Understanding and Mitigating the Impacts of Differentially Private Census Data on State Level Redistricting

Christian Cianfarani[1] and Aloni Cohen[1]

[1]Department of Computer Science, University of Chicago

Abstract

Data from the Decennial Census is published only after applying a disclosure avoidance system (DAS). Data users were shaken by the adoption of differential privacy in the 2020 DAS, a radical departure from past methods. The change raises the question of whether redistricting law permits, forbids, or requires taking account of the effect of disclosure avoidance. Such uncertainty creates legal risks for redistricters, as Alabama argued in a lawsuit seeking to prevent the 2020 DAS's deployment. We consider two redistricting settings in which a data user might be concerned about the impacts of privacy preserving noise: drawing equal population districts and litigating voting rights cases. What discrepancies arise if the user does nothing to account for disclosure avoidance? How might the user adapt her analyses to mitigate those discrepancies? We study these questions by comparing the official 2010 Redistricting Data to the 2010 Demonstration Data—created using the 2020 DAS—in an analysis of millions of algorithmically generated state legislative redistricting plans. In both settings, we observe that an analyst may come to incorrect conclusions if they do not account for noise. With minor adaptations, though, the underlying policy goals remain achievable: tweaking selection criteria enables a redistricter to draw balanced plans, and illustrative plans can still be used as evidence of the maximum number of majority-minority districts that are possible in a geography. At least for state legislatures, Alabama's claim that differential privacy "inhibits a State's right to draw fair lines" appears unfounded.

1 Introduction

In 2021, the state of Alabama sued the US Department of Commerce. Alabama sought to enjoin the use of a new disclosure avoidance system for the forthcoming release of 2020 decennial census data.

> "Forcing the State to redistrict with intentionally flawed data will impede Alabama's ability to draw representative districts with near-equal populations, which is what the Constitution and one-person, one-vote jurisprudence require. This will also impede Alabama's ability to draw districts to protect minority voting rights as required by the Voting Rights Act." (*Alabama v. Department of Commerce*, 2021)

Concerns persisted even after the data were published. In early 2022, one of us was contacted by lawyers at the ACLU in connection with racial gerrymandering cases in Georgia and Arkansas. They wondered whether the new disclosure avoidance system could affect the reported number of majority-Black districts in a demonstration redistricting plan.[1] If so, the States could try to take advantage of this uncertainty at trial.

Every ten years, the US Census Bureau conducts a decennial census and produces the *Redistricting Data Summary Files*—the basis for all redistricting for the following decade. These Redistricting Data are created by first aggregating data from the *Census Edited File* (CEF)—the confidential dataset that reflects the results of the completed census—and then applying the *Disclosure Avoidance System* (DAS). Through

[1]Ultimately, this issue was not raised by defendants. No expert testimony was needed.

2010, the DAS comprised traditional disclosure limitation techniques, primarily swapping [1]. These techniques are now understood as lacking formal guarantees against re-identification and individualized data disclosures [2, 3].

In contrast, the 2020 DAS is based on differential privacy, a modern framework for formally quantifying and limiting such disclosure risks [4, 5]. The 2020 DAS synthesizes and aggregates a new set of microdata census responses that have no one-to-one correspondence with the actual microdata in the CEF, but that closely approximate many of its statistics. While previous DASes have also introduced error into census tabulations, the 2020 DAS is the first to perturb total population counts in every geography smaller than states. This is inherent for the formal guarantees sought by the Census Bureau. Relative to the population, these errors can be quite large in very small geographies (24.79% in urban census blocks) but much smaller even in moderately-sized geographies (0.31% for counties of less than 1,000 people).

Many viewed the 2020 DAS as undermining the legal and political legitimacy of the the data's use for redistricting [6, 7, 8], calling for a rethinking of disclosure avoidance in the census. Others disagreed [9, 10]. The argument against the new DAS was made most forcefully by Kenny *et al.* [6]. They analyze the impact of the 2020 DAS using an ensemble-based redistricting analysis: a modern, computationally-intensive approach to exploring properties of the space of reasonable redistricting plans in a given geography. They conclude that "nonrandom local errors can aggregate into substantively large and unpredictable biases at district levels" and claim "the added noise makes it impossible to follow the principle of One Person, One Vote, as it is currently interpreted by courts and policy-makers" [6].

The concern is easy to understand: two districts whose populations are equal according to the official 2020 Redistricting Data will very likely have different populations according to the confidential 2020 CEF. If such discrepancies are large or biased, they could impact electoral power. And if the Supreme Court's *One-Person, One-Vote* (OPOV) caselaw requires balancing populations according to the CEF—as Alabama argued in its lawsuit—then redistricting plans created using the Redistricting Data could be illegal.

At the heart of this paper is a legal question. What data should redistricting law treat as the ground truth: the data as enumerated (the CEF) or as published (the Redistricting Data)? No court has ever faced this question, and we do not know how a court might decide. Alabama's lawsuit assumes that the enumerated data are treated as ground truth. Kenny *et al.* [6] consider this view to be the law as "currently interpreted by courts and policy-makers." Cohen *et al.* [9] argue the opposite: that the official Redistricting Data should be used as the ground truth. They see value in maintaining the "legal fiction" of perfect accuracy when the noise from disclosure avoidance is small relative to other data quality issues (quoting *Georgia v. Ashcroft* (2003)).

We won't answer the question. Instead, we ask how treating the enumerated data as the ground truth would affect a redistricter's ability to comply with the law. Our results may help redistricters better understand and mitigate the litigation risk arising from the legal uncertainty.

Research Questions This paper considers a data user who only has access to the officially published 2020 Redistricting Data (after disclosure avoidance), but is concerned that the new DAS will substantially alter their ability to comply with redistricting requirements as measured using the confidential CEF data (before disclosure avoidance).[2] We ask whether the user is able to achieve two goals. First, to produce plans balanced within One Person, One Vote population tolerances. Second, to determine how many majority-minority districts (MMDs)—where at least 50% of the population belongs to the relevant minority group—are possible in a given geography. We study both questions at the state-legislative level, where the legal standing of such plans depend on sharp thresholds (see Section 2). For OPOV, population deviations less than 10% are presumptively constitutional; larger deviations are presumptively unconstitutional. And showing that a political geography admits some number of majority-minority districts—districts with at least 50% of the population in the relevant minority group—is necessary to successfully overturn racially gerrymandered plans.

[2] As explained in Section 4, we do not have access to CEF and must use a different dataset as a stand in. See Section 7.1 for a discussion of how this does or does not affect our results.

Main findings Overall, our results suggest that one can use the 2020 Redistricting Data to meet two policy-relevant threshold requirements (OPOV, MMDs) on the unseen data. While discrepancies do arise, they often take a predictable form and are straightforward to account for. We support our claims with a large-scale ensemble analysis (Section 4) and discuss important limitations of our analysis in Section 7.1.

In the context of One-Person One-Vote, we observe large discrepancy rates in the fraction of plans satisfying a strict 5% population deviation limit[3] across many state legislative geographies (greatly extending a finding of [6]). Among plans sampled with at most 5% population deviation on the noised data, the median (across legislative geographies) discrepancy rate—the fraction of plans exceeding 5% deviation on the unseen data—was 9%. The max was over 80%. However, using a population deviation threshold in the redistricting sampling algorithm that is slightly tighter than the 5% policy target is an effective countermeasure. Sampling with a 4.8% limit makes the same discrepancy rate fall below 2% for 75% of geographies examined. Sampling with a 4% limit does so for all geographies examined. These results are presented in Section 5.1.

We focus our examination of MMDs on majority Black districts. We observe that discrepancy in the apparent number of majority Black districts in a plan can be large and biased. On average, the plans that result from optimizing the number of majority Black districts in the published data have fewer majority Black districts in the unseen data. Moreover, the magnitude of the discrepancy varies greatly by geography. Nevertheless, we find the published data still provide reliable evidence of the maximum number of majority Black districts achievable within a given state. This is the quantity most relevant to the *Gingles 1* test in racial gerrymandering cases. These results are presented in Section 6.1.

Along the way, we develop a richer qualitative understanding of the ways that noise from disclosure avoidance affects these policy-relevant quantities. We put these effects in context, showing that in the settings we consider the error introduced by the 2020 DAS is small relative to both total population and other sources of noise.

2 Legal context

2.1 One Person, One Vote

The One-Person, One-Vote principle holds that any two people living in the same state should have about the same voting power. In *Karcher v. Daggett*, the Supreme Court held that "congressional districts [must] be apportioned to achieve population equality as nearly as is practicable." Moreover, the decennial census "furnishes the only basis" for balancing populations, despite it's imperfections. In practice, this often leads to Congressional districts being balanced as closely as possible according to decennial census releases, despite the Court acknowledging the "inherent artificiality" of the data. After the 2010 Census, 26 states balanced their congressional districts to within a single person in total population, although larger deviations have been allowed [11]. State-level legislative districts must also be balanced in total population, although the requirement is much less strict. For state legislative districts, population deviations below 10% "will ordinarily be considered de minimis," while "disparities larger than 10% creates a prima facie case of discrimination" (*Brown v. Thomson* 1983). In practice, state legislative population deviations are often much smaller than 10%. One reason we do not study Congressional redistricting is that there is no amount of population deviation that is considered de minimis (*Karcher v. Daggett*). As such, there is no policy-relevant numerical target that a redistricter can hope to achieve after accounting for noise.

Note that courts define population deviation as the relative difference between the largest and smallest districts. We instead adopt the convention of [6] and others, using the maximum relative difference from the ideal district population. A 10% deviation under the former definition is approximately equal to a 5% deviation under the definition used in this paper (see Sec. 4.4.1).

[3]5% deviation as measured using the method of [6] and others, which approximates the 10% deviation threshold as measured by courts (see Section 4.1).

2.2 Voting Rights Act

Section 2 of the Voting Rights Act of 1965 outlaws vote dilution on the basis of race, including racial gerrymandering. The Supreme Court laid out a framework for judging racial gerrymandering claims in *Thornburg v. Gingles* (1986), and recently reaffirmed the framework in *Allen v. Milligan* (2023). To successfully challenge an enacted map, a minority group must first pass a three-part threshold test: the Gingles preconditions. The second and third preconditions together require voting to be polarized to such an extent that the minority group is prevented from electing its candidates of choice (*Thornburg v. Gingles* 1986).

Our paper focuses specifically on the first Gingles precondition, *Gingles 1*. Gingles 1 requires the minority group to be "sufficiently large and geographically compact to constitute a majority in a reasonably configured district" (*Allen v. Milligan* 2023). This is typically accomplished by producing for the court one or more demonstration maps (eleven in *Allen*), each with more *majority-minority districts* than the enacted map being challenged. Gingles 1 is satisfied if at least one demonstration map is reasonably configured ("if it comports with traditional districting criteria"). Whether a district is majority-minority is based on the voting-age population rather than the total population (*Bartlett v. Strickland*, 2009).

Throughout this paper, we specifically focus on Black voters. This reflects a common scenario in voting rights litigation and the largest racial minority in the USA according to Census data. Hence, districts will be considered majority-minority if a majority of the voting age population is Black (including respondents who selected multiple races).

