Figure 1. *The original Gerry-mander (editorial cartoon from Salem Gazette, 1813). (Courtesy of Cornell University Library, PJ Mode Collection of Persuasive Cartography. Creative Commons License.)*

Gerrymandering: Mathematics on Trial

On October 3, 2017, mathematics went on trial in the highest court of the United States. No, it was not accused of a crime—in fact, quite the opposite. The nine justices of the Supreme Court were trying to decide whether partisan gerrymandering, the practice of drawing election districts to favor one political party, is unconstitutional. And if so, can mathematics be used to determine when a map is too unfair?

The Supreme Court ultimately decided *Gill v. Whitford* on narrow procedural grounds, while leaving the central questions unresolved. Despite the anticlimactic outcome, the case galvanized mathematicians to take a much closer look at the problem of redistricting than they ever had before. The Geometry of Redistricting Workshop, organized in 2017 by Moon Duchin of Tufts University in Somerville, Massachusetts, drew crowds of up to 500 people. The interest was so great that four satellite workshops were organized in other states. And Duchin's invited address at the conclusion of the Joint Mathematics Meetings in January 2018 received a standing ovation—an almost unheard-of occurrence at a math meeting.

Moon is My Hero. *Photo from Moon Duchin's lecture on gerrymandering at JMM 2018.*

Gerrymandering first came to the public's attention in 1812, when the governor of Massachusetts, Elbridge Gerry, signed a bill creating new election districts that were designed to favor his party, the Democratic-Republicans. A Salem newspaper ridiculed one particularly odd-shaped district as a "Gerry-mander," combining the words "Gerry" and "salamander." (Figure 1) The term stuck, and gerrymandering has become, unfortunately, as American as apple pie.

Moon Duchin.

Gerrymandering is possible because of a combination of circumstances found almost exclusively in the U.S. The United States has a census every ten years that is used to apportion seats in Congress among the states. The states that have gained or lost seats have to re-draw their election districts (and many other states do so voluntarily). In most of the states, the legislature gets to re-draw the boundaries, which means in practice that the party in power does its best to ensure that it stays in power. “The rules of the game are part of the game,” says Gary King, a redistricting expert at the University of California at Irvine. “You get to move the goalposts.”

Here’s how it works: If Party A is in control of drawing the map, it can “pack” most of Party B’s voters into a small number of districts where they will have a supermajority. At the same time, it can “crack” the remaining voters among a large number of districts where Party A will have a small but hopefully safe majority. Both “packing” and “cracking” take advantage of the winner-take-all voting system, which makes a 51–49 margin in one district just as good as a margin of 80–20 in another.

A perfect example of “packing and cracking” came in the North Carolina congressional election in 2012. Although the Democratic party won more than 50 percent of the popular vote statewide, it won only 4 out of 13 congressional districts.

Could this result be just a run of bad luck? No, says Jonathan Mattingly, a mathematician at Duke University. He points to the huge gap in vote share between the ten “cracked” districts, in which the Democratic vote share was less than 50.1 percent, and the three “packed” districts, where the Democratic share was more than 74.4 percent. (Figure 2) If you chose a map at random, such a big gap would be incredibly unlikely: normally, the vote shares would increase in a gradual, more or less linear fashion. The gap is a tell-tale sign that the districts were carefully engineered. “The system is hard-wired to elect nine or ten Republicans,” says Mattingly. In statistical language, the map is an *outlier*. That is, if you take a random sample of legal maps, you will find that this one is at the far end of a bell-shaped curve on any test of fairness.

In the decade of the 2010s, three trends have come together to make gerrymandering a hot topic—and to place mathematics squarely at the center of the debate. First, computerized mapping tools have made it possible for legislators to draw gerrymandered districts with much more sophistication than before. Second, mathematicians have developed new methods to catch them at it. And third, the courts are taking notice.

In October 2017, the Supreme Court heard arguments in a redistricting case called *Gill v. Whitford*, an appeal of a Wisconsin district court decision that determined the electoral map in that state, drawn before the 2012 elections, to be unconstitutional. At the same time, appeals courts in Pennsylvania and North Carolina were hearing legal challenges to their states’ electoral maps. In North Carolina, where Mattingly testified as an expert witness, Judge James Wynn struck down the state’s electoral map in January 2018 as an unconstitutional gerrymander.

At the Supreme Court hearing, it seemed sometimes as if mathematics itself was on trial. Chief Justice John Roberts ridiculed the idea of using a formula to decide whether a map was constitutional, calling it “sociological gobbledygook” and

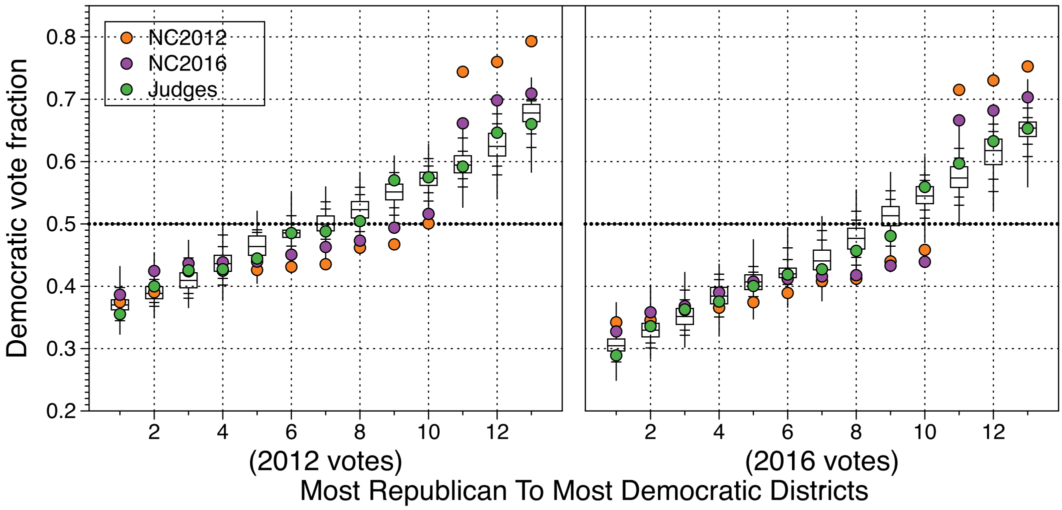

Figure 2. *Packing and cracking in the North Carolina elections, 2012 and 2016. The tell-tale jump in vote fraction between districts 10 and 11 is extremely unlikely to happen by chance. Note that there is no comparable jump when the electoral map is redrawn by a non-partisan commission of retired judges (green dots). (Figure courtesy of Greg Herschlag and Jonathan Mattingly.)*

