Base Rate Neglect and the Diagnosis of Partisan Gerrymanders*

Sanford C. Gordon† and Sidak Yntiso‡

November 2023

Abstract

We discuss conceptual issues arising from base rate neglect in the "Efficiency Gap" (Stephanopoulos and McGhee, 2015), a popular measure of unfairness in partisan gerrymandering. A measure of vote dilution asymmetry that adjusts for the base rate mitigates these; however, because it prioritizes a potentially unachievable degree of proportionality, its naive application may be impractical or infeasible. A comparison of measures using enacted and simulated plans clarifies the frequency with which the base rate issue may yield erroneous inferences if unaddressed. We show that a generalized version of dilution asymmetry that differentially weights harms associated with partisan gerrymandering permits and may even necessitate departures from pure proportionality. We also show via simulation that maps that approximate proportionality, and thus minimize dilution asymmetry, are feasible more often than previously thought.

Running Title: Base Rate Neglect and Diagnosis of Gerrymanders

Keywords: partisan gerrymandering; redistricting; efficiency gap; vote dilution

Word Count: 8,653

*For helpful suggestions, we gratefully acknowledge Hans Hassell, Doug Spencer, and seminar participants at the 16th Annual Conference on Empirical Legal Studies, the annual meeting of the Southern Political Science Association, New York University and the University of Chicago; and three anonymous referees.

†Department of Politics, New York University. 19 West 4th Street, 2nd Floor. New York, New York, 10012. Email: sanford.gordon@nyu.edu

‡Corresponding author. Department of Political Science, University of Rochester. Harkness Hall, 333 Hutchison Road, Rochester, NY 14627. Email: syntiso@ur.rochester.edu

1 Introduction

We consider some practical and jursprudential implications of base rate neglect in empirical assessments of vote dilution stemming from partisan gerrymandering. A common metric of partisan (un)fairness, the Efficiency Gap (McGhee, 2014; Stephanopoulos and McGhee, 2015), captures inter-party disparities in the empirical compound probability that a voter's vote is "wasted" ("cracked" into persistent minorities or "packed" inefficiently into a small number of districts) *and* that they belong to a specific political party. We argue in favor of a class of disparity metrics – which we call *dilution asymmetry* measures – derived from the probability a voter's vote is wasted *given that* the voter belongs to that party (cf., McGann et al., 2015; Nagle, 2017). Measures of unfairness derived from the compound probability fail to account for overall partisan imbalances in a state, whereas those based on the conditional probability do not.

Using enacted and simulated maps and actual and imputed vote returns, we show that correcting for base rate alters not only the magnitude of estimates of partisan disparity but also, frequently, the sign. We also conduct a longitudinal analysis to examine the evolution of the effects of partisan gerrymandering nationwide. An analysis using the Efficiency Gap would suggest that the practice evolved from one in which Republican and Democratic gerrymanders tended to generate little disparate treatment to one in which the preponderance of gerrymanders – including those in blue states – benefited Republicans. An analysis using our preferred Dilution Asymmetry metric, by contrast, suggests that partisan disparities benefiting Republican voters have increased over time in red states, whereas disparities benefiting Democratic voters, though somewhat muted in the early 2000s, have been a fairly consistent feature of blue states.

The Supreme Court's 2019 decision in *Rucho vs. Common Cause* would seem to foreclose federal judicial remedies for even the most egregious instances of partisan gerrymandering. But any conclusion that this obviates the need for continued refinement and analysis of measures of the extent and impact of partisan gerrymandering is misplaced: as we discuss below, the issue is very much alive in the states. Moreover, insofar as the majority in *Rucho* rejects the justiciability of partisan gerrymandering cases, it does so primarily by rejecting a constitutional right to fairness toward *parties*, and only secondarily toward *voters*. To sway a sympathetic state court, or a future, more favorable federal court, e.g., on equal protection grounds, it will likely be essential to place voters rather than parties at the center of the argument. A measure of asymmetric vote dilution that properly adjusts for the base rate does precisely this.

