Is the Science of Redistricting Possible in an Adversarial Legal System?

Wendy K. Tam[1]

[1]Departments of Computer Science, Political Science, Biomedical Informatics, and the Law School, Vanderbilt University, Nashville, TN, 37203, USA; email: *wendy.k.tam@vanderbilt.edu*

Annual Review of Law and Social Science 2026. AA:1–21

https://doi.org/10.1146/((please add article doi))

Keywords

Redistricting, Algorithmic Simulation, Law

Abstract

The incorporation of scientific methods into redistricting litigation has generated considerable optimism. However, for politically charged disputes such as redistricting, embedding scientific rigor within the judicial process faces substantial hurdles. The adversarial process, by design, is inclined to reward rhetorical force over epistemic integrity. While science may provide important insight, to do so, we must guard against its distortion and strategic misuse. Ultimately, we must also realize that the task of defining fairness in political representation is a deeply philosophical and normative exercise that belongs to the polity and its democratic institutions, where science can serve in a clarifying but not determinative role.

Contents

1. GERRYMANDERING, LAW, AND MEASUREMENT

The Elections Clause in Article I, Section 4 of the US Constitution states that the "[t]he Times, Places and Manner of holding Elections for Senators and Representatives, shall be prescribed in each State by the Legislature thereof." This includes redistricting or the drawing of electoral boundaries. In many states, the majority party of the state legislature assumes the primary role in creating these electoral districts. If these self-interested actors manipulate the composition of the districts to be highly skewed toward one party, the future election results are essentially pre-ordained. Manipulating district boundaries to unfairly favor one group is called gerrymandering.

This legislative power to determine electoral district composition, while broad, is not unfettered; and it is the role of the courts to guard against unconstitutional legislative acts. In *Vieth v. Jubelirer* (541 US 267), the Court made clear that "*excessive* injection of politics is *un*lawful"; and that the courts must intervene when "partisan competition has reached an *extremity* of unfairness." That is, there is no dispute that "partisan gerrymandering" is unconstitutional. The specific legal query in partisan gerrymandering cases is whether the injection of partisanship in the electoral map is excessive.

States are permitted to use partisan and non-partisan information in their mapdrawing. All are acceptable so long as partisanship is not used excessively. The resulting map must also adhere to other legal criteria; districts must be equipopulous and not violate the Voting Rights Act. Generally, districts must also be contiguous, geographically compact, and respect municipal boundaries such as cities and counties. Given that the process operates under these constraints, the central question then is when does the conglomeration of all of these mapmaking decisions result in the electoral map that is "excessive partisanship?" How to answer or what evidence would be probative for answering this legal question is not immediately obvious.

To render a decision, courts require a judicially manageable standard or some clear and coherent principle on which to base their constitutional judgment. While the Court has agreed that "excessive" use of political considerations in mapdrawing is unconstitutional, it has struggled to identify some *objective* way to measure or understand excessiveness. The difficulty was not in recognizing the harm of partisan gerrymandering, but in defining when ordinary politics crosses the line into the realm of the constitutionally impermissible. To judge these cases, the courts have increasingly looked to quantitative and scientific methods for assistance.

Against this backdrop, scholars have proposed a range of metrics and models designed to measure the level of partisanship in an electoral map. These methods purport to provide objectivity and rigor, but they also raise fundamental questions about how scientific measures should be evaluated and validated in legal proceedings.

2. SCIENCE IN THE COURTROOM

2.1. DNA Analysis

The incorporation of scientific methods into legal proceedings has the potential to improve the accuracy and objectivity of judicial outcomes. DNA analysis is perhaps the canonical success story. Famously in the O.J. Simpson trial, DNA analysis was used to match blood drops at the crime scene with Simpson's DNA profile; blood found in Simpson's Bronco was matched to the victims; and the glove found at his property was shown to have DNA that matched both Simpson and the victims (People of the State of California v. Orenthal James Simpson 1995). DNA technology was relatively new at the time. The first U.S. appellate decision to uphold a conviction based on DNA profiling had occurred only 7 years earlier (Andrews v. State 1988).

In the late 1980s, the courts struggled with the admissibility of DNA analysis (See, e.g., People v. Castro 1989, State v. Schwartz 1989, State v. Pennell 1989). In the courtroom, one side would present DNA evidence while the other side would sow doubt by questioning lab procedures and statistical significance. Judges were then tasked with determining if the DNA analysis was legitimate science. Today, DNA evidence typically meets courtroom admissibility standards. In general, the courts treat this type of evidence as highly probative when it is collected, tested, and interpreted under validated and transparent procedures. Challenges to this type of evidence now focus less on the science and more on the procedures for collecting samples and on the interpretation of the statistical claims.

DNA analysis, in the sense of identifying individuals from their genetic material, dates back to experiments performed in Alec Jeffreys's lab in the mid-1980s, years before the introduction of these techniques in the courtroom (Jeffreys et al. 1985). In a case often cited as the first real-world use of DNA profiling, a Ghanaian boy had been denied entry to the UK because it was unclear that he was biologically related to his claimed mother. The boy was admitted to the UK after Jeffreys's lab, using DNA testing, demonstrated that there was a very high likelihood of a parent-child match (Discs 2007). Importantly, the timeline shows that DNA profiling was first established in a laboratory for purposes other than courtroom disputes.

DNA analysis provides an illustrative case where science was able to enter the legal system in a relatively neutral and now, authoritative form.[1] The acceptance of DNA profiling

[1] We are not implying that every use of DNA analysis should be regarded as authoritative. Indeed, it is not, which is evident in that even the standard for its use has evolved over time with, say, the FBI recommending that more than a dozen matches between defendant's DNA and perpetrator's DNA would be necessary to establish credible evidence, when before, as few as three matches would suffice. As well, we are not implying that science of DNA analysis, or any other science, should

did not develop because judges somehow became more adept with their grasp of molecular genetics. Indeed, the legal system did not validate these methods directly, and are clearly not the correct institution to do so. Instead, the science was widely vetted in laboratories before entering the courtroom.

DNA analysis is relevant in the courtroom because it addresses a narrow *factual* question: whether the defendant's genetic material matches a sample recovered from the crime scene. That factual determination may, in turn, provide evidence bearing on a legal question. Although the factual inquiry may be highly probative, it does not, by itself, resolve the legal question. A judge determines the relevance of evidence, scientific or otherwise, within the framework of the applicable legal standard. Other considerations may also be relevant, including intent, whether the burden of proof, beyond a reasonable doubt, has been met, and whether legal defenses such as insanity or self-defense apply. DNA analysis is not intended to produce a definitive legal outcome, or even to address all of the questions a court must consider. Rather, it is relevant, not because of its scientific elegance, but because it helps answer a specific legal inquiry.

2.2. Science as Neutral Arbiter in Adversarial Proceedings

To be sure, despite the DNA success story, the aspiration that science might generally serve as a neutral arbiter of truth in the courtroom faces significant obstacles. Two particular hurdles are especially difficult to overcome. The first is the adversarial legal system, wherein attorneys are driven, not by a duty to seek objective truth, but by the imperative to advocate effectively for their clients (Kagan 2001). The adversarial structure encourages the strategic (rather than the objective) selection of expert witnesses, who may then be pressured to overstate the conclusiveness of their findings or to minimize uncertainties. In this situation, scientific testimony becomes less about pursuing objective truth than about reinforcing predetermined legal narratives (Kousser 1984, Meier 1986, Leigh 1991, Engstrom & McDonald 2011). Issues arising from the structural framework of the adversarial system have been of concern for some time. More than a century ago, Pound (1906), commenting on dissatisfaction with the administration of justice, wrote that our legal system "turns witnesses, and especially expert witnesses, into partisans pure and simple." Indeed, adversarial proceedings create critical tensions and risks for distorting scientific integrity, particularly when judges, juries, and lawyers lack sufficient scientific literacy to independently and rigorously evaluate scientific evidence.

