Chapter 2

Measuring partisan fairness

MIRA BERNSTEIN AND OLIVIA WALCH

CHAPTER SUMMARY

What does fairness in the context of redistricting look like? Can you identify a gerrymander based on election results alone? In this chapter, two mathematicians examine a variety of metrics that have been proposed in the courts as tools to detect partisan gerrymandering and to quantify its effects. The takeaway is that most of these metrics can lead to counterintuitive results. They are also unstable: slightly different conditions can yield markedly different outcomes. Fundamentally, these metrics share a common problem: you cannot interpret them without the context of what is "normal" for a particular state, based on the geographic distribution of its voters.

INTRODUCTION: WHAT IS A FAIR MAP?

In this chapter, we focus on **partisan gerrymandering**, in which boundaries of electoral districts are manipulated to give advantage to a political party. Given a **districting plan**, or electoral map, we'd like to be able to tell whether it has been gerrymandered and to quantify the advantage gained by the benefiting party. This issue arises frequently in gerrymandering lawsuits, and scholarship on this topic has been greatly influenced by decisions of the U.S. Supreme Court. While our focus will be on quantitative methods, we will include pointers to the necessary legal context along the way.

To begin with, any discussion of advantage requires a baseline: advantage relative to what? What would a districting plan that does *not* give any extra advantage to either side look like?

Finding the baseline turns out to be an extremely complex and controversial question. There are two fundamentally different ways one can define fairness in redistricting (as in many other contexts). One possibility is to focus on the map-making *process*: we could declare a map fair if it was drawn based on nonpartisan considerations, without the intent to hurt or benefit any party.[1] The other approach is to define fairness based on *results*: we could say that a districting plan is fair if it leads to fair electoral outcomes, i.e., outcomes that are consistent with some abstract notion of justice or equality.

In *Davis v. Bandemer*, the Supreme Court ruled that a districting plan is unconstitutional only if it violates both definitions of fairness: it must be motivated by "discriminatory intent" *and* have a "discriminatory effect" on voters of one party [1]. Over the years, the Court has considered several possible standards for assessing whether a plan's effect is discriminatory: **proportionality** (*Davis v. Bandemer*, 1986), **partisan symmetry** (*LULAC v. Perry*, 2006), and **low efficiency gap** (*Gill v. Whitford*, 2018). As we shall see, the Court had good reason to be skeptical of all three proposals. When elections are based on geographically defined districts, the outcomes depend on voter geography, and any standard that does not take this into account is bound to run into trouble.

It is only in the last few years that mathematical and technological advances have enabled researchers to evaluate the effects of a districting plan in a way that takes the geographic distribution of voters into account. This has led to the development of a new test of partisan gerrymandering, the **extreme outlier standard**, which takes on the task of disentangling the effects of gerrymandering from the effects of geographical districts. Unfortunately, even though this kind of work was in evidence in the last partisan gerrymandering case to come before the Supreme Court (*Rucho v. Common Cause*, 2019), the Court split along usual partisan lines to rule that partisan gerrymandering should not be adjudicated by the federal courts at all.

But of course, the fight against partisan gerrymandering is not over: it continues in state courts, state legislatures, and grass-roots efforts around the country. As we strive for fair maps, through both legal and political means, it is important to understand just how complex and contradictory the ideal of fairness can be. Quantitative methods can help us to contend with some of this complexity.

Ultimately, the question "what is a fair map?" is a philosophical one: it requires a normative choice (what *should* be done), and such choices cannot be dictated by any mathematical analysis. But mathematics can guide our decision-making by helping us to understand the implications of the choices we make; and, once we have chosen, it can help us to design more effective ways to implement our choices.

1 PROPORTIONALITY

When we give talks on gerrymandering to nonspecialists, we often start by asking the listeners a simple question:

[1] In this chapter, by "fairness" we mean exclusively *partisan* fairness. A map drawn without a partisan agenda could still be unfair in other ways—for instance, as a racial gerrymander.

> "Suppose Party *A* got 55% of the votes in your state. The state legislature has 100 seats. Ideally, how many of these seats should go to Party *A*?"

Invariably and unanimously, the answer is 55%. People are fine with small deviations like 52% or 58%, but by the time you get above 63% or so, pretty much everyone agrees that this is not ideal. In other words, most people's intuitive concept of fairness is **proportionality**: they want the number of seats that each party gets to be proportional to the number of votes it receives.

Because most people think of fairness in terms of proportionality, discussions of gerrymandering in the popular press are often framed in these terms as well. Here is a typical example from *The Washington Post*:

> "The 2012 election results give some sense of the extent of [Pennsylvania's] gerrymander. That year, Democratic candidates for the state's 18 U.S. House seats won 51 percent of their state's popular House vote. But that translated to just 5 out of 18, or a little more than one-quarter, of the state's House seats."[2]

There is, in fact, plenty of evidence that Pennsylvania's 2011 districting plan was a Republican gerrymander. Yet the disproportional results cited in *The Washington Post* cannot be used to prove this or to measure the advantage this gerrymander gave to Republicans. Counterintuitively, our system for electing representatives — dividing a state into districts, then choosing one representative per district — is not set up for proportionality. It is a system in which elections that are "fair" in the sense of process (no gerrymandering or partisan intervention of any kind) are highly *un*likely to produce proportional results.

To better understand what's going on here, we'll first introduce some formalism for discussing elections mathematically and then look at two real elections as case studies.

1.1 VISUALIZING VOTES AND SEATS

A districting plan (together with a procedure for choosing a representative in each district) forms an **electoral system**: a method for converting voter preferences into a choice of representatives. The results of every partisan election for a representative body (Congress, Parliament, state assembly, etc.) are often summarized in two sets of numbers: the fraction of **votes** that each party gets and the fraction of **seats** that each party wins. The fraction of votes can be calculated by simply aggregating the votes for each party across the state, but you might also want to calculate the average vote share across districts. These two values coincide when turnout is equal across all districts. (See Sidebar 2.1 for further discussion.)

For the remainder of this chapter, we will assume that every election involves only two parties. We arbitrarily choose one of the parties (in our real-world examples, it will always be Republicans) and look at the election from their point of view:

[2] *Ingraham, Christopher*, "How Pennsylvania Republicans pulled off their aggressive gerrymander", *The Washington Post*, 6 February, 2018

- We denote the Republican **vote share**—the fraction of the two-party vote that the Republican party received statewide in the election—by V.
- The Republican vote share in each of a state's N districts is represented by the vector $(v_1, v_2, \ldots, v_N)$, and the **average district Republican vote share** is denoted by $\overline{V}$.
- We denote Republican **seat share**—the fraction of available districts in which Republicans got more votes than Democrats—by S.

We can now visualize the results of a single election as a point in the **seats–votes plane**.[3] For instance, in Figure 1 we plot the results of the 2016 Congressional election in four states: Minnesota, Maryland, Ohio, and Michigan. It is natural to put vote share (V or $\overline{V}$) on the x-axis and S on the y-axis, because we usually think of vote share as an input into the electoral system that we are examining, while seat share is its output. Our goal is to understand how our system "converts" votes to seats.[4]

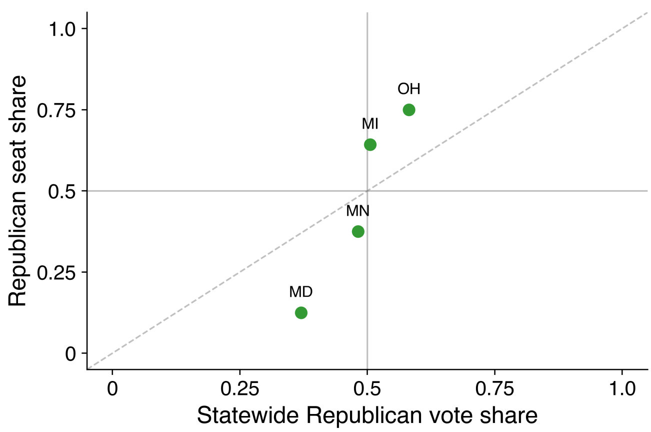

Figure 1: The 2016 Congressional election in four states, plotted on the seats–votes plane. A point in the upper-right quadrant ($V > \frac{1}{2}$, $S > \frac{1}{2}$) is an election where Republicans won a majority of the statewide vote and a majority of the seats (as in Ohio and Michigan). A point in the lower-left quadrant ($V < \frac{1}{2}$, $S < \frac{1}{2}$) is an election where Republicans won a minority of the statewide vote and a minority of the seats (as in Maryland and Minnesota).

If our electoral system promoted proportionality, we would expect most elections to cluster near the **line of proportionality**, defined by $S = V$. But this is not what happens! Figure 2 shows the fraction of Congressional districts in the 2012 and 2016 elections where Republicans won the majority of the Presidential vote share, for all states with six or more districts. The points are indeed clustered in a linear pattern, but not around the line of proportionality: the line of best fit has slope 2.6.

[3] All data, unless otherwise specified, taken from the MIT Election Science + Data Lab [2, 3].

[4] In reality, both V and $\overline{V}$ are affected by the districting plan, since districting lines can influence voter choices in many subtle ways (e.g., suppressing turnout in very safe districts, changing the availability of incumbents, etc). But as a first approximation, we can think of each party's vote share (V or $\overline{V}$) as an expression of the true underlying preferences of the voters.

The fact that the slope is greater than 1 is often termed a **winner's bonus**. For instance, in this case, you might say that there is an extra 2.6% of seat share for each additional percent in the winner's vote share. But reporting only the slope hides the fact that 11 out of the 52 elections plotted here buck the pattern completely by awarding more seats to the losing party. Such elections fall in the upper left and lower right quadrants.

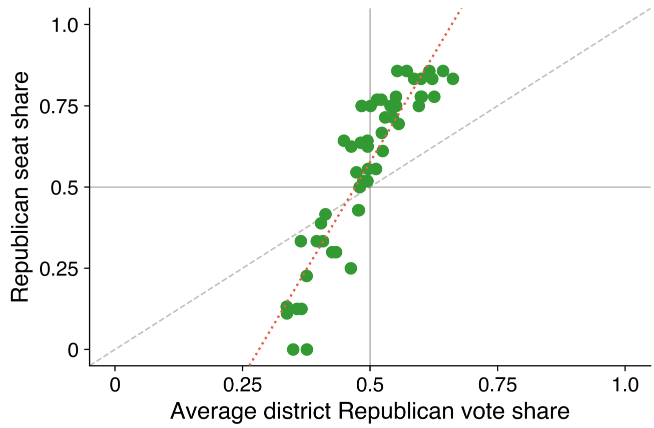

Figure 2: Data from the 2012 and 2016 elections for all 26 states with at least six Congressional districts. (Taken from The Daily Kos [4].) The x-axis shows average district Republican vote share ($\overline{V}$) in the Presidential race, while the y-axis shows the fraction of Congressional districts in which the Presidential election results had more Republican than Democratic votes. The red line is the line of best fit, slope 2.6. The reason for using Presidential election results is explained in Sidebar 2.2.

2.1 V VS. $\overline{V}$

You might wonder why we make the distinction between V, the statewide Republican vote share, and $\overline{V}$, the average district Republican vote share. The simple reason is that some authors use V, while others use $\overline{V}$. While V and $\overline{V}$ are typically close, they can be different in important ways. For instance, in Michigan's 2016 Congressional elections, V was very slightly greater than 0.5, while $\overline{V}$ was slightly less than 0.5. This changes the quadrant where Michigan appears on the seats–votes plane.

What is the conceptual difference between the two (beyond simply the way they are calculated)? Here's one way to think about it: V reflects the number of votes cast in the election, full stop. $\overline{V}$ is a version of V that's normalized by district turnout, making it so that every district is weighted the same, regardless of how many people turned out on election day.

In this chapter we use V for the majority of our figures, since we're often talking about "fairness" in an abstract sense, and V feels like the simpler and more elegant choice in this context. Others have argued for using $\overline{V}$ in all calculations of partisan metrics, to control for the effects of differential turnout across districts [5]. Here, we use $\overline{V}$ when the available data only contain district vote shares and not raw votes, and in each of the figures we've tried to be clear about which one we're using (or said "vote share" when the two quantities are the same).