arguing that the man on the street would not understand a court decision based on mathematics. On the other hand, Justice Stephen Breyer said that the formula is "not quite so complicated as the opposition makes [us] think." And Justice Elena Kagan several times mentioned the concept of outliers in her questioning.

"If the Supreme Court disappoints us by kicking the can down the road, as they did 30 years ago, I will personally not be devastated," said Duchin in a prescient comment before the Supreme Court decision was handed down. "There is a building consensus, and even if it doesn't come down from the Supreme Court, there are plenty of ways to try out these new standards case by case. I don't see the Supreme Court as the last word. We'll see a standard emerging from the states."

Searching for a Talisman

Of course, there have been efforts over the years to rein in the worst abuses of gerrymandering. For example, the Voting Rights Act prohibits racial gerrymanders, and the Supreme Court has not hesitated to enforce it. Most recently, in the spring of 2017, it ruled two of the "packed" districts in North Carolina to be unconstitutional racial gerrymanders. One of these, the 12^{th} district (see Figure 3, p. 7) was ruled unconstitutional by a unanimous 9–0 decision; the other was a narrower 5–4 decision. Both of these districts had to be re-drawn before the 2018 elections.

Electoral maps need to satisfy several other legal requirements, which vary from state to state. First, all districts must have roughly equal populations. Most states require districts to be contiguous, and they should respect existing legal boundaries of cities and counties as much as possible. Some states also require the map-drawers to keep "communities of interest" intact. This concept is as elastic as it sounds, encompassing various ethnic and cultural groups. For example, the Castro district in San Francisco, traditionally an area with a high concentration of gay and lesbian people, was designated a community of interest in California's 2010 redistricting process. The fact that mapmakers have to balance these competing and sometimes ill-defined interests means that there can be no one-size-fits-all mathematical solution to the redistricting problem; it is inherently a messy political process.

However, one legal requirement seems to beg for a mathematical interpretation. Thirty-seven out of fifty states require their election districts to be "compact," a rule that stems from the fact that many of the worst gerrymanders in the past have produced districts with absurdly convoluted shapes. The 12^{th} district in North Carolina is a good example; others are the 3^{rd} district in Maryland, which looks like a praying mantis, and the 7^{th} district in Pennsylvania, which looks like Goofy kicking Donald Duck (Figure 3).

However, one legal requirement seems to beg for a mathematical interpretation. Thirty-seven out of fifty states require their election districts to be "compact," a rule that stems from the fact that many of the worst gerrymanders in the past have produced districts with absurdly convoluted shapes.

The trouble with compactness is that nobody has been able to agree on what it means. Presumably the ideal district should not have too many narrow necks or long, meandering appendages. But it's hard to put into words what is meant by a "neck" or an "appendage" or even "too many." Perhaps math can come to the rescue?

One approach is simply to require the total perimeter of the districts to be as small as possible. If followed literally, this requirement would lead to a honeycomb-like arrangement of hexagonal districts. However, such a definition fails to respect natural and political boundaries. Also, because the perimeter depends in part on the size of the state, it is impossible to give a uniform standard that would be the same for all states.

From the mathematical point of view, the way to achieve comparability is to make the metric scale-invariant. A classical measure with this property is the isoperimetric ratio (known in political science as the Polsby-Popper index), defined as 4π times the area of the district divided by the perimeter squared. This number always lies between 0 and 1, with a score close to 1 indicating a plump, round district.

However, this measure too has some problems. Because it involves the perimeter, it is very sensitive to small-scale wiggles in the district boundaries. Such wiggles may be unavoidable when the boundary is a natural feature like a river or a mountain range. The Polsby-Popper index penalizes such "frilly" boundaries, even though they may make a lot of sense politically and practically.

Because perimeters are overly sensitive to small-scale features, some other measures of compactness avoid perimeters. For example, the "Reock score" is the ratio of a district's area to that of the smallest circumscribing circle, and the "convex-hull score" is the ratio of the area to the smallest convex region that contains the district. (Think of stretching a rubber band around it.) Other metrics take into account where people live, not just pure geometry. For example, a "dispersion index" would calculate the average distance between all pairs of people in the district, normalized by dividing by the district's diameter or the square root of its area. Such a metric would have been impossible to calculate before the computer era, but now it is easy.