The second obstacle lies in the dynamics of politically charged litigation like redistricting where, in addition to the challenges from the adversarial legal structure, the political stakes from the litigation are enormous. As such, there are even greater and intensified risks for scientific expertise to be distorted and weaponized by political factions. Statistical models, computational simulations, and data choices can become not merely tools of investigation, but instruments for legitimizing predetermined partisan objectives under the guise of scientific analysis (Cho & Cain 2024).

The *Frye* test (293 F. 1013 (D.C. Cir. 1923)) established a "general acceptance" test for scientific evidence in court proceedings. Under this test, a judge could evaluate scientific evidence by identifying a scientist who would state that particular evidence was acceptable

simply be regarded as value-free (Du Bois 1898). It is not. However, these are separate worrying issues that are outside of the scope of the present paper.

within the scientific norms of their field. Of course a judge could insist on general acceptance among a wider range of scientists, but given the more limited acceptable scope, it is easy to see that such a test would be especially problematic in highly partisan cases since it depends only on the self-assessment of a group about whether their own methods are valid.

In 1993, the Court replaced the *Frye* general acceptance test. Under *Daubert*, a judge could no longer abdicate his responsibility for assessing the reliability of scientific evidence by simply querying a scientist about whether a scientific technique is generally accepted (Daubert v. Merrell Dow Pharm., 509 US 579 (1993)). Instead, judges themselves were now required to play a gatekeeping role by assessing the reliability of scientific evidence. According to *Daubert*, "[i]n a case involving scientific evidence, evidentiary reliability will be based on scientific validity." How to make this determination could involve the assessment of whether the method had been peer reviewed, the associated error rates, the original *Frye* test, or other relevant criteria identified by the judge.

Daubert provided broad criteria for judges to evaluate scientific evidence, but it did not instantly imbue judges with scientific literacy (Breyer 1998). As such, judges often rely on the testimony of expert witnesses. Yet, judges struggle when they must choose between dueling experts, each cloaked in technical jargon, each wielding an aura of authority (Faigman 1989). Without proper scientific training, judges are forced to rely on superficial cues of credibility, such as an expert's credentials or demeanor rather than actual scientific rigor (Gatowski et al. 2001). It is not difficult to see how this system might elevate persuasive argumentation but methodologically weak analyses while sidelining more robust but technically complex evidence.

The cumulative effect is troubling. Scientific inquiry is intended to provide judges with a neutral and reliable basis for identifying evidence that is helpful for determining the proper outcome in legal disputes. However, when scientific testimony in politically charged cases is filtered through an adversarial legal system, the strategic deployment of experts often produces conflicting and highly technical claims that judges, lacking the relevant scientific training or expertise, are ill-equipped to adjudicate. In this situation, rather than clarifying facts, expert testimony just deepens uncertainty, leaving courtroom decisions more likely to reflect a judge's partisan predisposition than a reasoned parsing of the underlying evidence (Epstein et al. 2013, Glynn & Sen 2014). These concerns seem to have some backing in the empirical patterns of judicial decision-making in redistricting cases.

2.3. Superficial Evidence of Partisan Tendencies

Cox & Miles (2008) found that the likelihood that a federal judge will vote for the plaintiff in a racial gerrymandering case is highly correlated with the party of the President who nominated the judge. Judges nominated by Democrats were found to be significantly more likely than judges nominated by Republicans to vote for the plaintiff. McKenzie (2012) found a similar pattern in federal court judicial decisions on redistricting cases from 1981 to 2007. While the decisions did not align perfectly with the partisan bent of the federal judges, they were not insulated from them either. At the state court level, Wilhelm et al. (2025) likewise find that party loyalty affects judicial redistricting decisions, particularly when the litigation involves partisan gerrymandering claims. These politically charged cases, at least on their face, appear to present strong challenges to the political impartiality of judges.

These patterns that portend partisanship over judicial neutrality are strikingly similar among expert witnesses who testify in redistricting cases. Experts who testify repeatedly in these disputes almost always align with one partisan side. Those who testify for Republican clients do so consistently, while those aligned with Democratic clients likewise rarely cross over. These observed "partisan tendencies" of the expert witnesses hold true regardless of whether that political party is being accused of partisan gerrymandering or defending themselves from that accusation.

Imagine if this scenario played out in criminal cases involving DNA analysis. Suppose that certain experts consistently testified *only* to incriminate Black defendants while others consistently testified *only* to incriminate White defendants. Suppose further that this was true of virtually all expert witnesses, for all criminal cases. That is, some expert witnesses work exclusively to prosecute Black defendants while other expert witnesses work exclusively to prosecute White defendants. These same experts also sometimes argue that DNA analysis is unreliable, but only when the resulting outcome would favor their "preferred" racial group. It should be easy to see that, even if the underlying science was methodologically sound, such a pattern would be profoundly troubling. It strongly suggests that science is being systematically commandeered for predetermined outcomes. Even were these implications false, the optics themselves remain sufficiently problematic that they must be avoided because they strike at the very heart of public trust in science.

3. THE PARTISAN GERRYMANDERING QUESTION

We now turn specifically to the role of science for addressing partisan gerrymandering.[2] Here, as with DNA analysis, the role of science is not to directly answer the legal question, but to provide evidence for answering a factual question that may be relevant for resolving a legal question. Recall also that DNA testing was developed and validated outside the courtroom, which confers upon it much more credence as a trusted tool within the courtroom. For redistricting analysis to achieve the status that DNA analysis has earned as a neutral arbiter in the courtroom, it must undergo a similar validation process.

3.1. Scientific Methods for Redistricting Analysis

At oral arguments in *Rucho v. Common Cause* (588 US 684), plaintiff's attorney characterized the North Carolina map as "the most extreme partisan gerrymander to rig congressional elections that has been presented to this Court since the one-person/one-vote case." Justice Kavanaugh replied, "when you use the word 'extreme,' that implies a baseline. Extreme compared to what?" One way in which to help the justices form a judicially manageable standard for judging partisan gerrymandering is to identify a scientific method by which such a baseline could be established.

That baseline could arguably be the set of all maps that would satisfy the legal requirements for an electoral map (e.g. equal population, contiguity, the Voting Rights Act). That is, one potentially useful computational exercise would be to produce the entire set of electoral maps that satisfy a set of well-defined legal requirements (hereafter "legal maps"). By definition, any map in this set could legally be the state's electoral map, if the legislature so wished. For all of these possible maps, one could identify their partisan implications. This

[2] Our discussion does not address claims of racial gerrymandering. While racial gerrymandering may involve the same electoral boundaries, the jurisprudence around racial gerrymandering is more mature and distinctly different from that for partisan gerrymandering claims.

would allow Justice Kavanaugh and the other justices to assess whether the disputed map might be extreme in its partisan leaning. Such a baseline set of legal maps has the potential to help us gain a sense for the range of possible partisan outcomes given where the voters in the state reside. Factual knowledge about the range of possible partisan outcomes gives us a way to understand what an "extreme" outcome would be.