2.2 MODELING UNCONTESTED RACES

Figure 2 uses the votes cast in each district in the 2012 and 2016 *Presidential* elections to derive the average district vote share and seat share. Why didn't we use the actual results from the 2012 and 2016 Congressional races?

The problem is that almost half of the 26 states shown in Figure 2 (12 states in 2012 and 10 in 2016) had at least one **uncontested district**: a district in which one of the two parties did not field a candidate. In these districts, the Republican vote share in the Congressional election would have been either 0% or 100%. Clearly this number does not reflect the true partisan preferences of the district's voters, so we don't want to use it in computing V or $\overline{V}$ for the state. Instead, we use the Republican vote share in the Presidential election from the same year, which tells us how many voters in the district preferred Republicans to Democrats in a different context. Of course, there are many reasons why the Presidential vote share may differ from the Congressional vote share, so this is only a very rough estimate—but it's certainly better than 0% or 100%.

If we want a more precise estimate, we can look at recent election outcomes for all the districts in the state and use these data to construct a statistical model for how Republican vote share tends to vary from election to election. Ideally, the model would include all the major factors that might affect a district's Republican vote share, such as: the year of the election, which office it is for, whether one of the candidates is an incumbent, etc. For example, a simple model might indicate that, in our state, incumbents tend to get a ~5% bonus in Congressional elections; and that once you account for this bonus, the Republican vote share in each district tends to be ~2% higher for Congressional elections than for Presidential ones. Then if an uncontested district with a Democratic incumbent had a 30% Republican vote share in the 2016 Presidential election, we would predict that its Republican vote share in the 2016 Congressional election would have been about $30\% + 2\% - 5\% = 27\%$.

To check the quality of our model's predictions, we can perform a standard modeling test by splitting the data from contested districts (for which we know the true Republican vote share) into *training data* and a small amount of set-aside *test data*. We base our model only on the training data, then see how well it predicts the Republican vote share in the test data. If the model does reasonably well, then we can feel justified using it to impute what the Republican vote share would have been in uncontested districts as well.

Political scientists use these kinds of models all the time, both to estimate vote share in uncontested districts and to address more general questions like "Has incumbency advantage in the US increased or decreased over time?" See Gelman and King [6] for a broad overview of imputation strategies and Mayer [7] for a concrete example: an expert witness report in *Whitford v. Gill* with a detailed description of its imputation model.

1.2 WHY DISTRICTS DON'T PRODUCE PROPORTIONALITY (USUALLY)

The following two scenarios illustrate why we should not expect elections under our electoral system to produce proportional outcomes.

Example 1: Competitive districts. Competitive districts, where the two parties have approximately equal support, are generally considered good for democracy: since neither party can count on an easy win, both candidates have to work hard to gain their constituents' vote. However, competitive districts can be terrible for proportionality. In theory, if all the districts in a state are closely contested, then a small swing in overall preferences can drive big deviations from proportionality.

Consider Minnesota, where Republicans in 2016 won 38% (3 out of 8) of the Congressional seats with 48% of the overall statewide vote (V). Three of Minnesota's eight Congressional districts (#1, #2, and #8) were extremely competitive, won by a margin of less than 2% (Table 2.1). With so many competitive districts, Republicans in Minnesota could easily have gotten anywhere from 25% to 62.5% of the Congressional seats (2–5 districts) with essentially the same statewide vote share.

District	Total votes	% Republican	Winner	Margin of victory... in %	in votes
1	335,595	49.6%	D	0.8%	2,547
2	341,285	51.0%	R	1.9%	6,655
3	392,313	56.9%	R	13.7%	53,837
4	324,332	37.3%	D	25.4%	82,266
5	330,617	24.4%	D	51.2%	169,297
6	358,395	65.7%	R	31.4%	112,375
7	330,516	47.5%	D	5.0%	16,628
8	356,185	49.7%	D	0.6%	2,009
Statewide	2,769,238	48.2%	3R/5D		

Table 2.1: Results of the 2016 Congressional election in Minnesota, with the margin of victory in percentage points and votes.

Example 2: A dispersed minority. You may have already read about Massachusetts (Chapter 0), where Republicans consistently get about 35% of the vote in Presidential elections. If Congressional voting followed a similar pattern, then a proportional outcome would give the state 2–4 Republican representatives (out of 9 or 10). Yet Massachusetts has not sent a Republican to Congress in over 20 years. Why?

The reason why Massachusetts does not have any majority-Republican districts is not gerrymandering, but the geographic distribution of its voters. Republican voters in Massachusetts are a significant minority almost everywhere, but a majority almost nowhere. For instance, in 2012, Mitt Romney (R) outperformed Barack Obama (D) in only about 15% of the state's precincts, and even in these "red precincts," his average vote share was only 54.5%.[5] Moreover, the red precincts did not form

[5] Code at: `https://github.com/political-geometry/partisan-fairness`

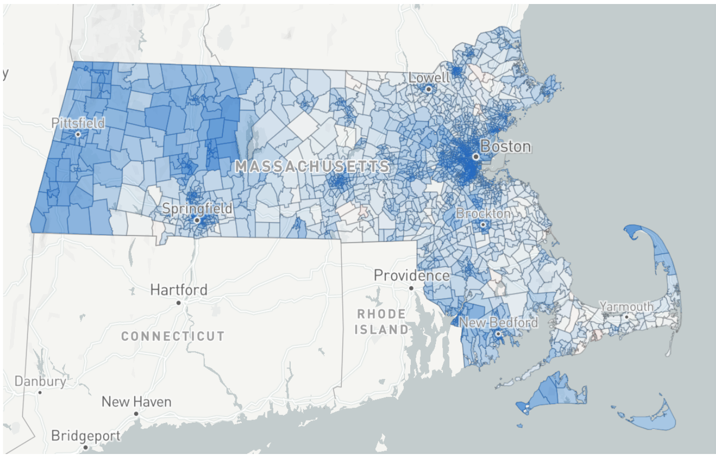

Figure 3: Voting patterns in Massachusetts precincts in the 2008 Senate election—one of the races shown in Duchin et al. [8] to produce no feasible Republican districts. Darker blues indicate a higher Democratic percentage of the two-way vote. Majority Republican precincts, shown in shades of red, are almost entirely absent from the plot, even though the two-way Republican vote share in the state as a whole was nearly 1/3 (31.9%). Image taken from the MGGG Districtr tool (`https://districtr.org/`).

a contiguous region: any Congressional district that included them would also have had to include some "blue precincts", many of which Obama won by a huge margin (getting 70% of the vote on average). It would have been challenging to create a majority-Republican district in Massachusetts in 2012, even if you were trying to gerrymander *for* the Republicans.[6]

Compare this to New York, where the two-party statewide Republican vote share tends to be only slightly higher than in Massachusetts (38.2% vs 35.3% in the 2016 Presidential election). Unlike Massachusetts, New York does have some majority-Republican clusters, and this has a noticeable effect on election results. From 2012 to 2020, Republicans consistently won about one-third of the state's Congressional seats (6 to 9 out of 27, depending on the year). New York's Congressional districts were drawn by the judiciary [9, 10], so are unlikely to have been gerrymandered for either party. Even if we suppose that the magistrate who worked on the maps was trying to "gerrymander" for proportionality, the fact that she was able to do so is in itself significant. The difference between New York and Massachusetts serves as a striking illustration of the extent to which a party's possible seat share in our electoral system is contingent on the geographic distribution of its voters.

If we want election outcomes to be proportional in a districted system, we generally would need to engage in what might be called "benign gerrymandering." In particular, to be confident in a proportional outcome, we would need to make many of our districts safe for one party or the other. We might allow ourselves a few competitive districts, but we definitely can't afford too many, since, in com-

[6]With some elections, it is *impossible* to make a majority-Republican district in Massachusetts out of small building blocks like towns or precincts, even if you abandon the requirement for contiguity. For a map corresponding to an election where a Republican district in MA would have been impossible to construct, see Figure 3. For a more detailed analysis of the obstacles to constructing a majority-Republican district in Massachusetts, see Duchin et al. [8].

petitive districts, small swings in voter preference can lead to large swings in the outcome. Thus, we arrive at the same conclusion that we have already hinted at in our Minnesota example: in the U.S. electoral system, competitiveness and (assured) proportionality are fundamentally incompatible.

It does not have to be this way. Almost half of the world's democracies elect their representative bodies using some method of **proportional representation**—that is, an electoral system that is designed to ensure proportionality with respect to political parties. (See Sidebar 2.3 and Chapter 20.) In such systems, competitiveness and proportionality do not undermine one another. Proportional representation does have some drawbacks (see, for example, King and Browning [11]), but one of its chief advantages is that it accords with most people's—including most Americans'—normative standard of fairness. If our goal is proportional outcomes, then we need to change not just the way we draw our districts, but our electoral system as a whole. In this chapter, however, our goal is to detect and quantify partisan gerrymandering within the context of our current system. And for these purposes, as we have just demonstrated, proportionality is simply the wrong metric.

THE SUPREME COURT WEIGHS IN

When partisan gerrymandering came to the Supreme Court in *Davis v. Bandemer*, the plaintiffs argued that intentional deviation from proportionality in redistricting violated the Equal Protection Clause. But the Court ruled decisively that "the mere lack of proportional representation [is not] sufficient to prove unconstitutional discrimination," citing essentially the reasons that we have outlined above:

> "If all or most of the districts are competitive...even a narrow statewide preference for either party would produce an overwhelming majority for the winning party... This consequence, however, is inherent in winner-take-all, district-based elections" [1].

Crucially, the Court in *Bandemer* did not say that proportional representation was *un*fair. In fact, in an earlier case (*Gaffney v. Cummings*, 1973), the Court had explicitly recognized proportionality as a legitimate goal that a state might pursue in designing its districting plan, even if this required making the district sizes slightly imbalanced. The *Bandemer* decision explains this apparent inconsistency as follows:

> "To draw district lines to maximize the representation of each major party[7] would require creating as many safe seats for each party as the demographic and predicted political characteristics of the State would permit... We upheld this "political fairness" approach in *Gaffney v. Cummings*, despite its tendency to deny safe district minorities any realistic chance to elect their own representatives. But *Gaffney* in no way suggested that the Constitution requires the approach...adopted in that case" [1].

[7] That is, to ensure proportionality by party (our footnote).

2.3 PROPORTIONALITY BY DESIGN

The Framers of the Constitution frequently used the phrase "proportional representation" in their deliberations. To them, it denoted the principle that "equal numbers of people ought to have an equal number of representatives" [12]. Writing during the American Revolution, John Adams put it slightly differently in *Thoughts on Government*: "equal interests among the people should have equal interests in [a representative assembly]" [13].

Nearly a century later, John Stuart Mill, in *Considerations on Representative Government*, pointed out what seems obvious in retrospect: equal representation by geographic region does not ensure equal representation by any other trait [14]. Mill called for true "proportional representation of all minorities," writing, "I cannot see... why people who have other feelings and interests, which they value more than they do their geographical ones, should be restricted to these as the sole principle of their political classification."

But how does one achieve equal representation of all interests? Must every group be represented in Congress according to its proportion in the population? If 24% of Americans are Catholic, must 24% of Congressional representatives be Catholic? And if 8% of Americans are Catholics who also believe in UFOs,[a] must 8% of representatives hold the same combination of beliefs?

This is where political parties come in. In principle, parties can form around any group of people with a common agenda who want their views represented in government. By choosing a party, a voter can explicitly designate which of her "feelings and interests" should form the primary basis of her "political classification." Instead of proportionality by geography, the Mill view then supports proportionality by party. Today, many people use the phrase **proportional representation (PR)** to denote an electoral system in which each party's seat share is structurally guaranteed to be (roughly) equal to its vote share.

There are many different PR systems in use around the world. The simplest one is **party-list PR**: each person votes for a party, and each party is allocated a number of seats proportional to the number of votes it receives. Party-list PR is the most common system across nations, but there are also more complex systems, which combine both proportional and local representation. For example, Germany uses a system called **mixed-member proportional representation (MMP)**, in which every person votes both for a local district representative and a (possibly different) political party. The district representatives account for about half the seats in the Bundestag, and the remaining seats are allocated by party in such a way as to make the overall results proportional. The system of electing one representative per district is used almost exclusively in Great Britain and its former colonies.

