Statistical detection of partisan gerrymandering in North Carolina Congressional elections

Samuel S.-H. Wang

March 3, 2017 – WORKING DRAFT

Contact information:

Prof. Samuel Wang
Program in Law and Public Affairs
Princeton Neuroscience Institute, Washington Road
Princeton University
Princeton, NJ 08544
Email: sswang@princeton.edu
Web: gerrymander.princeton.edu
Telephone: (609) 258-0388

Qualifications/Background: I am a full professor at Princeton University, with appointments in Neuroscience and Molecular Biology, and I am a Faculty Associate of the University's Center for Law and Public Affairs. I have published over seventy technical articles that use statistical testing, several of which include original contributions to statistical science. In the domain of partisan gerrymandering, I have written articles for the *Stanford Law Review* and the *Election Law Journal*. In addition, I have written about these standards for the *New York Times*.

My gerrymandering standards were awarded a prize in Common Cause's 2016 First Amendment Gerrymander Standard Writing Competition. The judges of this competition were leading scholars and practitioners of election law: Guy-Uriel Charles, Senior Associate Dean for Faculty & Research and founding director of the Duke Law Center on Law, Race and Politics at Duke Law School; Erwin Chemerinsky, Dean, Distinguished Professor of Law, and Raymond Pryke Professor of First Amendment Law at UC Irvine Law School; Allison Hayward, Board member of the Office of Congressional Ethics, former Vice President of Policy at the Center for Competitive Politics, and counsel to FEC Commissioner Bradley Smith; Michael Li, Democracy Program Senior Counsel, Brennan Center for Justice; and Derek Muller, Associate Professor, Pepperdine University School of Law.

Question: Here I review standards by which a statewide gerrymander may be identified using the principle of partisan asymmetry, and to apply those standards to the current Congressional district map of North Carolina. Using these standards, I answer the question of whether that map violates the principle of partisan symmetry, a concept cited by five Supreme Court justices in the 2006 case of LULAC v. Perry (548 U.S. 399).

Executive Summary: North Carolina Congressional election results fail a basic criterion of partisan symmetry in the following manner: for a given distribution of popular votes, if the parties switch places in popular vote, the numbers of seats would change in an unequal fashion. To probe this question further, I applied three statistical tests for partisan asymmetry (68 Stanford Law Review 1263). One test calculates the delegations that would result from a partisan-symmetric process based on nationwide districting patterns. The other two tests are based on century-old principles in the field of statistics, and can be done using a desktop computer or pencil and paper. None of the tests require a detailed consideration of maps. These tests also address weaknesses that could potentially be brought up in the use of other standards. I found that (1) The 2016 election yielded two excess seats for Republican Party compared with a party-symmetric process. (2) "Lopsided wins": The winning vote share in districts won by Democrats was greater than the winning vote share in Republican districts. (3) "Reliable wins": Statewide, the median Republican vote share exceeded the mean (i.e. average) party vote share, and Republican win margins were more uniform than expected from national patterns. Asymmetry favoring Republicans arose abruptly in 2012 and continued in 2014 and 2016, consistent with the effects of biased redistricting arising in the post-2010 cycle, at the same time that Republicans gained control of the process. These results support the hypothesis that North Carolina Republicans have executed an effective and persistent partisan gerrymander.

I. LEGAL BACKGROUND

The term "gerrymandering" describes the act of drawing district lines to make a legislator's victory overwhelmingly likely, by virtue of creating a district with predictable voting patterns. Such a pattern contradicts the saying that "voters should choose their representatives, and not the other way around."[1] One special case of gerrymandering has attracted particular attention from the Supreme Court: that of a partisan gerrymander. In this sophisticated form of gerrymander, individual legislators of both political parties may benefit by gaining safe seats, but the overall effect is to give specific net advantage to one party. In this situation the more-lopsided wins go, perhaps counterintuitively, to individual legislators of the party that does not control redistricting. The net effect is the creation of an overall districting scheme in which delegations do not naturally reflect the overall preferences of the state's voters.

Partisan gerrymandering's unconstitutionality rests on two rationales: the Fourteenth Amendment's Equal Protection Clause and "one person, one vote" principle, and the First Amendment-based protection of speech and association.[2] The justiciability of partisan gerrymandering arises from a series of Supreme Court cases starting with *Davis v. Bandemer* and continuing with *Vieth v. Jubelirer*[3] and *LULAC v. Perry*. In 1986, the Supreme Court established justiciability in *Davis v. Bandemer*.[4] The Court did not find a partisan gerrymander in *Bandemer*, but they did lay out a cause for action based on a two-prong test: 1) intent—an established purpose to create a legislative districting map to disempower the voters for one party; and 2) effect—proof that an election based on the contested districting scheme led to a distorted outcome.[5]

The *Vieth* case concerned whether Pennsylvania's Congressional districts constituted a partisan gerrymander. In that case, five justices voted to dismiss the claim – but five justices also expressed continuing support for the justiciability of partisan gerrymanders. Justice Anthony Kennedy was the only justice to be found in both of these groups. Justice Antonin Scalia wrote a plurality opinion for four justices. He wrote that "to the extent that our racial gerrymandering cases represent a model of discernible and manageable standards, they provide no comfort here [in the partisan context]"[6]. Justice Kennedy wrote a separate concurrence, and also declined to

---

[1] Mitchell N. Berman, *Managing Gerrymandering*, 83 TEXAS L. REV. 781, 781 (2005).

[2] *Bandemer*, 478 U.S. at 122-123; Vieth, 541 U. S., at 314 (J. Kennedy, concurring in judgment; "penalizing citizens because of their participation in the electoral process,. . . their association with a political party, or their expression of political views.", *citing* Elrod v. Burns, 427 U. S. 347 (1976) (plurality opinion)).