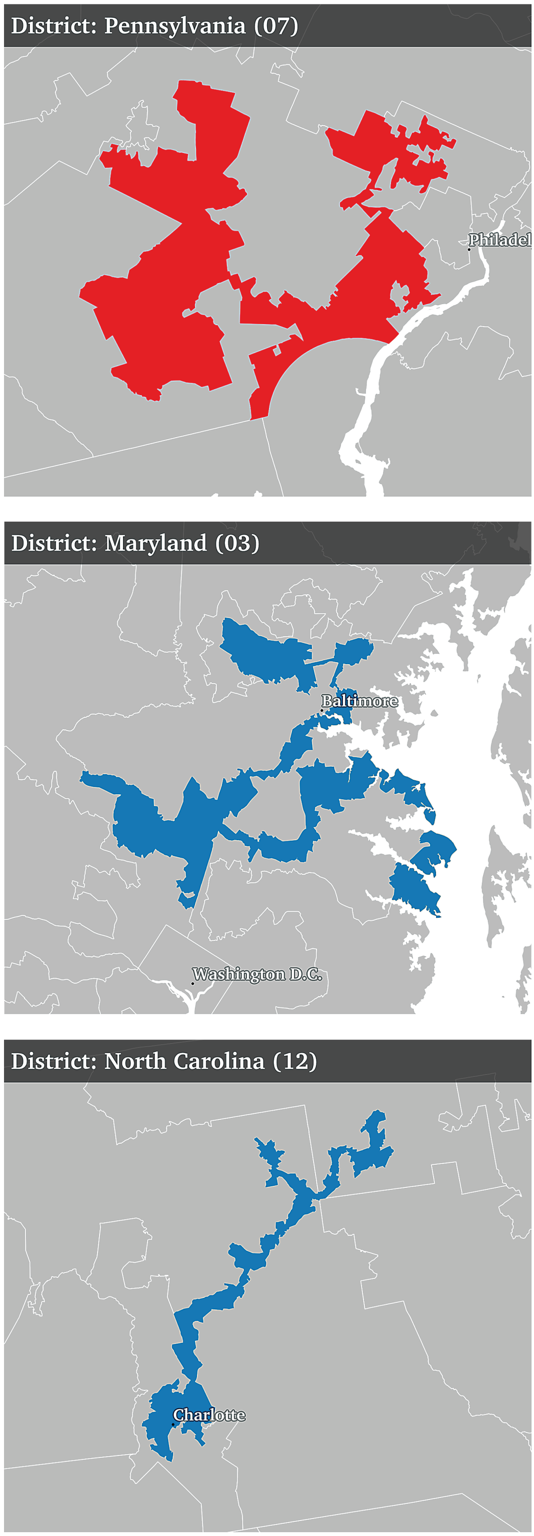

Figure 3. *Oddly shaped election districts (including NC-12, MD-3, amd PA-7). (Figure courtesy of Alasdair Rae.)*

All of these measures of compactness make a certain amount of sense, but the sheer number of ideas, and the lack of consensus on which one is best, has made the compactness requirement almost impossible to enforce. And in a broader sense, all of these metrics miss the point. “How does compactness promote fairness?” asks Duchin. Why should we believe a circular district to be inherently fair, while a horseshoe is intrinsically unfair?

In a 2004 case, *Vieth v. Jubilerer*, the Supreme Court wrote repeatedly that “no discernible and manageable standards for adjudicating such claims exist.” In the *Vieth* opinion, the justices drew a distinction between racial gerrymandering cases, where *individual districts* discriminate against certain voters’ rights, and partisan gerrymandering cases, in which the discussion concerns an entire *statewide map*. They concluded that for cases of the latter type, there are “no agreed upon substantive principles of fair districting.” Four of the justices argued that no such principles would ever be found, and therefore partisan gerrymandering should not even be under the Court’s oversight. The fifth justice, Anthony Kennedy, held out hope that “a limited and precise rationale” may yet be found.

But as computer redistricting programs become more sophisticated, it seems less and less likely that compactness tests offer a solution. Figure 4 shows the latest North Carolina map, adopted for the 2018 elections after the 2016 map was found to be an unconstitutional racial gerrymander. Notice that there are no longer any districts that look like egregious gerrymanders or that would fail a compactness test. Nevertheless, there are definitely signs of packing and cracking. The two largest cities, Charlotte and Raleigh, are mostly packed into a single district, while a number of smaller cities such as Asheville and Greensboro are cracked between two.

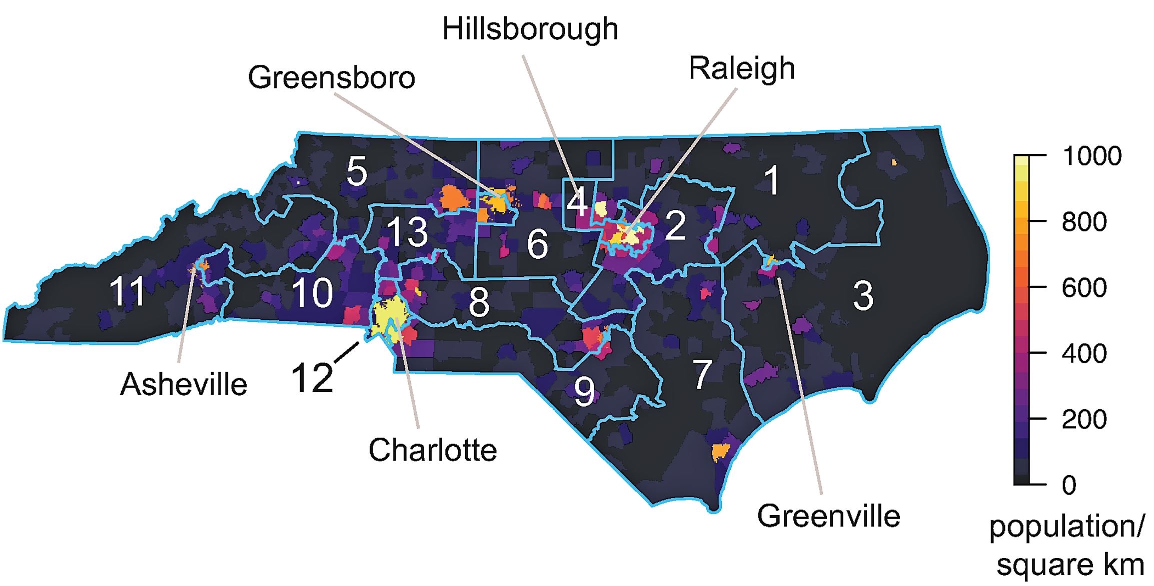

Figure 4. *Redrawn North Carolina districts for 2018 with population “heat map”. (Figure courtesy of Sam Wang, the Princeton Gerrymandering Project.)*

In a subsequent case, *LULAC v. Perry* (2006), Kennedy hinted at a kind of test that might satisfy him, called partisan asymmetry—a term that he borrowed from an *amicus curiae* brief by Gary King of UC Irvine and Bernard Grofman of Harvard University. In other words, lawyers should produce evidence that the electoral map treats the two parties differently. A measure of such asymmetry, Kennedy wrote, would be a "helpful (though not talismanic) tool" for detecting gerrymanders.