2.3 Redistricting with imperfect data

The question of which dataset is the proper basis for redistricting law is one facet how redistricting law deals with uncertainty or error more generally [12]. It has long been recognized that the Decennial Census is far from perfect. For example, significant overage error that varies race: the 2020 Census was measured to have a 3.30% Black undercount, a 4.99% Hispanic undercount, and a 0.66% White overcount [13]. Even if the Decennial Census was perfect, populations of areas can change significantly in the 10 years between censuses when districts are largely unchanged.

The Supreme Court recognizes these issues. Generally, the Redistricting Data is considered "the only basis for good-faith attempts to achieve population equality" despite being "less than perfect" (*Karcher v. Daggett* 1983). But the Court has also acknowledged that deviating from the Redistricting Data is sometimes appropriate. In *Mahan v. Howell* (1973), a plan was considered malapportioned when about 18,000 people were known to live off of the naval base in which they were counted in the 1970 Census.

3 Data and methods

3.1 Decennial Census and disclosure avoidance

We use data from the 2010 Census P.L. 94-171 Redistricting Data Summary Files (henceforth SWAP) and the Privacy-Protected 2010 Census Demonstration Data Vintage 2021-06-08 (DEMO) [14]. SWAP is the official redistricting data from the 2010 Census, produced by applying the swapping-based 2010 DAS to the 2010 CEF. DEMO was produced by applying the 2020 DAS to the same 2010 CEF. These data files contain tabulations of population by voting age, sex, race, ethnicity for each geographic unit on the census geographic hierarchy—census blocks, block groups, tracts, counties, states, and the nation as a whole. SWAP and CEF agree on the total and voting age populations of every census geography, as the 2010 DAS held these statistics as invariant, but other demographic counts may differ. In contrast, DEMO and CEF agree on the total population only at the state level, and all sub-state counts may differ. We link these data with TIGER/Line Shapefiles [15] and block equivalency files for legislative districts enacted after the 2010 Census [16] provided by the Census Bureau to create the input data for our MCMC redistricting algorithms.

3.2 Ensemble Analysis

Markov Chain Monte Carlo (MCMC) methods enable sampling ensembles of thousands or millions of redistricting plans from predefined distributions for analysis [17, 18, 19]. Our paper makes use of the ReCom algorithm, a merge-split MCMC algorithm which has been used by expert witnesses in gerrymandering cases to generate ensembles [20, *Harper v. Hall* 2022, *Allen v. Milligan* 2023]. These techniques are flexible, allowing for many constraints to be taken into account in the ensemble generation process. For example, [21] illustrate how ensembles can be used in the enforcement of the Voting Rights Act.

We use the implementation of the ReCom algorithm in the GerryChain software package [22]. In Section 6 we modify ReCom to select plans with more majority-minority districts (MMDs) using an optimization technique called *short bursts*, which has been shown to outperform other techniques including biased random walks and simulated annealing [23]. We split our short burst chains into a series of 10 sub-chains, each starting from the same initial partition. This was suggested to us by one of the authors of [20], and we found that it helped in discovering plans with more majority-minority districts.

Unless otherwise noted, we require population deviation at most 5% from the ideal, use block groups as the smallest unit of geography, and subsample our ensemble from a run of 1,000,000 steps. Using the Gelman-Rubin split-$\hat{R}$ diagnostic [24], we find that estimates of the metrics we study in Section 5 show clear signs of convergence in the large majority of the geographies we study ($\hat{R} \leq 1.01$, $\mathsf{ESS} \geq 400$). We see clear signs of convergence in a smaller majority ensembles in Section 6. In that context, we view the convergence tests as less important: we are using MCMC sampling more as a way to optimize quantity of interest (the number of MMDs) rather than to estimate one (the MMD discrepancy), in line with the *Gingles 1* test. See Table 3 in the Supporting Information for more information on chain convergence tests.

We generate ensembles for each of 93 state legislative geographies. This includes the upper and lower geographies for 45 bicameral states and one geography for Nebraska's unicameral legislature. It excludes 4 states where our MCMC chains do not find any plans that are population balanced to 5% in DEMO: Hawaii, New Hampshire, North Dakota, and Vermont. This limitation is due to our use of Census block groups rather than Census blocks as the unit of geography in our chains. Since the state legislative districts in these states are relatively small in population, it is computationally inefficient to find valid districts composed of block groups using MCMC methods.

4 Methods in brief and prior work

This paper employs *ensemble analyses*: we algorithmically sample large collections of possible districting plans and observe how statistics of our ensembles differ between our two data sources (see Section 3 for details). In the last five years, ensemble methods for studying gerrymandering and redistricting have made their way from academia [25, 26, 27, 28, 29] to the Supreme Court (*Gill v. Whitford* (2018), *Rucho v. Common Cause* (2019), *Harper v. Moore* (2022), *Allen v. Milligan* (2023)).[4]

To study the effect of the DAS, one would ideally compare the data with and without disclosure avoidance, namely the 2020 CEF and 2020 Redistricting Data. However, the CEF is confidential to those outside of the Census Bureau. We instead use two Census datasets which we call SWAP and DEMO as stand-ins for the 2020 CEF and Redistricting Data, respectively (see Section 3). The Census Bureau created DEMO created to help stakeholders study the impacts of the 2020 DAS by comparing it to SWAP. Both SWAP and DEMO are derived from the 2010 CEF, using the 2010 DAS and 2020 DAS, respectively. See Section 7.1 for further discussion on the limitations of these datasets.

At a very high level, our approach is simple. First, we generate an ensemble of 100,000 plans, say. These plans are drawn to satisfy some constraint—or to maximize some objective—on the *DEMO* dataset (standing in for 2020 Redistricting Data). Second, across plans in the ensemble, we see how often the constraint is violated—or how the value of the objective compares—on the *SWAP* dataset (standing in for 2020 CEF).

[4]Ensemble-based evidence has been more persuasive at trial courts than at the Supreme Court. It was received positively by the dissent in *Rucho v. Common Cause*, but less so in *Allen v. Milligan*: "[C]ourts should exercise caution before treating results produced by algorithms as all but dispositive of a §2 claim."

We call such disagreements in measurements between DEMO and SWAP *discrepancies*. We quantify discrepancies in two important quantities of redistricting plans: the population-deviation, and the number of majority-minority districts.

We adopt our basic setup from Kenny *et al.* [6], who also study the effects of the 2020 DAS on redistricting through an ensemble-based comparison of DEMO and SWAP. They show that a significant fraction of redistricting plans satisfying a given maximum population deviation τ on DEMO exceed the τ limit on SWAP. For example, for $\tau = 1\%$, about 75% of plans exceed the deviation limit on SWAP; for $\tau = 5\%$, about 5% of plans do [6, Figure S4.4]. Kenny *et al.* [6] conclude from this that "the added noise makes it impossible to follow the principle of One Person, One Vote, as it is currently interpreted by courts and policy-makers," pointing to state legislative districts as particularly challenging in this regard.

We disagree. In Section 5, we propose a simple remedy, while also greatly extending the above analysis—using 93 state legislative geographies (instead of 1), and ensembles of 100,000 plans (instead of 5,000). Kenny *et al.* [6] also briefly consider how the DAS affects the apparent number of majority-minority districts that can be drawn—the subject of Section 6 of this paper. Contrary to our results, they observe *fewer* majority-minority districts in DEMO versus SWAP. We suspect this is because their analysis is based on inferred race—using a technique called BISG on records of registered voters—rather than the race data in the redistricting files themselves.[5] We more closely follow the Gingles 1 test by using the race data in the Redistricting Data. See Section 3 for more details on ensemble generation.

4.1 Notation

Let data denote a reference dataset, either DEMO or SWAP. A state-legislative redistricting plan $\mathcal{P}$ is a partition of a state into k districts: $D_1, \ldots, D_k$. Each district D_i is a contiguous collection of census blocks. We denote a district's population in data as $\mathsf{pop}_{\mathsf{data}}(D)$, and it's voting-age population $\mathsf{VAP}_{\mathsf{data}}(D)$.

The ideal population of each district in a plan is $\overline{p}_{\mathsf{data}} = \frac{1}{k} \cdot \sum_{i=1}^{k} \mathsf{pop}_{\mathsf{data}}(D_i)$. It depends only on the total population of the geography and the number of districts. Because the 2010 and 2020 DASes do not affect a state's total population, $\overline{p}_{\mathsf{demo}} = \overline{p}_{\mathsf{swap}}$ for state legislative redistricting. As this setting is our focus, we will use $\overline{p}$ throughout. Fixing $\overline{p}$, we define the *population deviation* of a district and of a plan, respectively, as $\mathsf{dev}_{\mathsf{data}}(D) = |\mathsf{pop}_{\mathsf{data}}(D) - \overline{p}|/\overline{p}$ and $\mathsf{dev}_{\mathsf{data}}(\mathcal{P}) = \max_{i=1,\ldots,k} \mathsf{dev}_{\mathsf{data}}(D_i)$. This follows the convention of [6] of measuring deviation relative to the ideal population, and is a measure directly supported by our MCMC sampler.[6]

We denote by $\mathsf{pop}^{\mathsf{Black}}_{\mathsf{data}}(D)$ and $\mathsf{BVAP}_{\mathsf{data}}(D)$ the total and voting-age Black populations, respectively. We say a district is *majority-Black* if $\mathsf{BVAP}_{\mathsf{data}}(D)/\mathsf{VAP}_{\mathsf{data}}(D) > 0.5$. The number of majority-Black districts in a plan $\mathcal{P}$ according to data is denoted $\mathsf{MMD}_{\mathsf{data}}(\mathcal{P})$.

5 Balancing district populations with noise

A significant fraction of redistricting plans satisfying a given population deviation limit on DEMO exceeds that limit on SWAP, as first shown by Kenny *et al.* [6]. They conclude that as a result it is "impossible to follow the principle of One Person, One Vote," pointing to state legislative districts as particularly challenging in this regard.

This section challenges that conclusion. We propose a straightforward method for ensuring population balance under the unseen dataset, specifically focusing on state legislative redistricting. Our method—using a small tightening of the acceptable population deviation threshold—is very effective, and does not greatly increase the difficulty of finding valid plans nor require the adoption of new redistricting techniques.

[5] Other results from the same case study suffice for the effect they observe. Comparing DEMO and SWAP, the BISG-inferred "proportions of Black and Hispanic [registered voters] are much smaller, especially in the blocks where they form a majority group."

[6] Courts typically instead measure deviation as the largest population minus the smallest, divided by the smallest. To guarantee that this latter measure is at most τ^*, it suffices that $\mathsf{dev}(\mathcal{P}) \leq 2\tau^*/(2+\tau^*)$ (e.g., for $\tau^* = 10\%$, $\mathsf{dev}(\mathcal{P}) \leq 0.094$ suffices).

As noted in Section 2.1, the lack of a *de minimis* population deviation threshold in Congressional redistricting belies the uncertainty inherent in census tabulations; we can only speculate on whether the Court would deem privacy-induced uncertainty exceptional in that regard. Instead, we focus on state legislative districts, where tolerances are looser but precise thresholds are still observed. See Section 7 for a discussion of further limitations, especially the impact of using block groups as our geounit for redistricting.