2 Background

2.1 The Legal Status of Partisan Disparity Claims

Writing for a 5-4 conservative majority in *Rucho vs. Common Cause* (2019), Chief Justice Roberts wrote that claims of partisan gerrymandering were non-justiciable under the U.S. Constitution, holding that “The Constitution supplies no objective measure for assessing whether a districting map treats a political party fairly” (p. 20). In criticizing the view expressed in Justice Kagan’s dissent that manageable standards were in fact feasible, Roberts also wrote, “The Framers would have been amazed at a constitutional theory that guarantees a certain degree of representation to political parties” (p. 21). While the court had heard cases pertaining to partisan gerrymandering since the early 1970s, *Rucho* represented the culmination of a process that began in 2004 with *Vieth v. Jubelirer*, when Justice Scalia, in a plurality opinion, argued that partisan gerrymanders should not be justiciable owing to the absence of a discernible and manageable standard for adjudicating partisan gerrymandering claims. In a concurring opinion, Justice Kennedy did not rule out the possibility that such a standard might ultimately be articulated. Then, in *Gill v. Whitford* (2018), the Court took up Wisconsin’s 2011 redistricting plan for that state’s legislative assembly, following a a 2-1 ruling by a three-judge panel in the Western District of Wisconsin that the plan violated the 14th amendment’s equal protection clause. The Supreme Court remanded the case following a unanimous determination that the plaintiff had failed to demonstrate standing, but Chief Justice Roberts’ previewed the majority’s determination in *Rucho*, insisting, “[T]his Court is not responsible for vindicating generalized partisan preferences” (p. 21).

For our purposes, the Court’s determination in *Rucho* that there was no discernible and manageable standard for adjudicating constitutional claims concerning partisan gerrymandering is an important factor in considering the utility of diagnostic measures thereof. Stephanopoulos and

McGhee (2015), for example, in answering the implicit challenge in Kennedy's *Vieth* concurrence, suggest a threshold value of the efficiency gap that would constitute evidence for constitutional impermissibility. An additional point to which we return below is Roberts' reference to Justice O'Connor's argument in *Davis v. Bandemer* (1986), which raises concern about "a conviction that, the greater the departure from proportionality, the more suspect an apportionment plan becomes." The Court has repeatedly ruled that the Constitution does not oblige states to allocate seats with an eye toward proportionality.

While the prospects of the federal courts rejecting explicitly partisan gerrymanders on constitutional grounds are slim, this does not rule out such considerations in the state courts. In *League of Women Voters v. Commonwealth of Pennsylvania* (2018), the Pennsylvania Supreme Court ruled that the state's redistricting plan violated Article I, Section 5 of the Pennsylvania Constitution, which requires that "Elections shall be free and equal." In 2022, the high courts in New York (in *Harkenrider v. Hochul*), Kansas (in *Alonzo v. Schwab*), North Carolina (in *Harper v. Hall*), and Ohio (in *League of Women Voters of Ohio et al. v. Ohio Redistricting Commission* and subsequent cases) overturned districting maps as inconsistent with their state constitutions.

A closely related additional area of development is state law and state constitutional amendments. To reach its decision in the *Harkenrider* case, the New York Court of Appeals relied on a 2015 amendment to the New York Constitution, which states that "Districts shall not be drawn to discourage competition or for the purpose of favoring or disfavoring incumbents or other particular candidates or political parties." The Ohio Supreme Court relied on a 2015 amendment to that state's constitution stating that "No general assembly district plan shall be drawn primarily to favor or disfavor a political party," and another stating that the "statewide proportion of districts ... shall correspond closely to the statewide preferences of the voters of Ohio." In 2018, voters in Colorado, Michigan, Missouri, Ohio, and Utah approved ballot initiatives addressing the partisan composition of legislative districts in response to the *Gill* decision.[1]

[1] For a comprehensive overview, see Professor Doug Spencer's "All About Redistricting" website at `https://redistricting.lls.edu/`.