However, Kennedy explicitly rejected a test of asymmetry that King and Grofman had suggested, called the *partisan bias*. That test compares the number of seats that Party A won to the number of seats that Party B *would have won* if it had gotten the same number of votes. Any discrepancy would be evidence that the map was treating the two parties inequitably. A simpler version of this test is to compare how many seats Party A and Party B *would have won* if they had divided the vote 50–50. Grofman and King contend that pollsters can make this kind of estimate quite reliably: if the election actually ended 55–45, you can simply add 5 percent to Party B's vote in each district and subtract 5 percent to Party A's vote in each district. This "uniform swing hypothesis" tends to hold in real elections; a rising tide lifts all the ships by the same amount. King says, "We *can* make predictions. We know what would happen."

Nevertheless, the Court did not buy it. They ruled that partisan bias was not acceptable evidence because it rested on a "counterfactual" hypothesis—the "would haves" in the last paragraph. Thus, prior to *Gill v. Whitford*, the Supreme Court's position was that racial gerrymandering was definitely illegal, and that excessive partisan gerrymandering might be illegal, but they had never seen good enough ("talismanic") methods to identify when it is present. Two new ingredients emerged in *Gill v. Whitford*: one of them a new metric called the "efficiency gap," and the other a new method called outlier analysis.

...the Supreme Court's position was that racial gerrymandering was definitely illegal, and that excessive partisan gerrymandering might be illegal, but they had never seen good enough ("talismanic") methods to identify when it is present.

The Efficiency Gap

In 2014, Nicholas Stephanopoulos (a lawyer) and Eric McGhee (a political scientist) proposed a new way of measuring partisan asymmetry, which they called the efficiency gap. Stephanopoulos and McGhee argued that a redistricting plan is asymmetric if it causes Party B to waste more votes than Party A, or equivalently if it allows Party A to turn its votes into seats more efficiently. There are two ways a vote can be wasted: it can be cast for a losing candidate, or it can be a surplus vote for a winning candidate. Cracking creates wasted votes of the first type, and packing creates wasted votes of the second type. They proposed simply to add up the percentage of wasted votes cast for each party; the difference would be the efficiency gap. It's a very direct and simple measure of the total amount of packing and cracking. It also does not rest upon any counterfactual assumptions because the measure is based on votes actually cast.

McGhee and Stephanopoulos calculated efficiency gaps for elections dating back to 1972, and they found two interesting trends. First, there has been a long-term trend toward more Republican-skewed efficiency gaps. From 1972 to 1992 the efficiency gap favored Democrats, giving them an average of 0

Nicholas Stephanopoulos and Eric McGhee. *(Photo courtesy of Ruth Greenwood.)*

to 0.5 extra seats in Congress per state delegation.[1] But from 1992 to 2012 the efficiency gap swung over to favor the Republicans, and by 2012 the Republicans averaged 1.2 extra seats per delegation (26 extra seats overall). They also noted a spike in 2012, suggesting that there was something unusual about the redistricting that took place after the 2010 census.

Next, McGhee and Stephanopoulos took the bold leap of translating the historical data into a normative criterion. They noted that in about 15 percent of the historical maps, the efficiency gap was large enough to translate into two extra congressional seats for one party. This, they suggested, could become a legal test for partisan asymmetry: A state redistricting plan that created a two-seat advantage for one party would be *presumptively* unconstitutional. The burden of proof would then be on the state to prove that the asymmetry had legitimate reasons. McGhee and Stephanopoulos also proposed a percentage-based criterion for state legislatures: an efficiency gap would be presumptively illegal if it was 8 percent or greater. This 8 percent criterion became a bone of contention in the oral arguments at the Supreme Court.

The new test went to court with unexpected rapidity. Their article that introduced the efficiency gap came out in the winter of 2014. By that summer, Stephanopoulos had been invited to Milwaukee for talks with the lawyers who were putting together

[1]For technical reasons, they considered only the 22 states with 8 or more congressional seats.

the lawsuit that became *Gill v. Whitford*. The suit was filed in the summer of 2015. "In December of 2015 we became the first partisan gerrymandering lawsuit to survive a motion to dismiss, and in November 2016 we won at the trial," he says. That triumph broke a fifty-case losing streak for plaintiffs in partisan gerrymandering cases, and set the stage for the Supreme Court hearing in October 2017.

Because the efficiency gap was the most novel element of the legal case, it came under particular scrutiny from opposing lawyers and from the Supreme Court justices themselves. At one point, Chief Justice Roberts argued as follows:

> And if you're the intelligent man on the street and the Court issues a decision, and let's say the Democrats win, and that person will say: Well, why did the Democrats win? And the answer is going to be because EG was greater than 7 percent, where EG is the sigma of party X wasted votes minus the sigma of party Y wasted votes over the sigma of party X votes plus party Y votes. And the intelligent man on the street is going to say that's a bunch of baloney. It must be because the Supreme Court preferred the Democrats over the Republicans.

Remarkably, Roberts gave a correct formula for the efficiency gap and yet, in the same breath, said that the *intelligent* man on the street (not even the average man on the street) is incapable of understanding it, and therefore it should not be the law of the land.