5.1 Meeting population deviation limits using offsets

In this section, we are concerned with data user's ability to use DEMO to generate redistricting plans satisfying a maximum population deviation of τ under SWAP (equivalently, under the CEF). We propose a simple approach: draw the plan to satisfy a slightly tighter limit $\tau - \Delta$, for $0 \leq \Delta \leq \tau$. We call Δ the *offset*. To evaluate this approach, we measure the *discrepancy at τ with offset* Δ: the fraction of plans that exceed τ deviation under SWAP among an ensemble of plans with deviation at most $\tau - \Delta$ under DEMO. When $\Delta = 0$, we call this the *discrepancy at τ* or the *no-offset discrepancy rate*. Finally, we also consider what we call the *critical offset*: the smallest offset with discrepancy less than 2% (see Appendix A for more information on the computation of the critical offset) .

5.1.1 A case study: the LA state senate

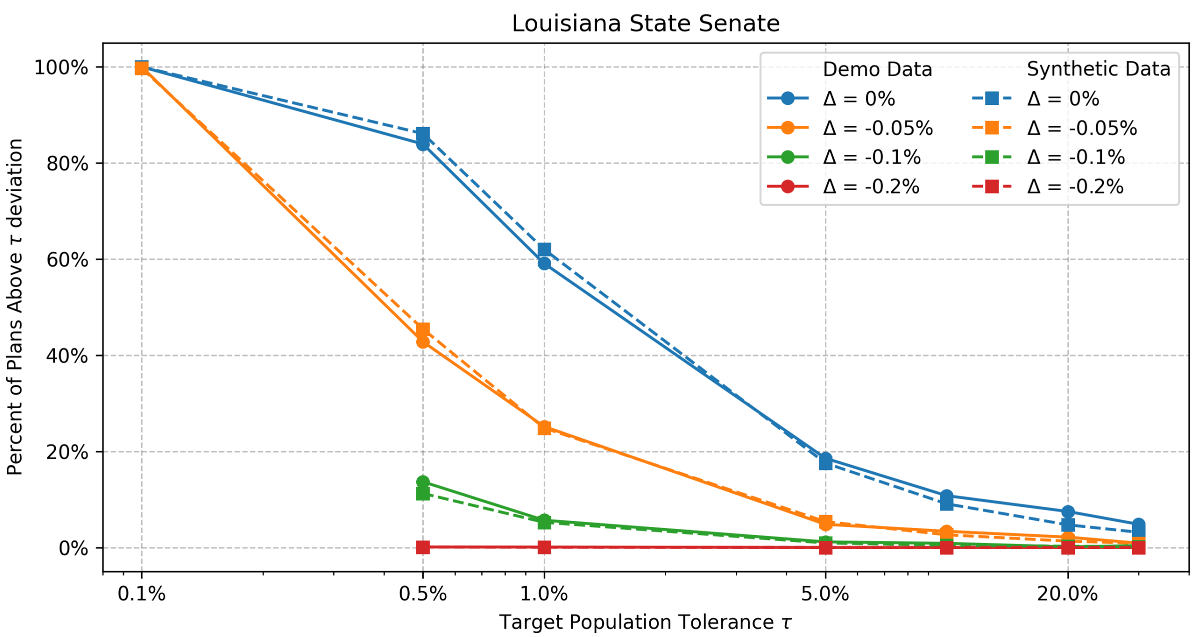

Figure 1: **Discrepancy with offsets in the Louisiana state senate.** We plot the fraction of plans exceeding intended population tolerance limit τ, with various offsets Δ for the Louisiana state senate (i.e., discrepancy at τ with offset Δ). Dots (solid lines) are computed using the DEMO and SWAP datasets with ensembles of 100,000 plans for each τ and Δ (see Sec. 5.1). Squares (dashed lines) are computed from 100,000 samples from the statistical model of district populations and disclosure avoidance noise described in Sec. 5.2.

We begin with a case study of the Louisiana state senate. The LA senate was used by Kenny *et al.* [6, Figure S4.4] to demonstrate the sort of population discrepancies we study. They measured the no-offset discrepancy for each of six population deviation thresholds τ between 0.1% and 30%. We replicate the analysis, summarized by the solid blue line in Figure 1. The overall conclusion remains the same: many

	Min	25^{th}	50^{th}	75^{th}	Max
Discrepancy rate	0.003	0.029	0.090	0.264	0.826
Critical offset (%)	0	0.1	0.1	0.2	1.0

Table 1: Summary of discrepancy rates ($\Delta = 0$) and critical offsets for 93 state legislative ensembles and $\tau = 5\%$ (extrema and quartiles). Note the ensemble maximizing discrepancy does not maximize the critical offset; likewise for the other order statistics.

plans drawn to a satisfy τ population deviation using DEMO exceed that threshold on SWAP. The exact numbers differ, perhaps because we use block groups as the building block (not precincts), a different MCMC sampler implementation, and larger ensembles.

To test the effectiveness of offsets as a mitigation, we measure the discrepancy at τ with twenty values of the offset parameter Δ between 0% to 1%, for the same six values of τ as before. The solid lines in Figure 1 illustrate the results for $\Delta = 0.05\%$, 0.1%, and 0.2%. Using $\Delta = 0.1\%$ reduces the discrepancy at $\tau = 5\%$ from 18.6% to 1.2%. Using $\Delta = 0.2\%$ reduces the discrepancy at $\tau = 1\%$ from 59.1% to 0.1%. Notably, the discrepancy at $\tau = 5\%$ with offset $\Delta = 0.2\%$ was 0. That is, of 100,000 plans generated using a 4.8% tolerance on DEMO, all satisfied a 5% tolerance on SWAP.

5.1.2 Offsets for state legislative redistricting

We perform a similar analysis for 93 state legislative geographies. For each geography, we measure the discrepancy at $\tau = 5\%$ with each of twenty offsets Δ between 0% and 1% (generating an ensemble of 10,000 plans for each offset). For each geography we record the no-offset discrepancy ($\Delta = 0\%$) and the critical offset (where the discrepancy first falls below 2%). Table 1 gives a coarse summary of the results. (See Table 4 in Appendix A for much more detailed results.) Observe that the no-offset discrepancy rates vary widely; in the worst case (Connecticut State House of Representatives) 83% of plans exceeded 5% population deviation under SWAP. But offsets provide a powerful mitigation: for most geographies, $\Delta \leq 0.2\%$ suffices to bring the fraction of problematic plans below 2%.

We do not know why different geographies exhibit different discrepancies and critical offsets, although district size relative to census geographies appears to play an important role. Qualitatively, we find that when districts are many times larger than census tracts, the no-offset discrepancy rate is low; when they are about the same size, it is high. In particular, in state lower house geographies where districts were at least 20 times the size of tracts, the critical offset is 0.1% or less. For the four geographies with fewer than twice as many tracts as districts, the critical offsets are greater than 0.5%. More generally, there is a moderate negative rank correlation between the no-offset discrepancies at $\tau = 5\%$ and the ratio of the number of tracts to number of districts (equivalently, the ratio of the average district population to average tract population). Spearman's rank correlation coefficient is $\rho = -0.645$, with 95% confidence interval (-0.789, -0.434). There is also a strong positive rank correlation between the critical offset and the same ratio ($\rho = 0.840$; 95% confidence interval (0.725, 0.909)). We found no evidence of rank correlations with the number of districts in a plan nor the average district population.

5.2 Why offsets work: most deviation isn't from disclosure avoidance

The effectiveness of our mitigation can be understood by disentangling the population deviation caused by disclosure avoidance from the deviation caused by other factors. The population deviation of an individual district $\mathsf{dev}_{\mathsf{swap}}(D)$ consists of two components:[7]

$$\mathsf{dev}_{\mathsf{swap}}(D) = \left| \underbrace{\frac{\mathsf{pop}_{\mathsf{demo}}(D) - \overline{p}}{\overline{p}}}_{\text{signed } \mathsf{dev}_{\mathsf{demo}}(D)} + \underbrace{\frac{\mathsf{pop}_{\mathsf{swap}}(D) - \mathsf{pop}_{\mathsf{demo}}(D)}{\overline{p}}}_{\mathsf{err}_{\mathsf{das}}(D)} \right|.$$

[7]This identity uses the fact that $\overline{p}_{\mathsf{demo}} = \overline{p}_{\mathsf{swap}}$ for state legislatures. Analyzing sub-state redistricting requires more care.

The first component is a signed version of $\mathsf{dev}_{\mathsf{demo}}(D)$—the apparent deviation in DEMO. This is entirely under the control of the mapmaker. The second component is the additional error from disclosure avoidance, which we denote by $\mathsf{err}_{\mathsf{das}}(D)$.

Plotting these two terms side-by-side clarifies their relative import. Figure 2 plots histograms of signed $\mathsf{dev}_{\mathsf{demo}}(D)$ and $\mathsf{err}_{\mathsf{das}}(D)$ for all the LA state senate districts in our $\tau = 5\%$ ensembles. Note the different scales on the horizontal axes. The signed $\mathsf{dev}_{\mathsf{demo}}(D)$ is roughly uniform across the acceptable range $[-5\%, 5\%]$. The DAS noise $\mathsf{err}_{\mathsf{das}}(D)$ is highly concentrated, with mean 0 and standard deviation 0.060%. Only a small fraction of districts have $\mathsf{dev}_{\mathsf{demo}}(D)$ close enough to 5% for the DAS noise to matter. Using an offset moves $\mathsf{dev}_{\mathsf{demo}}(D)$ away from the 5% threshold, further lowering the number of districts where the DAS noise matters.

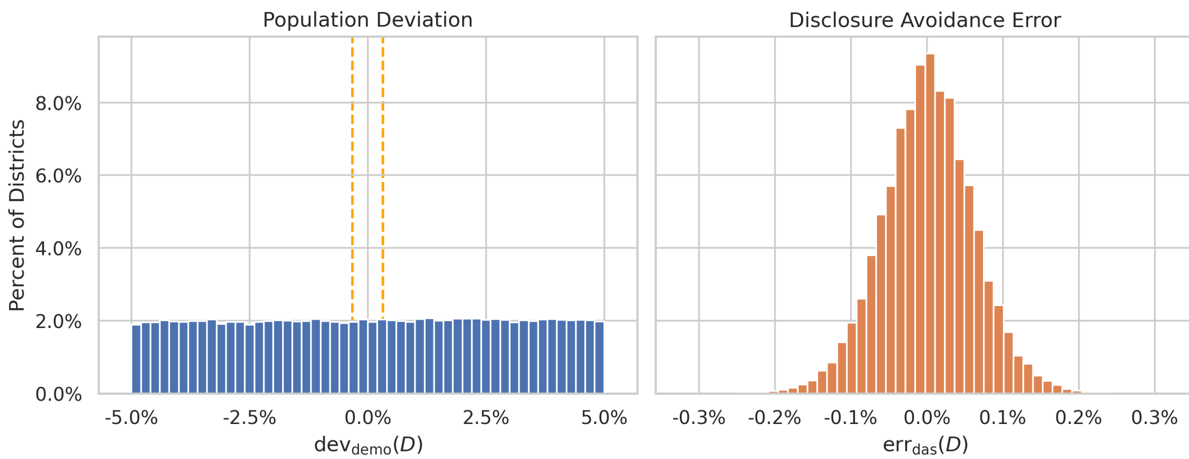

Figure 2: **The two components of population deviation in districts.** Histograms of apparent population deviation in DEMO ($\mathsf{dev}_{\mathsf{demo}}(D) = (\mathsf{pop}_{\mathsf{demo}}(D) - \overline{p})(\overline{p})$) and the additional error from disclosure avoidance ($\mathsf{err}_{\mathsf{das}}(D) = (\mathsf{pop}_{\mathsf{swap}}(D) - \mathsf{pop}_{\mathsf{demo}}(D))/\overline{p}$) for each unique district included in an ensemble of Louisiana state senate plans sampled with a 5% population tolerance on the DEMO data. Note the very different scales on the horizontal axis. Together, these terms make up the population of a state-legislative district's population deviation. The dashed vertical lines behind the left histogram represent the maximum and minimum error values observed in the right histogram.

