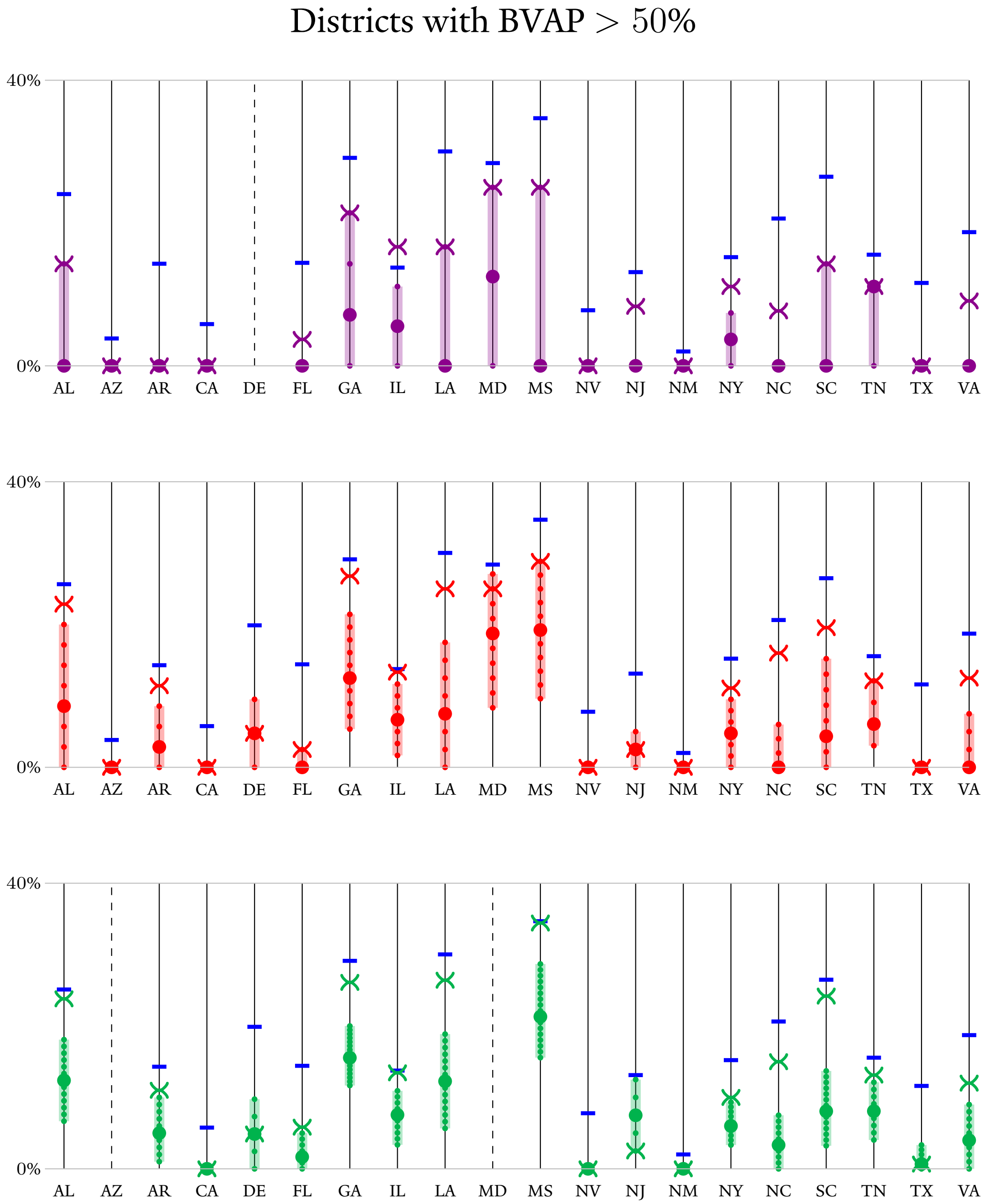


Figure 2: Shortfalls from proportionality, viewed with comparator ensembles of two million districting plans for Congress (top), state Senate (middle), and state House (bottom). Blue line (–): proportionality (BVAP share). Bracket (×): share of majority-Black districts (BVAP > 50%) in enacted plan. Colored dots and range: share of majority-Black districts in neutral ensemble plans, with large dot marking median. Note: Delaware has a single seat in the U.S. House of Representatives and is thus not included in the top panel. Arizona and Maryland employ multi-member districts in their state lower House and are thus not included in the bottom panel.

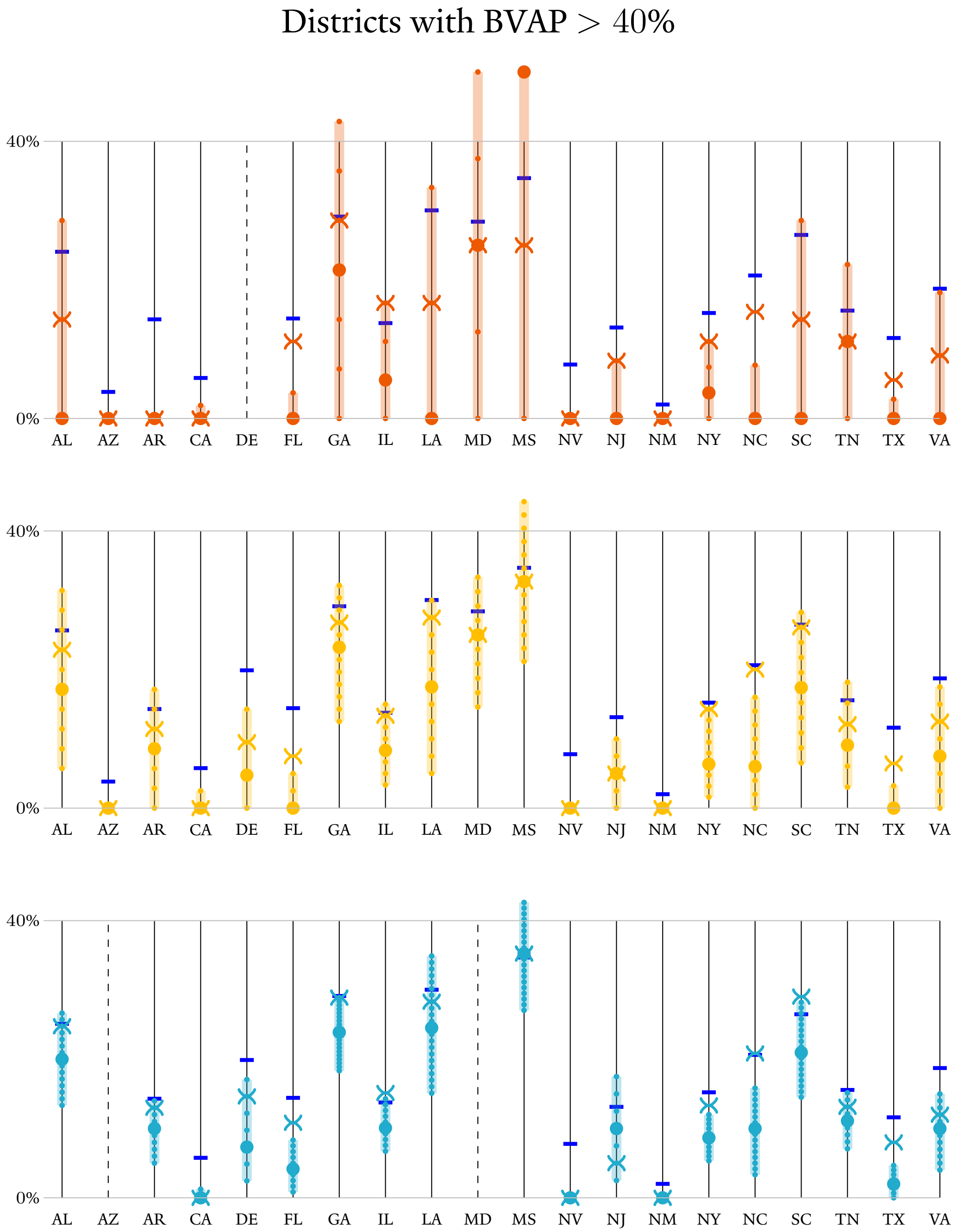


Figure 3: Congressional (top), state Senate (middle), and state House (bottom) districting and Black population. Blue line: proportionality (BVAP share). Bracket: share of districts with BVAP > 40% in enacted plan. Colored dots and range: share of districts with BVAP > 40% in neutral ensemble plans, with large dot marking median. Note: Delaware has a single seat in the U.S. House of Representatives and is thus not included in the top panel. Arizona and Maryland employ multi-member districts in their state lower House and are thus not included in the bottom panel.

made with no regard to race have a number of majority-Black districts that is proportional to the state's Black population share. And in fact for Alabama (7 districts), Louisiana (6 districts), Mississippi (4 districts), and South Carolina (7 districts), all with Black populations over 25%, the *median* number of majority-Black congressional districts is zero.[69] Strikingly, in Alabama, Louisiana, and South Carolina, the median is still zero even if we shift the frame to districts with 40% Black population (Figure 3). This is the sense in which random districts are punishing to minorities—they can often produce statistics not that different from the state overall, and will not happen on higher concentrations unless by design.

The design caveat is important: it is fairly easy to make majority-Black districts if one tries, and the figure shows that enacted plans are often right at the top of the ensemble, or higher still.[70] This comports with many observers' suspicion that states often use the crude device of demographic percentage as a substitute for a more nuanced VRA compliance.[71] It just doesn't happen by chance.

This showing is unsurprising—it has long been understood that randomness does not lend itself well to creating pluralities from minority populations,[72] as we continue to remind the reader—but the extent and consistency is remarkable. With BVAP $> 40\%$ districts in view instead of BVAP $> 50\%$, the story changes dramatically. Suddenly neutral ensembles can smash through the proportionality ceiling and the ensemble routinely includes plans that outmatch the enacted plans. But this is only if we refuse to maintain a laser focus on the median.

Demographics are not voting destiny and below, following the VRA itself, we will shift the focus to *electoral effectiveness* rather than raw demographics. But we will still have no more reason for believing that the ensemble median is ideal or fair than we do here.

even though it is frequently used in other domains of applied statistics. If we only sample every 10,000 plans visited by the random walk, we may miss rare events entirely and subvert the exploration features of Markov chain sampling.