In 2017, and again in 2019, Representative Don Beyer (D-VA) introduced a House resolution called the Fair Representation Act,[b] which would require Congressional elections to be conducted using a ranked choice voting system that promotes proportionality. Congressional representatives under the Fair Representation Act would be elected locally, but districts would be larger than they are now: each district would elect 3–5 representatives. A similar system for national legislative elections has been used in Ireland since the 19th century.

Ranked choice voting (including the multi-winner version that promotes proportionality) is introduced in Chapter 20. For more on the mechanics, advantages, and disadvantages of different electoral systems, a good place to start is the ACE Electoral Knowledge Network (`aceproject.org`).

[a] Gallup poll, August 2019 (https://news.gallup.com/poll/266441/americans-skeptical-ufos-say-government-knows.aspx)

[b] Don Beyer, *Let's change how we elect the House of Representatives*, The Washington Post, 27 June, 2017.

In other words, the Court's message was: if you think that fairness hinges on proportionality, that's your business — go ahead and gerrymander for proportionality. You can even pass a state or Federal law requiring it. However, proportionality by party cannot be a constitutional standard of electoral fairness, because our district-based electoral system, which was in use at the time of the Constitutional Convention, fails this standard by its very nature.

On the question of what does constitute a "discriminatory effect", the Court offered only some very general guidelines:

> "Unconstitutional discrimination occurs only when the electoral system is arranged in a manner that will consistently degrade a voter's or a group of voters' influence on the political process as a whole" [1].

The justices knew that this was too vague to be actionable, but the majority wanted to leave the door open for a more precise "arithmetic" standard that might be found in the future:

> "We are not persuaded that there are no judicially discernible and manageable standards by which political gerrymander cases are to be decided" [1].

Thus, the *Bandemer* decision, while inconclusive, was an invitation to keep looking.

2 PARTISAN SYMMETRY

As it happens, just around the time of *Gaffney v. Cummings* in the 1970s, political scientists had begun to develop a new set of statistical tools for analyzing elections in nonproportional systems [15]. By the 1990s, a cadre of top political scientists and statisticians— most notably, Andrew Gelman, Bernard Grofman, and Gary King—had built support for the idea that any reasonable definition of fairness in redistricting (or, more generally, in electoral systems that were not designed for proportionality) should be based on **partisan symmetry** [16].

The basic idea of partisan symmetry is that, in a fair voting system, if one were to swap the parties' vote shares, their seat shares should also swap. Of course, when we say "swap vote shares", we are not imagining that individual Republican voters would turn into Democrats overnight and vice versa. So to make this idea usable, we will need to approximate more realistic swings in overall voter preferences.

For instance, suppose that, in the first election held under a given districting plan, Democrats get 52% of the votes and 65% of the seats. In the next election, Republicans get 51% of the votes and 67% of the seats. These results are not proportional: in each case, the majority party secured a seat share well above its vote share. But the results are roughly *symmetric*: the size of the "winner's bonus" was approximately the same in both cases, so the districting plan does not appear to give either party an inherent advantage.

This kind of analysis appears to require comparing the results of more than one election. And this is a problem: we don't want to wait through several election cycles (and hope for just the right shift in the voters' preferences) before we can judge whether a given districting plan is sufficiently symmetric. So we need a modeling assumption that allows us to predict, from the results of a single election, what would happen if such a shift occurred.

2.1 UNIFORM PARTISAN SWING AND SEATS–VOTES CURVES

Uniform partisan swing (UPS) is a model for how voters' partisan preferences change over time: its core assumption is that the change is closely linked across different parts of the state. For example, if we know that Democrats are becoming more popular in one region, then, no matter what's driving this trend, we assume that it affects the rest of the state in the same way. Of course, we don't expect Republican strongholds to suddenly switch sides, but we do expect them to become a little *less* Republican.

To make this precise, let us formulate a linear UPS model.[8] Suppose we observe an election in a state with N districts, with statewide Republican vote share V and district Republican vote shares $v_1,\ldots,v_N$. In a second election, suppose the observed vote share has changed from V to $V' = V+\delta$. (For instance, $\delta = 0.05$ if the overall Republican share has gone up from 0.47 to 0.52.) Under the UPS model, we assume that all the individual district vote shares have also changed by the same amount:

$$v_i' = v_i + \delta \text{ for } i = 1,\ldots,N.$$ [9]

This allows us to predict the new Republican seat share S': it is just the fraction of the v_i' that are greater than 0.5.

Thus, under the UPS assumption, we can use the results of a *single* election under a districting plan D to predict the number of seats that Republicans would win for *any* Republican vote share V'. In other words, UPS allows us to think of a districting plan D, along with a single election outcome, as specifying a *function* $V \to S(V)$ for converting vote share to seat share.[10] It is important to remember that this is just a prediction based on an assumption of how voter preferences change. In most cases, however, the assumption turns out to be fairly accurate. (For a detailed look at how well the UPS model holds up compared to real data, see Katz et al. [5].)

The graph of the function $S(V)$ is called the **seats–votes curve**.[11] Figure 4 shows

[8] Why specify linear here? Because there is another way of formulating UPS that swings the odds, rather than the votes themselves. Swinging the votes, as in linear UPS, can yield unpleasant edge conditions, like having to add 2% Democratic vote share to a district that is 99% Democratic (which you would resolve by setting the vote share to 100%). Swinging the odds neatly avoids headaches like that. Unfortunately, nobody really uses it. We've included code for it in the GitHub repository that goes along with this chapter.

[9] If v_i' ends up outside the interval $[0,1]$, we round it to 0 or 1.

[10] Variations on this construction of the function $S(\overline{V})$, including a stochastic version of the UPS model, can be found in Katz et al. and King [5, 17].

[11] Confusingly, the term "seats–votes curve" has been used over the years to refer to many different relationships between votes and seats. In early works, it often denotes the best-fit curve for electoral data from multiple elections [11, 15]. These days, however, it almost always refers to the curve derived from a single election based on the UPS model, as described above. See Katz et al. [5] for more on the formulation we discuss.

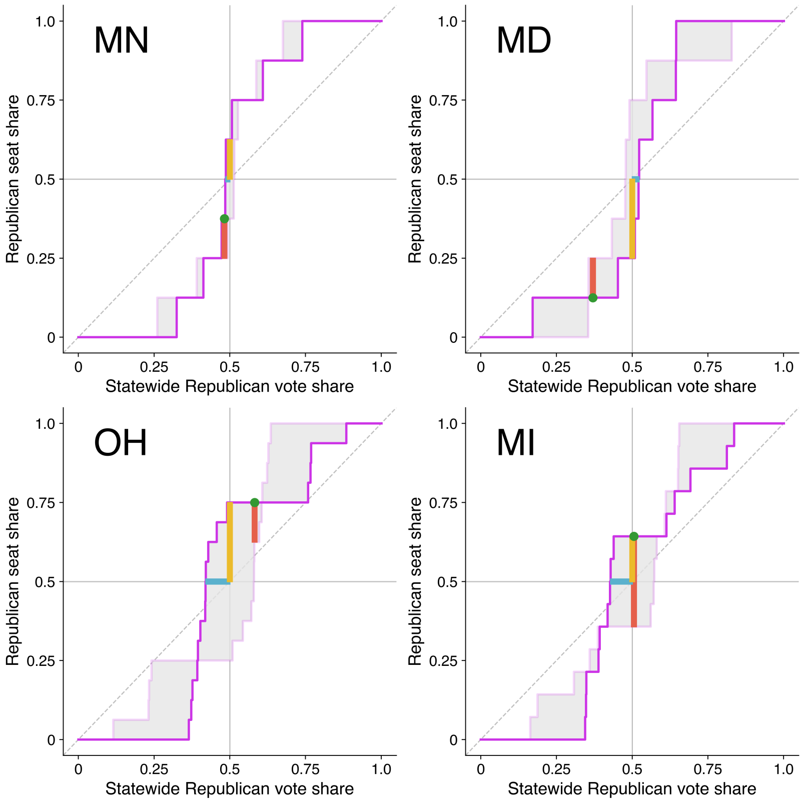

Figure 4: Seats–votes curves and symmetry scores for four 2016 Congressional elections. The seats–votes curve is dark magenta, and its 180° rotation is light magenta. The area of the gray region between the two curves is $\int_0^1 |S(V) - (1 - S(1-V))|\, dV = 2\int_0^1 |\beta(V)|\, dV$. The actual election is marked with a green dot, and the distance between the two curves at that point, $2\beta(V^\circ)$, is shown in red. $\beta(0.5)$ is shown in yellow and the approximate mean–median score in blue.

the seats–votes curves for the four elections in Figure 1, with the observed election outcome marked by a green dot. The distinctive stair-step shape of the curves is a consequence of the fact that seats are whole numbers, so seat share changes abruptly as rising δ pushes each $v_i + \delta$ past the 0.5 threshold.

We can now give a more precise definition for the concept of partisan symmetry. For every possible value of V, we want $S(V)$—the modeled Republican seat share—to equal the Democratic seat share when *their* vote share is V. Since Democrats get vote share V when the Republican vote share is $1 - V$, we get the following.

Definition 1. *A districting plan satisfies the* **partisan symmetry standard** *if* $S(V) = 1 - S(1 - V)$ *at every* V.

This definition has a nice geometric interpretation. Switching the role of Democrats and Republicans corresponds to reversing the directions of both the V-axis and the S-axis in the seats–votes plane. This double reflection is equivalent to a $180°$ rotation around the point $(0.5, 0.5)$. So the partisan symmetry standard is upheld in the political science sense if and only if the seats–votes curve is symmetric about $(0.5, 0.5)$ in the familiar, geometric sense. Note that every symmetric curve must pass through the point $(0.5, 0.5)$, which suggests this point as a good reference for assessing the symmetry of a districting plan.

2.2 MEASURING (A)SYMMETRY

Of course, we don't expect any districting plan to be exactly symmetric, so if we want to use partisan symmetry as a standard of fairness, we need to specify how much *asymmetry*, or bias, we are willing to tolerate. In other words, we need some way of quantifying a plan's deviation from the ideal of symmetry.

Following Katz et al. [5], we define the **partisan bias** of a districting plan at each vote share V by the formula

$$\beta(V) = \frac{S(V) - (1 - S(1 - V))}{2}.$$

If we regard the average of the curve and its $180°$ rotation as a symmetrization, then $\beta(V)$ measures the distance from the curve to its symmetrization at each value of V.

But how do we reduce the bias of the curve as a whole to a single number? Several different metrics, or **symmetry scores**, have been proposed in the political science literature:

- The seats–votes curve is derived from an actual election with Republican vote share $V°$. If we want to summarize a districting plan's partisan bias, one possibility is simply to use $\beta(V°)$.
- We can compute the average value of $|\beta(V)|$ over the entire interval $[0, 1]$: as an integral, this is

  $$\int_0^1 |\beta(V)|\, dV = \frac{1}{2}\int_0^1 |S(V) - (1 - S(1 - V))|\, dV,$$

  i.e., half the area between the curve and its rotated copy.[12]

  Since the extremes of the seats–votes curve usually correspond to unrealistic scenarios, we may choose to restrict the integral to a smaller interval around 0.5, such as $[0.4, 0.6]$. We could also remove the absolute value and compute $\int \beta(V) dV$, if we want the sign of the integral to reflect which party gains the advantage (and are OK with the possibility that positively and negatively counted areas might cancel). We refer to this summary score, and all its variants, as the **β-average** (signed or unsigned).
- Recall that any perfectly symmetric curve must go through the point $(0.5, 0.5)$. So one way to summarize a plan's deviation from symmetry is to measure

[12] The idea of using area to measure deviation from an ideal might remind you of Gini coefficients in economics, and some authors have called this score the "partisan Gini" [18, 19].