[3] *Vieth v. Jubelirer*, 541 U.S. 267 (2003).

[4] *Bandemer*, 478 U.S. at 110.

[5] *Bandemer*, 478 U.S. at 128 (upholding the District Court's finding that the Bandemer plaintiffs were required to prove discriminatory intent and effect).

[6] *Vieth*, 541 U.S. at 286.

join Justice Stevens's opinion stating that Stevens "would apply the standard set forth in the *Shaw* [race] cases" in "evaluating a challenge to a specific district" on partisanship grounds.[7]

Instead of the *Shaw* standard, Justice Kennedy suggested a basis for determining partisan gerrymandering under the First Amendment. Unlike ethnicity or socioeconomic status, identification with a political party can be changed with little effort. In this respect, partisan identification can be regarded as an act of speech or free association, both of which are protected by the First Amendment. In *Vieth*, Justice Anthony Kennedy has noted that the First Amendment can be interpreted as a mandate for "not burdening or penalizing citizens because of their participation in the electoral process, their voting history, their association with a political party, or their expression of political views.[8] Under general First Amendment principles those burdens in other contexts are unconstitutional absent a compelling government interest."[9]

Justice Kennedy did not articulate an exact standard to evaluate partisanship under the First Amendment. Since *Bandemer*, a central difficulty has been establishing a manageable standard, *i.e.* one that provides a reliable and usable determination of whether an offense has occurred. In *Bandemer*, the justices described the effects prong in general terms. Justice White advocated an analysis of an entire districting plan: "A statewide challenge, by contrast, would involve an analysis of 'the voters' direct or indirect influence on the elections of the state legislature as a whole," while also acknowledging that this was "of necessity a difficult inquiry."[10] But eighteen years later in *Vieth*, the plurality opinion stated that no acceptable standard had been established in the intervening time, and therefore it was time to abandon the search.[11] The Court in *Vieth* was notably divided, culminating in five separate opinions.[12] In a separate concurrence, Justice Kennedy provided a fifth vote against invalidating the districts in Pennsylvania, but left the door open for future remedies in other cases if a clear standard could be established.[13] The dissenting four justices voted in favor of a finding of partisan gerrymandering and offered several possible standards, but none was backed by a majority of

[7] *Id.* at 321.

[8] *Vieth*, 541 U.S. at 314 (J. Kennedy, concurring); *Elrod v. Burns*, 427 U.S. 347, 362 (1976).

[9] *Elrod*, 427 U.S. at 362.

[10] *Bandemer*, 478 U.S. at 143.

[11] *Vieth*, 541 U.S. at 279.

[12] *Vieth*, 541 U.S. 267, 271 (opinion. of J. Scalia, joined by C.J. Rehnquist, and O'Connor and Thomas, JJ.); *id.*, 306 (opinion of J. Kennedy, concurring in judgment); *id.,* 317 (opinion of J. Stevens, dissenting); *id.*, 343 (opinion of J. Souter, dissenting, joined by Ginsburg, J.); *id.*, 355 (opinion of J. Breyer, dissenting).

[13] *Vieth*, 541 U.S. at 306 ("I would not foreclose all possibility of judicial relief").

Justices.[14] *LULAC v. Perry* left this judicial stalemate unaltered, but it did contain various endorsements of the symmetry standard, spread across multiple opinions.[15]

In this report I present three tests, rooted in the symmetry principle, that address concerns expressed in the *Vieth* opinions of Justices Scalia and Kennedy. My method has advantages offered by mathematical rigor previously absent from the Court's opinions on partisan gerrymandering. By translating principles that have emerged from Constitutional jurisprudence into the language of classical statistics, these tests may plug a hole that has been left unfilled by the Court.

II. A FIRST LOOK AT PARTISAN ASYMMETRY

A basic test of partisan symmetry can be done by shifting the vote totals in each district by the same amount, enough to switch the total vote share for the two parties. The result of this hypothetical scenario is illustrated in **Table 1**. In the flipped scenario, the statewide vote is

Actual 2016 outcome in North Carolina			
District	**D vote**	**R vote**	**Winner**
1	70.3%	29.7%	D
2	43.3%	56.7%	R
3	32.7%	67.3%	R
4	68.2%	31.8%	D
5	41.6%	58.4%	R
6	40.8%	59.2%	R
7	39.1%	60.9%	R
8	41.2%	58.8%	R
9	41.8%	58.2%	R
10	36.7%	63.3%	R
11	35.9%	64.1%	R
12	67.0%	33.0%	D
13	43.9%	56.1%	R
Statewide	**46.3%**	**53.7%**	**10 R, 3 D**

Shifted by a constant amount to flip totals			
District	**D vote**	**R vote**	**Winner**
1	77.6%	22.4%	D
2	**50.6%**	49.4%	D
3	40.0%	60.0%	R
4	75.5%	24.5%	D
5	48.9%	51.1%	R
6	48.1%	51.9%	R
7	46.4%	53.6%	R
8	48.5%	51.5%	R
9	49.1%	50.9%	R
10	44.0%	56.0%	R
11	43.2%	56.8%	R
12	74.3%	25.7%	D
13	**51.2%**	48.8%	D
Statewide	**53.7%**	**46.3%**	**8 R, 5 D**

Table 1, A direct test for partisan asymmetry. *Left:* Results for the 2016 Congressional election. *Right:* Results are all shifted by a constant amount, 7.3% in each district, enough to exchange the statewide vote totals between the two parties. In this scenario, the size of the delegations is not switched, and still has a majority of Republicans.