[69] We also note that as the granularity of districting gets finer (more and smaller districts, like in state Houses), the range of seat-share outcomes observed in a neutral ensemble is reliably narrower, but the mean and median seat-share creep higher. Rodden and Weighill have a similar finding that increased granularity results in lower variance in their study of scale effects in Pennsylvania districting. However, fascinatingly, they find that in the specific case of Pennsylvania and a partisan measure rather than the racial measure considered here, the ensemble *average* is stable at every scale—there is no "sweet spot" of district size for Democrats in Pennsylvania. Jonathan Rodden & Thomas Weighill, *Political Geography and Representation: A Case Study of Districting in Pennsylvania*, *in* Political Geometry (Moon Duchin & Olivia Walch eds., forthcoming 2021). *See* mggg.org/GerryBook.

[70] It is crucial to remember that if race is considered among proactive redistricting goals, it is easy to outperform a neutral algorithm, and indeed it is often easy to outperform the enacted plans. For an automated search technique for majority-minority districts, see Sarah Cannon, Ari Goldbloom-Helzner, Varun Gupta, JN Matthews, and Bhushan Suwal, *Voting Rights, Markov Chains, and Optimization by Short Bursts*, Preprint (2020). *Available at* arxiv.org/abs/2011.02288.

[71] This is particularly sharply observed by Justin Levitt. Levitt, *supra* note 37, at 575-576 ("In some circumstances, the jurisdictions' reliance on crude demographic targets over-concentrates real minority political power; in other circumstances, it under-concentrates real minority political power. In still other circumstances, the real political effects are unclear, because the lure of the demographic assumption means that nobody has bothered to examine the real political effects.") (internal citation omitted).

[72] *See*, *e.g.*, Grofman, *supra* note 2.

3 Ensuring Representative Samples

3.1 Samples, not Simulations

Chen and Stephanopoulos repeatedly refer to their districting plans as "simulations," as they have in previous articles and litigation materials.[73] We start by reorienting the language to help highlight the task at hand.

When a measurement of a physical or agent-based event is not possible directly, when we wish to abstract out some inconvenient features that make measurement messy, or when we wish to repeat trials more times than there are available observations, we must make use of a simplified simulation event, often outsourced to a computer. The coinflip model from above is a *simulation*: the random number generator in Python is abstracting the physical flip of a fair coin. When you have a model of voter behavior and you run it many times, you are conducting a *simulated* election, since no votes were actually cast. When you use red and blue squares to model the states of magnets and set up a lattice of them to look for interactions, you are *simulating* a magnetic field.

On the other hand, a partition of Census blocks into connected pieces is not a simulated districting plan; it is an actual districting plan. If you generate many of these, you are *sampling* from the universe of possible districting plans. Calling this process simulation sets up a mistaken (if popular) analogy with statistical physics and agent-based modeling. Our proposed language shift comes with a salutary reminder: if the goal is representative sampling of plausible, valid plans, this brings with it a clear mandate to weight the observations appropriately so as to counteract various forms of sampling bias. We will see that the Article thoroughly conflates several conceptually distinct things that computers can do: provide examples, seek plans with better scores of some kind, or attempt representative sampling.

3.2 Random Walks and "Recombination"

Imagine the universe of all possible connected, population-balanced districting plans that satisfy a state's requirements. It turns out that this space of valid plans is quite large. Justice Alito memorably mused that there might be a hundred, or even thousands of alternatives.[74] In fact, the number of competing plans in a full-scale redistricting problem smashes past trillions of trillions and is likely in the range of googols (10^{100}), which means that a comprehensive survey of these plans is impossible, even for a quantum computer.

While it cannot be fully constructed, this vast space can still be explored. The mathematics literature provides an enormously useful tool called *Markov chains*: iterative processes that explore a *state space* (a universe of all possibilities) using a transition rule for moving from position to position.[75]

[73] *See* Jowei Chen, *The Impact of Political Geography on Wisconsin's Redistricting: An Analysis of Wisconsin's Act 43 Assembly Districting Plan*, 16 Election L.J. 443 (2017), Jowei Chen & David Cottrell, *Evaluating Partisan Gains from Congressional Gerrymandering: Using Computer Simulations to Estimate the Effect of Gerrymandering in the U.S. House*, 44 Electoral Stud. 329 (2016); Jowei Chen & Jonathan Rodden, *Cutting Through the Thicket: Redistricting Simulations and the Detection of Partisan Gerrymanders*, 14 Election L.J. 331 (2015).

[74] "So you've got – let's say you've got 100 maps or you might even have 25. I think you probably have thousands." Transcript of Oral Argument at 43. Rucho v. Common Cause, 139 S.Ct. 2484 (2019) (No. 18-422).

[75] By definition, a Markov chain is a random walk without memory, meaning that the position at time $n+1$ is governed by a probabilistic choice based only on the location at time n and not on the previous history. Many kinds of dynamical system

In scientific applications, there is a suite of practical Markov chain techniques going by the name of MCMC, or Markov chain Monte Carlo. In MCMC, scientists typically prescribe a desirable target distribution where some of the measurable attributes are weighted in a known way, then collect samples to observe the values of other attributes. This is ideal for the redistricting use case. We can survey the local rules of redistricting and design a distribution tailored to the requirements and preferences encoded in the rules. For instance, in our runs below, we will treat contiguity as a requirement: all plans must have connected districts. On the other hand, we will treat compactness as a preference: districts with more interior connectivity and shorter boundaries will be weighted higher than those with spindly limbs and bottlenecks.[76] To use MCMC for sampling, we run chains for a long time as we endeavor to collect samples that are representative of the target distribution. Eventually, the sample reaches stationarity: the "bell curve" stops changing and a representative sample is achieved.

Though this is fast becoming the leading method of generating plans for comparison, this was not always the mechanism of randomized redistricting – and representative sampling was not always the goal. In the 1960s, early computer redistricting packages did not seek representativeness, but optimization.[77] And even in the last ten years, quite a few political science publications[78] and expert reports[79] have been based on a very different style of district generation that we will name a "Petri dish" method: the small units of a state are given initial labels, and these proto-districts then merge and grow until they fill out the state with the right number of districts, like bacteria cultures growing in a plate. To create desired properties in the output plans, ad hoc adjustments are made to the merging rules. The resulting plans come with no theory describing their distribution and their authors present no account of the extent to which one kind of plan might tend to appear more often than another. For instance, a merging instruction meant to promote compactness could easily cause a certain two counties to be kept together in nearly every plan generated by the process, though their association has nothing to do with compactness per se.

A big jump in sophistication from the Petri dish ensembles came with the shift to MCMC, starting with the refinement of random walk methods based on a "Flip" step. A Flip chain begins with a complete districting plan and alters it slightly at each move, by reassigning one or a small number of its units. If carefully designed, Flip chains can have the property that they will converge in the

have steady states; Markov chains are remarkable because, when designed carefully, there is a unique steady-state distribution for the system, and the random walk process beginning at any initial configuration will always converge to it. This means that the empirical distribution drawn from a large enough sample of observations will converge to the same long-term shape, no matter what the initial position.

[76] For an overview of court approaches to compactness before and during the Shaw line of cases, see Richard Richard H. Pildes & Richard G. Niemi, *Expressive Harms "Bizarre Districts," and Voting Rights: Evaluating Election-District Appearances after Shaw v. Reno*, 92 MICH. L. REV. 483, 484 (1993) (noting that compactness violations are found "[w]hen physical geography is stretched too thin"). For a discussion of how ReCom compactness fits into the legal history, see Moon Duchin & Bridget Eileen Tenner, *Discrete Geometry for Electoral Geography*, ARXIV (Aug. 15, 2018). *Available at* arxiv.org/abs/1808.05860.

[77] Early work of this kind is cited by Chen & Stephanopoulos, *supra* note 5, at 884-886, though perhaps without realizing that these 1960s examples are from a different family of algorithms.

[78] *See supra* note 73.

[79] "Petri dish" methods have been used by Dr. Chen in numerous court cases, including LWV of Florida v. Detzner (Fla. 2d Judicial Cir. Leon Cnty. 2012); Romo v. Detzner (Fla. 2d Judicial Cir. Leon Cnty. 2013); Missouri NAACP v. Ferguson-Florissant Sch. Dist. and St. Louis Cty. Bd. of Elec. Comm'n (E.D. Mo. 2014); Raleigh Wake Citizens Association v. Wake County Board of Elections (E.D.N.C. 2015); Brown v. Detzner (N.D. Fla. 2015); City of Greensboro v. Guilford Cty. Bd. of Elec., (M.D.N.C. 2015); Georgia State Conference of the NAACP v. State of Georgia (N.D. Ga. 2017); LWV of Michigan v. Johnson (E.D. Mich. 2017).

long term to a steady state.[80] Unfortunately, at the scale of a real redistricting problem, Flip chains converge very slowly and encounter bottlenecks that make it difficult to be confident that they are exploring fully and not getting stuck in a corner of the search space. The Recombination (or ReCom) algorithm used by Chen and Stephanopoulos in the present Article was developed by the research team of Duchin, DeFord, and Solomon to get around these performance obstructions. It is a highly efficient graph algorithm that reassigns hundreds of units at each step, targeting a steady-state distribution in which the likelihood of drawing a particular valid plan is directly proportional to a certain explicit measure of compactness, with no dependence on hidden factors.[81]. This is a question that one should ask of all algorithms: in making distinctions (in redistricting no less than in assigning credit scores or recidivism risk), do the outputs depend only on the legitimate inputs in transparent ways?

At first blush, the distributional question – how to weight possible plans when sampling – might seem easy: simply take all valid plans and weight them equally. However, this does not work to produce good samples, because there are astronomically more non-compact plans than compact ones. And we can't just threshold the allowed compactness; if all are weighted the same, virtually the entire sample will be at the worst allowable level.[82] While other preference factors can be added to a ReCom run, having "eyeball" compactness fall in a reasonable range is built in. This means that compactness does not have to be manually thresholded as the authors do in the Article, where they reject plans in which the average Polsby-Popper score of a district is even the slightest bit worse than the enacted plan.[83] Indeed, the authors begin with the algorithmic engine of ReCom and add numerous flourishes that serve to negate its hard-won theoretical selling points.[84] In particular, it is completely unclear what distribution on districting plans they seek to sample from, and indeed there is no indication that they are attuned to the importance of that question.