State	$\int_{0.4}^{0.6} \beta(V)\,dV$	$\int_0^1 \lvert\beta(V)\rvert\,dV$	$\beta(V^\circ)$	$\beta(0.5)$	Mean–median score
MN	0.005	0.017	0.062	0.125	0.019
MD	−0.022	0.05	−0.062	−0.25	−0.018
OH	0.036	0.078	0.094	0.25	0.087
MI	0.021	0.051	0.143	0.143	0.082

Table 2.2: Symmetry scores for the 2016 Congressional election in four states. From left to right: β-average on the interval $[0.4, 0.6]$, with area favoring Democrats counted as negative; unsigned β-average on the interval $[0, 1]$, with all area counted as positive; $\beta(V^\circ)$, the bias at the actual vote share; $\beta(0.5)$, the bias at $V = 0.5$; and the mean–median score, the approximate horizontal distance from the curve to $(0.5, 0.5)$. For all scores except the unsigned β-average, positive values indicate a Republican advantage while negative values correspond to a Democratic advantage.

how far its seats–votes curve lies from this central point, either vertically or horizontally. The vertical deviation is just $\beta(0.5)$. This corresponds to the counterfactual that asks what the seat outcome would have been if the vote had been exactly evenly split.

- The horizontal distance between the seats–votes curve and the point $(0.5, 0.5)$ estimates how much Republicans could fall short of half of the vote while still winning at least half of the state's seats. Note that half of the seats is particularly important if you are analyzing a state legislature, where controlling the majority is significant.[13] This measure of partisan asymmetry is approximately equal to the difference between the median and the mean of the district vote shares, or the **mean–median score**. (See Sidebar 2.4.)

Symmetry scores for each of the 2016 Congressional elections in Figure 1 are shown in Table 2.2.

INTERPRETATIONS AND LIMITATIONS

Partisan symmetry scores take the seats–votes curve and reduce it to a smaller, ideally more illustrative set of numbers. But reducing the information in this way can have some undesirable consequences.

- **Inconsistency:** Having so many different ways of summarizing the partisan symmetry of a districting plan means that they can produce contradictory results. In particular, most of these are signed measures, where in our convention a positive result indicates a bias toward Republicans. For geometric reasons, $\beta(0.5)$ and the mean–median score always have the same sign, but they can have very different magnitudes.[14] And the three seat-based measures of symmetry—$\beta(V^\circ)$, $\beta(0.5)$, and β-average—can theoretically have any combination of signs. This can make it difficult to find a consistent interpretation of these scores.

[13] For states with an even number of districts, the seats–votes curve intersects the line $S = 0.5$ not at a single point, but in a line segment. In this case, we measure the distance from (0.5,0.5) to the midpoint of the line segment.

[14] Some have observed that under certain circumstances this sign can tell a counterintuitive story; see DeFord et al. [19].

- **Instability:** Two elections with similar results can lead to very different partisan symmetry scores for the same districting plan. For instance, Figure 5 shows the seats–votes curves for Congressional elections in Minnesota from 2012 to 2018. If the UPS assumption held perfectly, the curves would look exactly the same: only the green dot corresponding to the actual results would change position along the curve. In reality, the curves do look broadly similar, so the UPS assumption seems not too far off. Yet even slight shifts in the curve can result in very different summary scores: for instance, in Minnesota in 2012, all the metrics suggest a highly symmetric plan. In 2014, however, $\beta(V°)$ jumps from zero to -0.06 (suggesting a pro-Democratic bias), and then up to 0.06 in 2016 (suggesting a pro-Republican bias). So while the seats–votes curve as a whole is not sensitive to small deviations from UPS, the summary scores can change drastically, leading to qualitatively different conclusions.
- **Plateaus and "firewalls":** The scores we discussed are ill-equipped to capture the long flat plateaus that can arise in seats–votes curves. For example, look at Ohio and Maryland, where the actual election lies in the middle of a particularly long, flat part of the curve. In Maryland, as long as the Republican vote share stays between roughly 25% and 45%, the results will be the same — one seat for Republicans, seven for Democrats. This is certainly suggestive of gerrymandering, yet it is not captured by Maryland's small mean–median score and relatively small β-average on the interval $[0.4, 0.6]$.
- **Unrealistic counterfactuals:** It seems strange to declare a districting plan "fair" on the basis of a hypothetical situation that is extremely unlikely to occur. Again, we can think of Maryland. Why should it matter what the seats–votes curve looks like near $V = 0.5$ or $V = 0.6$, if the actual Republican vote share in Maryland hasn't gone above 0.4 in over a decade?

In fact, unrealistic counterfactuals pose a problem not just for partisan symmetry summary scores but for the symmetry standard as a whole.

In any situation where one party has a much higher statewide vote share than the other, a partisan mapmaker could start with the most ruthless districting plan they can think of, then make cosmetic adjustments to some of their own party's safe districts to game the scores, e.g., pulling $\beta(0.5)$ and mean–median to zero. It would cost them nothing, since all the adjustments would be in the unrealistic region of the curve. For this reason, advocates of the symmetry standard generally advise limiting it to states where both parties have a realistic chance of winning a majority [20].

Partisan symmetry scores reduce the information contained in the seats–votes curve, but the seats–votes curve itself is already a very reduced representation of the state: it depends only on the vector of votes in each district and contains no information about the state's geography and demography. Yet (as you'll hear again and again in this book), spatiality matters. We'll come back to this later in the chapter.

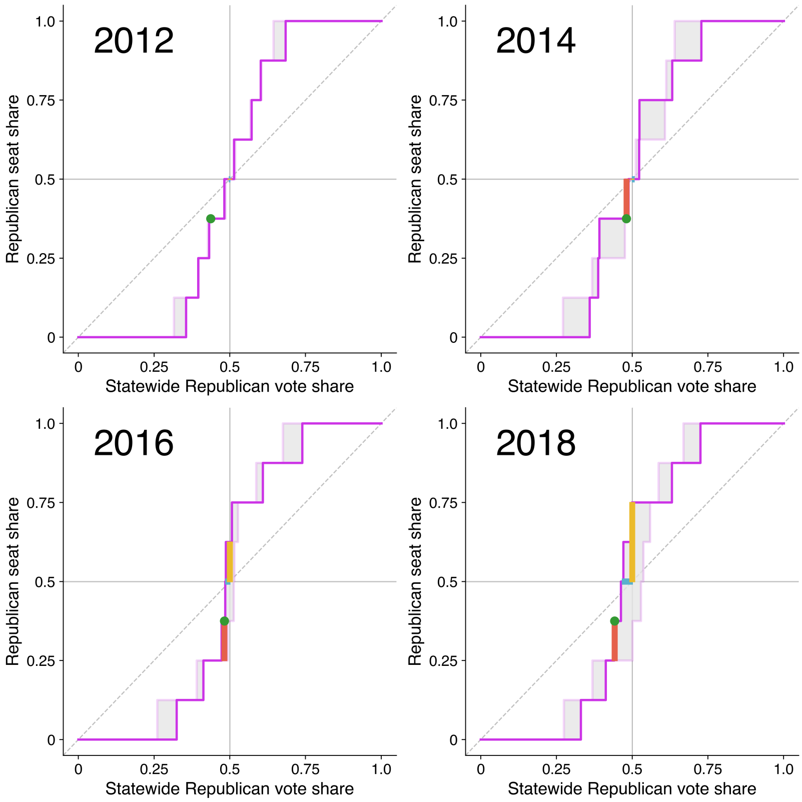

Figure 5: Seats–votes curves for Minnesota Congressional elections from 2012 to 2018, with summary scores as in Figure 4. As before, the magenta line is the seats–votes curve and the pale magenta line is the 180° rotation of that curve. The mean–median score is approximated by the blue line, $\beta(0.5)$ is the yellow line, and $2\beta(V^{\circ})$ is the red line. There is no red line in 2012, because $\beta(V^{\circ})$ is zero, and there are no yellow lines in 2012 and 2014 because the seats–votes curve in those years goes through the point $(0.5, 0.5)$. While the curves themselves are broadly similar, some of the scores change significantly from year to year: for instance, $\beta(V^{\circ})$ jumps from 0 in 2012 to -0.06 in 2014 to +0.06 in 2016.

THE SUPREME COURT WEIGHS IN AGAIN

After *Bandemer*, the next two partisan gerrymandering cases to reach the Supreme Court were *Vieth v. Jubelirer* (2004) and *LULAC v. Perry* (2006). In *LULAC*, several prominent political scientists submitted an *amicus* brief, proposing partisan symmetry as a new standard of fairness in redistricting [16].

The four liberal justices on the Court seemed content with the symmetry standard. The four conservative justices were not going to be convinced by any standard: they believed the Court should have nothing to do with partisan gerrymandering at all. So the decision was up to Anthony Kennedy, the swing justice. Kennedy agreed with the liberal justices that the Court should continue to hear partisan gerrymandering cases. In his *Vieth* opinion, he had expressed the hope 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" [21]. But in *LULAC*, he rejected all the standards of fairness under consideration by the Court, including partisan symmetry.

Kennedy's objection to the symmetry standard was not one outlined in the previous section. He had a more fundamental concern:

> "The existence or degree of asymmetry may in large part depend on conjecture about where possible vote-switchers will reside" [22].

In other words, what bothered Kennedy was that there was no way to measure partisan bias without relying on a statistical model (such as uniform partisan swing). It's not just that he doubted the validity of the UPS assumption; he was wary of using any model at all:

> "Even assuming a court could choose reliably among different models of shifting voter preferences, we are wary of adopting a constitutional standard that invalidates a map based on unfair results that would occur in a hypothetical state of affairs" [22].

So, after all was said and done, the Court ended up back where it started: without a Constitutional standard of fairness in redistricting, but still holding out hope that such a standard might be found.

2.4 THE MEAN–MEDIAN SCORE

Suppose we observe an election where the district Republican vote shares are $v_1, \ldots, v_N$. We define the **mean–median score** of the districting plan under which the election was held to be

$$M = \text{median}\{v_i\} - \text{mean}\{v_i\}.$$

The mean–median score has been proposed as a measure of partisan asymmetry [18, 23, 24, 25] because, under the equal turnout assumption ($V = \overline{V}$), it coincides with the horizontal distance between the point $(0.5, 0.5)$ and the seats–votes curve derived from the election. (See Sidebar 2.5 for a discussion of the equal turnout assumption.)

We show that the two measures are equal in the case when N is odd; the even case is proved similarly. Assuming equal turnout, each point on the seats–votes curve corresponds to an election with statewide Republican vote share $V + \delta = \overline{V} + \delta$ and district vote shares $v_1 + \delta, \ldots, v_N + \delta$. Since N is odd, there is some district m for which $v_m = \text{median}\{v_i\}$. When $v_m + \delta < 0.5$, Republicans will lose district m and at least half of the remaining districts; thus, $S(V + \delta) < 0.5$. Similarly, when $v_m + \delta > 0.5$, Republicans will win district m and at least half of the remaining districts; thus, $S(V + \delta) > 0.5$. We conclude that the seats–votes curve intersects the line $S = 0.5$ precisely when $v_m + \delta = 0.5$. The statewide Republican vote share at this point is

$$V + \delta \;=\; \overline{V} + 0.5 - v_m \;=\; 0.5 - M.$$

Thus, M is the horizontal distance from the intersection point to (0.5, 0.5), as required.

3 THE EFFICIENCY GAP

The *LULAC* case made one thing clear: if you wanted to fight gerrymandering in the Supreme Court, you had to devise a measure of fairness that would appeal to Justice Kennedy. "Symmetry" might be a good conceptual hook, since Kennedy didn't shut the door on it completely.[15] But whatever measure you came up with could not involve any counterfactuals or hypotheticals.

In 2014, law professor Nicholas Stephanopoulos and political scientist Eric McGhee came out with an influential article about a new metric, called the **efficiency gap (EG)**, which they believed would do the trick [26]. The efficiency gap proposes to measure unfairness by comparing the number of votes "wasted" by each party in an election. A fair map is defined to be one in which $\mathrm{EG} = 0$ because the parties waste the same number of votes—a kind of symmetry.[16] Once you have the full vote data, computing EG for any given election is very simple, and at first glance no hypotheticals are required.

In 2015, a team of lawyers including Stephanopoulos used the new EG measure to challenge the districting plan in Wisconsin. The case, *Whitford v. Gill*, made national headlines: for the first time since *Bandemer*, a federal court sided with the plaintiffs and declared a districting plan to be an unconstitutional partisan gerrymander. The Wisconsin plan's high efficiency gap was not the only basis for the decision, but it was a significant part of the plaintiffs' argument.

As *Whitford v. Gill* headed to the Supreme Court (becoming *Gill v. Whitford* in the process), there was a lot of excitement about the efficiency gap, both in the legal community and in the popular press. But many political scientists were skeptical. Indeed, as we shall see in this section, when you translate the requirement $\mathrm{EG} = 0$ into the language of votes and seats, you get some very uncomfortable results.