While this *might* be a reasonable way to conceptualize a judicially manageable standard, producing the set of all legal maps is not only a formidable task, even within our current computing capability, it is computationally infeasible. The sheer number of legal maps is simply too large. Since an exhaustive enumeration of all legal maps is not possible, a natural alternative is to generate a representative sample from this universe of legal maps. Sampling to produce a representative subset is a common strategy for understanding a large population without having to query each member of the population. Devising a computational method that produces a representative sample of legal maps, however, is itself a non-trivial problem.

Recently, a new generation of models have been introduced, with claims of providing the basis for the elusive judicially manageable standard for adjudicating partisan gerrymandering claims. The hope is that this new technology, which exists primarily in the form of algorithmic simulations, would be able to translate the murky world of political advantage into something the courts could measure, evaluate, and ultimately rule on. These algorithms (e.g., ReCom (DeFord et al. 2021), SMC (McCartan & Imai 2023), and EMCMC (Cho & Liu 2021)) purport to produce a baseline set of maps against which to assess whether a particular map constitutes a partisan gerrymander. How would we validate these methods, understand the output of these algorithms, and determine their relevance as evidence informing the legal inquiry?

We will start with gaining an understanding of the output of these algorithms and then move to discussing their relevance as evidence that informs a legal inquiry. These algorithms are designed to examine real-world redistricting applications. However, real-world instances are too large for us to assess the properties of these algorithms. Instead, we must assess their theoretical properties and behavior on smaller, but non-trivial, data sets where the ground truth is known. If the algorithms are theoretically sound and can be shown to produce the expected results on smaller data sets, then we can be more confident in their theoretical properties to produce reliable results in larger real-world applications where the truth is unknown.

3.2. Validating Scientific Methods

Fifield et al. (2020a) created a validation data set that consists of 25 census tracts from the state of Florida (hereafter "FL25"). To appreciate that this data set is non-trivial, consider that the number of ways to partition 25 precincts into 3 districts without constraints is a Stirling number of the second kind, $S(25, 3) = 141,197,991,025$. Despite how large this number is, with a bit of computational effort, all of the possible combinations of precincts into 3 districts can be enumerated, giving us the complete universe of possible districtings. When we impose a contiguity constraint, requiring each district to consist of contiguous precincts, the number of valid partitions reduces by several orders of magnitude to 117,688. Again, for this data set, we are able to identify, and thus closely examine, every one of the possible districtings. If we further impose a population constraint that requires the population deviation from the ideal population to be less than 10%, the number of valid

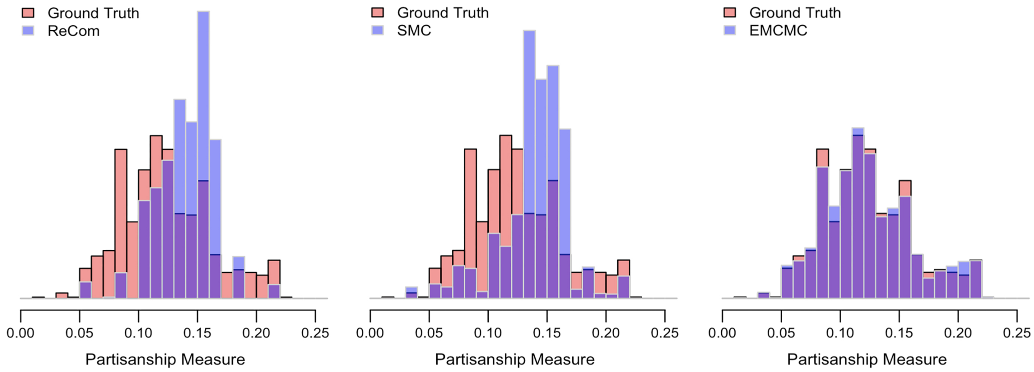

Figure 1

Validation Test using the FL25 Data Set

districtings drops to 927. While FL25 is "small" relative to real redistricting problems, it is well-suited for helping to validate redistricting simulation algorithms because it is large enough to be non-trivial and small enough that we can identify every legal map (i.e. the ground truth). Thus, since we have the true distribution of legal maps, we are able to examine whether these various algorithmic sampling methods are able to produce the true underlying distribution. Any proposed redistricting algorithm that purports to create a representative sample of legal maps should be able to recover, from the FL25 data set, the distribution of 927 maps that satisfy the contiguity and population constraints. Note that none of these algorithms are designed to identify every one of the 927 maps. That is not their purpose. Their purpose is to generate a representative sample of the 927 maps.

FL25 has already been used to demonstrate deficiencies in the Chen & Rodden (2013) as well as the Cirincione et al. (2000) algorithms. As demonstrated by several scholars, neither of those algorithms is able to produce a representative sample of the 927 valid maps in the FL25 data set (Fifield et al. 2020a, Cho & Liu 2018). Given their inability to identify a representative sample in this small data set, we have no reason to believe they could create a representative sample in larger, more complex, real-world redistricting contexts. While the Chen and Rodden algorithm was seen in court frequently at the end of the last decade, it has largely been supplanted by either ReCom or SMC, which now dominate courtroom applications. Yet, surprisingly, neither of these two algorithms has been validated on the FL25 data set.

We embark on that validation analysis here. Figure 1 shows our results with histogram frequency plots. The red histogram shows the *ground truth*, the distribution of a partisanship measure across all 927 maps that satisfy the constraints of contiguity and population equality. In the leftmost plot, the blue histogram shows the distribution of maps produced by the ReCom algorithm. In the middle, the blue histogram shows the results from the SMC algorithm. As we can see from the lack of overlap between the red and blue histograms in both figures, neither ReCom nor SMC produced a representative sample of the 927 legal maps. SMC seems to identify maps across all values of the partisanship measure, but not in the correct proportions. ReCom seems to miss certain types of valid maps all together. Both algorithms oversample certain map types while undersampling others, and thus fail to capture the actual distribution of legally valid possibilities. Accordingly, analyzing the output from either of these algorithms would produced a biased understanding of the possible redistricting outcomes.

We note, however, that two other algorithms have performed better on the FL25 data set. In the rightmost plot in Figure 1, the blue histogram shows the results from the EMCMC algorithm (Cho & Liu 2021). Here, the overlap is substantial, indicating that the algorithm produced a representative sample of the full set of 927 legal maps. In addition, the Fifield et al. (2020a) paper shows that while their basic MCMC algorithm fails to produce a representative sample for the FL25 data set, when they extended their basic algorithm with a "soft constraint" or used "parallel tempering," their algorithm produced a sample that resembles the ground truth.

A central challenge for these algorithms lies in the structure of the search space of legal maps. This space is often disconnected, meaning that it consists of multiple regions that cannot easily be reached from one another through small, incremental changes to a known legal map (Cho & Liu 2018). Each of these "regions" can be thought of as an island. Once an algorithm is on a particular island, it can move around that island easily and identify the other legal maps on that island by making small adjustments to a map. However, without some way to make large changes to a known legal map that result in another legal map, it is unable to identify legal maps that are on different islands. If the algorithm is not designed in such a way that it is able to traverse the separate map islands, it will not properly sample the entire set of legal maps.