Figure 2 suggests a very simple probabilistic model for population deviations under SWAP. The deviation of a plan is sampled as the maximum over district-level deviations: $\mathsf{dev}_{\mathsf{swap}}(\mathcal{P}) = \max_i \mathsf{dev}_{\mathsf{swap}}(D_i)$. Each $\mathsf{dev}_{\mathsf{swap}}(D_i)$ is independently sampled as $|X + E|$, where X is uniform over $[-(\tau - \Delta), \tau - \Delta]$ and $E \sim \mathrm{N}(\mu, \sigma^2)$. The parameters $\mu = 0.0\%$ and $\sigma = 0.060\%$ are the empirical mean and standard deviation of $\mathsf{err}_{\mathsf{das}}(D)$ across districts in our LA state senate ensembles.

The dashed lines in Figure 1 show the fraction of 100,000 values of $\mathsf{dev}_{\mathsf{swap}}(\mathcal{P})$ sampled as above that exceeded τ, for each τ and Δ. Qualitatively, the model closely approximates the empirical data for the LA senate, suggesting this is a useful mental model for understanding the effect of offsets. (Note that this is meant only as a simple model of the phenomenon; in particular, the observed values of $\mathsf{err}_{\mathsf{das}}(D)$ are not normally distributed.)

5.2.1 A closer look at errors from disclosure avoidance

To understand the effect of the 2020 DAS on OPOV questions, it appears that we must understand $|\mathsf{err}_{\mathsf{das}}(D)|$. That is, the magnitude of population error in a district D due to disclosure avoidance as a fraction of the ideal district population. This is perhaps unsurprising. We briefly discuss this error in enacted districts; as compared to other sources of error, and in relation to racial biases in the 2020 DAS.

Ideal population $\overline{p}$	<8k	8-16k	16-32k	32-64k	64-128k	128-256k	256-512k	$\geq$512k
Count	104	641	1,057	2,010	1,392	1,232	216	506
Max	0.0168	0.0214	0.0137	0.0069	0.0048	0.0051	0.0041	0.0011
98^{th} pct	0.0088	0.0111	0.0063	0.0039	0.0026	0.0017	0.0015	0.0006
90^{th} pct	0.0047	0.0064	0.0039	0.0023	0.0015	0.0010	0.0006	0.0003

Table 2: Error from disclosure avoidance in legislative districts enacted after the release of the 2010 census redistricting data, as a fraction of ideal population: $|\mathsf{err}_{\mathsf{das}}(D)| = |\mathsf{pop}_{\mathsf{demo}}(D) - \mathsf{pop}_{\mathsf{swap}}(D)|/\overline{p}$. Districts are grouped by ideal population.

Errors in enacted districts Real redistricting plans are drawn by people, not sampled from ensembles. We do not know how offsets fare in a real redistricting setting. But the distribution of $|\mathsf{err}_{\mathsf{das}}(D)|$ on districts created by the political process can give us some initial idea. Using DEMO and SWAP, we compute $|\mathsf{err}_{\mathsf{das}}(D)|$ for every congressional and state legislative district enacted after the 2010 Decennial Census (rather than for algorithmic ensembles of districts). We group the districts by ideal population, which doesn't depend on the redistricting plan. Table 2 reports the maximum, 98th, and 90th percentile values in each group. For the overwhelming majority of geographies—including all with ideal population above $32,000$—the relative error magnitude from disclosure avoidance was well under 1%. A Rhode Island State House of Representatives district saw the largest observed relative error: 2.14%, corresponding to just 301 people.

A redistricter can use these numbers to guide the use of offsets where population deviation under the CEF is a concern, though more research would be warranted. If the ideal district population is $\overline{p} = 200,000$, an offset of $\Delta = 0.1\%$ might be appropriate; if $\overline{p} = 20,000$, an offset of $\Delta = 0.63\%$ appears more appropriate.

Magnitude relative to population drift As a point of comparison for the population deviations so far, we analyze district population imbalances arising from population shifts in the 10 years between decennial censuses. This is only one acknowledged source of population deviation among districts, and ignores systematic undercounts and overcounts that differ by race and ethnicity [30]. Using data from IPUMS that standardizes congressional district geographies between decennial censuses, we compare total populations of congressional districts of the 110th-112th Congresses in both the 2000 and 2010 censuses [31].[8] When measured with the 2000 data used to draw the districts, the deviations are relatively small. Of states with more than one Congressional district, the average deviation is just 613 people (~0.1% of the ideal population), while the maximum is 6,698 (~1% of the ideal). By the 2010 Census, the average increased to 137,974 people (~19% of the ideal), while the maximum increased to 340,948 people (Texas, ~43% of the ideal).

Biases in population discrepancies Prior work found that, in precincts, TopDown's noise is correlated with racial/ethnic homogeneity [6, 32].[9] Precincts with higher Herfindahl–Hirschman index—a measure of racial and ethnic homogeneity—tend to have their populations inflated; precincts with lower index, deflated [6]. This motivates the following question: Does the distribution of district-level errors in our ensembles vary by the racial and ethnic makeup of the districts? If so, it might systematically shift voting power among groups even if district populations are still within the acceptable bounds. Fortunately, we do not observe any meaningful effect. Using of the 1,626,525 unique districts we sampled for the Georgia state house, a geography for which small changes in voting power of Black residents could have a meaningful impact on the makeup of the legislature, we find that the Pearson correlation

[8] Because Census geographies (i.e. block boundaries) are modified every 10 years, districts drawn using 2000 census data may split 2010 census blocks, complicating comparisons between years.

[9] We use the term precinct instead of the more correct "voting district" or VTD to avoid confusion with the much larger state and Congressional districts that are the main subject of this paper [33].

between per-district population error and Herfindahl-Hirschman index is minute: $r = 0.029$, 95% CI $(0.028, 0.031)$.

6 Counting majority-minority districts with noise

To challenge an enacted electoral map for racial vote dilution, plaintiffs must meet the Gingles 1 precondition. Gingles 1 requires showing that one can draw more majority-minority districts (MMDs) than exist in the enacted map being challenged. This is usually done by submitting expert reports with one or more such illustrative plans. As with population deviation, the apparent number of MMDs in a plan may differ by dataset. Absent clear law about which dataset is the appropriate benchmark, Alabama worried that this could make gerrymandering lawsuits more likely, while the ACLU worried this could make those lawsuits less successful (Section 1).

This section asks whether one can show that there are plans with more MMDs than an enacted plan, according to the SWAP dataset (standing in for the 2020 CEF) but using only the DEMO dataset (standing in for the 2020 Redistricting Data). We specifically focus on majority-Black districts. As is typical in VRA litigation, majorities are measured in the voting age population rather than the total population (*Georgia v. Ashcroft* (2003)).

We generate ensembles of plans using a technique called short-burst optimization, which is designed to output plans with many majority-Black MMDs. This is motivated by the observation that Gingles 1 incentivizes plaintiffs to maximize the number of MMDs in the illustrative plan. (See 6.2.1 for a comparison to the base ensemble.) It is generally intractable to determine the maximum possible number of MMDs in a plan at the scale of a US state. Therefore, we use the optimized chains as a proxy for a redistricter who is trying to draw a plan with a near-maximal number of majority-minority districts. While the human and the algorithm differ in their methods for selecting plans, this methodology sheds light on the qualities of plans that are drawn from a nearly-maximal distribution.

We examine the *MMD discrepancy* of plans generated using the DEMO dataset: $\mathsf{MMD}_{\mathsf{demo}}(\mathcal{P}) - \mathsf{MMD}_{\mathsf{swap}}(\mathcal{P})$. We measure the *mean MMD discrepancy* and the *non-zero discrepancy rate*—the fraction of plans with non-zero MMD discrepancy. Section 6.1 shows that MMD discrepancies can be significant, using the GA state house as a case study. It argues that, even so, the Gingles 1 test can continue to be used as is. Section 6.2 explores the phenomenon of MMD discrepancies more deeply.

6.1 Gingles 1 in the face of MMD discrepancy

Figure 3 illustrates the MMD discrepancies that arise in an ensemble of GA state house plans, which serves well as an illustrative geography due to its significant Black population and large legislature. The histogram breaks down the plans by number of majority-Black MMDs according to the DEMO dataset. Each histogram bin further broken down by MMD discrepancy (i.e. the difference between the number of MMDs as measured by DEMO and SWAP). Across the whole ensemble, the mean MMD discrepancy is 0.77, and 56% of plans have a non-zero MMD discrepancy. Observe that plans with more MMDs have greater discepancies (Pearson's $r = 0.21$). In particular, among plans with the maximum number of MMDs (51), 62.45% of plans have a non-zero discrepancy, which is to say most of the optimized plans contained more MMDs according to DEMO than according to SWAP. As shown in Section 6.2, similar patterns hold in many other state legislative geographies.

This do not bode well for a plaintiff bringing a vote dilution challenge (under the legal view that the CEF is the appropriate reference). Gingles 1 incentivizes plaintiffs to create demonstration plans with many MMDs. If the resulting plans behave like those in our short bursts ensemble—which optimizes the same quantity—then those plans may be expected to have fewer MMDs under the CEF, depending on the geography.