In short, not all algorithms are created equal, and it is quite surprising to read spanning-tree

[80] Since 2018, Chen has incorporated Flip chains into his expert work, but only in a hill-climbing manner which is designed for optimization, not representative sampling. *See* Common Cause v. Robert A. Rucho (M.D.N.C. 2018); Whitford v. Gill (W.D. Wisc. 2018). The work from the research teams of Duke's Jonathan Mattingly and Harvard's Kosuke Imai is particularly notable in targeting a prescribed distribution. For an extended discussion of challenges and sophisticated fixes for Flip chains, *see* Daryl DeFord and Moon Duchin, *Random Walks and the Universe of Districting Plans*, *in* POLITICAL GEOMETRY (Moon Duchin & Olivia Walch, eds., forthcoming 2021). *See* mggg.org/GerryBook.

[81] To be precise, the stationary probability of selecting a plan in the ReCom chain is approximately proportional to its spanning tree score, a measure of compactness that draws from clustering theory. *See* DeFord, Duchin & Solomon, *supra* note 9 and Duchin & Tenner *supra* note 76. A small adjustment to the Markov procedure makes the chain reversible and makes it target *exactly* the spanning tree distribution. *See* Sarah Cannon, Moon Duchin, Dana Randall, and Parker Rule, *A Reversible Recombination Chain for Graph Partitions*, Preprint (2020), *available at* mggg.org/ReCom. The simplicity for the modeler and the speed of heuristic convergence recommend ReCom over Flip-based Markov chains. ReCom "is more computationally costly than Flip at each step in the Markov chain, but this tradeoff is net favorable thanks to superior convergence and distributional design." This piece does not give us room for a full introduction to these methods, but we refer the reader to DeFord & Duchin, *supra* note 80

[82] Since non-compact plans are exponentially more numerous, the probability of selecting them approaches 100% as the problem size expands. And in addition to being undesirable for compactness reasons, sampling from a uniform distribution has been proven to be computationally intractable. Lorenzo Najt, Daryl DeFord, and Justin Solomon have shown that if you could create an algorithm that samples districting plans approximately uniformly, then you have solved a suite of problems long believed to be impossible. In particular, the solution would give you a way to crack internet encryption! Lorenzo Najt, Daryl DeFord, & Justin Solomon *Complexity and Geometry of Sampling Connected Graph Partitions*, Preprint (2019). *Available at:* arxiv.org/abs/1908.08881.

[83] No state law has a rule of this kind.

[84] In particular, they clearly break the key property that sample statistics converge to the same target distribution regardless of initial position. For details, see Appendix A.1.

ReCom described by the authors as "a refined version of the redistricting algorithm that one of us has developed in a series of expert engagements."[85]

4 Robustness and Stability

In this Part, we set aside questions of what ensembles can properly do, turning to a narrower investigation of whether the Article's particular ensemble design can produce reliable numerical findings that answer to their description.[86] We start by overviewing the ensemble protocol. Then we isolate some serious issues stemming from ecological inference – both how it is run and how it is used to define a "minority opportunity district." Finally, we investigate the possibility of incorporating richer electoral history rather than basing the whole analysis on Obama–Romney.

We will use the Texas state House as our case study throughout this Part of the Response. We began with a dataset containing dozens of statewide elections from the last Census cycle. From these, we selected nine elections to highlight (six generals and three Democratic primary or primary runoff elections). To emphasize probative elections, we ran ecological inference in the "preferred" manner described below in Part 4.2 and only considered elections in which the Black and Hispanic candidate of choice was very certain (identified in 100% of draws) and the two groups agreed. We additionally preferred more recent elections, and those with a Black or Hispanic candidate on the ballot.[87]

Figure 4 shows estimated polarization levels in those elections, illustrating that general elections have stark differences in White preferences compared to Black and Hispanic preferences, while Democratic primaries are far less polarized, even when candidates of color are on the ballot. This underlines the importance of incorporating primary elections into the analysis; their dynamics are quite different from generals and no candidate can ultimately be elected without first clearing a primary.

4.1 Anatomy of the Methods

Modeling redistricting calls for *operationalizing* the rules – transforming legal English into a form that can be handled by a computer – and this requires creativity and a suite of user choices. Working with electoral results often requires the use of inference techniques, and inference brings error. Considering both user choice and uncertainty, it is incumbent on the modeler to be vigilant to the ways that error and instability can propagate, snowballing in magnitude, through the steps of a workflow.

In Table 2 we set out the step-by-step procedure that leads from raw data to the findings that the authors call "overrepresentation" or "underrepresentation." The purpose of the table is to make invisible modeling choices visible. To be clear, every sophisticated modeling effort has many moving

[85] Chen & Stephanopoulos, *supra* note 5, at 884. Also "a modified version of a MCMC redistricting algorithm that one of us has previously employed in expert testimony." *Id.* at 893. The footnote in support of these claims says, of Chen's prior methods: "Under this related approach, a recombination MCMC algorithm developed by one of us was used to create a single map that satisfied the specified parameters. This process was repeated hundreds or thousands of times to generate a large number of maps. In other words, the maps were the endpoints of hundreds or thousands of separate Markov chains, not way-points along a single, very long Markov chain." There is simply no evidence of any setup that is capable of representative sampling in Chen's earlier work. Since expert witnesses can certainly update their methods when better ones become available, there is no need for this flagrantly misleading description. See Appendix A.1 for more on Markov chain theory.

[86] We do not attempt to elaborate a complete alternative approach here. For a fully implemented VRA protocol that works to avoid the issues we have flagged here, *see* Becker, et al. *supra* note 8.

[87] *Id.* (offering several methods of combining elections).

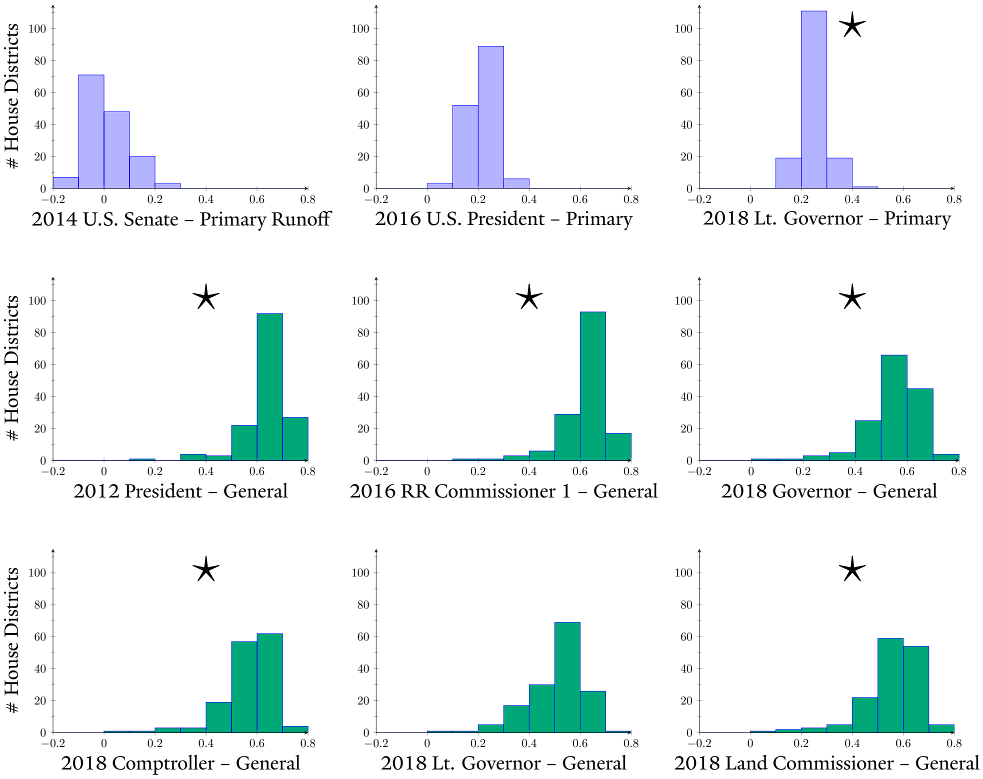


Figure 4: The racial voting gap in nine elections over the 150 districts of the Texas House of Representatives. Black and Hispanic voters agree on the candidate of choice in all nine elections. Each histogram plots the estimated difference between Black+Hispanic support for the candidate of choice and White support using the "Preferred EI" method described in the text. General elections show massive polarization of about 60 percentage points, while Democratic primaries and runoffs show broad agreement between White and minority voters. In six of these elections (marked with ⋆), the minority-preferred candidate is either Black or Hispanic.

parts, and the point here is not to critique the level of complexity, but rather to examine the decision junctures. Errors and arbitrary choices risk compounding throughout the entire workflow to accumulate in the final project. This is why responsible mathematical modeling always includes a sensitivity analysis, showing whether the findings are stable or variable as the settings are tweaked.

The Chen and Stephanopoulos definition of minority opportunity district has two components.

- (Obama win): Obama got more votes than Romney in 2012; and
- (group control): Obama garnered more votes from minority voters than from White voters;

with various additional cases when Black or Hispanic voters do not prefer Obama to Romney.[88] The

[88] To give their fuller description, "we define an opportunity district as one where (1) the minority- preferred candidate wins the general [Obama–Romney] election, and (2) minority voters who support the minority-preferred candidate outnumber white voters backing that candidate, provided that (3) minority voters of different racial groups are aggregated only if each group favors the same candidate." 901-902. The data demonstrations in this Part are based on applying their definition in full, after correcting small coding errors (see Appendix A). In their data, the groups considered are Black, Hispanic, and Other, a category that includes everyone else—White, Asian, two or more races, and so on. We found that the effects of replacing this super-category with only non-Hispanic White voters amount to about a one-seat difference in the Texas

Step in Article Workflow	Discussion
Obtain election results on a shapefile with block assignments.	Article sources this to DailyKos blog. Only Obama-Romney election is available. Unreliable shapefile will cause error.
Join demographic data using decennial voting age population (VAP) rather than citizen voting age population (CVAP).	Article uses VAP for EI but compares outcomes to CVAP proportionality.
Estimate voting behavior using ecological inference. Set up racial groups as Black, Hispanic, Other. Run EI on every county separately, in two phases: one phase to estimate turnout, then a second phase to estimate candidate preference.	Point estimates are recorded without error bounds. By-county method exaggerates EI problems with small counts. Many racial categories (Asian, Two or more races, etc) are combined with White vote, no matter the group voting preferences, which will have major impacts in some states.
Break down votes to census blocks.	Has potential to introduce substantial error.
Create units called "base polygons" for building plans.	Article uses Census Places shapefile to represent municipalities and reverse-engineers units from these. Unclear if ad hoc building blocks influence findings.
Designate tight threshold criteria: county splitting, place splitting, Polsby-Popper, and population deviation are better than or equal to enacted plan.	It is clear that other ways of operationalizing the rules, or the use of softer validity conditions, would lead to different findings.
Create seed plans that pass threshold tests.	Only one seed plan is used for each case with no method presented for its generation.
Attempt 10,000,000+ steps to recombine the "base polygons," only accepting new plans that meet threshold criteria.	No discussion of connectivity of search space or convergence of summary statistics.
Pass over 100,000 plans ("burn-in"), then save every 10,000th accepted step ("sub-sampling") to create an ensemble of size 1000.	Generous burn-in and sub-sampling combine to make the ultimate ensemble very small. 1000 observations is far too few to estimate a histogram on over 100 bins.
For each district, if Black voters preferred Obama to Romney and Hispanic voters preferred Obama to Romney, then count the district as a MOD if Black+Hispanic Obama voters outnumber White+Other Obama voters.	Defensible but not authoritative definition of a VRA-relevant district. Alternative definitions lead to different findings. Strict inequality applied to EI outputs leads to instability.
For each districting plan, count its number of MODs. Record the median number over the 1000 plans as the MNMOD.	A great deal of ensemble information is lost.
Say that minorities are "overrepresented" (resp., "underrepresentated") in a plan if the number of MOD is greater (resp., less) than the MNMOD.	Median is treated as a precise target and finally called "the race-neutral baseline."