---

[14] *Id.*

[15] *LULAC*, 548 U.S. at 468 (n.9) (opn. of Stevens, J. P., joined by Breyer, S) ("a helpful (though certainly not talismanic) tool"). *LULAC*, 548 U.S. at 473 (n. 11) (opn. of Stevens, J. P.; asymmetry as one of eight criteria he would use for determining effects-based violations). LULAC, 548 U.S. at 466 (opn. of Stevens, J.) ("Plan 1374C [the challenged plan] is inconsistent with the symmetry standard"). LULAC, 548 U.S. at 483 (opn. of Souter, J.)("do not rule out the utility of a criterion of symmetry"; "interest in exploring this notion is evident [on the Court]"). LULAC, 548 U.S. at 420 (opn. of Kennedy, J. joined by Justices Souter and Ginsburg)(indicating use as a standard based on election results, but not hypothetical future results).

53.7% Democratic, 46.3% Republican – but the delegation is still majority Republican, with 8 Republicans and 5 Democrats.

When this principle of symmetry is applied to earlier elections, a pattern becomes apparent (**Table 2**). From 2012 to 2016, the Congressional delegation is nearly unchanged by exchanging the statewide vote share, maintaining a Republican majority no matter which side gets more votes. In contrast, 2008 and 2010 showed an advantage favoring Democrats. In short, the current partisan asymmetry arose at the same time as post-2010 redistricting.

Year	Actual D% vote	Delegation	Flipped D% vote	Delegation
2008	55.2%	5 R, 8 D	44.8%	6 R, 7 D
2010	46.1%	5 R, 8 D	53.9%	5 R, 8 D
2012	51.4%	10 R, 3 D	48.6%	10 R, 3 D
2014	47.0%	10 R, 3 D	53.0%	10 R, 3 D
2016	46.3%	10 R, 3 D	53.7%	8 R, 5 D

Table 2: Net effect of exchanging the statewide popular vote, 2008-2016. Starting in 2012, a Republican majority became highly likely independent of which party wins a majority of the popular vote, or if the vote totals were exchanged.

III. STATISTICAL ANALYSES OF PARTISAN ASYMMETRY

Commonly, gerrymanders (racial, for example) are diagnosed by analyzing specific districts. However, partisan gerrymandering emerges from *patterns* of districting, and examination of a single district does not clearly identify partisan asymmetry. Indeed, any particular district may contribute to an overall advantage for its winner's party, to its loser's party, or to neither party, depending on the overall redistricting scheme. A partisan gerrymander can only be reliably diagnosed when considering a state's whole districting plan at once.

To be manageable, a numerical standard for gerrymandering has to answer two fundamental questions: (a) what is a reasonable cutoff beyond which a justiciable offense has occurred, and (b) what standard of fairness should be used for comparison? Statistical science, which is over a century old, may provide some help in addressing both questions.

For a cutoff value, I will use a longstanding criterion, that of "statistical significance." Statistical significance is quantified as the probability that a neutral process would have generated the observed outcome by chance.[16] In this report, I refer to the likely range of outcomes from a neutral process as a "zone of chance."

The second question is how to identify a neutral standard, around which this zone of chance would fall. To define that standard, I use national districting practices as expressed by

[16] In statistical scholarship, this probability is called p or alpha. For example, if an outcome would have arisen by chance 5% of the time, that would be called p=0.05 or alpha=0.05.

election outcomes (Test 1) and partisan symmetry (Tests 2 and 3). These tests are described in greater detail elsewhere[17].

Tests 2 and 3 detect patterns of partisan outcomes that are unlikely to have arisen by chance in a partisan-symmetric process. Such tests are well established in the scientific community as a way of testing for differences between two groups of observations (in this case, groups of districts), or overall asymmetry (in this case, the pattern of advantages gained by two political parties). The tests are taught to undergraduates and are accessible to anyone with an introductory statistics textbook and a spreadsheet program – or even pencil and paper. Judges may use these tests to analyze rapidly whether a pattern of election outcomes is likely to have arisen from a partisan process. In short, these tests are manageable.

Test 1 (the excess seats test): This test calculates, for a given statewide vote total, the range of number of seats that is likely to arise, assuming that the patterns of real-life national districting practice are used to apportion voters. Then calculate the difference between this range and the outcome of the actual election, and ask whether that outcome favors the redistricting party.

Test 2 (the lopsided outcomes test): This test compares the winning vote shares in Democrat-represented districts with the winning vote shares in Republican-represented districts. In a partisan gerrymander, the targeted party wins lopsided victories in a small number of districts, while the gerrymandering party's wins are engineered to be relatively narrow. To test whether the two groups of winning vote shares differ, use the two-sample t-test, a widely used statistical test.

Test 3 (the reliable-wins test): Systematic rigging of total statewide outcomes occurs by the construction of districts that offer secure wins for the party in control of the map. These wins would be wide enough to guarantee victory, but not so wide as to waste votes that could be used to shore up other districts. How this intent is detected depends on whether the state's partisan vote is closely divided, or whether one party is dominant.

In a closely divided state or when the redistricting party gets a minority of the votes: In a closely divided state, reliable wins occur when the average and median vote differ from one another. To perform Test 3, calculate the difference between a party's statewide average district vote share on the one hand, and the median vote share it receives on the other. In this situation a systematic gerrymander can be detected when the redistricting party's median vote share is substantially above its average vote share across districts.[18]

---

[17] Samuel S.-H. Wang, *Three Tests for Practical Evaluation of Partisan Gerrymandering*, 68 STANFORD LAW REVIEW 1263 (2016); Samuel S.-H. Wang, Three practical tests for gerrymandering: application to Maryland and Wisconsin. 15 *Election Law Journal* 367 (2016).