To the contrary, the efficiency gap (as Justice Breyer pointed out) is not complicated at all, and in fact it may be too simple for its own good. In particular, if all of the districts have equal populations—which is mandated in the U.S. by law—it turns out that the efficiency gap can be computed by a formula that doesn't involve counting "wasted votes" at all:

$$\text{Efficiency gap} = \text{Vote lean} - \frac{1}{2} \times (\text{Seat lean}).$$

Here, "vote lean" means the margin for Party A in the popular vote, and "seat lean" means the margin (as a percentage) for Party A in legislative seats. Thus, for instance, in a landslide district where Party A wins 75 percent of the vote, the efficiency gap is automatically 0, because the "vote lean" is 50 percent (75–25) and the "seat lean" is 100 percent (because there is only one seat, and it goes to the winner). Thus, any redistricting plan that creates nothing but landslide districts will have a perfect efficiency gap of 0. But is this what we really want in an electoral map? Probably not, because such a map would be very unresponsive to changes in voter opinion. As Wendy Tam Cho, a political scientist and statistician at the University of Illinois, writes, "Unfairness to the voters—the original sin of partisan gerrymandering—cannot be overcome by fairness to the parties."

Also, the formula above establishes as a matter of policy the notion of a "winner's bonus." To minimize the efficiency gap, each percentage point of vote lean should be translated into two percentage points of seat lean. This strikes many people as surprising at first; the naïve notion of fairness is that the number of representatives should be proportional to the popular vote, in

To the contrary, the efficiency gap (as Justice Breyer pointed out) is not complicated at all, and in fact it may be too simple for its own good.

Jonathan Mattingly. *(Photo courtesy of Greg Herschlag and Jonathan Mattingly.)*

which case the vote lean and seat lean would be identical. This is called proportional representation—and it has been explicitly rejected in the past by the Supreme Court as a criterion for fairness. Stephanopoulos and McGhee consider it an "elegant coincidence" that the 2:1 ratio of seat lean to vote lean approximately conforms to historical precedent. Perhaps they are right that some sort of winner's bonus is desirable, because it tends to enhance the responsiveness of the voting system to changes in voter opinion. But it is still far from clear that 2:1 is the "right" ratio or that the Supreme Court should endorse any particular number at all. (Incidentally, Roberts demonstrated a tin ear for mathematics on this point, too; he argued that the efficiency gap was the same thing as proportional representation, which would be true only if $2 = 1$.)

Finally, the efficiency gap is not very stable. It can change significantly from one election to the next—as indeed it has in Wisconsin. McGhee and Stephanopoulos addressed this point in their original paper, and they posited that the efficiency gap should only be used in combination with sensitivity testing—proof that reasonable changes (e.g., 5 percent) in the parties' percentage of the vote would not make the gap go away.

In the end, the Wisconsin court did not decide *Whitford v. Gill* on the efficiency gap alone. Instead, they used a three-pronged test: *intent* to discriminate, discriminatory *effect*, and *lack of justification* for the discriminatory effect. The efficiency gap applies only to determining the discriminatory effect. For this purpose, some mathematicians give it a lukewarm endorsement. "I'm not a fan of the efficiency gap, but when the gerrymandering is really gratuitous, everything works," says Mattingly.

Methods, not Metrics

While the chief justice and the lawyers for the Republicans focused their attention on the efficiency gap, Duchin and Mattingly say that they were missing a more important point. "I want to distinguish methods from metrics," Duchin says. The efficiency gap, like all the compactness indices, is a *metric*, a numerical measure of gerrymandering. But the really new ingredient in gerrymandering research is the *method* of outlier analysis.

"For at least 50 years, people have considered using computers, but it has been too complicated to sample from the space of possible [maps]," says Duchin. "Just in the last five years, some teams are able to do this, and get really robust sampling. That's the breakthrough we've been waiting for. You can compare your map to the histogram of all possible outcomes, and if you see that you have an extreme outlier you throw it out."

There are two reasons why the sampling problem has been so difficult. First, the legal constraints differ in each state, and they are not exactly clear-cut. Also, the sheer number of possible maps—even after taking into account all of these constraints—is so astronomical that no computer could possibly hold them all, or even any significant proportion. How, then, can you be sure you have obtained a representative sample from the space of all possible electoral maps?

Surprisingly, methods have been developed to address exactly this kind of problem. The Markov chain Monte Carlo (MCMC) method (See *What's Happening in the Mathematical*

Sciences, Vol. 8) is designed for the purpose of taking random samples from vast spaces with complicated constraints. It was developed in the context of nuclear weapons research, to explore the space of possible chain reactions inside a hydrogen bomb. The fact that the U.S. has depended on MCMC calculations for its national security for 60 years makes an effective rebuttal to anyone who wants to claim that they are "sociological gobbledygook" invented by political scientists.

Not only that, there are alternatives. Wendy Tam Cho uses a different approach called evolutionary algorithms, which is geared toward finding maps that exceed some fitness threshold (e.g., high compactness scores, or low efficiency gaps). Both Mattingly and Tam Cho can produce an ensemble of hundreds of thousands of maps that satisfy all the legal constraints.

Defenders of gerrymandered maps often say, "We couldn't help it," or "It's just geography." Indeed, it's true that voters for the Democratic party tend to concentrate into urban areas, and thus they pack themselves (without any help from the map makers) into a small number of geographic locations. However, as Figure 5 shows, this argument doesn't really hold water. In this map, the purple voters are concentrated into a high-density area in the center. On the left, a map that packs them into one district gives a 5–1 electoral advantage to the green party. On the right, a map that consists of six pie-shaped districts gives a 4–2 advantage to the purple party. How can we decide which of these maps is fairer? The answer, says Mattingly, is to ask how representative they are of all possible maps.