Table 2: Survey of methods in the Article for measuring whether minorities are overrepresented or underrepresented with respect to the race-neutral baseline.

logic of this particular construction is defensible but hardly authoritative. Even so, this definition of MOD could be useful in broad strokes for determining whether minority plaintiffs might have a legal right to additional representation. To see this we will investigate the robustness of the findings.

House, but the effect size would surely be higher in other parts of the country. In particular, three of the states treated in the Article (CA, NV, NY) are at or near 10% Asian population share, and Asian voters are far more likely to align with Black and Hispanic than White voters. *See* Christopher S. Elmendorf & Douglas M. Spencer, *Administering Section 2 of the VRA After* Shelby County, 115 Colum. L. Rev. 2143, 2210-2211 (2015).

4.2 The Racial Inference in the "Race-Blind" Protocol

A workflow in which the output of each step is fed, without calibration, into the next risks becoming a Rube Goldberg machine. Each part may be in working order, but the string of tenuous transitions creates precarity for the whole apparatus. And in this case, there is a method at the center of the machine that on its own has the capacity to destabilize the whole enterprise.

Ecological inference (EI) is the industry standard technique – or more properly, family of techniques – for relating demographics to voting history in geographic units. Since EI itself is stochastic,[89] one way to probe robustness is to simply re-run the code. For example, Chen and Stephanopoulos report that there are 28 Hispanic opportunity districts and 18 Black opportunity districts in the Texas state House, respectively. We ran their EI code for Texas two additional times exactly as written, using their own data.[90] The first re-run reported 27 and 18; the second found 26 and 20.[91]

The main driver of instability is group control, the second element in the authors' definition of MOD, which uses vote-by-race estimates in a hard-edged way (i.e., with a strict inequality). Besides disregarding uncertainty, the authors' protocol uses separate EI runs on every county, as opposed to a single statewide model. In Texas, for example, there are 254 counties, and 50 of them have 2010 Census population under 5,000 people. Ecological inference, like ecological regression and all other inference techniques used for this purpose, gives very unreliable estimates for small sub-populations. In their more-is-better approach, Chen and Stephanopoulos push their EI right to, and even past, its known limitations.[92]

Besides their choice to run **by county** rather than statewide, they use **VAP** (voting age population) and not CVAP (citizen voting age population), even though CVAP is clearly the litigation standard when working with Hispanic VRA claims in particular. They use a **two-phase** method, with a first run to estimate turnout and a second run (using only expectation and not uncertainty from the first) to estimate candidate preference. A conceptually cleaner and far more reliable approach is to create a dummy candidate called "Abstain" to account for non-voters. A single phase of EI then gives estimates for both turnout and candidate choice; the advantage of this approach is that all data is taken into account in the same model run. Finally, in precincts where the number of votes exceeds the VAP, they **scale the vote** down to match the VAP, as opposed to more intuitive options like scaling the population up to match the votes or creating a buffer column to avoid scaling at all.

As we illustrate in columns (2)-(4) of Figure 5, these innocuous-sounding choices can have a massive effect, especially in combination. These simple toggles can make the number of MODs measured *in the enacted plan* vary from 34 to 51.[93]

(2) Unstable EI: By county / VAP / two-phase / buffer

[89] Gary King, Ori Rosen, and Martin Tanner, Ecological Inference: New Methodological Strategies 7-10 (2004).

[90] Jowei Chen, Replication Code, *available at* www-personal.umich.edu/ jowei/race/.

[91] Compare this to Figure 5, in which their style of EI reports 44 or 45 MODs. As compared to their Article, which reports 46 MODs, our districting ensemble was generated using Texas Legislative Council precinct units rather than the custom units from the replication materials, which may account for the discrepancy.

[92] Supplemental Figure 8 shows that this makes a major difference! The by-county run reports implausibly lukewarm Black support for Obama. We further note that statewide EI, run in the hierarchical Bayesian style, produces estimates by precinct and is perfectly capable of detecting regional differences.

[93] As it happens, the highest estimate both uses the settings we think are conceptually preferred and is the best match to recent House district election patterns in Texas. "Ground truthing" the outputs will be discussed further in the next Part.

(3) Article EI: By county / VAP / two-phase / scale votes

(4) Preferred EI: Statewide / CVAP / one-phase / scale population

The × symbols in columns (2)-(4) mark the outputs to the question "How many minority opportunity districts in the enacted Texas House plan?" that are observed by simply running the identical EI script 20 times and applying the Chen–Stephanopoulos definition. Then, the same EI values that reported the highest and lowest static count are applied to count MODs in the ensemble of two million plans. This is shown as a test of the authors' hypothesis that any quirk that elevates the MOD count in the enacted plan would similarly elevate the ensemble.[94] As we see, this is not the case.

4.3 Turning the Knobs

The choice of EI is not the only proverbial "knob to turn" in the machine we are studying. We focus on one other crucial ingredient in this Part—the electoral history that must be central in any reasonable determination of effective "opportunity"—and then briefly discuss an evaluation of outputs against the ground truth of Texas House district performance. Another crucial modeling choice, how to model a preference for keeping counties and municipalities whole, is deferred to Appendix A.4.

Every column in Figure 5 uses the *same* large ensemble of 2 million alternative 150-district Texas plans, made with whole precincts as building blocks so that electoral data is handled with a minimum of error.[95] By holding the two-million-map sample of plans constant and only changing what kind of district is being counted, we can isolate the ways that altering the definition and measurement of minority opportunity districts can significantly shift the findings.

The left-most column shows that 54 out of the 150 districts in the enacted Texas House plan had more Obama votes than Romney votes in 2012, and that this is fully normal with respect to the comparator ensemble of alternative plans made with no partisan or racial data. Statewide, Obama received almost exactly 42% of the major-party vote share. The conversion of this spatial pattern of support to an outcome where Obama wins 36% of districts appears to be fully normal with respect to the randomized redistricting alternatives.

And it is this Obama electoral success on which the Chen–Stephanopoulos definition of minority opportunity is built. Every Texas district that they identified as an MOD qualifies because it has an Obama majority *and* has more estimated votes from Black and/or Hispanic voters than from White/Other voters. Additionally, though it is not visible in the figure, they affix a binary label of Black opportunity or Hispanic opportunity simply based on which group is estimated to have cast more votes for Obama.

As discussed above, the EI used to enforce "group control" for minority voters is then layered on top of the first counting question. A richer dataset could certainly be used as the basis of measuring electoral success, rather than the Obama re-election alone. Ensemble methods rely on statewide elections, but there is no reason to demand that the same elections be used across states; on the contrary, the best practice would clearly be to use as many statewide elections as possible.

[94] *See supra* note 15.

[95] ReCom always finds contiguous plans and places a heavy preference on compact plans. In these runs, we allow districts to deviate by no more than 5% from ideal population, and we collect every accepted plan into the ensemble (no burn-in, no sub-sampling).