[18] This is the mean-median test described in Wang, *supra* 5 and Wang, *supra* 26, and by Robin E. Best & Michael D. McDonald, *Unfair Partisan Gerrymanders in Politics and Law: A Diagnostic Applied to Six Cases.* 14 ELECTION LAW JOURNAL: RULES, POLITICS, AND POLICY 312 (2015).

In a state where the redistricting party is dominant: A hallmark of gerrymanders is the reliability of outcomes in individual races. In a state that is dominated by one party, reliable wins occur when that party's strength is spread highly evenly across districts. To perform Test 3, calculate the standard deviation of the redistricting party's vote share in the districts that it wins. Calculate the standard deviation of the party's vote share in the districts that it wins nationwide. Compare these two standard deviations using a well-established tool, the chi-square test for comparison of variances[19].

TEST 1: HOW MANY EXCESS SEATS?

The most obvious harm from partisan gerrymandering is representational. Partisan gerrymandering creates a situation in which the same overall statewide vote share would lead to a very different level of representation for the redistricting party and its opposing target. In the North Carolina congressional election of 2012, Democrats won only 4 out of 13 congressional House seats, despite winning slightly more than half of the statewide vote. Democratic winners were packed into districts where they won an average of 70.3 percent of the vote, while Republican winners won an average of 57.1 percent. In short, the districting scheme was associated with a strong representational asymmetry between the two major political parties.

However, an anti-majoritarian outcome (*i.e.* majority of popular vote, minority of delegation) is insufficient to prove deliberate distortion of electoral processes. Even if some imagined ideal of districting could maximize the likelihood of a majoritarian outcome, lack of congruence with this standard could still arise by chance and small variations in opinion. In 2012, if a few thousand voters in Arizona had cast their ballots for a Republican instead of a Democrat in the 1st or 2nd District, the delegation would have been, like the state's popular vote, majority Republican.[20] Conversely, the 2014 and 2016 elections in North Carolina showed a 46-47% vote for Democrats, and 3 out of 13 Democratic seats. How should this level of representation be regarded? In short, a simple majoritarian standard is not a surgical tool, but a blunt instrument.

A statistical approach is needed to distinguish what degree of inequity is allowable. This approach recalls Justice Kennedy's statement that "new technologies may produce new methods of analysis that make more evident the precise nature of the burdens gerrymanders impose on the representational rights of voters and parties. That would facilitate court efforts to identify and remedy the burdens, with judicial intervention limited by the derived standards."[21]

---

[19] Pearson, Karl (1900). *On The Criterion That A Given System Of Deviations From The Probable In The Case Of A Correlated System Of Variables Is Such That It Can Be Reasonably Supposed To Have Arisen From Random Sampling.* 5 50 (302) PHILOSOPHICAL MAGAZINE SERIES: 157–175. Snedecor, George W. and Cochran, William G. (1989), *Statistical Methods*, Eighth Edition, Iowa State University Press.

[20] State of Arizona, canvass of election results, http://apps.azsos.gov/election/2012/General/Canvass2012GE.pdf, 4-6.

[21] *Vieth*, 541 U.S. at 312-313 (J. Kennedy, concurring).

How many seats were gained by partisan gerrymandering? In my calculations for Test 1, I estimate the extent to which a party's elected number of seats exceeds an appropriate range that would arise naturally from national practices. This measure overcomes the central difficulty that representation is not necessarily proportional to public support. The idea that representation should be proportional is intuitive but wrong, and is violated in a system in which individual elections are winner-take-all[22].

A more sophisticated approach to quantifying the number of excess seats has relied on the detailed preparation of hypothetical maps[23] according to explicitly stated rules for how districts are drawn. However, such an approach may be criticized because it implicitly relies on the notion that automated standards for districting represent a benchmark of fairness. Such standards are susceptible to the critique that they may inadvertently contain hidden biases that actual legislative processes would not produce. My calculation of effects identifies a range of possibilities using national election results, which contain within them the standards used by real legislative processes, and by real legislators.

Computer simulations can be used to ask a simple question: if a given state's popular House vote were split into differently composed districts carved from the same statewide voting population, what would its Congressional delegation look like? The answer allows the definition of a range of seat outcomes that would arise naturally from districting standards that are extant at the time of the election in question.

It is possible to calculate each state's appropriate seat breakdown—in other words, how a Congressional delegation would be constituted if its districts were not contorted to protect a political party or an incumbent. This is done by randomly selecting combinations of districts from around the United States that add up to the same statewide vote total for each party. Like a fantasy baseball team, a delegation put together this way is not constrained by the limits of geography. On a computer, it is possible to create millions of such unbiased delegations in short order. In this way, one can ask[24] what would happen if a state had districts whose distribution of voting populations was typical of the pattern found in rest of the nation.

---

[22] Edward R. Tufte, *The Relationship Between Votes and Seats in Two-Party Systems*, 67 AMERICAN POLITICAL SCIENCE REVIEW 540 (1973).

[23] Jowei Chen and Jonathan Rodden, *Cutting Through the Thicket: Redistricting Simulations and the Detection of Partisan Gerrymanders*, 14 ELECTION LAW JOURNAL (forthcoming 2015); Jowei Chen and Jonathan Rodden, *Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures*, 8 QUARTERLY JOURNAL OF POLITICAL SCIENCE 239, 248 (2013).