This relates to the question of *intent* in the three-pronged test for gerrymandering. If a map is so drastically unfair that it

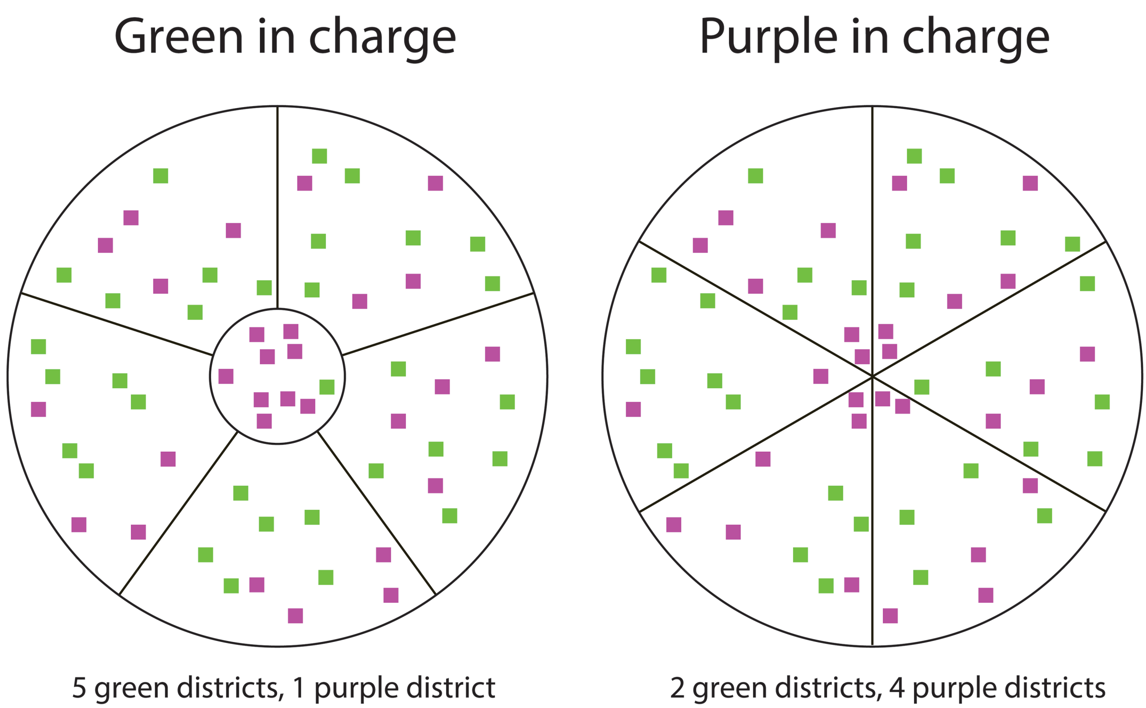

Figure 5. *Schematic illustration of packing and cracking in urban areas.. (Figure courtesy of Sam Wang, the Princeton Gerrymandering Project.)*

If you like the efficiency gap, you can use it; if not, you can use Mattingly's Gerrymandering Index or Tam Cho's biasedness index. The metric doesn't really matter, because the maps currently produced by politicians are extreme outliers for almost any metric.

would take a deliberate effort even to find such a bad map in the space of possible maps, one can reasonably conclude that it was chosen with intent.

That is exactly what both Mattingly and Tam Cho have found. Mattingly evaluated the 2012 and 2016 electoral maps in North Carolina on two metrics, a Gerrymandering Index (explained below) and a Representativeness Index. Out of 24,000 randomly generated electoral maps that satisfy all the legal constraints, not a single one had a worse Gerrymandering Index or a worse Representativeness Index than the 2012 electoral map. The 2016 map was barely any better. Clearly the Republican party (which controlled the redistricting process) had worked hard to identify these uniquely unfair maps.

In Maryland, the Democratic party was in charge, but the story was the same. Tam Cho compared the Maryland electoral map to a baseline of more than 250,000 "reasonably imperfect" maps, on a variety of indicators: responsiveness to changes in voter opinion, biasedness (using a test somewhat more complicated than the efficiency gap), as well as simply the number of Democratic legislators elected. Out of the ensemble of "reasonably imperfect maps," the actual map approved by the Maryland legislature was in the worst 5 percent for responsiveness; the worst 0.2 percent for bias; and the most extreme 0.003 percent in Democratic seats won. Again, only a deliberate effort could have produced such an extreme outlier.

As these examples show, the ensemble method is agnostic about the party in charge, and agnostic about the metric used to test for gerrymandering. If you like the efficiency gap, you can use it; if not, you can use Mattingly's Gerrymandering Index or Tam Cho's biasedness index. The metric doesn't really matter, because the maps currently produced by politicians are extreme outliers for almost any metric.

Even so, the Gerrymandering Index is intriguing because it includes the ensemble as part of its definition. For example, suppose that a state has three districts in which the vote share for Democrats was observed to be 0.4, 0.4, and 0.7. To test the hypothesis of gerrymandering, we compare the existing map to a representative ensemble of thousands of legal maps. Suppose we find that the average vote shares in that ensemble are 0.45, 0.5, and 0.55. Then the Gerrymandering Index would simply be the root mean square of the departures from average:

$$\sqrt{(0.4-0.45)^2+(0.4-0.5)^2+(0.7-0.55)^2} \approx 0.19.$$

But is this a large index or small? To answer this question, we go back to the ensemble and ask what percentage of the sample had an index this large or larger. If we find that less than 5 percent had a Gerrymandering Index of 0.19 or above, then we might conclude that it is abnormally large. (Here, 5 percent is an arbitrary cutoff, to be sure, but it is one with an easily understood meaning and a long tradition in scientific research.)

As Mattingly explains it, the Gerrymandering Index captures what is visually apparent in Figure 2: the orange and purple dots (observed vote shares) are not at all close to the centers of the box-and-whisker diagrams (average vote shares in the ensemble). "The index is an encapsulation of the picture," Mattingly says. "I thought that lawyers would like the numbers better, but actually they like the picture." Another picture they liked