3.1 DEFINITION AND EXAMPLES

For the purposes of defining EG, a vote is considered to be "wasted" if it does not contribute to the election of a candidate. This includes all votes cast for a losing candidate, as well as votes for a winning candidate in excess of the 50% required to win.

To see how this works in practice, we return to our running example of Minnesota's Congressional race in 2016 (Table 2.3).

Let's go through the calculation of wasted votes in District 1:

- First, we compute how many votes were required for a party to win in District 1. The total number of voters in the district (Republicans and Democrats) was 335,595. So to win, a party had to get just over half of those votes: 167,798.

[15] What Kennedy actually said about partisan symmetry was that he did not "altogether discount... its utility in redistricting planning and litigation" [22]. The four liberal justices were much more explicit in their support, but they were not the ones that needed to be convinced.

[16] Of course, this is a different use of the word "symmetry" than the technical sense described in Section 2 [5]. But it seems plausible that when Justice Kennedy wrote that "symmetry" may be useful in redistricting litigation, he was not wedded to the technical definition either.

District	Votes for R	Votes for D	Total votes	Needed to win	Wasted by R	Wasted by D
1	166,524	169,071	335,595	167,798	166,524	1,273
2	173,970	167,315	341,285	170,643	3,327	167,315
3	223,075	169,238	392,313	196,157	26,918	169,238
4	121,033	203,299	324,332	162,166	121,033	41,133
5	80,660	249,957	330,617	165,309	80,660	84,648
6	235,385	123,010	358,395	179,198	56,187	123,010
7	156,944	173,572	330,516	165,258	156,944	8,314
8	177,088	179,097	356,185	178,093	177,088	1,004
Total	1,334,679	1,434,559	2,769,238		788,681	595,935

Table 2.3: Vote data from 2016 Congressional election in Minnesota, with the number of votes wasted by each party. Importantly, with only small changes to the votes in Districts 1, 2, and 8, these numbers can change dramatically.

- The Democrats in District 1 got 169,071 votes and won the election. But as we just saw, only 167,798 of these votes were necessary to win. We say that the remaining 1,273 Democratic votes were wasted, since they did not contribute to the election of a Democrat.
- The Republicans lost District 1, so none of their votes contributed to electing a Republican. We therefore say that all 166,524 Republican votes in District 1 were wasted.

In this particular case, the losers happened to waste more votes than the winners, but it can also go the other way. In District 5, the Democrats wasted more votes than the Republicans, even though they won the election: they had such a large margin of victory that their "extra" votes outnumbered *all* the votes cast by Republicans.

Note that the language of wasted votes provides a useful way of quantifying the intuition that gerrymandering can be accomplished by "packing" and/or "cracking" voters of the opposing party. When many voters of Party A are packed into a few districts, Party A will win those districts by huge margins, causing it to waste more votes than Party B (as in Minnesota's District 5). On the other hand, when a block of Party A voters is cracked, this usually involves creating several districts with safe-but-low margins of victory for Party B. Now the winners waste fewer votes than the losers, once again giving Party B an advantage in terms of wasted votes.

After calculating the wasted votes in each district individually, we add up all the votes wasted by Republicans and Democrats in the entire state and call these quantities W_R and W_D respectively. We can see that, by this definition, many more Republican votes were wasted in Minnesota in 2016 (788,681 to 595,935). Another way of putting it is that Republican voters were unable to use their votes as *efficiently* as Democrats. This is where the name "efficiency gap" comes from.

Definition 2. *Suppose we have two parties, A and B. The efficiency gap favoring Party A, for a given pattern of votes, is defined as the difference in wasted votes, divided by the total number of votes:*

$$\mathrm{EG} = \frac{W_B - W_A}{T}.$$

Dividing by T means that we are measuring the difference in wasted votes relative to the total number of votes cast. This seems reasonable: for instance, a difference of 100 wasted votes would be quite significant in an election with only 1,000 voters, but barely noticeable in an election with millions of voters, like Minnesota's.

The sign of EG indicates the direction of the advantage. If $\mathrm{EG} > 0$, then Party B wasted more votes than Party A, so we conclude that the districting plan was tilted in favor of A. If $\mathrm{EG} < 0$, we conclude that the plan was tilted in favor of B. Of course, in reality, you never get EG exactly equal to zero. Based on their analysis of previous elections, Stephanopoulos and McGhee suggest 0.08 as a reasonable margin of error [26]. In other words, any plan with $|\mathrm{EG}| < 0.08$ should be considered fair enough, whereas a plan with $|\mathrm{EG}| > 0.08$, while not necessarily a gerrymander, should be seen as a cause for concern.

Working from Table 2.3, we find that the efficiency gap for the 2016 Congressional election in Minnesota is

$$\mathrm{EG} = \frac{W_D - W_R}{T} = \frac{595,935 - 788,681}{2,769,238} = -0.0696.$$ [17]

Since this is below the threshold of 0.08 proposed by Stephanopoulos and McGhee, the EG standard would lead us to conclude that Minnesota's districting plan is probably *not* a partisan gerrymander.

3.2 SOME ISSUES WITH THE EFFICIENCY GAP

Let's do a little sensitivity analysis. In Table 2.3, notice that the biggest difference in wasted votes between the winner and the loser occurs in the most competitive districts (#1, #2, and #8). What would have happened if a small number of voters in these districts had changed sides?

- If just 2400 Democrats in Districts #1 and #8 had switched their votes, it would have been enough to give the Republicans narrow victories in both districts. Most of the wasted votes in those districts would then belong to Democrats. We would get $\mathrm{EG} = 0.18$, signaling a huge Republican gerrymander.
- If 3400 Republicans in District #2 had switched their votes, the district would have gone to the Democrats. Most of the wasted votes in that district would then belong to Republicans, and we would get $\mathrm{EG} = -0.19$, now signaling an even more egregious Democratic gerrymander.

In other words, while the actual vote results do not flag Minnesota's districting plan as problematic, a shift of a few thousand votes (out of nearly 3 million) could make it look like *either* an egregious Republican gerrymander *or* an egregious Democratic gerrymander.

This example shows that, in the presence of competitive districts, EG is extremely volatile and thus arguably useless as a measure of unfairness. Stephanopoulos

[17] We identify Party A with Republicans and Party B with Democrats, so that, as usual, a positive score corresponds to a Republican advantage and a negative score to a Democratic advantage.

and McGhee are aware of this problem: they recommend not using high EG as an indicator of gerrymandering if small changes in competitive districts could make it fall below the 0.08 threshold. Still, it should give us pause that such a straightforward and reasonable-sounding metric turns out to behave so unreasonably in this case.

What about Massachusetts? For reasons that will become clear soon, we will use data from 2000 rather than 2016. Right away, we run into a problem: five of the state's ten Congressional districts were not even contested by Republicans in 2000. As we saw in Sidebar 2.2, we cannot reasonably include the votes from those districts in our calculation of EG, yet we can't just ignore these districts either.

The usual approach to uncontested districts is to use a statistical model to estimate what the election results in those districts *would have been* if both parties had fielded a candidate. This is the approach that Stephanopoulos and McGhee recommend as well. But notice that, if we take this route, we are basing our EG calculations on a *counterfactual*—exactly what Justice Kennedy protested against and what EG was explicitly designed to avoid. Thus, the claim that EG does not rely on statistical modeling turns out to be inaccurate for elections with uncontested districts (which are extremely common).[18] The truth is that almost all analyses of voting data have some statistical assumptions underlying them. You can't avoid "hypotheticals"; you can only de-emphasize them and hide them in the background.

Getting back to Massachusetts: to avoid dealing with uncontested districts, let's use data from the 2000 Presidential election instead of the Congressional one.[19] In 2000, a total of 1,616,487 people in Massachusetts voted for the Democrat (Gore) and 878,502 voted for the Republican (Bush), for $T = 1,616,487 + 878,502 = 2,494,989$ two-party votes in all.[20] Republicans were a minority in every district, so we know that *all* their votes were wasted: $W_R = 878,502$. Democrats needed just over half the votes in each district to win that district, Therefore, up to rounding, the total number of votes required to win all the districts was $T/2$. The rest of the Democratic votes were wasted: $W_D \approx 1,616,487 - \frac{2,494,989}{2} \approx 368,992$. Thus, the efficiency gap for Massachusetts under these assumptions is $\mathrm{EG} = \frac{W_D - W_R}{T} = \frac{368,992 - 878,502}{2,494,989} = -0.20$, indicating a massive tilt in favor of Democrats. This accords with our intuition that Republicans in Massachusetts were very inefficient at turning votes into seats: their 35% vote share got them no seats at all.

But if there is any unfairness to Republicans here, it has nothing to do with the districting plan. As we touched on in Section 1, there are some elections that could not yield a Republican district in MA no matter how the districts were drawn. It turns out that the 2000 Presidential election was one of them [8]. Thus, our calculation shows that, for this election, *every* possible districting plan in Massachusetts would have resulted in $\mathrm{EG} = -0.20$. The efficiency gap fails for the same reason

[18]For instance, the Wisconsin districting plan that was challenged in *Whitford v. Gill* (the court case in which EG was first introduced) had 99 districts, of which as many as 49 were uncontested in some of the elections under scrutiny.

[19]Note that this does not get around the issue of hypotheticals. We are just using an extremely rudimentary statistical model: "In a district where both parties field a Congressional candidate, the Republican vote share is roughly the same as it is in the concurrent Presidential election." This model would be too crude to use in a serious analysis, but it is good enough for our illustrative purposes here.

[20]We continue to ignore votes for other candidates, even though in the 2000 Presidential election, the third-party vote share in Massachusetts was unusually high — about 6.4% of the total.

proportionality does: it detects unfairness that is arguably real, but has nothing to do with gerrymandering.

Here is a more surprising result: what if Republicans in Massachusetts were not a 35% minority, but a 20% minority, still dispersed across the state? The Democrats would once again win all the seats, so for any districting plan, W_R would be 20% of T (since all Republican votes would still be wasted). W_D would be $80\% - 50\% = 30\%$ of T (since all Democratic votes above the 50% needed to win in each district would be wasted). Thus, the efficiency gap would be $\mathrm{EG} = \frac{W_D - W_R}{T} = 0.3 - 0.2 = 0.1$. Paradoxically, we would conclude that the districting plan is unfair to Democrats, even though they won all the seats in the state. In other words, the EG standard calls for a 35% minority to get more than 0 seats, but for a 20% minority to get *fewer* than 0 seats!

To be fair to Stephanopoulos and McGhee: they are aware of all these issues with EG, and they propose a number of safeguards, i.e., contexts in which a court should not interpret a high EG score as a sign of gerrymandering.[21] But when a method that seemed so simple turns out to have so many exceptions and counterintuitive results, you can't help but wonder: what is really going on here? Clearly we need a better understanding of what the efficiency gap is actually measuring.

SEATS AND VOTES AGAIN

Recall that we were able to calculate the efficiency gap for Massachusetts (and the hypothetical "20% Republican" version of Massachusetts) knowing almost nothing about the distribution of voters in districts. For the 20% version, all we used were the Republicans' (fictional) statewide vote share ($V = 0.2$) and seat share ($S = 0$). It's not hard to check that we could have used the same approach to calculate EG for the actual 2000 election, with $V = 0.35$ and $S = 0$. Of course, Massachusetts is a particularly straightforward case, but it turns out that, with some elementary algebra, we can approximate EG for *any* election with a simple expression in terms of V and S. We just need to make one simplifying assumption, **equal turnout**: that is, we assume that there are T/N voters in each of the N districts. (See Sidebar 2.5 for what happens when this assumption does not hold.)