[24] This can be done by using all 435 House race outcomes. For a state X with N districts, calculate the total popular vote across all N districts. Now pick N races from around the country at random and add up their vote totals. If their vote total matches X's actual popular vote within 0.5%, score it as a comparable simulation. See Sam Wang, *The Great Gerrymander of 2012*, N.Y. TIMES, Feb. 2, 2013, at SR1.

Because this approach uses existing districts, it uses as a baseline the asymmetries that are present nationwide in actual election results.[25] Redistricting is done by legislators who seek to satisfy many competing interests and motivations. The use of actual election results provides a desirable setting that naturally captures these interests and motivations. Indeed, the average result of these simulations approximates a "natural" seats/votes relationship that can be defined with mathematical rigor and exactitude. In short, these simulations detect distortions in representativeness in one state, relative to the rest of the nation.

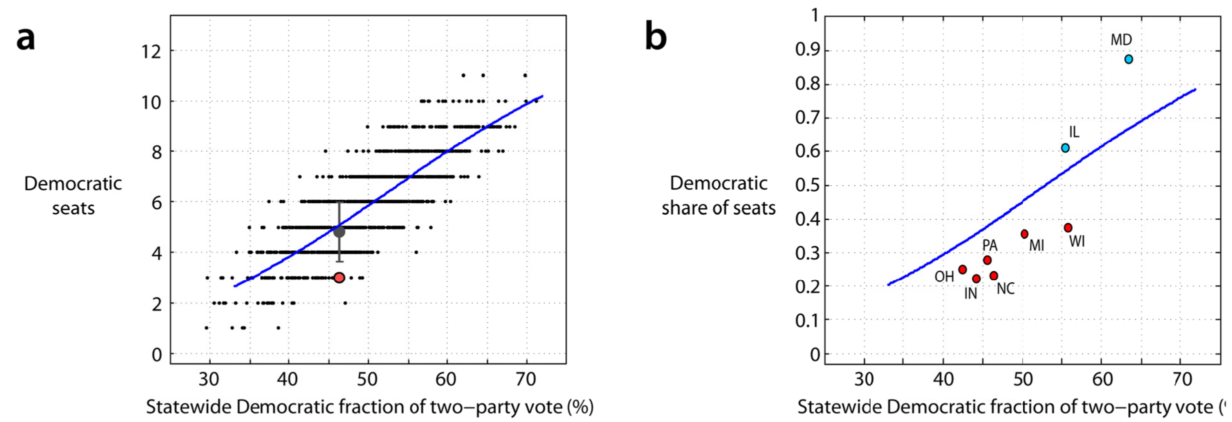

Figure 1: Simulated delegations for the 2016 Congressional election. (a) A graph showing 1000 simulated delegations, using national House results. The blue curve shows the average relationship of seats and votes for 1,000,000 delegations. The gray dot and bar show the average result and standard deviation for simulations whose vote totals were the same as the actual election. The red dot indicates the actual result. This simulation omitted partisan-asymmetric states (see text for details). (b) The same seats-votes curve as in (a), with the partisan-asymmetric states shown individually. Red dots indicate Republican-controlled redistricting. Blue does indicate Democrat-controlled redistricting.

A standard ThinkPad X1 Carbon laptop computer can perform one million simulations for a state in less than 20 seconds. **Figure 1** shows 1000 such "simulated delegations" for the state of North Carolina, along with the actual outcome in red. The blue curve defines a

[25] It is possible to explore the properties of this simulation procedure by giving it a variety of hypothetical nationwide distributions of districts as starting data. These hypothetical scenarios reveal that the "fantasy delegation" procedure has important features that are required of a descriptor of partisan asymmetry. First, for a symmetric distribution of Congressional districts, *i.e.* a scenario in which Democrat-dominated districts are no more packed than Republican-dominated districts, fantasy delegations are typically majoritarian, awarding more representatives to the party that receives more votes. Second, the fantasy delegations have the same natural variation in partisan composition as the nationwide distribution, as measured by standard deviation. Third, when the nationwide distribution of districts has asymmetry, for instance containing a number of districts that are very packed with one party (as is the case in real life for Democrats), the fantasy delegations show a bias toward the other party, a phenomenon that is well analyzed (reviewed in Jowei Chen and Jonathan Rodden, *Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures*, 8 QUARTERLY JOURNAL OF POLITICAL SCIENCE 239, 248 (2013).

mathematically expected seats/votes relationship, calculated by averaging 1 million simulations. In these simulations, I excluded states that are distorted as defined by one or more tests of partisan symmetry (Illinois, Indiana, Maryland, Michigan, North Carolina, Ohio, Pennsylvania, and Wisconsin). The average delegation had 5.2 Democratic seats. Only 4.2% of delegations had three or fewer Democratic seats. Therefore the probability that national partisan-symmetric districting standards would generate the actual outcome is 4.2%.

I tested how these results might change under a variety of scenarios. The results are shown in **Table 3**. For each row of the table, 1 million fantasy delegations were generated. In each case, between 9,000 and 12,000 fantasy delegations matched the real statewide vote totals. In all scenarios, simulated delegations had an average of at least 5 Democratic seats, two more seats than the actual election outcome. This average had a standard deviation of 1.0 seats, meaning that about two-thirds of the time, fantasy delegations had between 4 and 6 Democratic seats, or between one and three seats more than the actual outcome of 3 seats.

Omission of at-large states, which have only one representative and therefore do not undergo redistricting, did not affect the outcome. Omission of urbanized states, in which Democratic voters may be naturally packed, also had only a minimal effect on the outcome. Conversely, inclusion of all 49 states except for North Carolina also did not change the average simulated delegation substantially (5.0 Democratic seats on average). In this last case, 6.1% of simulations led to 3 or fewer Democratic seats; the larger fraction (compared with 4.2%) arises because when all states are used, the use of partisan-asymmetric states leads to more situations in which fantasy delegations are, like the real delegation, distorted.