2.2 The Efficiency Gap

The Efficiency Gap has featured prominently in recent scholarship (Stephanopoulos and McGhee, 2015; Caughey, Tausanovitch, and Warshaw, 2017; Cover, 2018; Jiang et al., 2020), litigation (including the cases cited above), and press coverage. The measure depends upon the distribution of a party's wasted votes, i.e., all votes cast in favor of the losing candidate or votes in excess of 50%+1 for the winning candidate. Formally, let $W_{s,t}^{p}$ be the number of wasted votes for party p in state s at time t, and $V_{s,t}^{p}$ the total number of votes for that party, with $p \in \{D, R\}$. Then the Efficiency Gap in state s, $(EG_{s,t})$ is defined as the difference in the number of individuals wasted by each major party divided by the total number of votes cast for the two major parties:

$$EG_{s,t} \equiv \frac{W_{s,t}^{D} - W_{s,t}^{R}}{V_{s,t}^{D} + V_{s,t}^{R}}. \tag{1}$$

At first blush, EG has several features that recommend it. Relative to previous partisan symmetry approaches, which depend upon counterfactual electoral shifts, EG can be calculated from real-world election results. The metric can be plainly interpreted as the party's surplus seat share over the seat share that would arise if both parties wasted an equal number of voters.

McGhee (2014) argues that another favorable feature of the efficiency gap is that it adheres to an *efficiency principle*: that any measure of efficiency must increase monotonically with seat share holding vote share constant.

In the years since it was first proposed, the efficiency gap has become a common diagnostic measure of partisan gerrymandering, extensively relied on by expert witnesses in gerrymandering cases and cited approvingly by the state supreme court majorities in Kansas and North Carolina, as well as the federal district court in *Gill*. A 2018 ballot initiative amended Michigan's constitution to require that the state's nonpartisan redistricting commission minimize disproportionate partisan advantage "using accepted measures of partisan fairness"; in subsequent litigation in that state, expert testimony has included analyses using the efficiency gap. Utah's Proposition 4 (2018) directed the legislature and redistricting commission to employ "best available data and scientific and statistical methods, including measures of partisan symmetry" to assess whether a proposed map conforms to a set of standards that includes partisan fairness. And while it does not use the term "Efficiency Gap," Missouri's Amendment 1 (2018) explicitly added calculation and consideration of that measure into the state's constitution.[2]

3 Conceptual Issues with Efficiency as a Criterion

3.1 Extant Critiques of the Efficiency Gap Metric

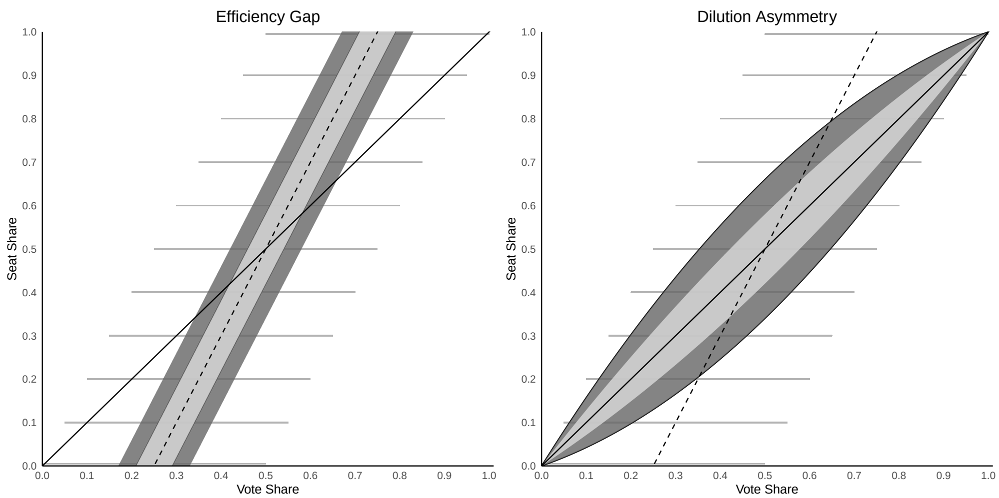
The Efficiency Gap has been the subject of a number of criticisms. First, the measure fails to differentiate between partisan inefficiencies that arise from the enacted plan and those inefficiencies that would be expected given the spatial distribution of voters, an issue to which we also return below. Second, EG fails to distinguish between types of wasted votes (Bernstein and Duchin, 2017), something we also discuss below.