To compute EG, we need to know W_R and W_D. To find W_R, let's first figure out how many votes the Republicans did *not* waste. By the equal turnout assumption, it takes $\frac{T}{2N}$ votes to win each district. Republicans won SN districts, which required a total of $SN \cdot \frac{T}{2N} = \frac{ST}{2}$ votes. The remainder of the VT Republican votes were wasted. This gives $W_R = VT - \frac{ST}{2} = \frac{2V-S}{2} \cdot T$. An analogous calculation shows that $W_D = \frac{1-2V+S}{2} \cdot T$. Thus,

$$\mathrm{EG} = \frac{W_D - W_R}{T} = \frac{1-2V+S}{2} - \frac{2V-S}{2} = S - 2V + \frac{1}{2}.$$

We can rewrite this as $EG = \left(S - \frac{1}{2}\right) - 2\left(V - \frac{1}{2}\right)$. In other words, the efficiency gap

[21] Unfortunately, these caveats are routinely ignored by people who cite EG in their writing. Journalists, in particular, often report EG scores uncritically, lending the appearance of scientific precision to many spurious claims about gerrymandering.

standard ($EG = 0$) locates fairness not on the diagonal $S = V$ but on the line with slope 2 passing through $(0.5, 0.5)$. The permissible zone where $|EG| < 0.08$ is simply a narrow band around this line.

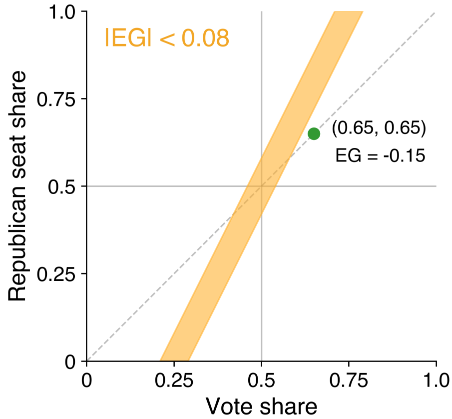

Figure 6: The yellow band marks the region in the seats–votes plane where $|EG| < 0.08$ under the equal turnout assumption. Note that the point $(0.65, 0.65)$ lies outside this region. Although such an election would accord with most people's intuitive definition of fairness (proportionality), its efficiency gap of -0.15 would indicate a Democratic gerrymander. (Figure adapted from Duchin [27].)

Notice that the permissible zone in Figure 6 leaves out most elections with proportional outcomes. For example, a districting plan under which Republicans got 65% of both votes and seats would be suspected of being a Democratic gerrymander. When we said in Section 1 that our electoral system tends not to produce proportional outcomes, that was a factual, empirical statement — but the efficiency gap standard has turned it into a normative judgment. Instead of saying that proportionality is not required for fairness, it actually implies that proportionality is unfair! [22]

THE SUPREME COURT DOES NOT WEIGH IN

We have discussed the efficiency gap at great length because, for a few years after its publication in 2015, it was all the rage in anti-gerrymandering circles. Due to its central role in the early stages of *Gill v. Whitford*, the first major partisan gerrymandering case since *LULAC*, it got a lot of attention in the popular press, and numerous scholarly articles were published analyzing, critiquing, and tweaking it [28, 29, 30, 31, 32].

[22] In practice, as the proponents of EG point out, a plan that produced proportional results would be safe in court no matter what its efficiency gap was. It takes both "discriminatory effect" and "discriminatory intent" for a redistricting plan to be declared unconstitutional. Since the Supreme Court has already declared proportionality a legitimate goal in *Gaffney v. Cummings*, no one could accuse the designers of a proportional plan of discriminatory intent.

But this does not address the deeper issue. The purpose of a quantitative standard of fairness is not to replace our intuitive notion, but to formulate it more precisely. Given how strongly our intuition associates fairness with proportionality, a measure that labels proportionality *unfair* can hardly be said to be a measure of fairness at all.

In the end, the Court did not issue a formal opinion on using EG as a measure of fairness. Instead of being decided on the merits, *Gill v. Whitford* was sent back to the lower courts to deal with issues of legal standing.

2.5 THE EQUAL TURNOUT ASSUMPTION

In any real election, turnout across districts will certainly not be exactly equal. The districts across a state have very nearly equal total population at the beginning of the Census cycle, but the number of people who actually turn out to vote is bound to vary. For instance, as you can see in Table 2.3, district turnout in Minnesota in 2016 ranged from a minimum of 324,332 (District 7) to a maximum of 392,313 (District 3).

How well does the simplified EG expression $S-2V+\frac{1}{2}$ approximate the original EG formulation in real elections? In theory, it could be very far off, but in practice, the two values are usually quite close. The table here shows the comparison for the 2016 Congressional elections, for all states with 8 or more Congressional districts in which every district was contested by both parties.

State	**EG**	$S-2V+\frac{1}{2}$
MD	−0.11	−0.12
MI	0.15	0.13
MN	−0.07	−0.09
MO	0.06	0.04
NC	0.19	0.2
NJ	0.06	0.0
OH	0.11	0.09
TN	−0.01	−0.02

The largest discrepancy, 0.06, occurs in New Jersey, where the variation in district turnout in 2016 was extraordinarily high: minimum 167,070, maximum 334,038.

In general, turnout tends to be lower in Democratic areas than in Republican areas [33]. Correspondingly, in recent elections, most states had lower average turnout in Democrat-won districts than in Republican-won districts, sometimes by a significant ratio [30]. This observed tendency causes the equal turnout assumption to overestimate W_R, underestimate W_D, and therefore (usually) underestimate EG, as can be seen in the table above.

While the simplified $\mathrm{EG}=S-2V+\frac{1}{2}=0$ formula calls for half the votes to secure half the seats, this is not true of $\mathrm{EG}=\frac{W_D-W_R}{T}=0$. The implications can be rather counterintuitive. Ellen Veomett in [30] derives a more sophisticated formula for EG in terms of S, V, and the turnout ratio, and then applies this formula to a hypothetical 50-50 election with the same turnout ratio as the actual 2016 Congressional election in Texas. It turns out that, to receive a score of $\mathrm{EG}=0$, such an election would have to award Democrats 60% of the seats!

4 ENSEMBLES AND OUTLIERS

So far, we have considered three standards of fairness in redistricting: proportionality, partisan symmetry, and equality of wasted votes (EG). Each of these standards has its advantages and disadvantages, but they all share the same fundamental flaw: they attempt to set an *absolute* baseline of fairness, without taking the specifics of a state's distribution of voters into account.

Proportionality and EG both require a particular relationship between seat share and vote share ($S = V$ and $S = 2V - \frac{1}{2}$ respectively). But as we have seen again and again, in our district-based electoral system, the same vote share can legitimately lead to many different seat shares. The more sophisticated partisan symmetry standard avoids this pitfall, but it too fails the test of geography: as we will demonstrate later in this section, certain kinds of asymmetry in the geographic distribution of voters will naturally lead to plans with asymmetric seats–votes curves.

Given the myriad potential arrangements of voters in a state, it is hard to imagine any absolute standard, no matter how sophisticated, that would be satisfied by all neutrally drawn districting plans under *any* population distribution. And as long as a measure of fairness is unable to distinguish between the effects of gerrymandering and the effects of geography, mapmakers will use this as an excuse, rendering the standard toothless.

THE SUPREME COURT WEIGHS IN (FOR THE LAST TIME?)

How do we deal with this seemingly intractable problem? One possibility is to give up and declare the problem unsolvable. It is impossible to decide whether a districting plan is fair unless we first agree on a definition of fairness. And since there does not appear to exist an absolute, all-purpose standard of fairness, we are stuck. This is the view of the Supreme Court in its most recent (and perhaps final) partisan gerrymandering decision, *Rucho v. Common Cause* (2019):

> "Federal courts are neither equipped nor authorized to apportion political power as a matter of fairness. It is not even clear what fairness looks like in this context. Deciding among... different visions of fairness poses basic questions that are political, not legal. There are no legal standards discernible in the Constitution for making such judgments. And it is only after determining how to define fairness that one can even begin to answer the determinative question: 'How much is too much?' [34]"

Based on this reasoning, the Court in *Rucho* declared partisan gerrymandering to be **nonjusticiable**: "outside the courts' competence and therefore beyond the courts' jurisdiction" [34].

The *Rucho* decision emphasizes that the Court "does not condone excessive partisan gerrymandering" and acknowledges that "[e]xcessive partisanship in districting leads to results that reasonably seem unjust" [34]. It urges the states and

Congress to address this problem via the political process, the first step of which is to decide what kind of fairness they want.

4.1 INTRODUCING ENSEMBLES

Everything we have seen in this chapter so far seems to confirm the Court's opinion in *Rucho*: our electoral system does not lend itself to a one-size-fits-all standard of fairness. But recent advances in statistics and computer science have given us a new way of tackling the challenges of geography: **ensemble-based analysis**. Here, the baseline of fairness is defined empirically, by creating a collection (or **ensemble**) of districting plans for the state and looking at its properties in the aggregate.

The first step is to have a computer randomly generate a large number of "eligible" nonpartisan districting plans: plans that satisfy all state and Federal laws (population equality, contiguity, compactness, the Voting Rights Act, etc.) and use no partisan data in their construction. For each eligible plan, we then use precinct-level results from a recent election to figure out which of its hypothetical districts Republicans would have won. Thus, we can report what the Republican seat share S would have been under each of our eligible plans.

This is where the state's voter geography enters the picture. For instance, in a Massachusetts election where it is provably impossible to create a Republican district [8], all the randomly generated plans will have $S = 0$. But in New York, which has a similar Republican vote share but has some more concentrated Republican areas, the number of Republican seats under each plan will depend on how the plan happens to chop up those areas. In general, our ensemble of randomly generated plans will produce a range of values for S, reflecting the possibilities and the probabilities for this particular state, with this particular geography.

Finally, we compare the plans in our ensemble to the enacted or proposed plan. Was the seat share within the reasonable range for S across the ensemble of alternatives or is it a statistical outlier? If it is outside the reasonable range, does it deviate in the direction of proportionality, or does it exacerbate the controlling party's advantage? This leads to the **extreme outlier standard**: unless the mapmakers can provide some other explanation for their plan's deviation from the norm, we can reasonably conclude that the plan was the result of partisan gerrymandering. We can then quantify its effect on the disadvantaged party by comparing it to the nonpartisan plans in our ensemble.[23]

As an example of ensemble analysis, consider North Carolina's 2016 Congressional districting plan (the subject of *Rucho v. Common Cause*). The official set of criteria used in the plan's construction included the usual desiderata of population

[23] What if the mapmakers do offer an alternative explanation? Suppose they claim that their plan is different from all the plans in our ensemble because they were trying to achieve some additional innocuous goal X. Then we can simply restrict our ensemble to only those plans that also achieve goal X (and create more such plans if necessary). If the districting plan under examination is still an outlier, then we know it's not because of goal X. Once all other possible explanations are gone, the only one that remains is partisan gerrymandering: we know the intent and we can quantify the effect. The court can decide how much is too much, but at least we have identified precisely the quantity that we were looking for.

equality, contiguity, and compactness, minimizing the number of counties split between districts, and adhering to the Voting Rights Act. But the list of criteria also included “partisan advantage”: the mapmakers were instructed to try to “maintain the current partisan make-up of North Carolina’s congressional delegation” [35].

To analyze how much advantage Republicans actually got from the resulting plan, researchers at Duke University randomly generated 66,544 nonpartisan plans that performed at least as well as the enacted plan on the criteria of population equality, compactness, and minimizing split counties. To comply with the Voting Rights Act, they ensured that each plan in their ensemble had at least two districts that were majority African-American.[24] The histogram in Figure 7 shows the number of seats that Republicans would have won in the 2016 Congressional election under each of these 66,544 random plans [36].

The statewide Republican vote share in this election was 53%. Most of the randomly generated plans give Republicans eight seats, but even if Republicans had won nine seats, they could still reasonably claim that their plan did not excessively disadvantage Democrats: after all, a random nonpartisan plan would have a 30% chance of producing the same result.[25]

But under the North Carolina districting plan, Republicans in 2016 won 10 of the state’s 13 seats. Among the computer-generated nonpartisan plans, only 1% produced such an extreme result.

This has two implications. First, it gives strong evidence that North Carolina’s districting plan was, in fact, an intentional partisan gerrymander: the odds of obtaining such a plan using only nonpartisan considerations are tiny. In the specific case of North Carolina, this evidence is unnecessary, since we already knew that “partisan advantage” was one of the official criteria for creating the plan. However, in situations where the gerrymanderers are better at hiding their tracks, the fact that ensemble analysis can produce evidence of *intent* to gerrymander could prove very useful.