Simulated fantasy delegations for North Carolina, 2016

States used to construct fantasy delegations	Average number of Democratic seats	Fraction of delegations with 3 or fewer Democratic seats
Using states that pass all three tests*	5.2	4.4%
Additionally omitting at-large states**	5.2	4.1%
Additionally omitting urbanized states***	5.3	3.3%
Additionally omitting both at-large and urbanized	5.4	2.6%
Using all states except North Carolina itself	5.0	6.1%
Additionally omitting at-large states**	5.0	5.6%
Additionally omitting urbanized states***	5.2	3.9%
Additionally omitting both at-large and urbanized	5.3	3.7%

*For these simulations, Illinois, Indiana, Maryland, Michigan, North Carolina, Ohio, Pennsylvania, and Wisconsin were not used.

**At-large states, with only one Representative: Alaska, Delaware, Montana, North Dakota, South Dakota, Vermont, and Wyoming.

***Urbanized states: having more than 1 million people in communities of 250,000 people or greater, are Arizona, California, Colorado, Florida, Illinois, Indiana, New York, North Carolina, Ohio, Pennsylvania, Tennessee, and Texas. See Wang (2016) at 1302.

Table 3: Simulated delegations for Test 1 under alternate assumptions. Each row of the table summarizes the result of 1 million simulations.

TEST 2: DID ONE PARTY WIN BY MORE LOPSIDED MARGINS?

The success of a gerrymandering scheme depends on the ability of the redistricting party to create safe margins of victory for both parties, with larger margins for their opponents. This pattern of outcomes can be quantified by sorting the districts into two groups, by winning party. Each party's winning vote shares can then be compared by what is said to be "the most widely used statistical test of all time"[26]: the t-test for comparing the averages of two groups of observations. In this way, the difference between each party's winning margins is used to test for intensive packing of the opposing party's voters.

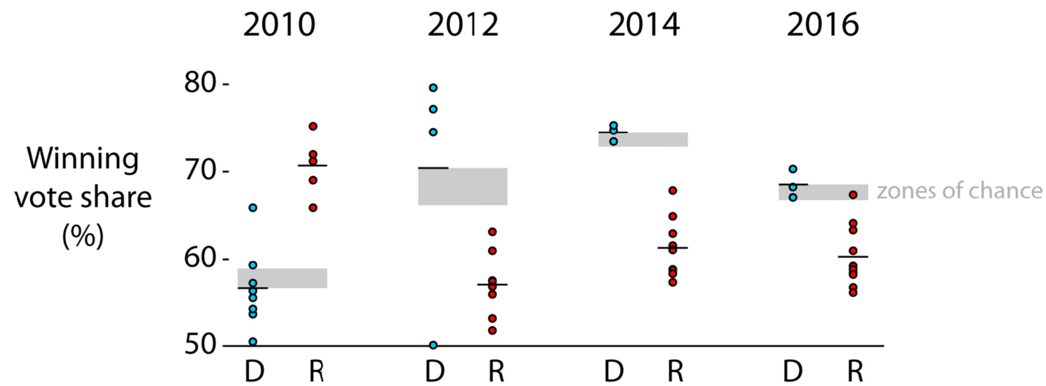

Figure 2: Comparison of winning margins in North Carolina Congressional elections, 2010-2016. Each point indicates the vote share of the winning Democratic (D) or Republican (R) candidate. The short horizontal lines indicate the average vote share. The shaded gray bands indicate zones of chance within which both averages would be expected to fall if the results had arisen from a partisan-symmetric process.

The results of Test 2 are shown in **Figure 2**. In 2016, Democrats won three districts with an average of 68.5% of the two-party vote, while Republicans won ten districts with an average of 60.3% of the two-party vote. According to the t-test, the probability of such a large difference is 0.31% (i.e. "p=0.0031")[27]. With near-certainty, this outcome did not arise from a party-neutral process.

Similar differences, all demonstrating narrower Republican wins, occurred in 2014 (74.5% in Democratic districts and 61.2% in Republican districts; p= 0.0032) and in 2012 (70.3% in Democratic districts and 57.1% in Republican districts; p= 0.015). However, in 2010

[26] Richard Lowry, Chapter 11. t-Test for the Significance of the Difference between the Means of Two Independent Samples, VASSARSTATS, http://vassarstats.net/textbook/ch11pt1.html (last visited Dec. 5, 2015).

[27] For comparing two groups of results, the t-test depends on both the average, the standard deviation of each group (i.e. the degree of spread in the data), and the number of samples. It can be calculated using the Excel function "ttest" or using an interactive tool available at gerrymander.princeton.edu. The significance values here are two-tailed.

the advantage was reversed (56.6% in eight Democratic districts and 70.6% in five Republican districts). In short, post-2010 districting schemes conferred an electoral advantage on Republicans that has lasted through three Congressional elections.

TEST 3: DID ONE PARTY HAVE MORE RELIABLE WINS?

In a partisan gerrymander, district outcomes are distributed to favor the redistricter's party across a large number of districts. If the average vote is closely divided between the parties, such an advantage can be probed using the mean-median difference. If one party is dominant, such an advantage can be probed using the chi-square test.

The mean-median difference. In a closely divided state or in a state where the redistricting party is wins a minority of the vote, reliable wins can be tested using the difference between the average (also called the mean) and the median vote share[28] for contested[29] districts. The median serves as a measure of the overall behavior of a state's district-level elections. The goal of a gerrymander is to maximize the number of districts won, which occurs when the median outcome is more unfavorable to the opposing party than that party's share of the vote. The mean-median difference is therefore a simple measure of asymmetry or skewness, and when it is allowed to develop without partisan acts, it has well-defined mathematical properties[30].