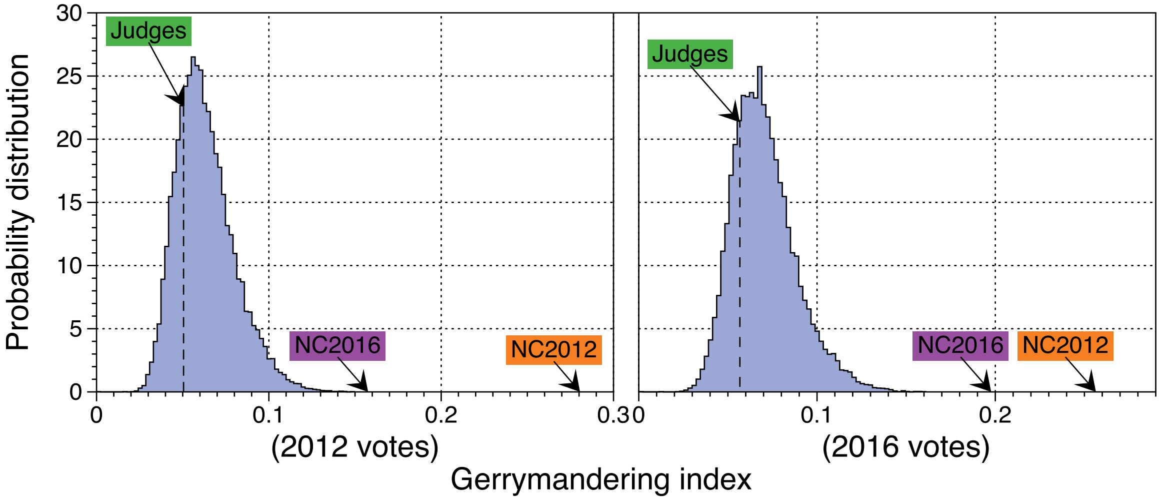

Figure 6. *Comparing judges' maps to actual electoral maps on a histogram of Gerrymandering Index. (Figure courtesy of Greg Herschlag and Jonathan Mattingly.)*

was Figure 6: a histogram of the Gerrymandering Index for the 24,000 maps in Mattingly's random sample. Both the 2012 and 2016 actual maps lie extremely far out in the right-hand end of the histogram.

The green dots in Figures 2 and 6 are also worthy of some explanation. In 2016, former University of North Carolina president Thomas Ross gathered a commission of ten former judges to see if they could draw a more impartial electoral map than either the 2012 or 2016 maps in North Carolina. The green dots represent the vote shares in the judges' districts, arranged from the most to the least Republican. These vote shares follow a gradual, nearly linear progression (see Figure 2), instead of making a giant leap from 50 percent to 74 percent. Not surprisingly, the judges' map scored in the top 25 percent on the Gerrymandering Index (Figure 6). According to Mattingly's standards, it would be an excellent choice for accurately representing the will of the people.

The ensemble method can also reveal other non-obvious characteristics of a given electoral map. In Figure 7 (page 17), Mattingly uses Wisconsin voting data to show how many Republicans would be elected to the state legislature, given the actual Democrat-Republican vote splits from several different elections in this decade (including presidential and gubernatorial elections). In certain elections, when the Republicans win a majority of the popular vote, the map does not produce unusual results; the number of Republican seats lies in the center of the histogram. But when the Democrats have the majority, a hidden "firewall" pattern emerges. In those years, the number of Republican seats lies on the far right-hand side of the histogram. "It's not that the map is always bad—it's designed so that it's only an outlier when the Republicans need a kick," Mattingly says.

The Proper Use of Mathematics

Mattingly and Duchin get asked sometimes whether it wouldn't be better simply to automate the procedure of drawing elec-

A switch from partisan to non-partisan or bipartisan commissions would address the causes, rather than the symptoms, of the gerrymandering disease.

toral maps. They give very similar answers. "Let us never do that," Duchin says. "The process is all about people, the relations between them, their culture and their history. The role of computers is to place guard rails on the process." In very similar fashion, Mattingly says, "I do not want to insert myself into the map-drawing process. We live in a country of checks and balances. I'm only saying that we should clip the tails of the probability distribution."

Rather than proposing maps, Duchin thinks that there are other ways for mathematicians to have an impact. "Much work needs to be done to benchmark these new methods," she says. For example, what is a typical efficiency gap? Is a 2:1 winner's bonus the right ratio in practice, or is the story more complicated than that? Also, Duchin says, "We need to study the geometry and topology of the universe of possible maps. We need to design good sampling algorithms. In particular, our goal at Tufts is to have ensembles of redistricting plans available for all 50 states within two weeks of the 2020 census results."

Mattingly's example of the judges' map in North Carolina brings up another very natural question: Wouldn't it be better to have judges, or at least some kind of impartial commission, draw up the electoral map in all states? In fact, as of 2017, seven states (Alaska, Arizona, California, Colorado, Idaho, Montana and Washington) have independent redistricting commissions. Ross's experiment shows that a non-partisan panel of judges can clearly outperform the current partisan process. There is also evidence that maps drawn by bipartisan commissions score higher on traditional measures of compactness.

A switch from partisan to non-partisan or bipartisan commissions would address the causes, rather than the symptoms, of the gerrymandering disease. However, it would require a sea change of political willpower to enact such laws in the other 43 states. Until that happens, we will continue to need mathematicians to patrol the sidelines and keep the politicians from going too far out of bounds.