Second, we can now quantify the *effect* of the gerrymander on the outcome of the 2016 election. Nine Republican seats would have been a high but still plausible outcome under a nonpartisan plan. Since the Republicans actually got ten seats, we can say with high confidence that they gained at least one seat, or 8% in seat share, as a direct consequence of the gerrymander.

The beauty of the extreme outlier standard is that it does not require us to agree on a normative definition of fairness. Its baseline of fairness is just the absence of clear partisan gerrymandering, i.e., the requirement that a plan’s partisan metrics should not look radically different from an ensemble of neutral, nonpartisan plans.

[24] The actual requirements of the Voting Rights Act are more complex, but since North Carolina has had two majority-minority districts for years, it is reasonable to assume that a plan with two such districts would be in compliance. For more on the Voting Rights Act in the context of redistricting, see Chapter 6 and Chapter 7.

[25] Note that a plan with 9 Republican seats would have an efficiency gap of

$$S - 2V + 0.5 = 0.69 - 2(0.53) + 0.5 = 0.13,$$

if we used the equal turnout assumption formula. So the EG standard would end up flagging 30% of randomly generated plans as Republican gerrymanders!

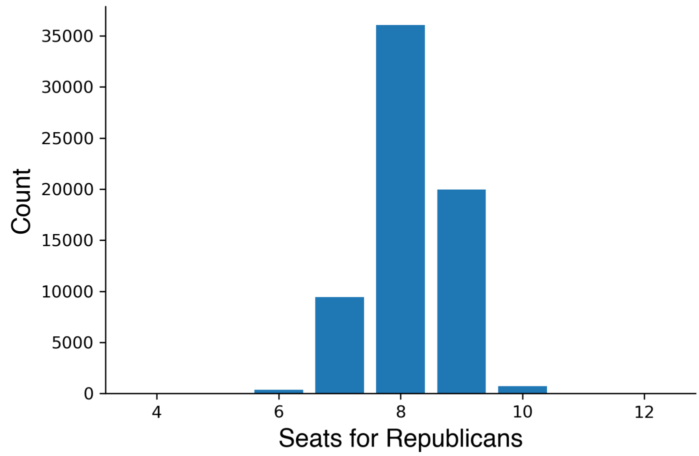

Figure 7: Number of Congressional seats (out of 13) that would have been won by Republicans in North Carolina's 2016 Congressional election under 66,544 randomly generated neutral plans. The actual number of seats won was 10. Figure based on data from Herschlag et al. [36].

This sounds simple, but it relies on our ability to produce a representative sample of eligible plans satisfying all the legal requirements. Until researchers had developed the statistical and computational tools to attempt this, we had no way of knowing what "absence of gerrymandering" looks like with a particular state's geography.

Ensemble analysis is an active area of current research. There are many important technical details to work out about how the set of eligible maps is generated; different research groups use somewhat different approaches. You can find more information on the technical aspects of ensemble analysis in Chapter 16 and Chapter 17.

4.2 APPLICATIONS OF ENSEMBLES

We can use ensemble analysis to verify empirically some of the theoretical results mentioned earlier in this chapter.

First, consider proportionality (or lack thereof). In Figure 7, most of the plans in the ensemble give Republicans 62% of the seat share (8 out of 13 districts) even though their vote share was only 53%. The proportional outcome would have been 7 districts ($S = 0.53$), but this happens in less than 14% of the plans in the ensemble.

The introduction to this book, as well as several other chapters, feature similar histograms for many other elections.[26] In most of them, proportional outcomes are unlikely. The expected winner's bonus differs from state to state and from election to election: sometimes the most likely seat share is extremely far from proportionality, sometimes quite close. This bears out our claim that there is no one "correct" seat share for any given vote share.

[26]Note that the histograms in Chapter 0 show seat share and vote share for Democrats rather than Republicans.

In all the histograms in Chapter 0, the party that got the majority of the statewide vote also gets a majority of the seats under most plans in the ensemble. But in general, even this basic result is not guaranteed! Figure 8(A) reproduces the histogram for Pennsylvania's 2016 Presidential election, where the statewide Republican vote share was 50.4%. Figure 8(B) shows what would happen if we shifted the vote by 0.5% toward Democrats, using uniform partisan swing. In this hypothetical election (which is extremely close to the real one), Republicans would have gotten less than half the vote, yet all 50,000 plans in the ensemble still lead to $S \geq 0.5$ (9+ seats out of 18), and 96% of the plans lead to $S > 0.5$.[27]

The 2016 election in Pennsylvania also poses a problem for the partisan symmetry standard. Figure 8(C) shows the seats–votes curves for all the plans in the ensemble superimposed on each other. The mean value of S for each $\overline{V}$ is shown in black. This "average seats–votes curve" for the ensemble is highly asymmetric by any measure of symmetry, as are almost all the individual seats–votes curves. Something about the geographic distribution of voters in Pennsylvania gives the Republicans a huge advantage, making it very unlikely that a random neutral districting plan will pass the symmetry standard.

What is going on here? One possible explanation is the frequently cited demographic trend of "Democrats packing themselves." Figure 8(D) illustrates this phenomenon at the level of precincts, the smallest geographic units for which election data are available. (Intuitively, you can think of each precinct as a neighborhood.) The histogram shows many Pennsylvania voters living in extremely Democratic precincts (Republican vote share < 10%), but almost none living in extremely Republican precincts (Republican vote share > 90%). Similar trends have been observed in other elections across the country, even in states that lean strongly Republican overall [37].

The same asymmetry that we see at the extremes of the histogram also exists in the state as a whole. In 2016, in Pennsylvania's majority-Republican precincts, the overall margin of victory for Republicans was 16%. In majority-Democratic precincts, the overall margin of victory for Democrats was 21%. Let's call this difference in the margins of victory **differential packing**, and denote it by d_{precinct}. For Pennsylvania in 2016, $d_{\text{precinct}} = 0.21 - 0.16 = 0.05$, indicating that, at the precinct level, Democrats were somewhat more packed than Republicans.

Does differential packing in precincts extend to districts? Not necessarily: if the precincts were distributed completely at random across the state, all such local differences would get diluted. But intuitively, nearby precincts tend to be similar to each other, so we would expect the trend to persist at larger spatial scales as well. This might plausibly lead to a tendency for Democratic districts to be more packed,

[27] This is a good opportunity to revisit the quote from *The Washington Post* in Section 1, regarding Pennsylvania's 2012 Congressional election. That year, Republicans won 13 out of 18 seats, with a statewide vote share of just over 49%. These results were taken by the author of the article as proof of "an aggressive gerrymander". Now compare this outcome to the hypothetical election in Figure 8(B), with a similar statewide Republican vote share. A plan that gave Republicans 13 seats would indeed be flagged as a gerrymander, but a plan with just one fewer Republican seat would not be. Intuitively, a seat share of 67% (= 12/18) for a minority party would probably still look suspicious to the author of the article and to most readers. But ensemble analysis shows that, given Pennsylvania's voter geography, it would not be an indication of gerrymandering.

on average, than Republican ones.

Ensemble analysis confirms this empirically. For a given districting plan, we can define d_{district} to be the difference between the average margin of victory in Democratic districts and the average margin of victory in Republican districts. Across all the districting plans in our Pennsylvania ensemble, the mean of d_{district} is 0.044. Differential packing at the district level in Pennsylvania is, on average, less extreme than at the precinct level, but not by much.

In a state with $\overline{V} = 0.5$, a positive value of d_{district} automatically implies $S > 0.5$: if Democratic districts are won by larger margins, then there must necessarily be fewer of them. In both Figure 8(A) and (B), the overall Republican vote share is extremely close to 0.5 and d_{district} is significantly greater than 0 for most plans in the ensemble. From these two facts alone, we can conclude that most of the plans will give the majority of Pennsylvania's seats to Republicans, though to what extent, we can't say. (For a much deeper dive into the geography and vote spatiality of Pennsylvania, see Chapter 5.)

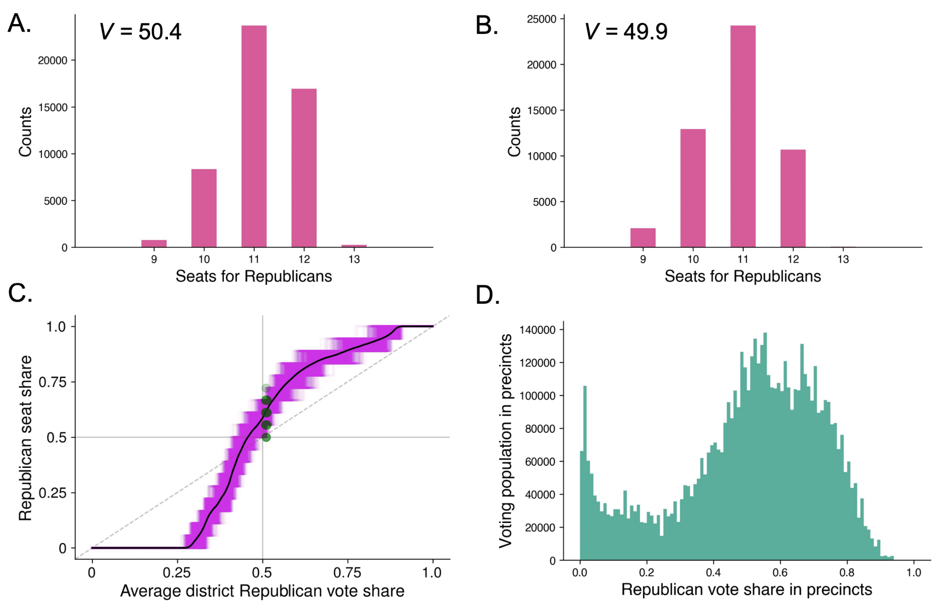

Figure 8: Analysis of the 2016 Presidential election in Pennsylvania. (A) Number of Congressional seats (out of 18) that would have been won by Republicans under 50,000 randomly generated plans. (B) Number of seats that would have been won by Republicans if the statewide Republican vote share decreased by 0.5% (assuming uniform partisan swing). (C) Seats–votes curves for 50,000 randomly generated plans. The black curve shows the mean value of S for each vote share. (D) Total number of voters in precincts at each Republican vote share.

Finally, let us revisit the efficiency gap through the lens of ensembles. For their report on redistricting criteria in Virginia [38], researchers at MGGG generated two ensembles of neutral plans: one with 11 seats (for Virginia's Congressional delegation) and one with 40 seats (for the state Senate). Figure 9 shows the number of seats won by Democrats and the distribution of EG for each plan of ensemble.[28]

We see immediately that the EG distributions look like noisier versions of the seat distributions. This is unsurprising, given the formula $EG \approx S - 2V + \frac{1}{2}$ that we derived in Section 3. Here V is fixed: it is the actual Republican vote share in the 2017 election for Attorney General. So under the equal turnout assumption, each plan's EG is just a function of its seat share S. The noise in the EG histograms reflects slight variations in district turnout across the plans in each ensemble, but it is negligible.

In general, we expect a plan that is an outlier in seats to be an outlier in EG and vice versa. Thus, in the context of ensemble analysis, the efficiency gap offers very little new information.