A third concern emerges from the fact that EG does not just tolerate, but may actually *require* departures from a 1:1 proportionality between voteshare and seatshare in a state (Duchin, 2018). The left panel of Figure 1 illustrates this by plotting the seat share of one party against its vote share in a hypothetical allocation of a 10-seat congressional delegation and assumed equal turnout across districts. The solid 45° line denotes perfect proportionality. The horizontal line segments denote the logical bounds on vote share associated with a particular seat share: so, for example, a party with two seats out of ten could theoretically have received anywhere between 10% and 60% of the statewide vote share.

The pale gray region in the left panel denotes the set of (vote share, seat share) combinations consistent with a standard of $|EG| < 0.08$, a standard proposed by Stephanopoulos and McGhee (2015) (for state legislatures). The darker region (inclusive of the light one) doubles the set of permissible values for EG. As is evident from the figure, outside of a range of vote shares in the neighborhood of 50%, a standard based on EG would be presumptively invalid. To understand why this is, suppose turnout is equal in each district. Let $\sigma_{s,t}$ be the share of seats held by the

[2] Specifically, Section 3(c)(1)(b) instructs that a non-partisan state demographer calculate wasted votes based on an index of electoral performance, and mandating that the demographer "shall ensure the difference between the two parties' total wasted votes, divided by the total votes cast for the two parties, is as close to zero as practicable."

Figure 1: Features of Efficiency Gap and Dilution Asymmetry Standards (10-Seat Delegation)



In each panel, the horizontal line segments denote the range of vote shares that could theoretically give rise to a particular seat share. The solid $45°$ *line corresponds to perfect proportionality of votes to seats; the dashed line corresponds to double proportionality. Pale gray regions denote (seat share, vote share) combinations consistent with* $|EG| < 0.08$ *and* $|DA^P| < 0.16$*. Dark gray regions (inclusive of lighter regions) denote combinations consistent with* $|EG| < 0.16$ *and* $|DA^P| < 0.32$*.*

Democratic Party in state s at time t, and $\nu_{s,t}$ the Democratic share of the two-party vote. Then, as shown by McGhee (2014), the efficiency gap given equal turnout across districts (suppressing the s and t subscripts for clarity) may be expressed as

$$EG = 2\nu - \sigma - \frac{1}{2}. \tag{2}$$

For any fixed value of $EG_{s,t}$, the implied seat-vote relationship has a slope of 2. This "double-proportionality" is depicted by the dashed line in the figure. 1:1 proportionality minimizes the absolute value of the Efficiency Gap if and only if $\sigma_{s,t} = \nu_{s,t} = \frac{1}{2}$, i.e., if the state's voting population is perfectly evenly divided between Republicans and Democrats (and the solid and dashed lines in the figure intersect).

It has long been observed (e.g., Tufte, 1973) that *empirically*, the slope obtained from a regression of seats on votes is closer to 2 than 1 (and may even exceed 2). As Warrington (2019) notes, the realized slope will tend to depend on the geographic distribution of partisanship within a state. More recently, Barton (2022) points out that electoral systems with single member districts and plurality rule are inherently biased in favor of the majority party, rendering proportionality frequently infeasible more generally. That being said, it would seem peculiar that a diagnostic approach would rule out perfect proportionality as inconsistent with a norm of fairness except in knife's edge cases. Put another way, the repeated holding by the federal courts that the constitution does not require proportionality is surely not equivalent to a *prohibition* thereof.[3]