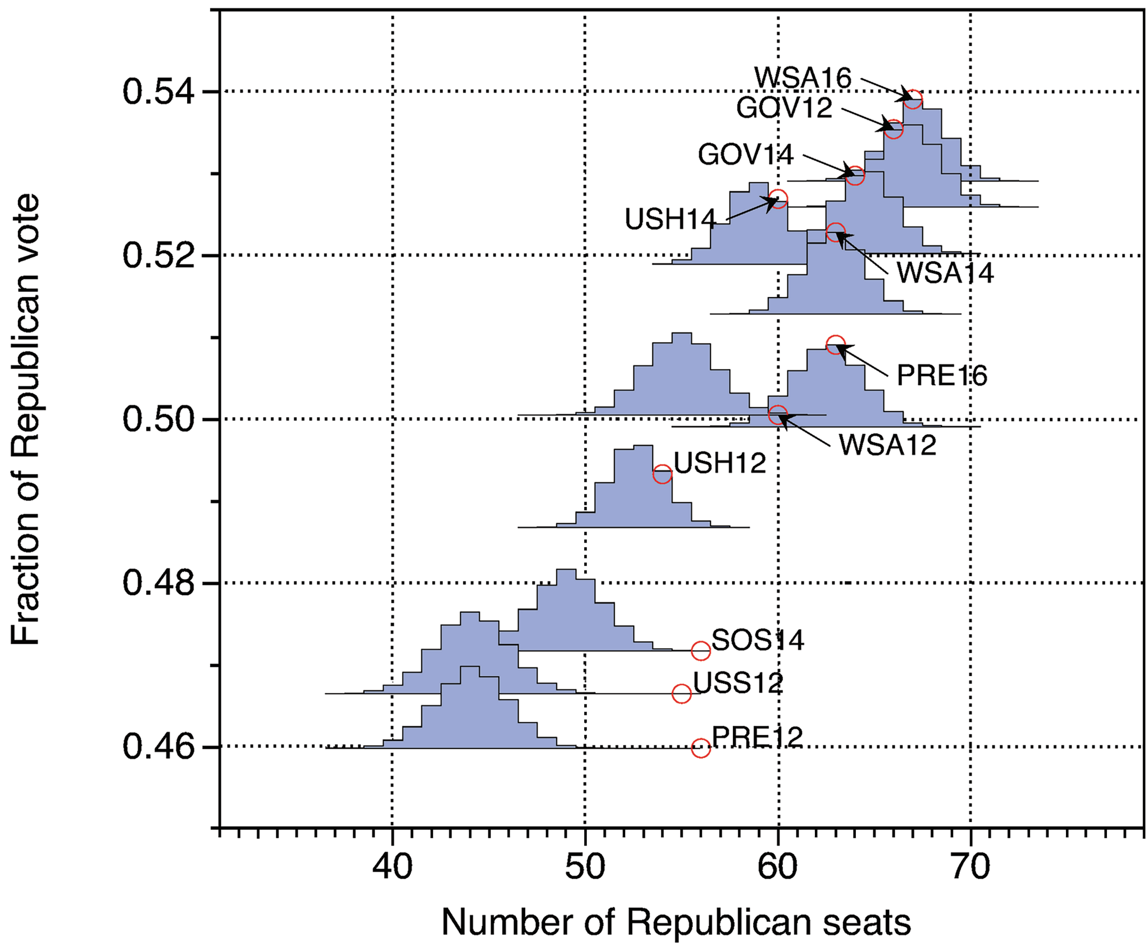

Figure 7. *Wisconsin "firewall". Actual seats won by Republicans in several elections are shown with red circles. (For example, PRE16 refers to the presidential election of 2016.) When the Republican vote share is high, their seat share lies near the peak of the histogram showing the results for a random ensemble of maps. By contrast, when the Republican vote share is below 48 percent, their seat share is an extreme outlier (bottom three graphs), giving them several more seats in Congress than a typical map from the ensemble. (Figure courtesy of Greg Herschlag and Jonathan Mattingly.)*

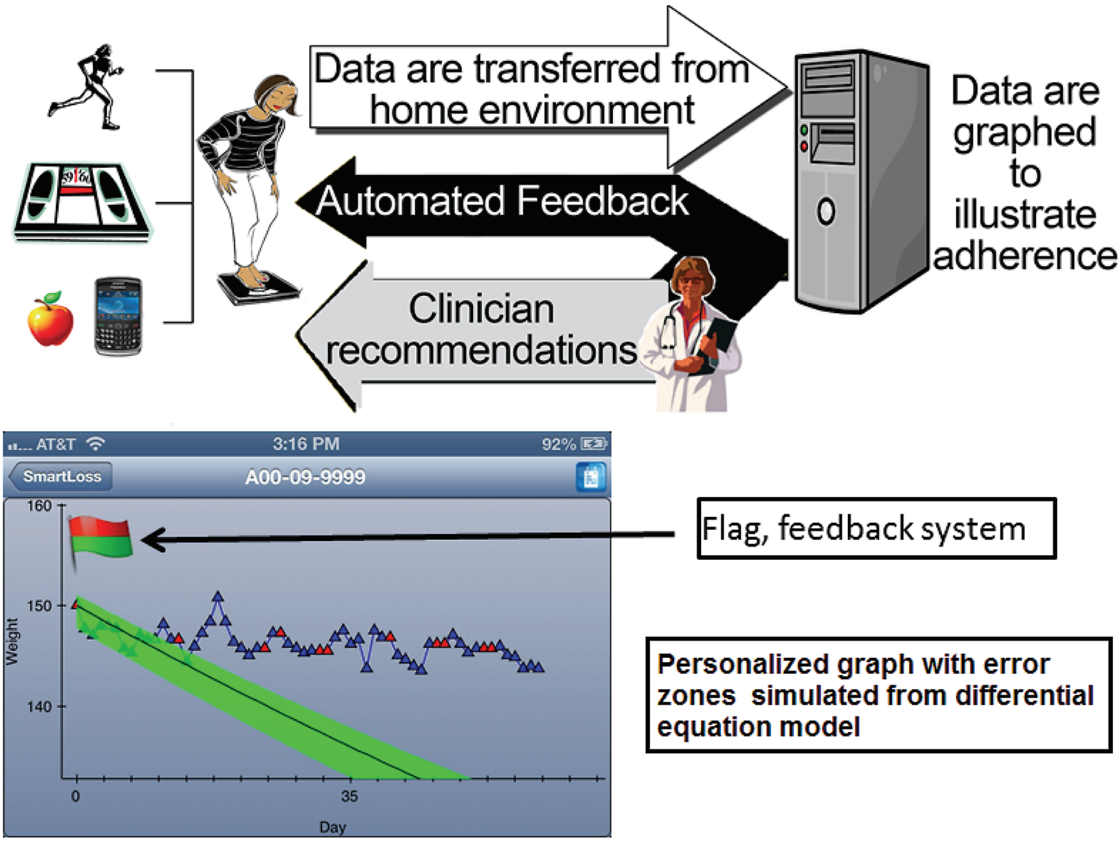

Figure 1. *Using the smartphone app to predict weight loss (and diagnose non-compliance). (Figure courtesy of Diana M. Thomas.)*