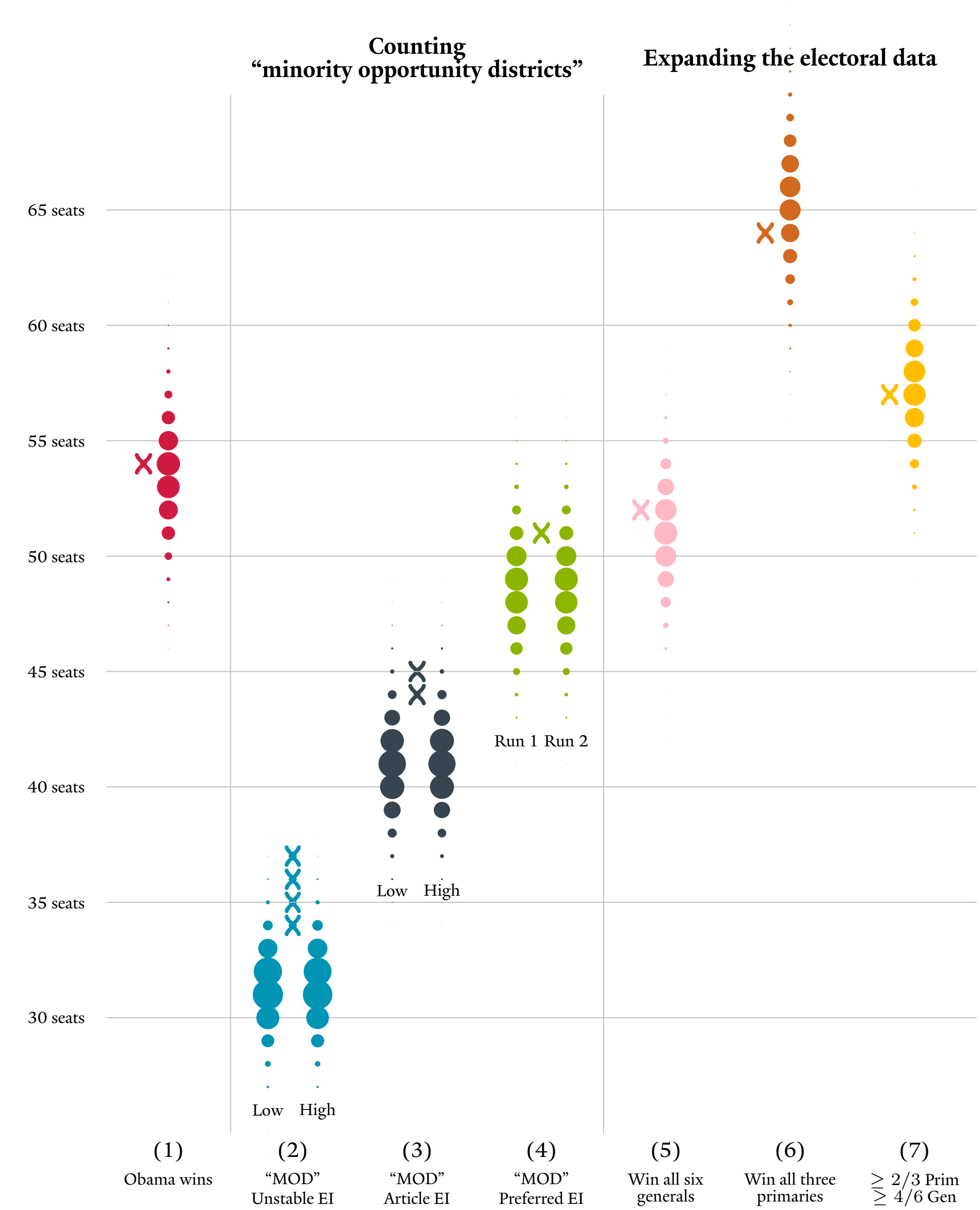


Figure 5: Comparison for TX House (150 districts) based on applying various counting questions to a single precinct-level ensemble of two million plans. Radii of colored disks are proportional to the frequency with which an outcome was observed in the ensemble; bracket (✗) marks the number of each kind of district in the currently enacted plan. The definition of minority opportunity district (MOD) begins with an Obama win as in (1) and layers a group control requirement on top. (2)-(4) show that this group control condition depends heavily on the way EI is run. (5)-(7) compare starting points with broader electoral history as an alternative to relying on the vote pattern from a single election.

Columns (5)-(7) of Figure 5 explore the number of districts won by minority candidates of choice for different mixes of election contests. We see in column (5) that an Obama win is highly predictive of a win for the Democrat (who in each case is the candidate of choice for both Black and Latino voters) in the five other general elections considered here. However, primary elections behave quite differently. Column (6) shows that the Black and Latino candidate of choice has the most votes in the Democratic primary in 64 districts in the currently enacted plan, which is still in the normal range but no longer falls above the ensemble average. Finally, since electoral opportunity requires that candidates of choice first advance from primaries and then prevail in general elections, we count how many districts have seen success for the minority candidate of choice in at least two of three primaries *and* at least four of six generals.[96]

So, how should we ultimately set the knobs on our machine? The decision should consider stability and replicability, but must also be made in view of the available ground truth provided by recent district performance. We can examine exactly which current enacted districts are not labeled MODs by the Article's method, but have a rock-solid recent history of opportunity for minority candidates of choice. One clear category is urban-proximal districts with significant White crossover support, such as HD 46, 49, 50, and 51 in Travis County, home to Austin. District 46 is currently represented by Sheryl Cole, who is Black, and the only candidates who have won or even received strong vote support in the full ten-year cycle are Cole, Jose Vela, and Dawnna Dukes—all clearly minority-preferred.[97] The fact that the Chen–Stephanopoulos definition of MOD systematically disqualifies districts of this kind should be a signal that it would be wise to soften its handling of "group control."

Conclusion: The Algorithmic Future

From predictive policing to smart cars to medical diagnosis, algorithmic assistance is becoming ensconced in every area of public and private life. As the science that is relevant to law and governance gets closer to the research frontier, skillful mathematical modeling will become indispensable for policymakers.

As an example, consider the disclosure avoidance measures being advanced by the Census Bureau. Title 13 of the U.S. Code requires the Bureau to take measures to protect the privacy of respondents' data. In the 2010 Census, this was achieved by an ad hoc mechanism: a Bureau employee manually swapped data between small census blocks to thwart identifiability. In 2020, this is no longer adequate to protect from increasingly sophisticated reidentification attacks.[98] With this threat in mind, the Bureau has turned to a "differentially private" hierarchical noising algorithm called TopDown, following a concept recently introduced by academic computer scientists.[99] To analyze the

[96] Here we are only trying to illustrate that more electoral results can easily be incorporated. *See* Becker, et al., *supra* note 8 for several workable methods of combining many statewide elections to create an overall index of electoral success.

[97] *See* Ballotpedia ballotpedia.org/Texas_House_of_Representatives_District_46.

[98] In such an attack, an adversary uses a simple computational technique to reconstruct the Census person-by-person data file and then pairs it with commercially available data to match names, phone numbers, and addresses with all the information included on the Census form. *See* U.S. Census, "Disclosure Avoidance and the 2020 Census," at www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html; Michael Hawes, "Differential Privacy and the 2020 Census," NCSL Webinar (Mar. 5, 2020). *Available at* www2.census.gov/about/policies/2020-03-05-differential-privacy.pdf.

[99] John Abowd, Robert Ashmead, Simson Garfinkel, Daniel Kifer, Philip Leclerc, Ashwin Machanavajjhala,

potential for differential impacts of these privacy strategies on marginalized populations—and indeed on VRA enforcement!—there is no reasonable alternative to collaboration. Computer scientists and mathematicians who are at the research forefront, geographers who understand Census data, social scientists and litigators who use the Census data, and organizers who mobilize Census response will need to work together, since the technique itself is novel and the use case is full of special complexities. We would go so far as to suggest that it would be a serious mistake for litigators who take on issues around privatized data to rely exclusively on established networks of experts.

This moment, when cutting-edge scientific computation is becoming unavoidably implicated in many domains of law, requires sweeping reforms to legal training, legal publication, and the recruitment and development of litigation experts. Law students of the twenty-first century should be conversant with probability, statistics, and an introduction to algorithms. Legal publications that draw on technical material should be refereed by competent domain experts, and expert reports should be held to high modeling standards, including scientific norms of well-documented software and replicable findings. Courts are unlikely to remain satisfied by the mere invocation of "an algorithm" or even just "a computer program" that takes the appropriate criteria into account in some way.

By the same token, it is essential to create pathways for STEM researchers to get serious training in the humanities and social sciences. This will enable modelers and software engineers to build tools that answer to the needs of law and policy—and to understand their limitations. The scientific publication ecosystem, which in many fields has ossified around the major application domains of the mid-twentieth century, would benefit enormously from new journals that take social, legal, and civil rights applications seriously. And above all, we should all be collaborating more.

In the end, we find ourselves in resounding agreement with Chen and Stephanopoulos when they describe ensemble methods as the "most important development in recent memory" in election law.[100] But rather than handing us single-statistic indicators, ensembles are better used to find ranges and quantify tradeoffs, highlighting the properties entailed or promoted by the rules and helping to flag extreme outliers. For ensembles to gain traction in VRA litigation and beyond, it is crucial that researchers are transparent about their design choices: modeling methods must be discussed and justified "above the line," and ideas for operationalizing vernacular rules and priorities should share billing with the ultimate findings. From a voting-rights perspective, the stakes could hardly be higher, as a newly invigorated conservative Court gears up to take on the Voting Rights Act at the turn of a new census and redistricting cycle.

By hiding complexity and contingency, *The Race-Blind Future of Voting Rights* creates an appearance of definitiveness for its account of a race-neutral baseline suited to a minimalist Voting Rights Act. The risks of this illusion are serious. When the Court needs a standard to meet its aims, it has been known to reach right into the academic literature; in fact the *Gingles* factors that completely reconfigured the landscape of VRA enforcement were lifted not from a legal brief but straight from the 1982 law review article of Blacksher and Menefee.[101]

Brett Moran, William Sexton, Pavel Zhuravlev, *Census TopDown: Differentially Private Data, Incremental Schemas, and Consistency with Public Knowledge*, Working paper (2019), github.com/uscensusbureau/census2020-das-2010ddp/blob/master/doc/20191020_1843_Consistency_for_Large_Scale_Differentially_Private_Histograms.pdf.

[100] Chen & Stephanopoulos, *supra* note 5, at 868.

[101] Thornburg v. Gingles, 478 U.S. 30, 50-51 (1986) (citing Blacksher & Menefee, *supra* note 58, for each of the three threshold factors, and thirteen times overall.).

Because the definitions of district effectiveness track contested normative ideals, some of the work to be done is conceptual. At the same time, the data and statistical requirements for this ambitious modeling project—studying the interactions of human geography and plurality districts in the context of voting rights law—will require many hands on deck and significant further work before researchers converge on a robust and widely applicable protocol that reflects the complexity of the use case. We are optimistic that Chen and Stephanopoulos's bold and provocative proof of concept will inspire just that kind of follow-on work.

A Technical Appendix

A.1 Markov Chain Principles

The Fundamental Theorem of Markov Chains says that any Markov chain that is "ergodic" has a unique stationary distribution. The Markov Chain Central Limit Theorem ensures that if you collect a large enough sample from a suitably designed chain, you will get a reliable estimator for statistics on the state space.[102] These two results are the core theoretical selling points for applying MCMC to benchmark the baseline behavior for neutral districting plans.

Ergodicity, a hypothesis needed to secure these fundamental convergence and estimation results, requires that your elementary move—in our case, a recombination move that merges two districts and partitions them a new way—can reach any plan in your state space if run for long enough from any starting location. In other words, your state space must be path-connected. Sometimes the moves are designed to target a pre-set distribution, and other times a description can be attempted post hoc.

ReCom is designed to approximately target a particular distribution on districting plans, namely the "spanning tree distribution," in which the probability of choosing a given plan is proportional to a certain compactness score.[103] That's it—if you want to know how much more likely one plan is to be selected than another, you compute this spanning tree compactness score for both plans. If Plan A has a score twice as high as Plan B, it is twice as likely to be selected.[104] To be clear, none of this theory applies to "Petri dish" methods of district generation. Petri dish plans are perfectly respectable for example generation, but support no statistical claims. Likewise, hill-climbing algorithms (those that only accept a map with a score better than or equal to the previous one) will fail to be ergodic because they get stuck at local maxima. They are designed for heuristic optimization, not for representative sampling.[105]

In the construction of the ReCom algorithm, we treat population balance and contiguity as basic and non-negotiable requirements of redistricting. Contiguity does not have to be enforced by ReCom because it is an automatic consequence of its merge and split procedure. But we address population balance with a validity check: when a move is attempted, it can only be accepted if it has population deviation no greater than some user-chosen threshold. Any hard requirement like this that prevents the random walker from advancing can be called a "rejection filter" in the procedure.