Most redistricting algorithms are built on Markov chains, where every Markov chain represents one electoral map. For a Markov Chain Monte Carlo (MCMC) sampling algorithm to properly sample the full set of legal maps, the Markov chain must be irreducible. That means the algorithm, at least in theory, is able to, starting from any legal map, swap geographical units in such a way that it can produce any other legal map. A Markov chain is reducible if its state space can be partitioned into subsets such that, once the chain enters one subset, it cannot reach certain other subsets. Intuitively, the chain becomes confined to one "island" of states and is unable to travel to the others. When this happens, the algorithm fails to explore the full state space and therefore cannot produce representative samples from the intended distribution. Determining whether a Markov chain is irreducible, so that every state is reachable from every other state with positive probability, is a nontrivial task. In the computationally sophisticated models needed for redistricting, the transition structure is highly complex, making the identification of an irreducible Markov chain a difficult, yet essential, task. Since ReCom employs a reducible Markov chain, it fails to sample, or even explore, the entire universe of feasible maps.[3]

EMCMC, by contrast, embeds an optimization routine that enables it to make larger, targeted moves that permit jumps between disconnected map regions. Likewise, while we could not replicate the Fifield et al. (2020a) results, they likely benefited from initiating multiple Markov chains from distinct initial maps, on different islands, running in parallel. For FL25 where the complete set of legal maps is known a priori, such initialization is straightforward to implement. One can simply select different known legal maps from

[3]SMC is not an MCMC algorithm, but is, instead, built on the theory of Sequential Monte Carlo methods. For these methods, successful sampling does not rely on an irreducible Markov chain, but the implementation of a suitable Markov chain nonetheless remains non-trivial for redistricting applications. This particular SMC implementation, which uses spanning trees to determine its transitions, also does not produce a representative sample of the legal maps in FL25.

different regions. For an actual redistricting instance, however, where the universe of legal maps is unknown, this strategy would not work. Indeed, it is non-trivial but essential to embed some type of optimization algorithm or other mechanism capable of discovering and traversing new regions in the map search space. Fifield et al. (2020a) do not discuss such a mechanism as part of their algorithm.

To be sure, there are other "validation data sets." For instance, Fifield et al. (2020b) created a 70 precinct, 2 district data set. While this data set includes more precincts, they only enumerate the maps that have 2 districts. The reduction to two districts substantially simplifies the technical problem, making the search space considerably easier to navigate despite the increase in the number of precincts. For the SMC algorithm, the authors created yet another validation data set with 36 equipopulous units in a six-by-six grid for which they enumerated the possible sets of 6 contiguous districts. A grid structure with uniform units is far more regular than precinct geography so imposing constraints is much less likely to produce the same disconnected search space that is a certainty in real-world redistricting data sets, and also present in FL25. Hence, it is possible that SMC provides a representative sample for the 6×6 grid but still fails on FL25. Each data set has its own structural idiosyncrasies, with some proving to be far more difficult to traverse and sample than others. More validation data sets can and should be developed, but a theoretically sound algorithm ought to produce valid samples across *all* such validation data sets. If an algorithm succeeds in some cases but fails in others, we should understand why and how this might manifest in a real-world application.[4]

4. MATHEMAGICS WITHOUT LEGAL THEORY

The legal system is built around legal theories, not mathematical abstractions. While mathematics can be a useful tool for illustrating arguments or supporting factual claims, it does not automatically map onto legal reasoning. A party cannot evade the need for a coherent legal theory simply by presenting complex formulas or statistical models that leave judges bewildered. Mathematical displays, no matter how sophisticated, should not distract from the central task of applying and interpreting legal standards. Ultimately, it is judges, not mathematicians, who define the governing legal principles. As such, arguments must be framed in terms of those legal standards rather than in the language of mathematics alone. Moreover, though mathematical notation is sometimes difficult to parse, the role of mathematics is to illuminate, not to obfuscate. Accordingly, it behooves those using mathematics to be especially clear with their communication of those mathematics and to situate it appropriately within a legal framework.

We have discussed how, if an algorithm is intended to produce a representative sample

[4] We note that McCartan & Imai (2023) produced yet another validation data set from Florida with 50 precincts and 4 districts. We do not have access to this data set and have not examined it. Indeed, Fifield et al. (2020b) produced even more validation data sets. However, as we have stated, while more data sets is useful, an algorithm that can sample one data set ought to be able to sample any and all of them. FL25 is widely available as part of the CRAN redist package, and developed by variations of the authors of McCartan & Imai (2023) and Fifield et al. (2020a), so it is a bit of a curiosity why there is no longer study of FL25, their original validation data set, or discussion of the idiosyncrasies of the newer validation data sets, the differences between them, and why new methods are accompanied by new validation data sets without examining previously created data sets. Such discussion would be helpful in understanding the nuances for designing these types of sampling algorithms.

of *all* legal maps, this *might* result in a relevant baseline against which judges might assess whether a particular map is an unconstitutional partisan gerrymander. However, it is judges, not mathematicians, who decide if the results from an algorithm are relevant to the legal claim. Moreover, there is not only one way to conceive of a partisan gerrymandering test. Indeed, other conceptualizations are possible.

4.1. The "Carefully Crafted" Test

Chikina et al. (2017) propose a different partisan gerrymandering test that identifies "local outliers," which they term ϵ-outliers. Their logic is that these ϵ-outlier maps are "carefully crafted," meaning that the map drawer could have chosen a large number of other similar maps, but chose not to for partisan reasons. The Court may accept their argument, but to do so, two things must happen. First, their "carefully crafted" phenomenon must be understood within a constitutional theory of partisan gerrymandering (Cho & Rubenstein-Salzedo 2019b). Second, for the court to meaningfully and properly determine the legal relevance of their test, they must at least grasp the intuition of the mathematical abstraction.

In court, one of the authors, in explaining how to understand the results of the ϵ-outlier test, stated that

> when I report that Pennsylvania's 2011 Congressional districting is gerrymandered, I mean not only that there is a partisan advantage for Republicans and that districtings with less partisan bias were available to mapmakers, but indeed that among the entire set of available districtings of Pennsylvania, the districting chosen by the mapmakers was an extreme outlier with respect to partisan bias, in a statistically rigorous way (Pegden 2017).

This statement was not qualified in some way. He did not, for instance, qualify this statement to say that a map might be an ϵ-outlier among a *subset* of legal maps, even if its level of partisan bias is not unusual at all in the set of *all* legal maps. Instead, the language used in court is unequivocal in that the comparison was to "the entire set of available districtings," which is inconsistent with the underlying mathematics in their technical paper where they state that "our method cannot answer the question of how many seats Republicans and Democrats should have in a typical districting. . . because we are still not mixing the chain" (Chikina et al. 2017). Mixing, here, refers to the algorithm reaching a point where it has produced a representative sample of all possible maps. Their algorithm, however, is not designed to do so and examines only set of "similar maps." The relationship of that set of "similar maps" to the set of all possible legal maps is unknown.