3.2 Consequences of Base Rate Neglect and a Simple Correction

Unless a state is close to evenly split between Democrats and Republicans, meeting the constitutional standard envisioned by Stephanopoulos and McGhee (2015) may necessitate departing from perfect proportionality to pass muster, even if a proportional plan were feasible. We trace this issue to the fact that EG as a measure of disparity suffers from *base rate neglect*: a failure to account for the overall partisan composition of voters in a state. To illustrate the problem, consider the following example from the study of discrimination. Suppose there are one hundred motorists on the highway: 80 white, and 20 black. The police pull over three white and three black motorists. Clearly, the black motorists experience a disproportionately higher likelihood of being stopped: the rate at which a motorist is stopped given that they are Black is $3/20 = 15\%$, compared to the analogous quantity for white motorists of $3/80 = 3.75\%$. One summary measure of the disparity is the difference between these two quantities: $15 - 3.75 = 11.25$ percentage points.[4] But a measure of disparity equivalent to EG, $\frac{3-3}{20+80}$, would yield a difference of zero percentage points. What gives?

Consider the expression for the efficiency gap in equation (1), and let $V_{s,t} = V^D_{s,t} + V^R_{s,t}$. Then (1) is equivalent to

$$\frac{W^D_{s,t}}{V_{s,t}} - \frac{W^R_{s,t}}{V_{s,t}}.$$

The fraction $\frac{W^p}{V}$ is the empirical probability that a vote is wasted *and* that it was in favor of party p.[5] The efficiency gap is the difference between these estimated compound probabilities for

[3] Moreover, insofar as *Rucho* renders partisan gerrymandering cases nonjusticiable in federal courts, determining the extent to which deviations from proportionality are permissible becomes the province of state courts, legislatures, districting commissions, and (via the initiative process) electorates.

[4] Another would be the ratio: $15/3.75 =$ a $4\times$ greater risk of being stopped.

[5] By definition, the probability of two events A and B happening equals the conditional probability of A occurring

the two parties. It is immediate that equalizing W^D and W^R minimizes $|EG|$. However, just as equalizing the number of black and white motorists who are stopped does not equalize the expected cost imposed on the motorists in our example, equalizing the number of wasted Democratic and Republican (or, for that matter, black and white or urban and rural) voters does not equalize the expected risk of vote dilution between those populations when their numbers differ.[6]

Rather, we are interested in the difference in the empirical probability that a vote is diluted/wasted *given that* it was for one party and the same quantity for the other. This yields an appropriate measure of partisan *Dilution Asymmetry* at time t in state s, ($DA^P_{s,t}$):

$$DA^P_{s,t} \equiv \frac{W^D_{s,t}}{V^D_{s,t}} - \frac{W^R_{s,t}}{V^R_{s,t}}. \tag{3}$$

The superscript P is used to denote the fact that partisanship is just one dimension on which disparities may exist in the dilution of votes.

While our rationale is novel , we wish to be clear that we are not the first to propose this metric: McGann et al. (2015) argue that comparing party's shares of wasted votes is more appropriate than comparing absolute numbers. Nagle (2017) develops a measure of the difference in shares, labeling it a "voter-centric" approach in contrast to the "party-centric" approach of Stephanopoulos and McGhee (2015). McGhee (2017) terms the measure EG^{VC} (for "Efficiency Gap, Voter-Centric"), while Tapp (2019) calls the measure the "relative efficiency gap" (REG). Nagle, perhaps anticipating the court's argument in *Gill*, justifies the measure with reference to the normative goal of equalizing the average effectiveness of like-minded voters (p. 201), with Cover (2018) citing a similar justification.[7] We build on these arguments by exploring empirical properties of the measure and by tying differences between observed and simulated values of EG and DA to, and isolating the origins of counterintuitive values of EG in, base rate neglect. A focus on the normative founda-

given B multiplied by the probability of B occurring i.e., $\Pr(A \wedge B) = \Pr(A|B)\Pr(B)$. So, $\Pr(\text{vote for } p \text{ and wasted}) = \Pr(\text{vote wasted}|\text{vote for } p)\Pr(\text{vote for } p)$, which is estimated as $\left(\frac{W^p}{V^p}\right)\left(\frac{V^p}{V}\right) = \frac{W^p}{V}$.