Rejection filters should be used with great caution. A threshold set too tight can disconnect the state space entirely, making it impossible to transition between some two plans by a sequence that

[102] Charles J. Geyer, *Introduction to Markov Chain Monte Carlo*, *in* Handbook of Markov Chain Monte Carlo (Steve Brooks, Andrew Gelman, Galin L. Jones and Xiao-Li Meng, eds. 2011), *available at* mcmchandbook.net/HandbookChapter1.pdf.

[103] *See* Duchin & Tenner, *supra* note 76.

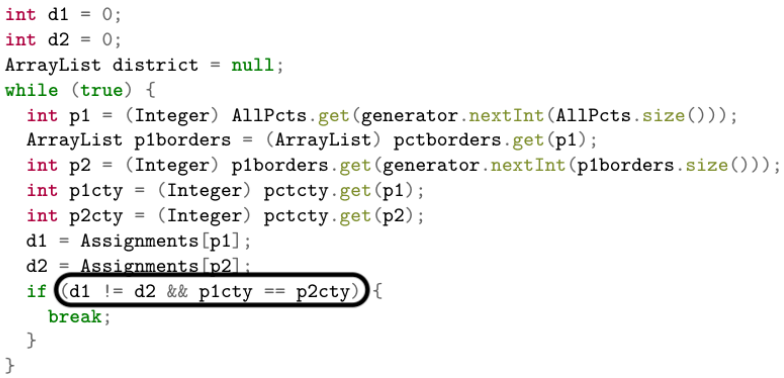
[104] *See* Cannon et al., *supra* note 81 (finding that by making a small change to ReCom, a reversible chain is obtained that *exactly* targets this distribution).

[105] As to the rebranding of earlier methods as "a recombination MCMC algorithm," every expert report and publication of Dr. Chen's uses either a Petri dish method or, from 2018 onwards, a series of hill-climbing Flip runs. Indeed, the word "optimize" appears repeatedly; there are no occurrences of "recombination" or "spanning tree" and there is no discussion of convergence or the relative weighting of plans. *See, e.g.*, *supra* note 73; Expert Report of Jowei Chen, Common Cause v. Rucho, 279 F. Supp. 3d 587 (M.D.N.C.) (No. 1:16-CV-1026), vacated, 138 S. Ct. 2679 (2018); Expert Report of Jowei Chen, Ph.D., Whitford v. Gill, No. 15-cv-421-jdp (W.D. Wis. Oct. 15, 2018).

passes the threshold test at all intermediate steps.[106] In the disconnected case, the sample that is collected can only tell you statistics of the component containing your starting point, which may be a small and non-representative corner of the state space. And even when you don't fully disconnect the space, imposing strict conditions can create "bottlenecks" that make it hard to transit the space. (For instance, it could be possible to get between some two parts of the space, but only by choosing some very unlikely sequence of steps in a particular order.)

In scientific applications, it can be hard to know whether you have imposed conditions that cause your random walker to get stuck. A standard trial used to raise confidence that the random walker is exploring effectively is called the *multi-start heuristic*: run the chain from very different starting positions, and see if you collect comparable statistics. If not, you can be sure that your runs are too short or your space is disconnected.

The Chen–Stephanopoulos protocol imposes numerous stringent and arguably superfluous requirements: a hard limit requiring that the average Polsby-Popper score be less than or equal to the enacted plan (layered on top of the ReCom preference for compactness); a hard limit requiring that any new plan has at least as many fully intact counties as the enacted plan; and a requirement that any two districts to be merged by the procedure must share a county between them. Imposing the last requirement (see Figure 6) unquestionably disconnects the search space; any district made of whole counties can never be altered. It also clearly ensures that the spanning tree distribution is no longer a steady-state.



```
int d1 = 0;
int d2 = 0;
ArrayList district = null;
while (true) {
  int p1 = (Integer) AllPcts.get(generator.nextInt(AllPcts.size()));
  ArrayList p1borders = (ArrayList) pctborders.get(p1);
  int p2 = (Integer) p1borders.get(generator.nextInt(p1borders.size()));
  int p1cty = (Integer) pctcty.get(p1);
  int p2cty = (Integer) pctcty.get(p2);
  d1 = Assignments[p1];
  d2 = Assignments[p2];
  if (d1 != d2 && p1cty == p2cty){
    break;
  }
}
```

Figure 6: The code snippet that selects districts to be merged, which we have run through a Java formatter for legibility. Two districts can only be merged if a county is split between them.

In addition, they choose a very high *subsampling parameter*, waiting for 10,000 accepted steps before adding any new plan to the ensemble. Subsampling is essential with MCMC methods invoking physics-inspired techniques like temperature variation, as in the past redistricting work of Mattingly and Imai,[107] because it skips over "hot" plans that may not pass the validity requirements. But we know of no argument for subsampling with a ReCom chain, where all plans pass validity checks.

[106] If (reversible) ReCom is run with rejection filters that maintain the ability to transit between any two states, then the resulting stationary distribution will be the spanning tree distribution on the restricted state space.

[107] *Supra* note 80.

In this setting, subsampling needlessly throws information away. To quote from the Handbook of MCMC, “Subsampling cannot improve the accuracy of MCMC approximation; it must make things worse.”[108]

To summarize the most serious problems we find with Article’s sampling protocol:

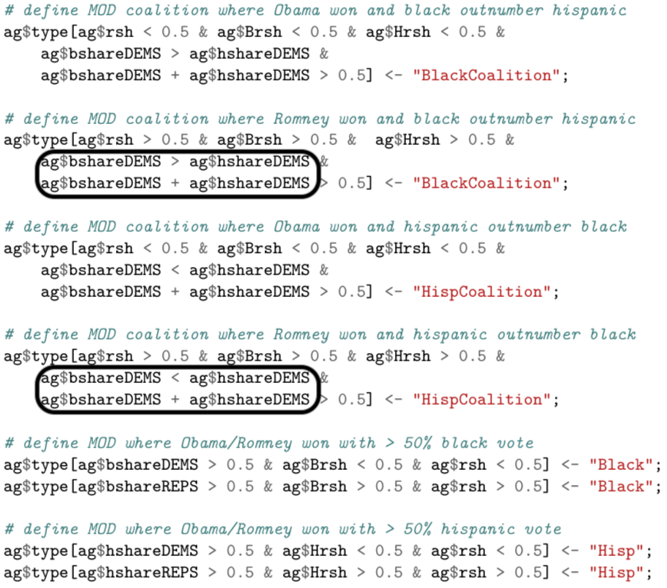
- Overzealous subsampling (every 10,000th map) leads to samples that are far too small (1000) to estimate a full distribution.
- Numerous rejection conditions (Polsby-Popper, intact counties, etc.) and no multi-start heuristics raise concerns of bottlenecks and disconnection.
- Only allowing merging for districts that share a county between them skews the sampling distribution in an uncontrolled manner and creates strong dependence on the initial starting point for a chain.

Finally, it is not clear what the authors intended to demonstrate with the “alternative methodology” offered in their Appendix D. There, they present an experiment in which even 100 rather than 1000 points pulled from a ReCom distribution will give the same median value, whether created along one long chain or 100 individual ones. If this tells us anything about the present application, it lends (weak) support to the idea that ReCom converges quickly to a stationary distribution on its component of the state space. That is, assuming the individual runs were shorter than the single run, this demonstration supports the long-standing claims about the efficiency of ReCom but only amplifies, rather than assuages, worries that the random walker in the Chen–Stephanopoulous experiments might be stuck in a small component of the state space. As we will see below in Figure 10, those worries are warranted.

[108] *Supra* note 102 at 27.

A.2 The Code Itself

The authors' codebase uses a mix of R (for EI) and Java (for ensembles). As we were repeatedly reminded during our replication work, even one line of buggy code can compromise an entire data operation. As one small indication, Figure 7 shows a minor error in their MOD definition logic. In principle, a glitch like this could throw the analysis way off, though in practice this particular error is buried in a rare case (where minority voters prefer Romney to Obama) and may have little to no impact on findings.



```
# define MOD coalition where Obama won and black outnumber hispanic
ag$type[ag$rsh < 0.5 & ag$Brsh < 0.5 & ag$Hrsh < 0.5 &
    ag$bshareDEMS > ag$hshareDEMS &
    ag$bshareDEMS + ag$hshareDEMS > 0.5] <- "BlackCoalition";

# define MOD coalition where Romney won and black outnumber hispanic
ag$type[ag$rsh > 0.5 & ag$Brsh > 0.5 &  ag$Hrsh > 0.5 &
    ag$bshareDEMS > ag$hshareDEMS &
    ag$bshareDEMS + ag$hshareDEMS > 0.5] <- "BlackCoalition";

# define MOD coalition where Obama won and hispanic outnumber black
ag$type[ag$rsh < 0.5 & ag$Brsh < 0.5 & ag$Hrsh < 0.5 &
    ag$bshareDEMS < ag$hshareDEMS &
    ag$bshareDEMS + ag$hshareDEMS > 0.5] <- "HispCoalition";

# define MOD coalition where Romney won and hispanic outnumber black
ag$type[ag$rsh > 0.5 & ag$Brsh > 0.5 & ag$Hrsh > 0.5 &
    ag$bshareDEMS < ag$hshareDEMS &
    ag$bshareDEMS + ag$hshareDEMS > 0.5] <- "HispCoalition";

# define MOD where Obama/Romney won with > 50% black vote
ag$type[ag$bshareDEMS > 0.5 & ag$Brsh < 0.5 & ag$rsh < 0.5] <- "Black";
ag$type[ag$bshareREPS > 0.5 & ag$Brsh > 0.5 & ag$rsh > 0.5] <- "Black";

# define MOD where Obama/Romney won with > 50% hispanic vote
ag$type[ag$hshareDEMS > 0.5 & ag$Hrsh < 0.5 & ag$rsh < 0.5] <- "Hisp";
ag$type[ag$hshareREPS > 0.5 & ag$Hrsh > 0.5 & ag$rsh > 0.5] <- "Hisp";
```

Figure 7: Code snippet to identify "minority opportunity districts," separated and commented for readability. Black and Hispanic opportunity districts are defined for object "ag" (an aggregate of precinct-level vote estimates) by multivariate compound statements; in the circled expressions, DEMS should be replaced by REPS to match the description in the Article.