Consider the 2016 North Carolina Congressional election. The Democratic two-party share of the total vote in all 13 districts was, in terms of percentages and sorted in ascending order:

32.7, 35.9, 36.7, 39.1, 40.8, 41.2, <u>41.6</u>, 41.8, 43.3, 43.9, 67.0, 68.2, 70.3

When vote percentages are sorted in this way, the median can be calculated simply by striking out pairs of values at the ends of the list repeatedly, until the middle is reached. If two values remain, then take the midpoint of those two values.

---

[28] The mean-median difference has also been suggested by Robin E. Best & Michael D. McDonald, *Unfair Partisan Gerrymanders in Politics and Law: A Diagnostic Applied to Six Cases.* 14 ELECTION LAW JOURNAL: RULES, POLITICS, AND POLICY 312 (2015). In the present paper I give mathematically rigorous confidence intervals on that statistic and describe the circumstances under which it is applicable.

[29] The presence of uncontested races reduces the value of the mean-minus-median statistic. In those cases, the partisan breakdown is not known with accuracy. Consider the example of a 20-district state, residents of an uncontested district would have voted at a rate of 80% for their party, instead of the nominal 100%. If their district were drawn differently, the appropriate mean for comparison would be based on the 80% figure, and shift the overall mean by 1%.

[30] The mean-median difference is a simple and old measure of "skewness," a statistical term for asymmetry. David P. Doane & Lori E. Seward, *Measuring Skewness: A Forgotten Statistic?*, 19 JOURNAL OF STATISTICS EDUCATION (2011), www.amstat.org/publications/jse/v19n2/doane.pdf; Karl Pearson, *Contributions to the Mathematical Theory of Evolution, II: Skew Variation in Homogeneous Material*, TRANSACTIONS OF THE ROYAL PHILOSOPHICAL SOCIETY, SERIES A, 186, 343-414 (1895). G. UDNY YULE AND MAURICE G. KENDALL, AN INTRODUCTION TO THE THEORY OF STATISTICS 162-3 (3d ed. 1950).

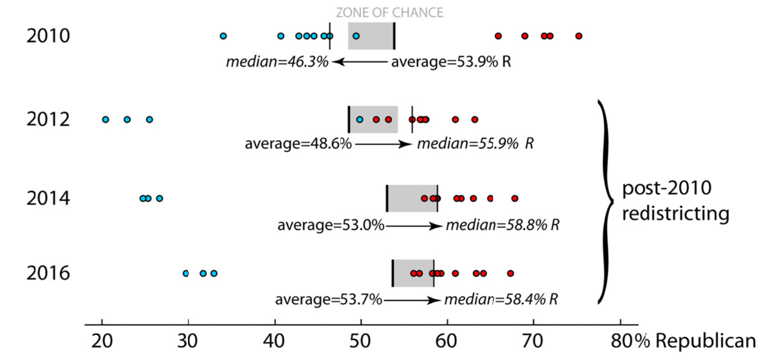

Figure 3: The mean-median difference in vote shares in North Carolina Congressional elections, 2010-2016. Each point indicates the Republican vote share. The shaded gray bands indicate zones of chance within which both the median and mean would be expected to fall if the results had arisen from a partisan-symmetric process.

Races won by Republicans are indicated in *italics* and the median percentage, 41.6%, is underlined. The mean (*i.e.* average) Democratic vote share is 46.4%. The difference between the median and the mean is 4.8%. This difference reflects the fact that counterintuitively, Republican vote shares were above average in considerably more than half of the districts: 77% (10 out of 13), to be exact.

In other words, North Carolina's Democratic voters were empowered as if they comprised 41.6% of voters, even though they actually comprised 46.4%. The difference, 4.8%, is the number of voters who were effectively disenfranchised. Since approximately 4,600,000 North Carolinians cast votes in the 2016 Congressional election, redistricting achieved an effect equivalent to over 220,000 Democratic voters casting their ballots for Republicans. Based on known properties of the mean-median difference, the probability that the mean-median differences for 2012, 2014, and 2016 arose by chance in a partisan-symmetric process are 2.9%, 6.8%, and 6.8%, respectively[31].

The chi-square test. A redistricting party gains an advantage when its wins are highly uniform across districts. This is especially important in a state where the redistricting party wins a large majority of the vote; an example is Maryland, which is currently dominated by Democrats[32]. Here, I apply the test in North Carolina.

One way to identify a more uniform advantage than expected by chance is the chi-square test[33]. In 2012, 2014, and 2016, the vote share of Republican Congressional winners had a standard deviation of 11.1%, 13.8%, and 13.3%, nationally; and 3.5%, 3.3%, and 3.5% in North

[31] One-tailed probabilities calculated under the assumption, based on Tests 1 and 2, that an anomalous outcome would favor the majority party.

[32] Wang, *Election Law Journal*, supra note 17.

[33] Also called the two-sample F-test for equal variances, or the chi-square test. Variance is defined as the square of the standard deviation. *Supra* at 19.

Carolina. The probabilities that these differences would have arisen by chance in 2012, 2014, and 2016 is 0.14%, 0.007%, and 0.02%, respectively. Thus Test 3 reveals that in North Carolina, Republican wins were unusually uniform across the state compared with national patterns.