[6] See also footnote 14 in Judge Griesbach's dissent in *Whitford vs. Gill*, 218 F. Supp. 3d 837 (W.D. Wis. 2016).

[7] As Stephanopoulos and McGhee (2018, p. 1520) demonstrate via simulation, unless the proportion of votes going to a third party candidate is quite large (20% or more), this measure also satisfies their "efficiency principle" – that is, the notion that holding a party's statewide voteshare constant, an increase in that party's seatshare should be reflected in greater advantage for that party as captured by the measure.

tions of the measure initially identified by these previous scholars also helps us clarify and explore differences between the impact of packing and cracking on voters, which we do below.

Our preference for the term Dilution Asymmetry over terminology employing "efficiency" stems from the reason hinted at by Nagle: efficiency refers to an optimization problem for *parties*, rather than a harm imposed on *voters.* For similar reasons, we prefer "dilution" to "waste" – waste is something that a (partisan) gerrymanderer seeks to minimize by efficiently allocating voters to districts, whereas dilution is a harm imposed on a voter (or group of voters).[8]

The issue of harm to parties vs. harm to voters is invoked in the jurisprudence on partisan gerrymandering. For example, in her dissent in *Rucho*, Justice Kagan (favorably citing Kennedy's concurrence in *Veith*) bases her defense of party-based standards on two interpretations of the rights protected by the first amendment. One would seem to pertain to parties as corporate entities with associational rights (to "band together in support of candidates whose espouse their political views") while the other explicitly invokes the possibility of partisan gerrymanders subjecting certain voters to "disfavored treatment" (12). And in *League of Women Voters of Pennsylvania v. Commonwealth*, the Pennsylvania Supreme Court held that the state's constitution mandates that the electoral process be conducted "in a manner which guarantees, to the greatest extent possible, a voter's right to equal participation in the electoral process for the selection of his or her representatives in government," and "all voters have an equal opportunity to translate their votes into representation" (804).

While the normative justification for the waste disparity measure is attractive, equally compelling is the fact that it reestablishes proportionality as permissible for any base rate. To see why, again let $\sigma_{s,t}$ be the share of seats held by the Democratic Party in state s at time t, and $\nu_{s,t}$ the Democratic share of the two-party vote. Suppose the average turnout in districts won by Republicans equals the average turnout in districts won by Democrats. Suppressing state and time

[8] Cf., Kagan's dissent in *Rucho*: "Partisan gerrymandering operates through vote dilution—the devaluation of one citizen's vote *as compared to others*" (11, emphasis ours). Moreover, vote dilution is a common legal term of art in a range of cases beyond partisan gerrymandering cases, including one-person-one-vote and racial gerrymandering cases.

subscripts, it is straightforward to demonstrate that, DA^P may be expressed as

$$DA^P = \frac{\nu - \sigma}{2\nu(1-\nu)}, \tag{4}$$

so $|DA^P|$ is minimized when the seat share equals the vote share.

The right panel of Figure 1 reproduces the left panel using Dilution Asymmetry instead of Efficiency Gap. The lens-shaped regions surrounding the 45° line indicate absolute values of Dilution Asymmetry consistent with a hypothetical standard. For example, any combination of voteshare and seatshare that falls in the light gray region has an absolute Dilution Asymmetry less than 0.16, while any combination that falls in the darker lens (inclusive of the light one) is consistent with an absolute vote dilution less than 0.32. (Efficiency Gap ranges from -0.5 to 0.5, whereas Dilution Asymmetry ranges from -1 to 1; hence, for direct comparability, it is helpful to compare DA^P thresholds equal to twice those of the EG thresholds used in the left panel of the figure.) As is immediate from the figure, and in contrast to the corresponding figure above for EG, proportionality is preserved as permissible for any base rate vote share and threshold level of permissible absolute dilution asymmetry.