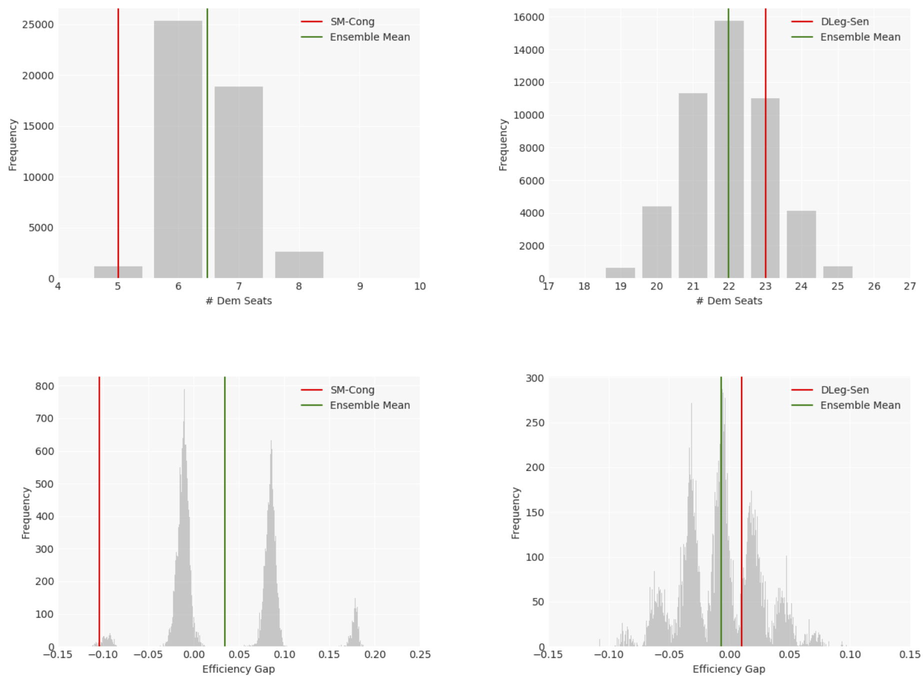

Figure 9: Histograms of seats and efficiency gaps from ensembles in Virginia, using data from the 2017 election for Attorney General. The plots on the left correspond to ensembles with 11 seats, and the plots on the right correspond to ensembles with 40 seats. The red lines show results from Virginia's actual districting plans: the special master's Congressional map (SM-Cong, left) and the legislature's state Senate map (DLeg-Sen, right). The green lines show the ensemble means. Taken from Figure 5 of DeFord and Duchin [38].

THE STATE SUPREME COURTS WEIGH IN

So far, the extreme outlier standard has been validated by the Supreme Courts of Pennsylvania and North Carolina, in *League of Women Voters v. Commonwealth of*

[28] Since all analyses in DeFord and Duchin [38] are in terms of Democratic rather than Republican vote share, $EG > 0$ here corresponds to an advantage for Democrats.

Pennsylvania (2018) and *Common Cause v. Lewis* (2019) respectively. In each case, the Court heard from several expert witnesses who had used different methods to generate their ensembles, but arrived at the same conclusion: the plan in question was an extreme outlier, constructed with clear partisan intent and conferring a significant, quantifiable advantage on the mapmakers' party. Both Courts concluded that the plan under consideration caused sufficient damage to the disadvantaged party to violate the Free Elections Clause of their respective state Constitutions.

Neither Court chose to set any specific numeric threshold for how extreme an outlier a plan has to be in order to be deemed unconstitutional. Perhaps this was because, in both cases, the Republican seat share of the enacted plans was so far out of the range of the nonpartisan ensembles that it left no room for doubt. In the next redistricting cycle, there are likely to be many more cases brought in different states, and some of them might be less clear-cut. We may even see proponents of different versions of the extreme outlier standard arrive at different conclusions because of their differences in methodology. A lot of legal and mathematical details remain to be worked out. Nonetheless, based on the experience of the last few years (and the resounding endorsement of U.S. Supreme Court Justice Elena Kagan, albeit writing for the dissent [39]), it seems clear that the extreme outlier standard is well on the way to becoming the principled, widely accepted quantitative standard of fairness in redistricting that the Supreme Court in *Rucho* has assured us cannot possibly exist.

5 CONCLUSION: DEBATING FAIRNESS

We began this chapter by asking a seemingly simple question: "What is a fair map?" We discovered that this is not a simple question at all: the idea of fairness is extremely complex and multifaceted, and our voting system isn't set up to accommodate any absolute definition of it. This is what makes it so hard to detect gerrymandering and to measure its effects. Did an election have a disproportionate result because of gerrymandering… *or* because the districts were highly competitive, as in Minnesota? Is the seats–votes curve asymmetric because of gerrymandering… *or* because of differential packing, or both?

Ensemble analysis circumvents this problem by establishing a baseline grounded in the actual geography of a state. Instead of measuring deviation from an abstract ideal, it allows us to measure deviation from neutrality. As a new and powerful way to measure the actual effects of gerrymandering, it should prove invaluable in redistricting reform around the country.

But as the Supreme Court reminds us in *Rucho*, "deciding among… different visions of fairness poses basic questions that are political, not legal" [34]. While absolute standards of fairness are not good for deciding whether gerrymandering has actually occurred, they are still very much relevant at the political level. The people of a state (or of the United States) can decide on any definition of fairness that reflects their values (such as proportionality or partisan symmetry) and pass a law accordingly. Of course, voter geography might make it difficult for mapmakers to comply with this law. But drawing maps that strive to comply with such laws is something that the Supreme Court has previously declared constitutional, and it is up to the people of each state to decide whether careful partisan tuning is in accordance with their values.

A more serious problem is that many of the properties that voters might want from a districting plan are incompatible with one another. Foremost among these, as we have seen, are competitiveness and proportionality. Arizona's State Constitution requires mapmakers to strive for competitive districts [40]; Ohio's Constitution requires mapmakers to strive for proportionality by party [41]. Most likely, Arizona voters would not declare themselves to be against proportionality, nor Ohio voters against competitive districts. Yet as we saw in Section 1, under our current electoral system, competitiveness and (assured) proportionality are fundamentally incompatible.

This might be an argument for change that goes beyond the way we draw our district lines. If proportionality—or some other ideal that is elusive to districters—is truly what we want, maybe it's the voting system that should change, and not the standard we use to judge it.

ACKNOWLEDGMENTS

The authors thank Doug Spencer for his help on the law content of the chapter.

REFERENCES

[1] Davis v. Bandemer. 478 U.S. 109 (1986).

[2] MIT Election Data and Science Lab. U.S. House 1976–2018, 2017.

[3] MIT Election Data and Science Lab. U.S. President 1976–2016, 2017.

[4] Daily Kos Elections' presidential results by congressional district for 2016, 2012, and 2008. Retrieved from https://www.dailykos.com/stories/2012/11/19/1163009/-Daily-Kos-Elections-presidential-results-by-congressional-district-for-the-2012-2008-elections. Data also linked at https://tinyurl.com/dailykos2018.

[5] Jonathan N. Katz, Gary King, and Elizabeth Rosenblatt. Theoretical foundations and empirical evaluations of partisan fairness in district-based democracies. *American Political Science Review*, 114(1):164–178, 2020.

[6] Andrew Gelman and Gary King. A unified method of evaluating electoral systems and redistricting plans. *American Journal of Political Science*, pages 514–554, 1994.

[7] Analysis of the Efficiency Gaps of Wisconsin's Current Legislative District Plan and Plaintiffs' Demonstration Plan. Kenneth R. Mayer, at 46 tbl.10, Whitford v. Gill, 218 F. Supp. 3d 837 (W.D. Wis. 2016) (No. 3:15-cv-00421-bbc), 2015 WL 1009101.

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

[9] U.S. Judge is named to help guide redistricting. *The New York Times*, Feb. 22, 2012. Retrieved from https://www.nytimes.com/2012/02/22/nyregion/federal-judge-is-named-to-help-oversee-redistricting-in-new-york.html.

[10] New Congressional Lines Imposed by Federal Court. *The New York Times*, Mar. 12, 2012. Retrieved from https://www.nytimes.com/2012/03/20/nyregion/judges-impose-new-congressional-map-for-new-york.html.

[11] Gary King and Robert X Browning. Democratic representation and partisan bias in congressional elections. *American Political Science Review*, 81(4):1251–1273, 1987.

[12] James Madison. Notes of Debates in the Federal Convention of 1787, 9 June, 1787.

[13] John Adams. Thoughts on Government, Applicable to the Present State of the American Colonies. In a Letter from a Gentleman to His Friend. Philadelphia: John Dunlap, 1776. Retrieved from `https://www.masshist.org/publications/adams-papers/index.php/view/PJA04dg2`.

[14] John S. Mill. Considerations on Representative Government. London: Parker, Son, Bourn. 1861. Retrieved from `https://www.gutenberg.org/files/5669/5669-h/5669-h.htm`.

[15] Edward R. Tufte. The relationship between seats and votes in two-party systems. *American Political Science Review*, 67(2):540–554, 1973.

[16] Brief for Professors Gary King, Bernard Grofman, Andrew Gelman, and Jonathan N. Katz as Amici Curiae in Support of Neither Party. League of United Latin American Citizens v. Perry, 548 U.S. 399 (2006) (no. 05-204). Retrieved from https://gking.harvard.edu/publications/brief-amici-curiae-professors-gary-king-bernard-grofman-andrew-gelman-and.

[17] Gary King. Representation through legislative redistricting: A stochastic model. *American Journal of Political Science*, pages 787–824, 1989.

[18] Bernard Grofman. Measures of bias and proportionality in seats-votes relationships. *Political Methodology*, 9(3):295–327, 1983.

[19] Daryl DeFord, Natasha Dhamankar, Moon Duchin, Varun Gupta, Mackenzie McPike, Gabe Schoenbach, and Ki Wan Sim. Implementing partisan symmetry: Problems and paradoxes. *Political Analysis*, to appear, 2021.

[20] Bernard Grofman and Gary King. The future of partisan symmetry as a judicial test for partisan gerrymandering after *LULAC v. Perry*. *Election Law Journal*, 6(1):2–35, 2007.

[21] Vieth v. Jubelirer. 541 U.S. 267 (2004) (Kennedy A. concurring opinion).

[22] League of United Latin American Citizens v. Perry. 548 U.S. 399 (2006).

[23] John F. Nagle. Measures of partisan bias for legislating fair elections. *Election Law Journal*, 14(4):346–360, 2015.

[24] Michael D. McDonald and Robin E. Best. Unfair partisan gerrymanders in politics and law: A diagnostic applied to six cases. *Election Law Journal*, 14(4):312–330, 2015.

[25] Brief for Heather K. Gerken, Jonathan N. Katz, Gary King, Larry J. Sabato, and Samuel S.-H. Wang as Amici Curiae in Support of Appellees. Gill v. Whitford, 585 U.S. ___ (2018) (no. 16-1161). Retrieved from https://gking.harvard.edu/publications/brief-heather-k-gerken-jonathan-n-katz-gary-king-larry-j-sabato-and-samuel-s-h.

[26] Nicholas O. Stephanopoulos and Eric M. McGhee. Partisan gerrymandering and the efficiency gap. *University of Chicago Law Review*, 82:831, 2015.

[27] Moon Duchin. Gerrymandering metrics: How to measure? what's the baseline? *American Academy of Arts and Sciences*, Retrieved from https://www.amacad.org/news/redistricting-and-representation.

[28] Kristopher Tapp. Measuring political gerrymandering. *The American Mathematical Monthly*, 126(7):593–609, 2019.

[29] Mira Bernstein and Moon Duchin. A formula goes to court: Partisan gerrymandering and the efficiency gap. *Notices of the AMS*, 64(9):1020–1024, 2017.

[30] Ellen Veomett. Efficiency gap, voter turnout, and the efficiency principle. *Election Law Journal: Rules, Politics, and Policy*, 17(4):249–263, 2018.

[31] Benjamin Plener Cover. Quantifying partisan gerrymandering: An evaluation of the efficiency gap proposal. *Stanford Law Review*, 70:1131, 2018.

[32] Christopher P. Chambers, Alan D. Miller, and Joel Sobel. Flaws in the efficiency gap. *Journal of Law and Politics*, 33:1, 2017.

[33] Jan E. Leighley and Jonathan Nagler. *Who votes now?: Demographics, issues, inequality, and turnout in the United States*. Princeton University Press, 2013.

[34] Rucho v. Common Cause. 588 U.S. ___ (2019).

[35] Common Cause v. Rucho. 318 F. Supp. 3d 777, 807–808 (MDNC 2018).

[36] Gregory Herschlag, Han Sung Kang, Justin Luo, Christy Vaughn Graves, Sachet Bangia, Robert Ravier, and Jonathan C. Mattingly. Quantifying Gerrymandering in North Carolina. *Statistics and Public Policy*, 7(1):30–38, 2020.

[37] Jowei Chen. Unintentional Gerrymandering: Political Geography and Electoral Bias in Legislatures. *Quarterly Journal of Political Science*, 8(3):239–269, June 2013.

[38] Daryl DeFord and Moon Duchin. Redistricting reform in Virginia: Districting criteria in context. *Virginia Policy Review*, 12(2):120–146, 2019.

[39] Rucho v. Common Cause. 588 U.S. ___ (2019) (Kagan E. dissenting opinion).

[40] Arizona Const. Art. IV, pt II, §1, cl. 14F.

[41] Ohio Const. Art. XI, §06.