The confusion is exacerbated by a conflicting description of the ϵ-outlier test by a different expert for the same redistricting case. The plots in Figure 2 were generated by Moon Duchin, in her role as an expert hired by the Governor. The plot on the left is the result of her analysis of the Pennsylvania's current plan (hereafter "CURR"). Duchin explains that CURR (shown with a red line) is an ϵ-outlier "when it is compared only to plans that closely resemble it which were found by a neutral algorithmic search encoding the stated principles set out by the Court." However, in the plot on the right, the Governor's plan (shown with a purple line, and hereafter "GOV") "falls squarely within the ensemble of similar plans [depicted by the gray area] created using nonpartisan criteria, and therefore gives no reason at all to believe that it was drawn with Democratic-favoring partisan intent" (Duchin 2018). Her language makes evident that the comparison set for both maps is only to "plans that

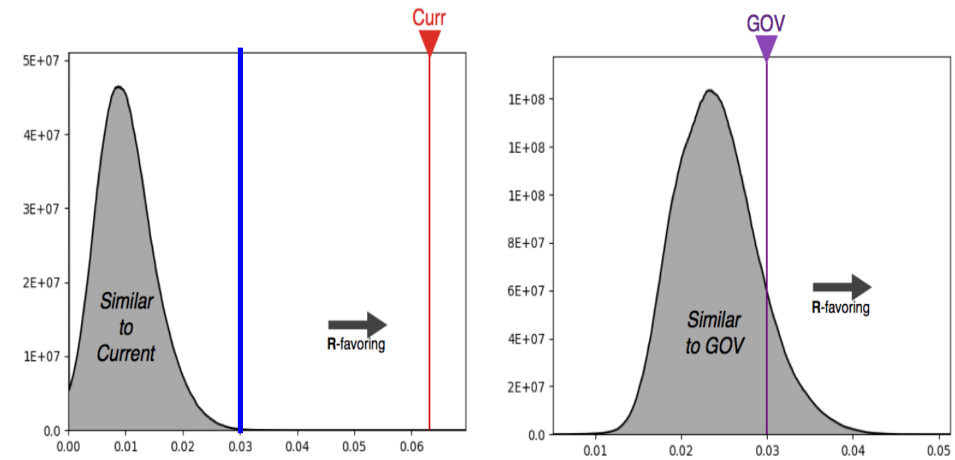

Figure 2

These plots are re-produced from Duchin (2018). The plot on the left is an analysis of the current plan. The plot on the right is an analysis of the Governor's proposed plan. The blue line has been added to the original figure to indicate where the GOV plan would fall on the plot on the left.

closely resemble" the plan in question and not "the entire set of available districting" as Pegden claimed.

Although what "closely resembles" means, either empirically or legally, remains unclear, strong substantive claims about intent and *legal* claims about constitutionality are made based on this analysis. For CURR, Duchin states that "the currently enacted plan show[s] highly significant levels of partisan gerrymandering, with p values more than 10 times as extreme as the tighter common standard." She claims that the GOV plan, "by contrast, falls squarely within the ensemble of similar plans created using nonpartisan criteria, and therefore gives no reason at all to believe that it was drawn with Democratic-favoring partisan intent." Her claims extend far beyond mere factual assertions about the mathematical abstraction.

However, in the same way that DNA analysis addresses only a narrow factual question, mathematical models likewise only speak to factual matters. The judge may rely on this evidence to inform a legal judgment, but only within the proper legal framework and with a clear understanding of what the evidence does and does not establish. DNA analysis describes only whether genetic material matches a certain person. While a judge may find this evidence highly probative, he must still evaluate it through the lens of the law. For redistricting, that legal framework would include, for instance, the Elections Clause that grants states broad authority to set their electoral boundaries. The ϵ-outlier test might provide evidence that a state has exceeded this constitutional discretion, much in the same way that DNA analysis may help a judge determine whether the legal standard of "beyond a reasonable doubt" has been met, but the test, itself, would not establish a legal violation.

To make these legal determinations, the scientific evidence must be clearly understood. Aside from different experts providing conflicting accounts of what the output from this algorithm represents, the judge might be rightly confused on other points as well. For instance, the GOV plan's partisan metric is 0.03, which is a description of how biased the plan is toward one party. This value is the mean-median score for a map, which is described as "a reliable and uncontroversial score with a long pedigree." Notice, however, that in the plot on the left, a plan with such a score would be an extreme outlier because it falls at the very tail end of the distribution she uses for comparison. In the figure on the right, however, the gray distribution that is used for comparison has moved so that a map with this exact same score, and thus partisan leaning, would no longer be an outlier. That is, for a particular fixed value that measures the partisan bias of a map, the ϵ-outlier test is "flexible" in its "judgment" such that that value is extreme among one set of maps created by the algorithm (and thus a partisan gerrymander) while deemed entirely ordinary among a different set of maps created by the algorithm and thus regarded as not a partisan gerrymander. Neither comparison set encompass the entire set of possible maps. And, there is no way to understand either comparison set with regard to any other comparison set.[5]

While it is plausible that the proper comparison set should include only maps that "closely resemble" the map in question, the meaning of "closely resemble" must be elucidated for the court. According to the Duchin report, "trillions of trials" were conducted, each generating a comparison map by shifting "a precinct on the boundary between two districts." If that is indeed the procedure, it seems odd that in every one of trillions of trials, the resulting map would have dramatically shifted the mean-median score from 0.06 to no larger than 0.03. That is, in the CURR comparison set, there are *no* similar maps with a similar partisan metric, even when the only difference between the maps can be only a single precinct swap. Logically, this is not be credible if the underlying algorithm is to be understood as described. The GOV comparison set makes more intuitive sense since the partisan effect of its similar maps does not diverge so dramatically.

The description of these algorithms and how to interpret their results are sufficiently ambiguous and logically inconsistent that it is not possible to use them to make reasoned legal judgment. Although the reports include copious mathematical notation, math alone does not justify a legal outcome. How to incorporate a mathematical formula for an ϵ-outlier in a legal setting remains uncertain at best. Even the factual question being addressed by this particular mathematical exercise is obscure. An ϵ-outlier test may very well be accepted by the courts, but judges, not mathematician, make that decision. To do so, judges must grasp the underlying mathematical concepts (even if not the technical details) to render a reasoned judgment. Critically, while science can inform law, it cannot be conflated with it. The roles of science and law remain distinct even if they interact (Cho & Rubenstein-Salzedo 2019a).

4.2. The "Spanning Tree Distribution"

We showed above that the ReCom algorithm does not produce a representative sample of the set of all legal maps that satisfy a contiguity and population equality constraint in the FL25 data set. The authors never claimed that it would and explicitly acknowledge that their algorithm "cannot (and should not) target the uniform distribution", which would be

[5]Notice as well that in the GOV comparison set, there are well over 100 million maps that have a mean-median score of about 0.025 (though it not clear exactly how many since "1E+08" is listed 3 times on the y-axis), but there appears to be only 500,000 such maps in the CURR comparison set. For CURR, there are almost 5 million maps with a mean-median score of 0.01, but *zero* such maps in GOV comparison set. The GOV comparison set and the CURR comparison set differ dramatically.

the set of all legal maps given a fixed set of legal criteria. Instead, they contend that for their algorithm, "[a]cross a range of small and large experiments with synthetic and observed data, we find that a run assembled in hours to days on a standard laptop produces large, diverse ensembles of plausible districting plans" that "approximately targets a closed-form expression called the spanning tree distribution."[6]

The authors do not, however, rigorously prove convergence to the spanning tree distribution or explain how this distribution might be related or relevant to any particular legal theory.[7] Their claim is that their method "favors plans that look plump and compact to the eye," but, this particular notion of visual compactness is neither a legally defined standard nor a measurable criterion recognized by the courts. Are all legally relevant plans those that appear "plump and compact," or should other visually compact plans also be considered? While the Court has regarded compactness as a neutral traditional districting principle, "plump and compact to the eye" has never been prescribed by the Court as a defintion or benchmark for compactness. Once again, we encounter a mathematical abstraction that simply presumes legal relevance.

To be sure, leaving aside whether the mathematics is correct, judges must be able to discern the relevance and legal significance of what is being presented. Even if ReCom produced the spanning tree distribution, it is unclear why this would matter. In the same way that we needed to understand the implications of only generating "similar maps" for the proposed ϵ-outlier test, we need to understand why alignment with the spanning tree distribution is legally relevant. We know that the spanning tree distribution is not the distribution of all legal maps. As we saw with FL25, ReCom identifies some of the maps that satisfy specified legal constraints, but not a representative sample of them. The legal interpretation of the spanning tree distribution is far from self-evident. Whether these maps in the "spanning tree distribution" are the only relevant ones is a legal decision to be made by the courts, not a mathematical one, and it cannot masquerade as such.