Rescuing Gingles 1 We argue that the above interpretation misunderstands Gingles 1: *demonstration plans can continue to be used as evidence in support of Gingles 1 claims*, despite the MMD discrepancies. The relevant legal question is not whether the *specific* plans adduced have so many MMDs. It is whether

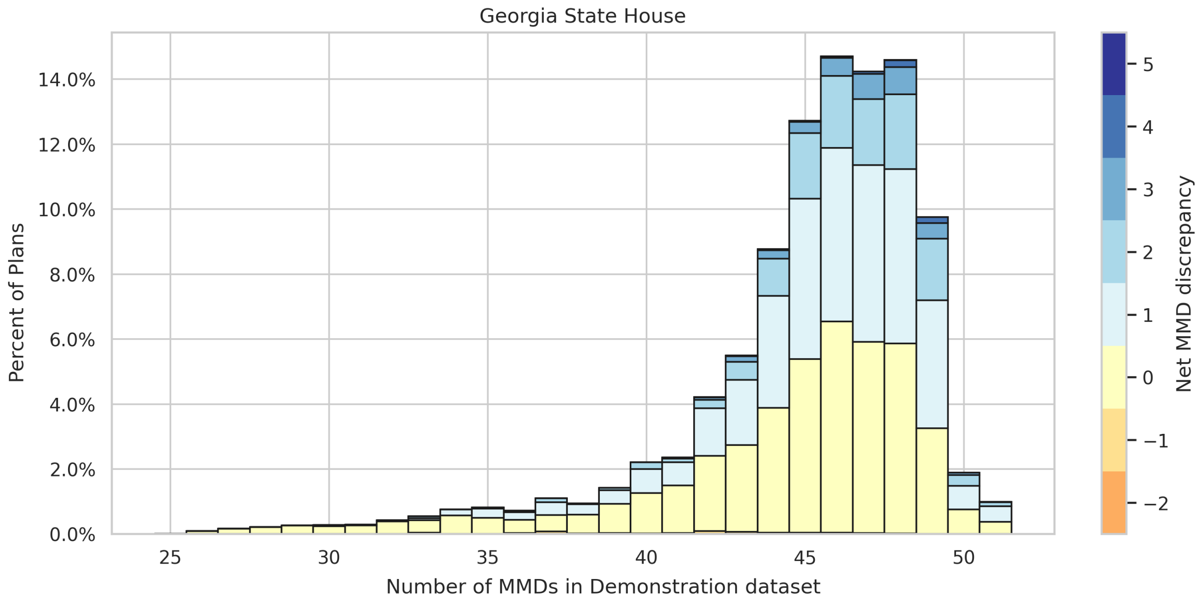

Figure 3: **MMD discrepancies in the Georgia state house.** We plot the number of majority-Black districts in plans sampled in our short bursts ensemble, which is designed to produce plans with many MMDs. Plans are grouped according to the number of majority-Black districts as measured using DEMO. Bars are colored to indicate the fraction of those plans which have the corresponding MMD discrepancy. Each ensemble contains 100,000 plans sampled with a 5% population deviation limit using the DEMO data.

they support a conclusion that there *exist* plans with more MMDs than the enacted plan. (These questions are typically conflated, as they are equivalent when ignoring disclosure avoidance.)

To address this reframed Gingles 1 question, we ask whether the maximum observed number of MMDs within an ensemble is affected by the discrepancy between the two datasets. Concretely, among plans with the most MMDs under DEMO, do any of them have zero MMD discrepancy? *The answer is yes* — in every short-bursts ensemble we sampled. Our ensembles generated using DEMO always yielded plans that are also maximal when measured using SWAP.

The other side of the Gingles 1 equation is the number of MMDs in the enacted plan. It is conceivable that the enacted plan $\mathcal{P}^*$ has more MMDs according to the CEF than according to the Redistricting Data. Let m be the maximum observed number of MMDs in the demonstration plans. A defendant might argue, then, that the enacted plan might maximize the number of MMDs, even if the number measured in the noisy data is less than m. This issue would arise only if $\mathsf{MMD}_{\mathsf{demo}}(\mathcal{P}^*) < m$ and $\mathsf{MMD}_{\mathsf{swap}}(\mathcal{P}^*) \geq m$. In our ensembles, this is unlikely to happen, as the MMD discrepancy is biased in the other direction. In particular, in 52 of 54 ensembles, the fraction of those plans where $\mathsf{MMD}_{\mathsf{demo}}(\mathcal{P}) = m - 1$ that also satisfy $\mathsf{MMD}_{\mathsf{swap}}(\mathcal{P}) > \mathsf{MMD}_{\mathsf{demo}}(\mathcal{P})$ is less than 1%. In the Massachusetts State House, it is 2.0%, and in the California State House, it is 3.9%.

Taken together, if the Gingles 1 precondition is satisfied in the DEMO data—shown using an illustrative plan—then the precondition is likely satisfied in the SWAP data as well, for our ensembles. This holds even when the illustrative plans are more likely than not to have fewer MMDs in SWAP than in DEMO.

So far, we imagined a plaintiff who has found an illustrative map with more MMDs than an enacted map. Does the 2020 DAS make finding such illustrative maps harder? We consider this question only briefly. For each state legislative geography, we generate one ensemble using DEMO and one using SWAP. For each pair, we compare the observed maxima. They agreed in 42 of the the 54 legislative geographies where any majority-Black districts were observed. Of the remaining 12, the maximum was greater (by one) in the SWAP ensemble in 7 cases. The only geography with a difference greater than one district was the Louisiana House of Representatives, with 37 majority-Black districts in DEMO ensemble and 35 in the SWAP ensemble.

6.2 Exploratory analysis of MMD discrepancies

6.2.1 Discrepancies vary greatly by geography and ensemble

We examine the MMD discrepancies in the 54 geographies where our ensembles found at least one MMD (28 lower house, 26 upper house). Different geographies experience very different levels of MMD discrepancy. Some geographies experience very minor discrepancies; others very significant (e.g., 2.68% of OH upper house plans have non-zero discrepancies, compared to 63.45% of MS lower house plans).

Still, some patterns we see in the GA house generalize to almost all geographies. First, positive MMD discrepancies—more MMDs according to DEMO than SWAP—are both more frequent and larger in magnitude than negative discrepancies. Second, plans with more MMDs have greater discrepancies. Moreover, we observe that MMD discrepancies are generally greater in lower houses than in the (typically smaller) upper houses within the same state.

We briefly compare the MMD discrepancies observed in in our short bursts ensembles to those observed in our base ensembles. The short bursts ensemble generally has greater MMD discrepancies than base ensemble. In the base ensemble, the non-zero discrepancy rate is 9% and the mean discrepancy is 0.057. In the short bursts ensemble, those numbers are 56% and and 0.77, respectively. Even so, the short bursts ensemble always succeeds in finding a plan with more MMDs than the base ensemble in the same geography, even accounting for the greater MMD discrepancies.

6.2.2 Discrepancies appear related to BVAP margin

To try to better understand how MMD discrepancies arise, we briefly look at discrepancies in individual districts in the GA house. We observe that districts with small positive BVAP margins are more likely to experience discrepancies. Note this is a preliminary finding based on just a single geography.

For each of the individual districts D generated in the GA state house ensembles, we record the *BVAP margin*: $\mathsf{BVAP}_{\mathsf{demo}}(D) - 0.5 \cdot \mathsf{VAP}_{\mathsf{demo}}(D)$. Figure 4 groups these districts by BVAP margin, and plots the observed fraction of each group with an *MMD discrepancy*: majority Black according to one of the DEMO or SWAP, but not both (i.e., $\mathsf{MMD}_{\mathsf{demo}}(D) \neq \mathsf{MMD}_{\mathsf{swap}}(D)$). Note that no district with BVAP margin greater than ± 150 had an MMD discrepancy. Unsurprisingly, discrepancies are more common as the BVAP margin approaches zero from either side. More interestingly, discrepancies are much more likely when the BVAP margin is positive—just over 50% Black—than negative. This is consistent with the positive bias in *net* discrepancies of the plans.

The correlation between small positive BVAP margins and discrepancies helps explain the greater discrepancies observed in the short bursts ensemble than in the base ensemble. (The observed likelihoods of discrepancy by BVAP margin appears similar across the two ensembles.) Districts with small Black majorities are much more common in the short bursts ensemble than the base ensemble (Figure 5 in Appendix B). This is intuitive. Just as packing and cracking produces maps with few MMDs, we would expect maps with many MMDs to look neither packed (large positive margins) or cracked (small negative margins).

7 Discussion

At the heart of this paper is a legal question. What data should redistricting law treat as the ground truth: the data as enumerated (the CEF) or as published (the Redistricting Data)? We don't answer the question. Instead, our results suggest that the answer may not greatly affect a redistricter's ability to comply with the law. While noticeable discrepancies do arise as a result of the new DAS, it is straightforward to account for those discrepancies when reasoning about plans. Though we are not experts in this area of law, we think it is unlikely that disclosure avoidance would cause courts to reject the long

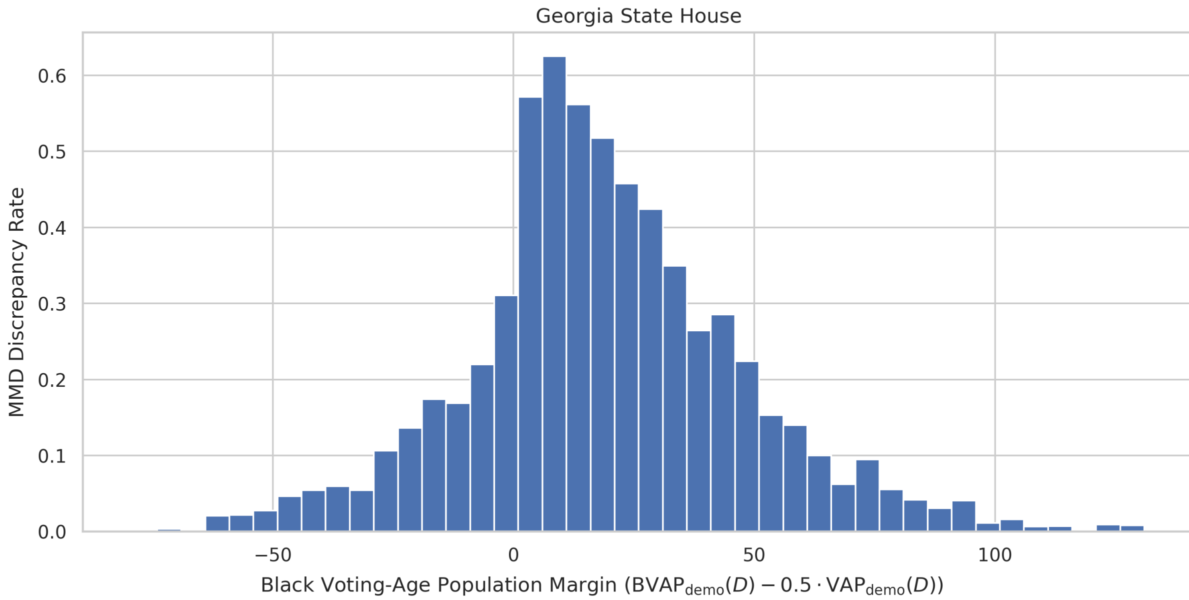

Figure 4: **MMD discrepancy rate by BVAP margin for the Georgia state house.** We group the distinct districts short bursts ensemble according to the Black voting-age population margin. The height of each bar depicts the district-level MMD discrepancy rate for the group. Namely, the fraction of plans measured as majority Black according to one of DEMO and SWAP, but not both. Districts with small Black majorities in DEMO are much more likely to experience an MMD discrepancy than those with small non-Black majorities.

acknowledged "legal fiction" that the Decennial Census exactly reflects the underlying population. But others disagree. The case for statistical adjustment has only strengthened with the Census Bureau's recent release of the so-called Noisy Measurement Files: an intermediate byproduct of the 2020 DAS that can be used to derive statistically better estimates of population than the official Redistricting Data [34].