But what about the difficult part of the code, where the graph algorithm is implemented? Because the code is sparsely commented and the replication materials come with no unit tests or examples, it becomes formidably difficult to analyze it in detail. This is even true when a high-level data team has several months to develop a replication study like the present one; now imagine a litigation context, where a similar codebase is turned over to opposing attorneys who have only two weeks to examine it. The chances that the code can be confirmed to perform as advertised are essentially zero.

The MGGG Redistricting Lab offers open-source, publicly accessible implementations of ReCom in Python and Julia with extensive documentation;[109] the contributors and users include students, faculty, redistricting practitioners, and professional developers. This is a case in which it should be a relatively easy decision to use widely reviewed code with a multi-year track record rather than a homemade alternative.

[109] *Available at* github.com/mggg/GerryChain and github.com/mggg/GerryChainJulia.

A.3 Running EI on Each County

In Part 4, we noted that running EI separately on each of the 254 counties in Texas might sound powerful, but is actually inadvisable because it leans into EI's known difficulties dealing with small sub-populations. For instance, Andrews, Dawson, Martin, and Gaines are a cluster of demographically similar counties in West Texas with a combined population of around 50,000 people. It stands to reason that EI will handle them better together than individually, because there will be more varied precinct data on which to base the model inferences.

And, perhaps surprisingly, this particular toggle in the settings matters quite a bit. Figure 8 shows that changing from "Preferred EI" in only this way causes a huge change in the inferred Black support for Obama across precincts. Instead of estimates of precinct-level Black support for Obama uniformly near 90%, by-county runs report a significant share of precincts with 40-60% support. As a point of comparison, the 2012 Cooperative Congressional Election Survey found that overall support for Obama among non-Hispanic Black voters in Texas was over 90%.[110]

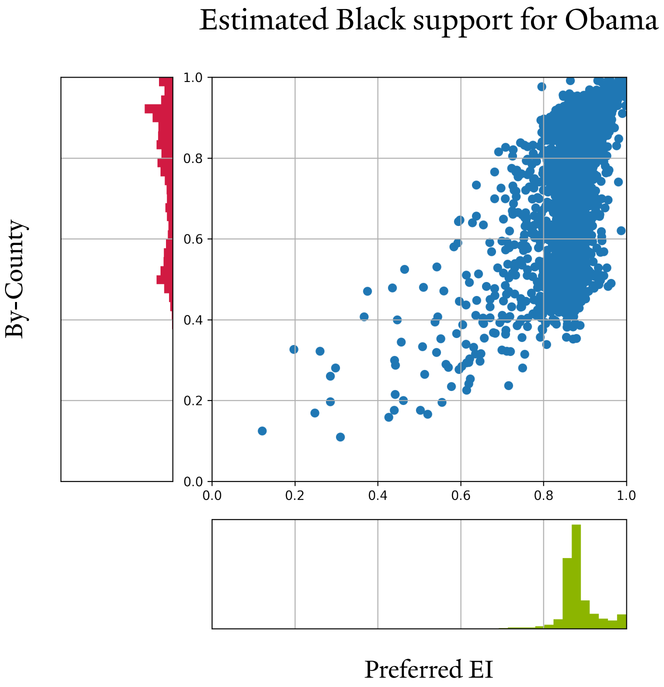


Figure 8: The estimate of Black voters' support for Obama in each precinct as we toggle only the statewide/by-county setting for EI. The scatterplot shows the 9082 precincts of Texas. Statewide EI, as in the "Preferred" style, reports high levels of Obama support; running the same script looped over the individual counties, as in the Article, reports a large share of precincts where Black voters are roughly evenly split between Obama and Romney, which seems fairly implausible.

[110] Stephen Ansolabehere and Brian Schaffner, *CCES Common Content, 2012*, Harvard Dataverse, doi.org/10.7910/DVN/HQEVPK (vote validated dataset, variables = *race*, *hispanic*, *inputstate* and *CC410a*). Estimated support is 94.7%, with a 95% confidence interval of [90.8, 97.3] based on exact binomial test.

A.4 Interpreting Criteria in Ensemble-Generation

In Part 4 we discussed the impact of "turning the knobs" related to EI and use of additional statewide electoral data to define minority opportunity districts. Another crucial modeling choice centers interpreting the traditional criterion of respecting political boundaries, such as counties and municipalities, by trying to minimize the extent to which those units are split by district lines.

County-conscious sampling

County preservation is a reasonable priority for House districting in Texas because a corresponding rule is found in Article 3, §26 of the state's constitution. County preservation is *not* a named priority in several other states in the Article.[111] Nevertheless, the authors implement a uniform county filter across every state. And the filter is extremely strict; we will see that it categorically blocks the random walker from making changes in most of rural Texas, so nearly all variation is in the urban counties.

There are many ways of handling county splits, and several are compared in Figure 9. We generate seven different ensembles for this figure with one million plans each, using different methods that either use weighting or rejection filters to accomplish greater county integrity. Column (1) shows an ordinary ReCom ensemble made with no attention to county lines. Column (2) implements the requirement that any merged districts must share a county (as in Figure 6). Column (3) implements the rejection filter that blocks new plans with even one more county split than the enacted plan. And column (4) combines both (2) and (3), as in the Article. Columns (5) and (6) use a softer method to weight in favor of county integrity: after two districts have been merged, ReCom draws a spanning tree of the double-district to cut it in two in a new way. One can assign random weights to the edges and use a fast algorithm to find a minimum spanning tree (MST). A soft way to favor county integrity is to put slightly higher random weights on the edges within counties, so that a minimum spanning tree is likely to include between-county edges and therefore likely to divide along a county line. Column (5) shows the outcome if the within-county edges are given weights in $(1.1, 2.1)$ while the between-county edges have weights from $(1, 2)$; column (6) bumps the within-county weights to the $(2, 3)$ range for a stronger effect. Finally, column (7) applies a different rejection filter, restricting the number of county *pieces*. That is, the intact-county filter in (3) only asks whether a county is divided, but does not measure the number of divisions. In large counties like Dallas County and Harris County, which are larger than a congressional district, the intact-county filter would make no distinction between touching three districts or 30. The pieces filter restricts the number of fragments in the plan overall to no more than the number in the enacted plan.

Scanning the outputs, it is striking that the approaches that most skew the Obama count are the ones that impose rejection filters alone, (3) and (7), and they push the count in opposite directions. That they are different makes sense because the ability to cut urban counties into an unlimited number of pieces will bear on "packing and cracking." This offers a strong reminder that each state's rules should be handled with care and reinforces a theme that emerges throughout our study of the Article's methods: graduated effects should be preferred to binary yes/no effects when dealing with matters of degree.

[111] *See supra* notes 12-14 and accompanying text.

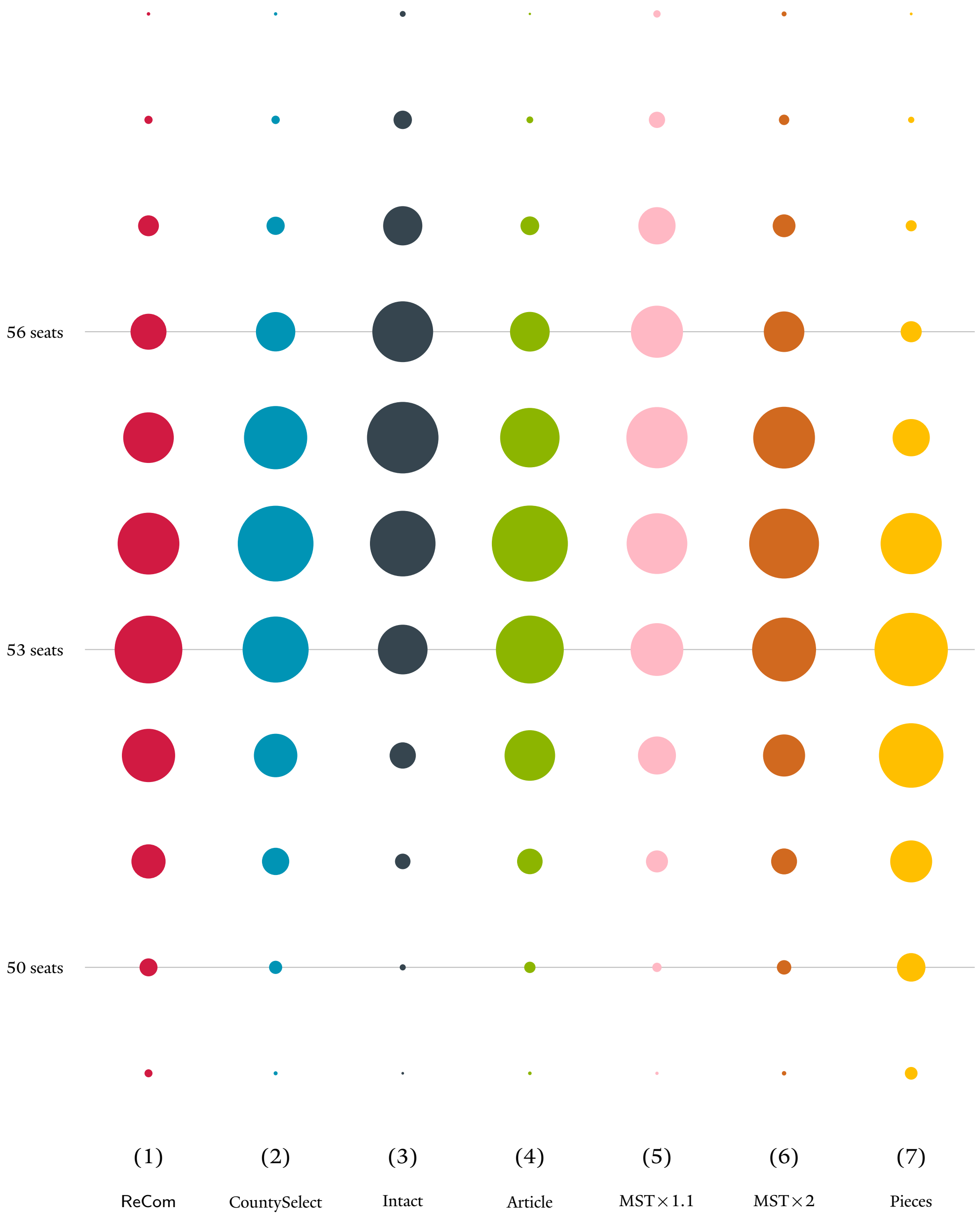


Figure 9: Many ways of handling a county criterion. Each column shows an ensemble of 100,000 maps, counting the number of House districts with more Obama than Romney votes. Given this variability, it would be reasonable to declare that 51–57 Obama districts, or even 50-58, is the normal range. It is far less reasonable to declare 54 as the median — and declare a plan with 55 such seats to "overrepresent" minorities — which would be the conclusion from the method in the Article.