Interestingly, the spanning tree formulation is incorporated not only in ReCom, but in SMC as well as the Autry et al. (2021) algorithms.[8] Why might one be drawn to such a mathematical formulation? One reason is that it is mathematically convenient because it is well-defined, simple to compute, and dramatically reduces the search space so that the computational task is greatly simplified. At the same time, there is no way to measure "plump and compact to the eye." Moreover, given any particular map, one cannot determine

[6]In graph theory, a spanning tree of a connected graph is a subgraph that includes all vertices and contains no cycles (Harary 1972).

[7]The closest they come to a "proof" is running their algorithm on a small 7×7 grid that they divide into 7 districts. For this toy example, they state that "ReCom succeeds at approximating the exact spanning tree distribution on a small grid with respect to the cut edges count. This explicit comparison is not possible on a full-scale problem because of the lack of a complete enumeration of plans." FL25 provides a complete enumeration, so it would be a fine data set for such a test.

[8]The reliance on the spanning tree likely explains why the distributions drawn by ReCom and SMC, while not identical, look quite similar. It would be surprising, in fact, if either came close to the true underlying distribution. If they did, one might rightly wonder how that could even happen given that, theoretically, it should converge to the spanning tree distribution, not the distribution of all "legal maps." Another difficulty not easily surmountable is the previously discussed disconnected nature of the space of "legal maps." Even if a Markov chain based on spanning trees could sample a connected portion of the space, there still needs to be a mechanism to identify the various disconnected regions. For FL25 where the full enumeration is known, one could start separate chains in separate regions, but in an actual redistricting instance, this shortcut is not possible, and so a mechanism to identify disconnected regions is a necessity.

whether it is a member of the "spanning tree distribution." In fact, for any actual electoral map, there is no particular reason to believe it would belong in the spanning tree distribution at all. If not, then comparing a real-world map to a distribution of which it is not even a member makes little sense.

The spanning tree distribution may be a convenient computational simplification, but computational simplicity is not a substitute for legal justification, and pointedly, does not confer legal relevance. If otherwise lawful maps are excluded, then aligning with the spanning tree distribution is not a sign of legal soundness but of a conceptual drift away from the domain of law. Courts must therefore be wary of treating such mathematical formulations as if they carry inherent legal meaning. The question is not whether an algorithm converges quickly, but whether the set of maps it produces reflects the full scope of what the law deems permissible.

4.3. Legal Theory

Even for the set of *all* legal maps, presuming we can identify that distribution, we nevertheless can not simply assume its legal relevance. The legal question is not whether a map is unusual or extreme in a statistical sense, but whether it is unconstitutionally partisan. Here, mathematical analysis can be informative, but these are inherently questions of law. The challenge, therefore, is to identify how a mathematical abstraction can meaningfully assist the Court in resolving the relevant legal questions.

During oral arguments in *Rucho*, Justice Kavanaugh's questioned the idea of how one would establish a baseline for judging excessiveness. Justice Alito asked why, given that all of the baseline maps are non-partisan, one map in this set would be regarded differently than another map in the set? Justice Gorsuch raised doubts about how to interpret the proposed baseline set of maps without introducing a proportional representation (PR) standard.[9] Moreover, how do we incorporate the constitutional authority given to the states through the Elections Clause for determining their electoral maps, including the use of partisan information? These are valid quandaries, and the success of a mathematical exercise does not answer them. Science and law are interdependent, but they are not interchangeable.

First, let us consider Justice Gorsuch's concern that the baseline set of maps embeds a notion of PR. It does not. The baseline set of maps is constructed without the use of partisan information. Because partisan information is not used, proportional representation cannot be built into the resulting distribution (Cho 2019). The baseline set of maps is affected by how partisan voters are residentially clustered in the state. If Democrats are concentrated

[9] The Court has long recognized that our form of political representation is explicitly not proportional representation (PR), so any judicially manageable standard for assessing unconstitutional gerrymandering cannot rest on an expectation that the legislative seats would correspond proportionally to vote shares. See for instance, in *League of United Latin American Citizens v. Perry* (548 US 399 (2006)), where the Court stated, "[t]o be sure, there is no constitutional requirement of proportional representation, and equating a party's statewide share of the vote with its portion of the congressional delegation is a rough measure at best." In *Davis v. Bandemer*, the Court said "[o]ur cases, however, clearly foreclose any claim that the Constitution requires proportional representation or that legislatures in reapportioning must draw district lines to come as near as possible to allocating seats to the contending parties in proportion to what their anticipated statewide vote will be." In *Davis*, the Court referred to *Whitcomb v. Chavis*, 403 US 124 (1971) and *White v. Regester*, 412 US 755 (1973) for how proportional representation analysis has "major deficiencies" when determining whether electoral districts exhibit invidious discrimination in a racial gerrymanering case.

in cities, for instance, adherence to municipal boundaries (an objective the Court accepts as neutral and not partisan per se) will tend to produce "packed" Democratic districts. Such clustering reflects geographic and residential patterns, not partisanship per se. Thus, the baseline captures what emerges from neutral decision rules alone, not from partisan design. Perhaps the resulting distribution shows PR as a natural outcome, perhaps not, but the notion of PR plays no part in the construction of the baseline set of maps.

Next, Justice Kavanaugh questions how a baseline for judgment might be formed. Note that every decision in the mapmaking process, whether motivated by partisan or nonpartisan objectives, has a partisan effect. Each boundary shift moves potential voters across districts, altering the overall partisan composition of the map. The central question, then, is when does the conglomeration of all of these decisions amount to "excessive partisanship?" Here, the set of legal maps forms a baseline that is analytically valuable. The set of legal maps helps us understand the range of partisan outcomes that might occur even when no partisan information is considered. This type of neutral decision-making, while not politically inert, will, on average, balance out. That is, some decisions will favor Democrats while other decisions will favor Republicans, even if the decision is non-partisan at heart. However, if these neutral, "non-partisan decisions" seem to push the final outcome excessively in one partisan direction, the further it pushes it in one partisan direction, the stronger the evidence that the underlying decisions were actually motivated by partisanship.

This logic parallels how we understand tossing a fair coin. Suppose we flip a fair coin 100 times. We expect roughly equal numbers of heads and tails, allowing for natural variation. An outcome of 50 heads and 50 tails would give us no reason to doubt the coin's fairness. Even 55 heads and 45 tails would not raise serious suspicion. But a result of 100 heads and 0 tails would strain credulity in the fairness of the coin. Instead, the rational inference would be that the coin is biased, not that we happened upon an extraordinarily unlikely run of luck. Though, of course, an extraordinary lucky run is theoretically possible.

The same reasoning applies to the distribution of baseline maps generated to satisfy legal constraints without partisan information. If one non-partisan decision happens to make a map more Republican, that is unremarkable. If a second such decision does the same, there is still little cause for concern. But if the next hundred "non-partisan" decisions all move the map in the Republican direction, one is rightfully incredulous that there were no underlying partisan intentions behind these decisions.

As with coin tossing, the further the partisan effect of a legislature's enacted map lies in the tails of the baseline distribution of partisan outcomes, the stronger the evidence of partisan motivation. At some point, the Court might regard the evidence of partisanship to be excessive. This reasoning addresses Justice Alito's concern. He correctly observes that a fair coin could, in theory, land heads 100 times in a row. What he overlooks is that, faced with such behavior and asked to wager any amount of money on whether the coin is fair or biased, the only rational bet would be that the coin is biased.