Our hypothetical data user in this paper is concerned with redistricting plans valid under the CEF. One can instead view the data user as wanting to produce results that would have been valid had the 2010 DAS been used in 2020 (as advocated by critics of the new DAS). For our OPOV analyses, there is no difference between these two views, because the 2010 DAS holds total population invariant. But as the 2010 DAS perturbs race, our analyses of minority representation apply more directly to the latter view, though we believe our overall conclusions to apply in both settings (see Section 3).

7.1 Limitations

An important limitation is that we study algorithmically sampled plans, incorporating only a few redistricting criteria (contiguity, population balance, and compactness). But real districts are drawn by people as part of a political process, taking account of many factors.

Our sample frame does not contain all permissible plans: it excludes plans that subdivide block groups, for instance. Nor does it necessarily contain only permissible plans: ignoring political subdivisions and communities of interest in states where required, for example. It is hard to predict the net effect of these differences our findings. As we explain next, breaking up block groups would tend to strengthen the effects of the noise. On the other hand, the common practice of keeping counties intact would tend to weaken those effects. Future work could use more sophisticated ensemble analyses to account for individual states' redistricting requirements [21, 35], or perhaps a dataset of hand-drawn plans.

We were limited in our computational resources. In order to widen the geographic scope of our analysis, we chose to use block groups, rather than blocks, as the geounit from which our districts were composed.

This greatly reduced the computational complexity, allowing us to run more chains and study larger geographies. But in real-world redistricting, block groups are often split between neighboring districts. To better understand the extent to which this choice impacted our results, we computed the critical offset and no offset discrepancy rate for three small-enough state geographies using blocks as our base geounit (Table 5 in Appendix A). We observe a notable increase in both metrics in all three geographies, indicating that our results may underestimate the impact of the new DAS on real redistricting. We hypothesize the difference is greatest where districts are very small, as in the three examined geographies. This aligns with guidance from the Census Bureau suggesting that data users should aggregate block-level data to increase the accuracy of their analyses.

Another limitation is that comparing the DEMO and SWAP datasets does not isolate the impacts of the 2020 DAS [36]. It reflects the impact of both the 2010 and 2020 DASes. This is inherent. Outside the Census Bureau, it is impossible to directly compare the public 2020 Redistricting Data and the confidential 2020 CEF. DEMO was created to aid stakeholders in understanding the effects of the 2020 DAS by comparing it to SWAP. If we were somehow able to run our analyses using DEMO and the 2010 CEF, nothing in Section 5 would change except possibly in the last paragraph. By using only total population—unaffected by the swapping—we observe the impact of the 2020 DAS alone [37]. On the other hand, Section 6 examines the BVAP as a fraction of the VAP in districts. In 2010, swapping did not affect VAP. But it could affect the BVAP if households with different BVAP were swapped between districts. We cannot correct for this, as the 2010 swapping rates are secret. This is an inherent limitation any work based on comparing the 2010 demo and redistricting data [38].

Other limitations stem from our scope, rather than our methods. Most obviously, we focused on state-legislative redistricting, excluding both Congressional and sub-state redistricting, and on *Gingles 1*, excluding the other parts of the *Gingles* framework and racial-gerrymandering litigation more broadly. So while our results give us hope that discrepancies from disclosure avoidance are tolerable, more serious discrepancies may still lurk elsewhere. As a point of comparison, we provide a limited set of results for experiments on Congressional district geographies in Table 6 in Appendix A. Future work could further study Congressional and sub-state redistricting, or extend [9] by quantifying discrepancies the second and third *Gingles* tests.

Finally, we do not contend with the potential use of the Noisy Measurement File (NMF) for redistricting. While recent work has leveraged the NMF to analyze the error introduced by the 2020 DAS [39], it introduces novel difficulties when applied to redistricting. For example, since hierarchical consistency is not maintained, the population of a district as measured by the NMF can change depending on how its constituent geounits are aggregated. We see addressing this issue and evaluating whether the NMF is legally suitable for this use case as an avenue for future work.

7.2 Additional Related Work

Cohen *et al.* [9] also use ensemble methods to study the effects of the 2020 DAS on redistricting. Instead of using the Demonstration Data Set, they analyze samples of noised data generated using an early version of the 2020 DAS, albeit with imperfect input data. They conclude that the "the practical effects ... [do] not materially threaten any intended uses we considered." For population balance and *Gingles* 1, this is based largely on showing the magnitude noise is small relative to the populations considered and to known sources of error. For *Gingles* 2 and 3, they propose a tweak to the standard method making it more robust to noise. This in part inspired our study of offsets in Section 5.

Among non-ensemble analyses of the impact of the 2020 DAS on redistricting, the most relevant is a study examining the reliability and consistency of measurements of demographic characteristics in geographies of different sizes [40]. These are important for VRA enforcement, but don't do away with issues presented by the sharp legal thresholds we consider. Other works study the impact of the 2020 DAS on different policy issues, including minority representation [41], misallocation of federal funds [42], and public health monitoring [43], for example.

Acknowledgements

We thank Micah Altman, John Abowd, and Moon Duchin for valuable feedback. CC and AC were supported by the DARPA SIEVE program under Agreement No. HR00112020021. Preliminary work on this project was supported by the Boston University Center for Antiracist Research.

References

[1] Laura McKenna. Disclosure avoidance techniques used for the 1970 through 2010 decennial censuses of population and housing. Technical report, U.S. Census Bureau, 2018.

[2] Simson Garfinkel, John M Abowd, and Christian Martindale. Understanding database reconstruction attacks on public data: These attacks on statistical databases are no longer a theoretical danger. *Queue*, 16(5):28–53, 2018.

[3] Travis Dick, Cynthia Dwork, Michael Kearns, Terrance Liu, Aaron Roth, Giuseppe Vietri, and Zhiwei Steven Wu. Confidence-ranked reconstruction of census microdata from published statistics. *Proceedings of the National Academy of Sciences*, 120(8):e2218605120, 2023.

[4] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pages 265–284. Springer, 2006.

[5] John Abowd, Robert Ashmead, Ryan Cumings-Menon, Simson Garfinkel, Micah Heineck, Christine Heiss, Robert Johns, Daniel Kifer, Philip Leclerc, Ashwin Machanavajjhala, Brett Moran, William Sexton, Matthew Spence, and Pavel Zhuravlev. The 2020 Census Disclosure Avoidance System TopDown Algorithm. *Harvard Data Science Review*, (Special Issue 2), Jun 24 2022. https://hdsr.mitpress.mit.edu/pub/7evz361i.

[6] Christopher T Kenny, Shiro Kuriwaki, Cory McCartan, Evan TR Rosenman, Tyler Simko, and Kosuke Imai. The use of differential privacy for census data and its impact on redistricting: The case of the 2020 us census. *Science advances*, 7(41):eabk3283, 2021.

[7] Mike Schneider. People, homes vanish due to 2020 census' new privacy method, Oct 2021.

[8] Gus Wezerek and David Van Riper. Changes to the census could make small towns disappear, Feb 2020.

[9] Aloni Cohen, Moon Duchin, JN Matthews, and Bhushan Suwal. Private numbers in public policy: Census, differential privacy, and redistricting. *Harvard Data Science Review*, (Special Issue 2), 2022.

[10] Ron S. Jarmin, John M. Abowd, Robert Ashmead, Ryan Cumings-Menon, Nathan Goldschlag, Michael B. Hawes, Sallie Ann Keller, Daniel Kifer, Philip Leclerc, Jerome P. Reiter, Rolando A. Rodríguez, Ian Schmutte, Victoria A. Velkoff, and Pavel Zhuravlev. An in-depth examination of requirements for disclosure risk assessment. *Proceedings of the National Academy of Sciences*, 120(43):e2220558120, 2023.

[11] Justin Levitt. Where are the lines drawn? `https://redistricting.lls.edu/redistricting-101/where-are-the-lines-drawn/`. Accessed: 2023-09-11.

[12] Jeff Zalesin. Beyond the adjustment wars: Dealing with uncertainty and bias in restistricting data. *Yale Law Journal Forum*, 130:186, 2020.

[13] Shadie Khubba, Krista Heim, and Jinhee Hong. *National Census Coverage Estimates for People in the United States by Demographic Characteristics: 2020 Post-Enumeration Survey Estimation Report*. US Department of Commerce, US Census Bureau, 2022.

[14] David Van Riper, Tracy Kugler, and Jonathan Schroeder. IPUMS NHGIS Privacy-Protected 2010 Census Demonstration Data, version 20210608 [Database]. Minneapolis, MN: IPUMS, 2020.

[15] US Census Bureau. TIGER/Line Shapefiles.

[16] US Census Bureau. Redistricting Data Program.

[17] Daryl DeFord and Moon Duchin. Random walks and the universe of districting plans. In *Political Geometry: Rethinking Redistricting in the US with Math, Law, and Everything In Between*, pages 341–381. Springer, 2022.

[18] Eric A. Autry, Daniel Carter, Gregory Herschlag, Zach Hunter, and Jonathan C. Mattingly. Multi-scale merge-split markov chain monte carlo for redistricting, 2020.

[19] Benjamin Fifield, Michael Higgins, Kosuke Imai, and Alexander Tarr. Automated redistricting simulation using markov chain monte carlo. *Journal of Computational and Graphical Statistics*, 29(4):715–728, 2020.

[20] Daryl DeFord, Moon Duchin, and Justin Solomon. Recombination: A Family of Markov Chains for Redistricting. *Harvard Data Science Review*, 3(1), mar 31 2021. https://hdsr.mitpress.mit.edu/pub/1ds8ptxu.

[21] Amariah Becker, Moon Duchin, Dara Gold, and Sam Hirsch. Computational redistricting and the voting rights act. *Election Law Journal: Rules, Politics, and Policy*, 20(4):407–441, 2021.

[22] Parker Rule, Matthew Sun, and Bhushan Suwal. mggg/gerrychainjulia, March 2021.

[23] Sarah Cannon, Ari Goldbloom-Helzner, Varun Gupta, JN Matthews, and Bhushan Suwal. Voting rights, Markov chains, and optimization by short bursts. *Methodology and Computing in Applied Probability*, 25(1):36, 2023.

[24] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian Data Analysis, Third Edition*. CRC Press, 2013.

[25] Richard L. Engstrom and John K. Wildgen. Pruning thorns from the thicket: An empirical test of the existence of racial gerrymandering. *Legislative Studies Quarterly*, 2(4):465–479, 1977.

[26] Carmen Cirincione, Thomas A Darling, and Timothy G O'Rourke. Assessing south carolina's 1990s congressional districting. *Political Geography*, 19(2):189–211, 2000.

[27] Jowei Chen, Jonathan Rodden, et al. Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science*, 8(3):239–269, 2013.

[28] Moon Duchin, Taissa Gladkova, Eugene Henninger-Voss, Ben Klingensmith, Heather Newman, and Hannah Wheelen. Locating the representational baseline: Republicans in massachusetts. *Election Law Journal: Rules, Politics, and Policy*, 18(4):388–401, 2019.

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

[30] US Census Bureau. Census Bureau Releases Estimates of Undercount and Overcount in the 2020 Census, March 2022.

[31] Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 17.0 [dataset]. Minneapolis, MN: IPUMS, 2022.

[32] Samantha Petti and Abraham Flaxman. Differential privacy in the 2020 us census: what will it do? quantifying the accuracy/privacy tradeoff. *Gates open research*, 3, 2019.