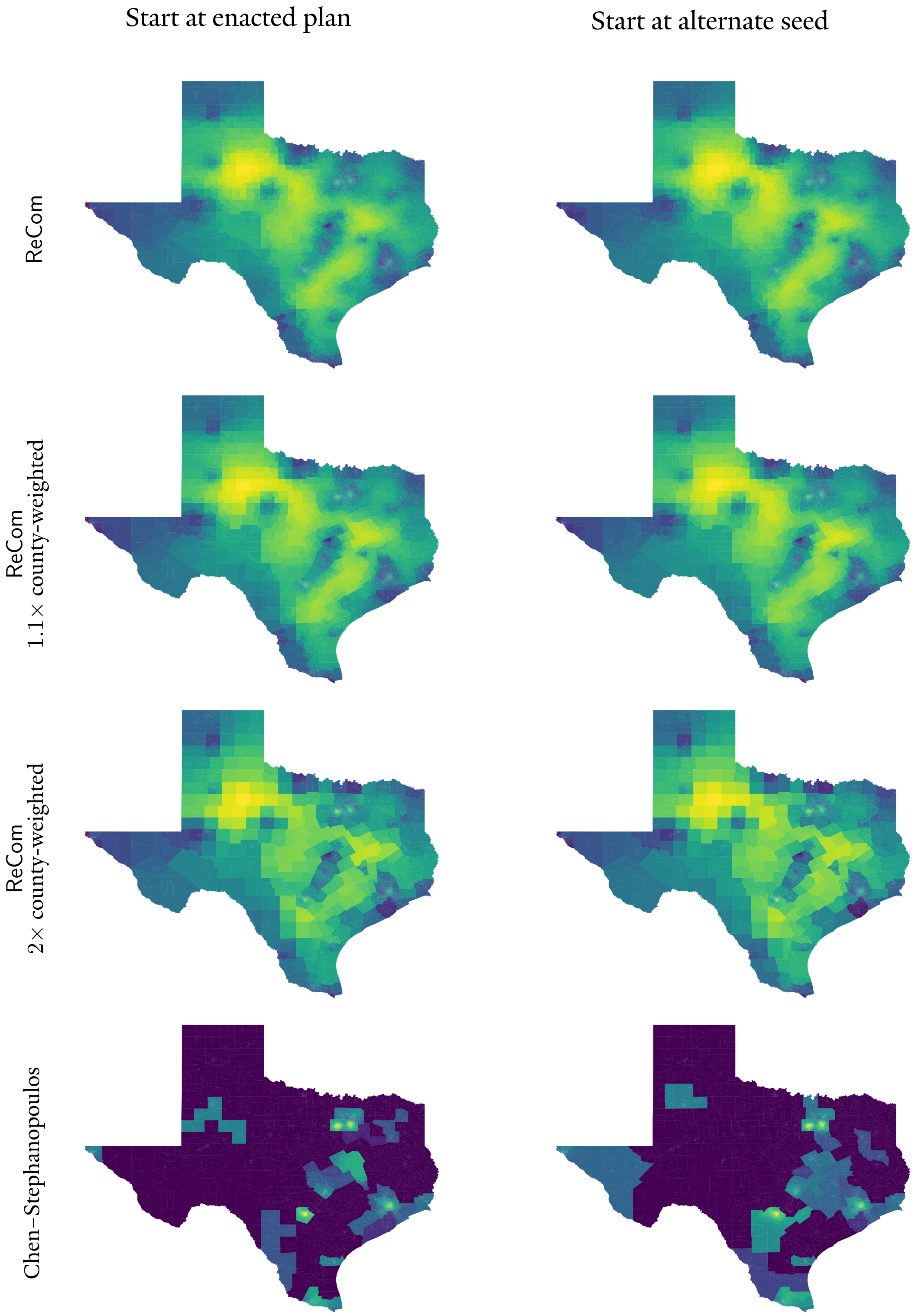


Figure 10: Heatmap of Texas precincts, showing how many times they change district assignment over a run of 1 million steps. The Chen–Stephanopoulos method does not explore effectively.

Impacts on ensemble geography

In Figure 10 we present a series of heatmaps that show how many times each precinct in Texas changes its district assignment across a run of one million steps with varying criteria, with yellow as the highest frequency and blue as the lowest.

In the top row we run ReCom with no county filter as in Figure 9(1) and observe that precincts flip to different districts across the state. In the next rows we impose the tree-weighting scheme described above, as in columns (5)–(6) from the last figure. The effects are quite visible compared to ordinary ReCom, especially in North Texas where counties follow a grid pattern—now we can see whole rural counties flipping together. We note that a stronger weighting factor produces crisper county lines, but still allows widespread changes. Each of the weighted methods, like ordinary ReCom, leads to a visually indistinguishable heatmap whether the chain was run from the enacted plan or from an alternate random seed.[112]

Finally, we compare to ensembles created with the strict double-layered county filter just as Chen and Stephanopoulos use it in their Article (Figure 9(4), combining (2) and (3)). The results are quite stark: most of the state is completely untouched by their randomization of districts, even after a million steps, staying exactly as it was in the initial plan. The most-flipped precincts are hard to see in a map of the state because they are so concentrated in Dallas/Fort Worth, San Antonio, and Houston. And even after one million steps, the chains that were run from different initial starting points are producing visibly different patterns.

This finding presents conclusive evidence that there are large areas in the universe of legally valid districting plans for the Texas House that their ensembles never visit.[113]

Municipalities

A municipality preservation rule is also imposed in the Article, again with a hard threshold. This does not match up with the *ex ante* rules for redistricting in Texas or in most other states in the authors' sample. Of the nineteen states in the study, only three (AZ,[114] CA,[115] SC[116]) mention cities as such in their redistricting rules, and four (DE,[117] IL,[118] NV,[119] VA[120]) have no rule at all regarding counties, municipalities, or *any* political boundaries.

[112] Multiple seed plans are available in our replication materials.

[113] We confirmed that this is true for their actual ensembles used in their analysis as well as for our reconstructions of their method. Demonstrations can be found in our replication materials.

[114] Az. Const. art. 4, pt. 2, § 1(14)(E)("To the extent practicable, district lines shall use visible geographic features, city, town and county boundaries, and undivided census tracts").

[115] Ca. Const. art. 21, § 2(d)(4)("The geographic integrity of any city, county, city and county, local neighborhood, or local community of interest shall be respected in a manner that minimizes their division to the extent possible without violating the requirements of any of the preceding subdivisions.").

[116] "2011 Redistricting Guidelines," S.C. Senate, Apr. 13, 2011, redistricting.scsenate.gov/Documents/Redistricting-GuidelinesAdopted041311.pdf; Election Law Subcommittee Report, *2011 Guidelines and Criteria for Congressional and Legislative Redistricting*, S.C. House Judiciary Committee, redistricting.schouse.gov/6334-1500-2011-Redistricting-Guidelines-(A0404871).pdf.

[117] *See supra* note 13.

[118] *See supra* note 14.

[119] Justin Levitt, *Nevada*, All About Redistricting, redistricting.lls.edu/state/nevada (last visited Jan. 21, 2021).

[120] Justin Levitt, *Virginia*, All About Redistricting, redistricting.lls.edu/state/virginia (last visited Jan. 21, 2021).

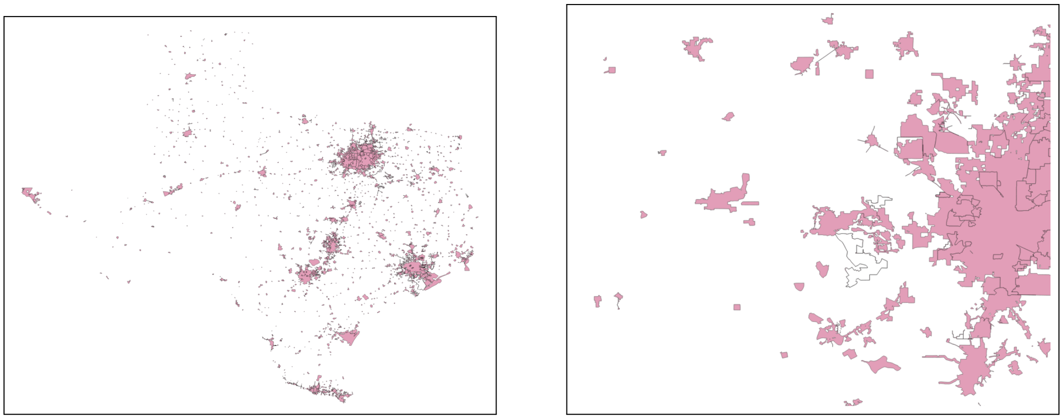


Figure 11: Left: the Places shapefile in Texas. Right: A close-up in the Fort Worth area.

The authors' style of operationalizing municipality preservation is interesting enough to merit discussion. In many states, there is no authoritative source to find boundaries for relevant municipal geographies. In order to build an approach across states, the authors turn to a Census data product called *Census Places*.[121] These include not only "Incorporated Places" like cities and towns, but also "Census-Designated Places" like Native American reservations and various land use areas that are chosen by the Census Bureau, not the state, as being appropriate for statistical tabulation.

Figure 11 shows Census Places statewide and in a Fort Worth inset, illustrating that the Places can include strands and spurs and empty loops. The authors make their technique municipality-conscious in two ways, both extremely strong. One is to impose another rejection filter that requires accepted plans to have at least as many intact Places as the enacted plan. The second is a fundamental shift whose impacts are hard to understand completely. They do not build their plans out of whole precincts, as we do in our replication runs. Instead, they create novel geographic units that they call "base polygons," defined as intersections of block groups and Places.

These choices—new building blocks, yet another rejection filter—certainly could have a major impact on the findings, and they are not justified in the Article or well-tailored to state law.

We stand to learn a great deal from continued investigations that meet the highest standards of data science while staying grounded in the details and the meaning of the law.

[121] Census Places are described in Chapter 9 of the Geographic Areas Reference Manual, *available at* www2.census.gov/geo/pdfs/reference/GARM/Ch9GARM.pdf.