The ability to create and analyze a baseline set of maps helps us answer a factual question about the degree of partisanship that credibly underlies the decisions that determined the electoral districts. This evidence combined with a sound legal theory allows us to respect the Elections Clause that grants wide latitude to the states to prescribe the times, places, and manner of its elections, preserve our system of geographically based single-member districts, and be divorced from notions of proportional representation. At the same time, it allows for continued judicial oversight by providing a judicially manageable standard for assessing whether legislative decisions are excessively partisan. This framework marries constitutional principles with empirical rigor, with logic that is grounded in law, not in the assertion that a mathematical exercise alone can determine a legal outcome.

5. RECONCEPTUALIZING "FAIRNESS"

While it is technically complex to devise how one might create the baseline representative set of valid maps, the fundamental challenge is not mathematical, but rooted instead in how to conceptualize fair representation. Our discussion to this point presupposes that single member districts are integral to our electoral system. They are not. Single-member districts are a value-laden institutional design choice with representational consequences (Rae 1967, Lijphart 1994, Carey & Hix 2011, Amy 2002). They are not fundamental to democracy. Yet, they shape and constrain how we define fairness. If we wish to seriously engage with representational fairness, we must be willing to consider alternative frameworks, with a nuanced understanding that is grounded in the historical evolution of American society.

For instance, the requirement that electoral districts are contiguous seems natural and not particularly controversial in and of itself, but the logic of this requirement harkens back to a time when representatives were expected to "walk their district" (Engstrom 2013, Kousser 1974). Even before social media and the Internet, this notion that representatives walked door-to-door in their districts had already become antiquated. Moreover, how communities form and how city lines are drawn are historically intertwined with poverty, politics, and racism. So, when we choose to favor districts that are "compact" or to preserve cities and other political subdivisions, while these criteria may seem neutral in their defintion, they also implicitly inhere historical inequalities.

If voters are geographically clustered in ways that make representation inefficient, that pattern may imply that our best notion of fairness would not be tied to these empirical realities. As such, fairness may require a legal standard that is untethered from geography—perhaps even one incorporating some form of rough proportionality. For instance, we might define fairness as the outcome that is as close to rough proportionality as possible within the set of "legal maps," where "legal" may not include geographic constraints such as compactness or the preservation of political subdivisions. Such judgments are fundamentally philosophical or political in nature, not the product of mathematical exercises (Pitkin 1967).

As society evolves, our understanding of fairness and legality should adapt accordingly, informed by the history that has led us to this point, but not historically constrained. If we recognize that compactness requirements can lead to minority exclusion as a result of historic housing discrimination, we might reasonably choose to loosen the compactness criteria. Likewise, city and county boundaries have been shaped by historical processes of exclusion where annexations may have been undertaken to exclude nonwhite areas, motivated either by racial prejudice or by the desire of wealthier communities to avoid financing services for poorer populations (Sugrue 1996, Massey & Denton 1993, Trounstine 2018). These racial and economic spatial patterns were further entrenched by political choices and redlining practices. Against this backdrop, setting numerical thresholds for acceptable number of city or county splits is not a mathematical problem, but a value-laden and normative one that demands serious engagement, not blind adherence to inherited practices.

These considerations all inform how we should define a baseline set of "legal maps." Even widely accepted scientific knowledge must be translated into legal judgments that involve normative decisions about fairness. Such questions are not technical in nature. Instead, they require human deliberation and democratic consensus, not algorithmic development. Ultimately, agreeing on what fairness demands for a richly diverse society is far more difficult than the development of any of the mathematical exercises we have discussed. Mathematics may inform our choices, but it cannot supply our values. Mathematical analysis is relevant only after we have confronted the more difficult task of articulating what fairness means in our society.

6. CONCLUSION

While science has the potential to enhance fairness and accuracy in legal proceedings, the adversarial process often distorts its role, particularly in politically sensitive cases. Lawyers curate experts to fit their narratives; judges and juries struggle to interpret complex evidence; scientific uncertainty is often misrepresented; and science, itself, is not a value-free enterprise. The problem is not simply that the legal system is ill-suited to accommodate complex scientific analyses; it is that the institutional structure actively rewards selective and strategic uses of science by interested parties. The adversarial process does not aim to discover scientific truth but to win disputes, and in this contest, the role of science is reduced to a rhetorical weapon in the hands of the most persuasive advocate. Compounding these challenges, judges, who ultimately determine which scientific testimony is admissible, are not scientists themselves. They frequently rely on unreliable heuristics to assess expert credibility, further muddying the relationship between law and science.

Nowhere are these distortions more dangerous than in politically charged cases, where scientific testimony is often wielded not as a neutral tool for truth-seeking but as a weapon to justify predetermined political outcomes. When courts rely on manipulated or selectively framed science to decide redistricting cases, they risk lending an undue veneer of objectivity to what are actually, and simply, political choices.

At least in the past, political ambitions were more naked, and power plays were bluntly honest about their political intent. That candor has been replaced with the guise of scientific authority that obscures the true underlying political motivations. The pretense of scientific justification, however, does not move us closer to justice or fairness. If we cannot fix the structural issues that allow the adversarial process to subvert scientific integrity, then perhaps we are better off without "science" in politically charged cases at all. A legal system that uses science as a mere tool for advocacy rather than a means of objective understanding does more harm than good.

Institutional reforms can help mitigate these distortions. For instance, court-appointed neutral experts, stricter admissibility standards, and enhanced scientific training for judges are among the measures that could help insulate the courtroom from the political manipulation of science (Cho & Cain 2022). But even such reforms will fall short if the underlying scholarship itself lacks independent validation. The acceptance of DNA evidence in court illustrates how scientific methods can gain legitimacy after they withstand rigorous, transparent scrutiny by the broader scientific community. If we want science to play a constructive role in democratic decision-making, we need new institutional models that are not organized around the adversarial clash of partisan advocates. We need structures that reward epistemic integrity rather than rhetorical power. Regardless, it makes little sense to resist the process of external validation or to substitute courtroom persuasion for genuine scientific consensus.

Finally, a coherent legal theory must connect any scientific or mathematical method to the governing legal standards. Unlike DNA analysis, where the relationship between the scientific evidence and the legal question was more direct and organic, the use of mathematical models in redistricting cases has lacked a similarly transparent connection. It is not that such a connection cannot be made, but there must genuine collaboration and synergy between the mathematical and legal communities. Without a sound legal theory linking mathematics to constitutional or statutory principles, even the most elegant formulas risk becoming instruments of rhetoric rather than tools of reasoning.

Ultimately, we must have the fortitude to build the society we claim to value—one grounded in integrity, justice, and truth—without taking shortcuts to get there. We cannot turn a blind eye to scientific distortions when the distortions favor our political side and then express outrage when the same abuses are wielded against us. The impulse to "fight fire with fire" is, of course, understandable. When one perceives the opposing side as engaging in unrestrained power consolidation, responding in kind can appear not only strategically necessary but even morally justified. Yet, however compelling this reaction may be, it ultimately perpetuates the very cycle of degradation it seeks to counteract, drawing all parties into an ever-lower equilibrium of norms and expectations. This truth is neither easy nor comfortable, and may strike some as naively idealistic. But the discomfort does not diminish its truth. Indeed, enabling science to function as a neutral arbiter within the adversarial process requires strength of character when that discipline is least convenient.