[33] United States Census Bureau. *Geographic Areas Reference Manual*, chapter Voting Districts. United States Department of Commerce Washington DC, 1994.

[34] United States Census Bureau. 2020 Census Redistricting Noisy Measurement File (NMF), jun 15 2023.

[35] Cory McCartan, Christopher T Kenny, Tyler Simko, George Garcia III, Kevin Wang, Melissa Wu, Shiro Kuriwaki, and Kosuke Imai. Simulated redistricting plans for the analysis and evaluation of redistricting in the united states. *Scientific Data*, 9(1):689, 2022.

[36] danah boyd and Jayshree Sarathy. Differential Perspectives: Epistemic Disconnects Surrounding the U.S. Census Bureau's Use of Differential Privacy. *Harvard Data Science Review*, (Special Issue 2), jun 24 2022. https://hdsr.mitpress.mit.edu/pub/3vj5j6i0.

[37] Christopher T. Kenny, Shiro Kuriwaki, Cory McCartan, Evan T. R. Rosenman, Tyler Simko, and Kosuke Imai. Comment: The Essential Role of Policy Evaluation for the 2020 Census DisclosureAvoidance System. *Harvard Data Science Review*, (Special Issue 2), jan 31 2023. https://hdsr.mitpress.mit.edu/pub/6ffzuq19.

[38] United States Census Bureau. Disclosure Avoidance for the 2020 Census: An Introduction, 2021.

[39] Christopher T Kenny, Cory McCartan, Shiro Kuriwaki, Tyler Simko, and Kosuke Imai. Evaluating bias and noise induced by the us census bureau's privacy protection methods. *Science Advances*, 10(18):eadl2524, 2024.

[40] Tommy Wright and Kyle Irimata. Empirical study of two aspects of the topdown algorithm output for redistricting: Reliability & variability. Technical report, U.S. Census Bureau, 2021.

[41] Miranda Christ, Sarah Radway, and Steven M Bellovin. Differential privacy and swapping: Examining de-identification's impact on minority representation and privacy preservation in the us census. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 457–472. IEEE, 2022.

[42] Ryan Steed, Terrance Liu, Zhiwei Steven Wu, and Alessandro Acquisti. Policy impacts of statistical uncertainty and privacy. *Science*, 377(6609):928–931, 2022.

[43] Nancy Krieger, Rachel C. Nethery, Jarvis T. Chen, Pamela D. Waterman, Emily Wright, Tamara Rushovich, and Brent A. Coull. Impact of differential privacy and census tract data source (decennial census versus american community survey) for monitoring health inequities. *American Journal of Public Health*, 111(2):265–268, 2021. PMID: 33351654.

A Convergence and Robustness Tests

MCMC Convergence tests To provide evidence of the convergence of our chains, we report the Gelman-Rubin split-$\hat{R}$ metric and the effective sample size (ESS) for each legislative geography we study in Table 3. Traditionally, an $\hat{R}$ of below 1.1 has been accepted as evidence of convergence, although more recent work recommends that using a threshold of 1.01 for $\hat{R}$ along with a rank-normalized ESS of at least 400 for more robust results. We test for the convergence of our measurement of population balance (i.e. whether our estimate of the mean of the function $f(\mathcal{P}) = \mathbb{1}\left[\mathsf{dev}_{\mathsf{swap}}(\mathcal{P}) > 0.1\right]$ has converged, where $\mathcal{P}$ is a plan sampled from the stationary distribution of our Markov chain) and our measurement of MMD discrepancy (i.e. $f(\mathcal{P}) = \mathsf{MMD}_{\mathsf{demo}}(\mathcal{P}) - \mathsf{MMD}_{\mathsf{swap}}(\mathcal{P})$). For the MMD discrepancy measurement, we use chains that were optimized with short bursts to include more majority-Black districts. We say that a chain shows signs of convergence after a certain number of steps when $\hat{R} < 1.01$ and $\texttt{ESS} > 400$.

All geographies except for the Wyoming state lower house ($\hat{R} = 1.06$, $\texttt{ESS} = 43.3$) showed signs of convergence after 1,000,000 steps when measuring population balance. Seven geographies did not show signs of converging after 1,000,000 steps under the MMD discrepancy measurement. All six of these geographies have relatively few majority-Black districts; the highest number of majority-Black districts found in any chain from these geographies was three.

	State Lower House				State Upper House			
	Population Balance		MMD		Population Balance		MMD	
	$\hat{R}$	ESS	$\hat{R}$	ESS	$\hat{R}$	ESS	$\hat{R}$	ESS
AL	1.00002	26913.7	1.00077	1133.32	1.00006	133104	1.00204	2142.72
AK	1.00009	31080			1.00002	119936		
AZ	1.00001	154550			1	156374		
AR	1.00042	21480.9	1.0218	197.11	1.00004	119013	1.00302	739.608
CA	1.00004	56517.1	1.00672	1860.24	1.00000	126913		
CO	1.00003	51174.8			1.00004	123734		
CT	1.0005	15570.6	1.00736	480.295	1	111337	1.00018	21121.8
DE	1.00016	30695	1.00174	5329.73	1.00002	137713	1.00275	590.067
FL	1.00022	29016.6	1.00202	2909.33	1.00003	118191	1.00483	3174.69
GA	1.00022	16842.7	1.0081	815.738	1.00006	81420	1.00707	808.596
HI								
ID	1.00003	73215.4			1.00007	70425.9		
IL	1.00004	31295	1.00518	2259.8	1.00007	82855	1.0018	3537.89
IN	1.00022	30818.3	1.00506	1888.19	1.00001	86611.6	1.00803	501.225
IA	1.00016	22155.3			1.00006	65249.7		
KS	1.00025	15885.9	1.04704	114.65	1.00002	90061.6		
KY	1.00003	29711.9	1.00137	5543.87	1.00003	120805	1.00601	547.725
LA	1.00011	24147.8	1.00934	823.373	1.00004	100337	1.00512	659.69
ME	1.0017	2570.13			1.00003	88116.7		
MD	1.00014	50966.3	1.00222	1357.65	1.00003	84614.4	1.00316	1113.23
MA	1.00025	20335.7	1.00194	2972.42	1.00002	113115	1.00565	841.854
MI	1.00007	33451	1.00053	6806.22	1	132017	1.00458	2165.22
MN	1.00013	21548.3	1.02956	159.673	1.00009	57560.7		
MS	1.00018	14836.4	1.00573	689.779	1.00002	59905.3	1.00237	1153.06
MO	1.0003	17052.9	1.00417	1209.08	1.00002	152340	1.00798	471.46
MT	1.00101	4000.78			1.00017	27990.2		
NE					1.00009	57554.3	1.02018	229.16
NV	1.00008	73842.8			1	202695		
NH					1.00002	157558		
NJ	1.00001	120961	1.01255	228.897	1	118854	1.00098	13138.7
NM	1.00011	26332			1.00003	65050.3		
NY	1.00014	26070.7	1.0021	810.3	1.00007	77228.2	1.00249	1700.83
NC	1.00008	28524.7	1.00607	709.641	1.00004	97880.5	1.00678	2020.03
ND								
OH	1.00008	38567.1	1.00784	1555.93	1.00003	158958	1.01815	205.732
OK	1.00026	24267.1	1.00641	434.785	1.00004	73844.2	1.00066	9354.15
OR	1.00012	50909.7			1.00002	142515		
PA	1.00043	14389.1	1.00824	847.099	1.00006	101015	1.00063	4371.36
RI	1.00013	15854.8			1.00009	58597.5		
SC	1.0002	19387.5	1.00863	793.865	1.00007	87268.3	1.00558	1167.13
SD	1.00012	46681.3			1.00006	55289.7		
TN	1.00017	33200.2	1.0081	491.748	1.00001	156826	1.00449	648.003
TX	1.00012	25229.2	1.00654	1943.14	1	93990.8		
UT	1.00002	27696.1			1	131608		
VT					1.00001	272274		
VA	1.00006	34841.9	1.00915	1230.04	1.00001	115771	1.0045	770.027
WA	1.00002	78639.5			1.00012	79849.2		
WV	1.00003	34632			1.00001	244574		
WI	1.00006	31502.7	1.00152	3732.32	1.00001	143569	1.01361	323.169
WY	1.05908	43.2695			1.00033	28393		

Table 3: Convergence tests for state legislative district ensembles. Cells that are highlighted in red represent chains that do not meet the threshold of $\hat{R} \leq 1.01$, $\mathsf{ESS} \geq 400$, suggesting possible non-convergence. Cells that are highlighted in gray represent ensembles for which the $\hat{R}$ computation is undefined, either because no plans were found that were balanced in population in our experimental setup, or because every plan within an ensemble contained the same number of MMDs. Cells that are highlighted in black represent legislative geographies that do not exist.

Robustness of Critical Offset Computation In Table 4 we compute the critical offset for state lower house geographies by repeatedly running chains with a higher offset Δ from the base acceptable τ. In our case, we default to $\tau = 5\%$, and Δ is increased in increments of 0.5%. We repeat this computation ten times for each geography to observe the variation in critical offset between runs. In 17 of the 45 geographies, every test yielded the same critical offset. The largest deviation we observed between the smallest and largest critical offset for a geography was 0.3% for the Wyoming state lower house, with all other geographies having deviations no greater than 0.2%.

Robustness to Geounit Type All results in the paper were reported for chains using the census block group as the base unit of district composition. In Table 5 we report results for our critical offset computations using chains that are composed districts from census blocks, the smallest unit of geography defined in the census hierarchy which is commonly used as the smallest geographic unit in redistricting. Due to computational constraints, we only report results for three legislative geographies: RI state lower house, RI state upper house, and CT state lower house. In all three geographies, we found that using blocks instead of block groups led to an increase in both no-offset discrepancy rate and critical offset.

Robustness to District Type We provide metrics from chains run on Congressional districting plans for each state that was apportioned more than one seat in the House of Representatives after the 2010 census. While there is no population deviation that is generally considered de minimis in Congressional redistricting, we adopt the same $\tau = 5\%$ used in our other experiments for the sake of consistency. We find that in every state, the critical offset is 0%, meaning that fewer than 2% of plans in each ensemble measured a population deviation in excess of τ under SWAP. Similarly, in every ensemble we found that there existed plans which maximized the number of majority-Black districts in DEMO while having an MMD discrepancy of 0.

Robustness to Minority Group We compare our results from Section 5 to state lower house chains optimized to increase the number of majority-Hispanic districts in a given plan, rather than majority-Black districts. Among the ensembles where majority-Hispanic districts were found, all had no-offset discrepancy rates of less than 50%, whereas the Mississippi state house had a no-offset discrepancy rate of 63.45% when optimized to promote majority-Black districts. Only in the Texas state lower house were there no plans found that were maximal in DEMO while having 0 MMD discrepancy.

B Distribution of BVAP Population in Standard and Optimized Ensembles.

Figure 5 plots histograms of the %BVAP for all plans sampled in the base ensemble and the short bursts ensemble. The short burst optimization technique samples heavily from plans that are just over 50% BVAP.