To get a sense of how base rate neglect in the Efficiency Gap can lead to strange conclusions that Dilution Asymmetry avoids, consider a hypothetical state consisting of 100,000 voters divided evenly across ten districts. (For purposes of exposition, we set aside geographic constraints for the moment.) 70,000 always vote Republican and 30,000 always vote Democratic. Consider the following hypothetical plans:

- Seven (perfectly uncompetitive) districts with 10,000 Republicans each, and three with 10,000 Democrats each; and
- All ten districts have 7,000 Republicans and 3,000 Democrats each.

The first scenario yields an Efficiency Gap of $\frac{3\times 4{,}999 - 7\times 4{,}999}{100{,}000} \approx -0.20$, indicative of substantial unfairness *against Republicans*. The second yields an EG equal to $\frac{10\times 3{,}000 - 7\times 1{,}999}{100{,}000} \approx 0.1$, indicative of bias against Democrats, but *lower* bias than that measured against Republicans in the first scenario. In other words, the efficiency gap implies that proportionality harms the majority *more* than winner-take-all harms the minority.

To get a sense of how base rate neglect in the Efficiency Gap can lead to strange conclusions that Dilution Asymmetry avoids, consider a hypothetical state consisting of 100,000 voters divided evenly across ten districts. 70,000 always vote Republican and 30,000 always vote Democratic. In the absence of any geographic constraints, the electorate can be partitioned to create anywhere from zero to five districts with Democratic majorities (six if ties went to the Democrat). Now consider the following hypothetical plans:

1. Seven districts with 10,000 Republicans each, and three with 10,000 Democrats each; and

2. All ten districts have 7,000 Republicans and 3,000 Democrats each.

In the first scenario, each district is maximally uncompetitive, but the partition produces a *proportional allocation of seats* (70-30) in the state's delegation. The Efficiency Gap for Scenario 1 is $\frac{3\times 4{,}999-7\times 4{,}999}{100{,}000} \approx -0.20$, indicative of substantial unfairness *against Republicans.* The second scenario is a *perfect Republican gerrymander–* it allows the majority party to capture 100% of the seats. In Scenario 2, EG is equal to $\frac{10\times 3{,}000-7\times 1{,}999}{100{,}000} \approx 0.1$, indicative of bias against Democrats, but *lower* bias than that measured against Republicans in the first scenario. In other words, for our hypothetical state, the efficiency gap implies that proportionality harms the majority *more* than winner-take-all harms the minority.

What is driving large negative values of the Efficiency Gap measure in the first example isn't actually disparate harm to Republicans, but rather that there are simply more Republicans available to waste than Democrats. Adjusting for the base rate corrects for this issue: in the map with seven Republican seats, the Dilution Asymmetry is zero (no bias), whereas in the map with ten it is 0.71, reflecting the much higher likelihood that Democratic votes are diluted.

Stephanopoulos and McGhee (2015, p. 863) note that in a hypothetical state where the majority enjoys a statewide voteshare of greater than 75%, EG will indicate a bias against that majority even if it holds all of the seats. They acknowledge this feature as a limitation of the measure, while noting that the occurrence of such lopsided statewide majorities is exceedingly rare. The example above adds two clarifications: first, the origins of this property of the measure in neglect of the base rate; and second, that one can construct examples that yield strange inferences with voteshares less lopsided than 75%.