Without the moral resolve to create the conditions under which science can serve justice, the entire enterprise is a pretense that should be abandoned. Politics has always been about power grabs. At the moment, we are witnessing it in its rawest form through an unprecedented barrage of mid-decade redistrictings. Yet, as troubling as these open power grabs may be, their motives are at least transparent. More dangerous is the sleight of hand that disguises partisanship through the language of science, laundering political manipulation into something that looks like objectivity and justice. Exploiting scientific authority for short-term political wins is not merely bending the truth for a noble end, it undermines the value that rigorous science can provide to society and sets the foundation for long-term institutional damage that ultimately hurts us all.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Thanks to Bruce Cain, George Chacko, Kathy Cordeiro, and Emily Zhang for helpful comments.

LITERATURE CITED

Amy D. 2002. Real choices/new voices. Columbia University Press, 2nd ed.

Andrews v. State. 1988. Fla. Dist. Ct. App. 533 So.2d 841

Autry EA, Carter D, Herschlag GJ, Hunter Z, Mattingly JC. 2021. Metropolized multiscale forest recombination for redistricting. *Multiscale Modeling & Simulation* 19(4):1885–1914

Breyer S. 1998. The interdependence of science and law. *Science* 280(5363):537–538

Carey JM, Hix S. 2011. The electoral sweet spot: Low-magnitude proportional electoral systems. *American Journal of Political Science* 55(2):383–397

Chen J, Rodden J. 2013. Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science* 8:239–269

Chikina M, Frieze A, Pegden W. 2017. Assessing significance in a Markov chain without mixing. *Proceedings of the National Academy of Sciences* 114(11):2860–2864

Cho WKT. 2019. Technology-enabled coin flips for judging partisan gerrymandering. *Southern California Law Review Postscript* 93:11–27

Cho WKT, Cain BE. 2022. AI and redistricting: Useful tool for the courts or another source of obfuscation? *The Forum: A Journal of Applied Research in Contemporary Politics* 21:1–14

Cho WKT, Cain BE. 2024. Deploying trustworthy ai in the courtroom: Lessons from examining algorithm bias in redistricting ai. *The University of Chicago Legal Forum* 2023:87–114

Cho WKT, Liu YY. 2018. Sampling from complicated and unknown distributions: Monte carlo and markov chain monte carlo methods for redistricting. *Physica A* 506:170–178

Cho WKT, Liu YY. 2021. A parallel evolutionary multiple-try metropolis markov chain monte carlo algorithm for sampling spatial partitions. *Statistics and Computing* 31:Article 10

Cho WKT, Rubenstein-Salzedo S. 2019a. Rejoinder to "understanding our markov chain significance test". *Statistics and Public Policy* 6(1):54

Cho WKT, Rubenstein-Salzedo S. 2019b. Understanding significance tests from a non-mixing markov chain for partisan gerrymandering claims. *Statistics and Public Policy* 6(1):44–49

Cirincione C, Darling TA, O'Rourke TG. 2000. Assessing south carolina's 1990s congressional districting. *Political Geography* 19(2):189–211

Cox AB, Miles TJ. 2008. Judging the voting rights act. *Columbia Law Review* 108(1):1–54

DeFord D, Duchin M, Solomon J. 2021. Recombination: A family of markov chains for redistricting. *Harvard Data Science Review* 3(1)

Discs DI. 2007. Desert island discs with alec jeffreys. BBC Radio 4

Du Bois W. 1898. The study of negro problems. *The Annals of the American Academy of Political and Social Science* 11(1):1–23

Duchin M. 2018. Outlier analysis for pennsylvania congressional redistricting

Engstrom EJ. 2013. Partisan gerrymandering and the construction of american democracy. Ann Arbor: University of Michigan Press

Engstrom RL, McDonald MP. 2011. The political scientist as expert witness. *PS: Political Science & Politics* 44(2):285–289

Epstein L, Landes WM, Posner RA. 2013. The behavior of federal judges: A theoretical & empirical study of rational choice. Harvard University Press

Faigman DL. 1989. To have and have not: Assessing the value of social science to the law as science and policy. *Emory Law Journal* 38:1005–1095

Fifield B, Higgens M, Imai K, Tarr A. 2020a. Automated redistricting simulator using markov chain monte carlo. *Journal of Computational and Graphical Statistics* 29(4):715–728

Fifield B, Imai K, Kawahara J, Kenny CT. 2020b. The essential role of empirical validation in legislative redistricting simulation. *Statistics and Public Policy* 7(1):52–68

Gatowski SI, Dobbin SA, Richardson JT, Ginsburg GP, Merlino ML, Dahir V. 2001. Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-daubert world. *Law and Human Behavior* 25(5):433–458

Glynn AN, Sen M. 2014. Identifying judicial empathy: Does having daughters cause judges to rule for women's issues? *American Journal of Political Science* 59(1):37–54

Harary F. 1972. Graph theory. Basic Books

Jeffreys AJ, Wilson V, Thein S. 1985. Individual-specific "fingerprints" of human dna. *Nature* 316:76–79

Kagan RA. 2001. Adversarial legalism: The american way of law. Harvard University Press

Kousser JM. 1974. The shaping of southern politics: Suffrage restriction and the establishment of the one-party south, 1880–1910. New Haven: Yale University Press

Kousser JM. 1984. Are expert witnesses whores? reflections on objectivity in scholarship and expert witnessing. *The Public Historian* 6(1):5–19

Leigh LJ. 1991. Political scientists as expert witnesses. *PS: Political Science & Politics* 24(3):521–524

Lijphart A. 1994. Electoral systems and party systems: A study of twenty-seven democracies, 1945–1990. Oxford University Press

Massey DS, Denton NA. 1993. American apartheid: Segregation and the making of the underclass. Harvard University Press

McCartan C, Imai K. 2023. Sequential monte carlo for sampling balanced and compact redistricting plans. *The Annals of Applie Statistics* 17(4):3300–3323

McKenzie MJ. 2012. The influence of partisanship, ideology, and the law on redistricting decisions in the federal courts. *Political Research Quarterly* 65(4):799–813

Meier P. 1986. Damned liars and expert witnesses. *Journal of the American Statistical Association* 81(394):269–76

Pegden W. 2017. Pennsylvania's congressional districting is an outlier: Expert report. Expert Report filed in *League of Women Voters of Pennsylvania et al., v. the Commonwealth of Pennsylvania et al.*, in the Commonwealth Court of Pennsylvania (Civ. No. 261 MD 2017)

People of the State of California v. Orenthal James Simpson. 1995. Cal. Super. Ct., Los Angeles Cnty. No. BA097211, jury verdict

People v. Castro. 1989. N.Y. Sup. Ct. 545 N.Y.S.2d 985

Pitkin HF. 1967. The concept of representation. University of California Press

Pound R. 1906. The causes of popular dissatisfaction with the administration of justice, In *Annual Convention of the American Bar Association*

Rae D. 1967. The political consequences of electoral laws. New Haven: Yale University Press

State v. Pennell. 1989. Del. Sup. Ct. 584 A.2d 513

State v. Schwartz. 1989. Minn. 447 N.W.2d 422

Sugrue TJ. 1996. The origins of the urban crisis: Race and inequality in postwar detroit. Princeton University Press

Trounstine J. 2018. Segregation by design: Local politics and inequality in american cities. Cambridge University Press

Wilhelm T, VandeKamp G, Petrie A, Ray KP. 2025. Redistricting and party loyalty in the the state supreme courts. *Journal of Law and Courts* 13(2):664–679