IV. DISCUSSION

In this report I have presented three tests for rapid identification of partisan gerrymanders. Test 1 requires a computer program[34], and Tests 2 and 3 can be done with resources that are likely to be available on a judge's or clerk's desktop computer. All three tests rely on well-established statistical principles. The tests measure different aspects of partisan asymmetry, and therefore fall within the scope of principles that have been expressed by the Supreme Court. I suggest that these tests may constitute a manageable standard for courts to evaluate the impact of a state's districting scheme on its residents' Equal Protection and First Amendment rights.

The broader implications of this report are twofold. First, I have used statistical science to express the idea that a pattern of election results might have arisen by chance, and therefore not warrant judicial intervention. By establishing "zones of chance" in which the partisan impacts of a districting plan are ambiguous, the three tests presented here can help a judge evaluate whether an identifiable injury has occurred in the first place. Second, a standard based on the tests is unambiguous, and may even mitigate the need to demonstrate predominant partisan intent. For these reasons, these statistical tests comprise a valuable and timely addition to the judge's toolkit for rapid and rigorous identification of partisan gerrymanders.

My statistical analysis of the effects of gerrymandering may be of particular relevance to First Amendment analysis, which "allows a pragmatic or functional assessment that accords some latitude to the States."[35] By allowing for a normal amount of statistical variation, the three tests proposed in this Article build in zones of chance where any of a range of outcomes would lead to an acceptable amount of asymmetry.

The results of the Tests are highly correlated with one another. Therefore, in situations where one test is unsuitable, another can be used instead. In this way the tests can be used separately – or combined to reduce the risk of falsely identifying a gerrymander where none occurred. Conversely, the use of multiple tests also reduces the risk of failing to detect a gerrymander where one did occur. Finally, because the three tests do not use geography, they can easily be combined with other standards which may require circuitous geographic boundaries, such as state-mandated requirements[36], Section 2 of the Voting Rights Act, and other precedents that exist in federal law.

---

[34] A version of this software is available on GitHub at https://github.com/. It is also available for use at http://gerrymander.princeton.edu.

[35] *See e.g.*, *Eu v. San Francisco County Democratic Central Comm.*, 489 U. S. 214. (1989).

[36] The three tests proposed here address the overall apportionment plan, but do not cover the case of individual self-dealing in single districts. Local laws may provide additional constraints. For example, the current Congressional

ADVANTAGES AND DISADVANTAGES OF THE THREE TESTS

Before the judge chooses which test to apply, he or she should take the following advantages and disadvantages into account.

Test 1 quantifies the representational consequences of a gerrymander. Its most powerful use is to obtain an exact range for the appropriate number of seats for a given vote share. It addresses whether a redistricting scheme leads to an elected delegation that deviates from national districting norms. Test 1 can always be calculated for any set of election returns. Because it uses data from other states, it has the advantage of taking into account the overall nationwide demographic character of districts. Therefore it has the virtue of measuring effects that go beyond the natural effects of population clustering. However, because it requires computer simulation, it requires the use of a computer program, a version of which can be accessed at gerrymander.princeton.edu, or obtained separately by contacting the Author.

Of the three tests, Test 1 is the most like the efficiency gap devised by Stephanopoulos and McGhee. Both Test 1 and the efficiency gap are calculated using the statewide vote share and seat share. The formula for the efficiency gap is designed to detect departures from partisan symmetry, whereas Test 1 uses simulations to capture the natural relationship that arises from national districting practices. A possible weakness of both approaches is that they use the number of seats as an explicit input, and it could be argued that such a standard establishes an implicit, normative level of representation.

Test 2 has the advantage of simplicity: it can be worked out using a spreadsheet program such as Microsoft Excel (using the function "ttest2") that can perform a two-sample t-test. If such a program is not available, it can be done using a hand calculator and a table of statistical values. It directly tests for noncompetitive races, a mainstay of gerrymandering. It identifies partisan asymmetry. Test 2 has the disadvantage that it can only be used if both parties win at least 2 seats each, since this is required to calculate standard deviations, a necessary step of the test.

Test 3 measures the reliability of wins for the redistricting party. Test 3 can always be performed, since it is calculated using most or all of a state's district-level results. In the case of the mean-median difference, Test 3 does not rely on any data from other states, and is therefore self-contained. The mean-median difference measures a quality that statisticians call "skew." Other, tests of skew have more statistical power to detect anomalies[37], but the mean-median difference is easier to calculate.

In the case of the chi-square test, national data must be used to provide a standard for comparison. The chi-square test measures whether a set of electoral victories is more uniform than expected from national patterns, and is especially useful when one party is dominant (as is

---

districts in Florida do not violate the three tests presented here. Nonetheless, the Florida Supreme Court has found the map to violate the Florida Constitution redistricting provisions (article III, section 20(a) that reads "No apportionment plan or district shall be drawn with the intent to favor or disfavor a political party or an incumbent"). *Detzner*, 2015 WL 4130852. This stricter standard extends a mandate for competitive races to the level of single districts.

[37] The standard measure of skewness can be calculated using the Excel function SKEW.

the case of Maryland, in Shapiro v. McManus). In addition to detecting such a partisan gerrymander, the chi-square test may also be useful in detecting "bipartisan gerrymanders," in which unusually secure wins are constructed for both parties in a symmetric manner.

V. CONCLUSION

The 2016 Congressional election in North Carolina was marked by the following features of partisan asymmetry: (1) one to three more Republican seats were won than would be expected by national districting standards, (2) Democratic wins were more lopsided than Republican wins, (3) the median district in North Carolina was more Republican-voting than the statewide average, and (4) Republican wins were unusually uniform compared with national wins. These advantages arose in 2012 and persisted in 2014, indicating that they arose as a result of partisan control over redistricting. Any one of these tests is available to judges as a potentially manageable standard for partisan gerrymandering.