State	Mean Δ	StDev
AL	0.200	0.000
AK	0.420	0.024
AZ	0.000	0.000
AR	0.250	0.000
CA	0.050	0.000
CO	0.115	0.023
CT	0.495	0.027
DE	0.250	0.000
FL	0.100	0.000
GA	0.245	0.015
HI		
ID	0.100	0.000
IL	0.100	0.000
IN	0.145	0.015
IA	0.220	0.024
KS	0.465	0.032
KY	0.155	0.015
LA	0.195	0.015
ME	0.570	0.068
MD	0.100	0.000
MA	0.310	0.030
MI	0.100	0.000
MN	0.250	0.000
MS	0.340	0.049
MO	0.230	0.024
MT	0.595	0.052
NE		
NV	0.150	0.000
NH		
NJ	0.045	0.015
NM	0.350	0.050
NY	0.100	0.000
NC	0.150	0.000
ND		
OH	0.075	0.025
OK	0.300	0.000
OR	0.145	0.015
PA	0.160	0.020
RI	0.825	0.025
SC	0.235	0.023
SD	0.155	0.015
TN	0.110	0.020
TX	0.100	0.000
UT	0.205	0.015
VT		
VA	0.145	0.015
WA	0.060	0.020
WV	0.165	0.023
WI	0.150	0.000
WY	0.585	0.081

Table 4: Mean and standard deviation of critical offset (Δ) computed for each state lower house plan. For each plan, starting at $\Delta = 0$, we sample an ensemble using a population tolerance threshold in demo of $5\% - \Delta$, and compute the percentage of plans that exceed a 5% population tolerance. If this percentage is under 2%, the critical offset is Δ, otherwise we increase Δ by 0.05% and repeat the procedure. The average and standard deviation are computed over 10 experimental runs. Cells that are highlighted in gray represent states where no plans were found that were balanced in population in our experimental setup. Cells that are highlighted in black represent legislative geographies that do not exist.

	Block Group		Block	
	Discrepancy Rate	Critical Offset (%)	Discrepancy Rate	Critical Offset (%)
RI SH	0.767	1.0	0.914	¿1.0
RI SS	0.217	0.3	0.405	0.55
CT SH	0.826	0.45	0.976	0.7

Table 5: Comparison between no offset discrepancy rate and critical offset for three geographies using either block groups or blocks as the smallest unit of district construction.

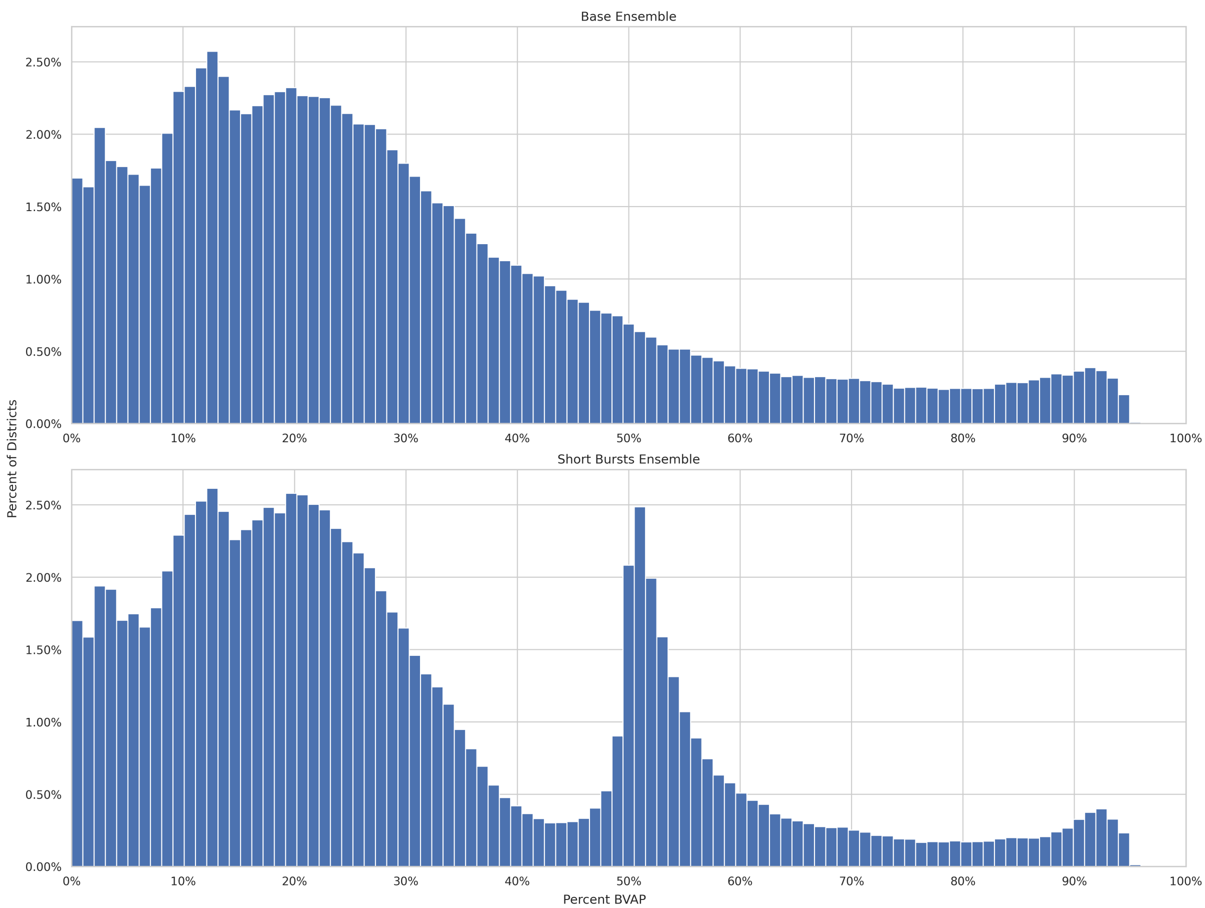

Figure 5: **Histograms of % BVAP population for districts in two Georgia state house ensembles.** The data include all distinct districts in our base (top) and short bursts (bottom) ensembles. The short bursts ensemble, which is designed to produce plans with many MMDs, samples districts with small Black majorities much more often than the base ensemble which ignores race.

	Population Balance		Majority-Minority Districts		
	Standard Constraints		Short Bursts		
State	Δ	DR ($\Delta = 0$)	Mean Discrep.	Discrep. Rate	Max Agreement
AL	0.0	0.00018	0.01519	0.01519	✓
AK					
AZ	0.0	0.00026	0.0	0.0	✓
AR	0.0	0.0	0.0	0.0	✓
CA	0.0	0.01114	0.0	0.0	✓
CO	0.0	0.00022	0.0	0.0	✓
CT	0.0	0.00028	0.0	0.0	✓
DE					
FL	0.0	0.00274	0.07734	0.07734	✓
GA	0.0	0.00118	0.20043	0.19532	✓
HI					
ID	0.0	0.00105	0.0	0.0	✓
IL	0.0	0.00141	0.09004	0.08689	✓
IN	0.0	0.00014	0.0	0.0	✓
IA	0.0	4.0e-5	0.0	0.0	✓
KS	0.0	6.0e-5	0.0	0.0	✓
KY	0.0	0.00012	0.0	0.0	✓
LA	0.0	9.0e-5	0.00068	0.0007	✓
ME	0.0	0.0011	0.0	0.0	✓
MD	0.0	0.00022	0.05793	0.05834	✓
MA	0.0	0.00027	0.0	0.0	✓
MI	0.0	0.00075	0.17567	0.16873	✓
MN	0.0	0.00016	0.0	0.0	✓
MS	0.0	1.0e-5	0.0021	0.0021	✓
MO	0.0	8.0e-5	0.0	0.0	✓
MT					
NE	0.0	4.0e-5	0.0	0.0	✓
NV	0.0	7.0e-5	0.0	0.0	✓
NH	0.0	0.00145	0.0	0.0	✓
NJ	0.0	0.00054	0.02155	0.02155	✓
NM	0.0	3.0e-5	0.0	0.0	✓
NY	0.0	0.00246	0.04771	0.04796	✓
NC	0.0	0.00099	0.00978	0.00978	✓
ND					
OH	0.0	0.00019	0.00491	0.00491	✓
OK	0.0	7.0e-5	0.0	0.0	✓
OR	0.0	9.0e-5	0.0	0.0	✓
PA	0.0	0.00043	0.07582	0.07582	✓
RI	0.0	0.00284	0.0	0.0	✓
SC	0.0	4.0e-5	0.00811	0.00813	✓
SD					
TN	0.0	0.0002	0.11336	0.11336	✓
TX	0.0	0.00657	0.0	0.0	✓
UT	0.0	9.0e-5	0.0	0.0	✓
VT					
VA	0.0	0.00037	0.05446	0.05446	✓
WA	0.0	0.00053	0.0	0.0	✓
WV	0.0	2.0e-5	0.0	0.0	✓
WI	0.0	0.00035	0.0	0.0	✓
WY					

Table 6: Metrics for ensembles of Congressional district plans. States highlighted in black were apportioned a single congressional representative after the 2010 census, and states highlighted in gray had formatting issues in the underlying boundary data preventing ensembles from being generated. For each ensemble sampled with standard constraints, we calculate the critical offset (Δ) and the no-offset discrepancy rate (i.e. the proportion of plans exceeding a 5% population balance threshold under SWAP). For each ensemble sampled using short burst optimization, we calculate the mean MMD discrepancy, the discrepancy rate (i.e. the proportion of plans with a non-zero MMD discrepancy), and whether there exist plans in the ensemble that contain a maximal number of MMDs under DEMO and have an MMD discrepancy of zero.

	Short Bursts		
State	Mean Discrep.	Discrep. Rate	Max Agreement
AL			
AK			
AZ	0.292	0.26553	✓
AR			
CA	0.23149	0.21587	✓
CO	0.16204	0.16143	✓
CT	-0.0043	0.07111	✓
DE			
FL	0.2314	0.2328	✓
GA	0.16762	0.15927	✓
HI			
ID			
IL	0.2395	0.22991	✓
IN			
IA			
KS	0.15886	0.15109	✓
KY			
LA			
ME			
MD	0.02444	0.02674	✓
MA	0.30746	0.27896	✓
MI			
MN			
MS			
MO			
MT			
NE			
NV	0.32114	0.28643	✓
NH			
NJ	0.05054	0.07931	✓
NM	0.58207	0.48976	✓
NY	0.18247	0.17124	✓
NC			
ND			
OH			
OK	0.08246	0.11259	✓
OR			
PA	0.26186	0.24954	✓
RI	0.26918	0.30381	✓
SC			
SD			
TN			
TX	0.36716	0.31295	✗
UT			
VT			
VA			
WA	0.56812	0.49774	✓
WV			
WI	0.16091	0.15871	✓
WY			

Table 7: Metrics for ensembles optimized to increase the number of majority-Hispanic districts. We sampled an ensemble of 100,000 plans for each state lower house geography using short burst optimization, calculating the mean MMD discrepancy, the discrepancy rate (i.e. the proportion of plans with a non-zero MMD discrepancy), and whether there exist plans in the ensemble that contain a maximal number of MMDs under DEMO and have an MMD discrepancy of zero.