4 Comparisons in Practice

4.1 Severity of Disparities in Enacted Plans

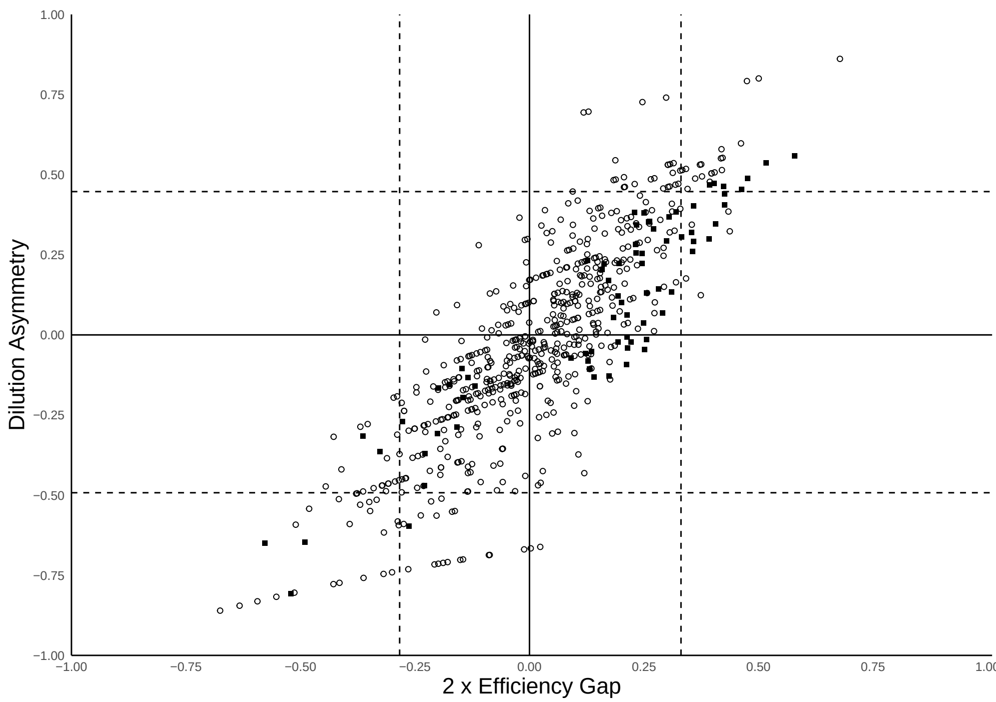
A natural question that arises is the frequency with which analysis using an Efficiency Gap standard would lead the analyst to conclude that a districting plan was impermissibly biased against one party when adjusting for the base rate would lead one to assess that the plan was biased *in its favor*.

To answer this question, we first compare the Efficiency Gap and Dilution Asymmetry calculated using congressional elections returns data from 1976-2020 (MIT Election Data & Science Lab, 2017). We exclude states with fewer than five congressional districts; Louisiana (owing to its runoff system); and California since 2010, and Washington since 2004 (owing to their top-two systems). This leaves 662 state-year observations, which we plot in Figure 2.[9] (Note that the x-axis doubles the efficiency gap so that it appears on the same scale (-1 to 1) as dilution asymmetry.)

We should not be surprised to see that the graph reveals the two measures to be highly correlated ($\rho = 0.84$): this is to be expected given that Democratic voteshare enters positively, and seatshare negatively, into both measures, and that in many cases the base-rate effect is not severe due to

[9] Uncontested elections pose a potential challenge to estimation. Namely, the proportion of wasted votes cannot be determined when only one candidate appears on the ballot. To incorporate uncontested elections into our analysis of wasted votes in congressional elections, we estimate a Bayesian imputation model based on previous elections, as well as district- and election-cycle- factors (e.g., Stephanopoulos and Warshaw, 2020). We estimate the two-party vote share as follows: $V_{it} \sim Binomial(N_{it}, \pi_{it})$ where V_{it} is the vote tally for Democratic candidates in district i at time t and N_{it} is the turnout. V_{it} and N_{it} are set to zero for uncontested elections. The target parameter, $\pi_i \in (0, 1)$, is modeled based on linear predictors: $\pi_{it} = g^{-1}(x_{it}\beta)$ where g^{-1} is the inverse logit, and x_i includes lagged congressional vote share; district-level presidential vote share collected from Jacobson (2015) (for the period 1976-2014) and from Daily Kos (for 2016, 2018, and 2020)); incumbency status (whether an open-seat, Democratic incumbent or Republican incumbent), sourced from Hodges (2016) and online resources; region indicators based on the economic regions defined by the Bureau of Economic Advisors; and election-cycle indicators. We employ weak priors on the intercepts and slopes for each covariate: $\beta \sim N(0, 100)$. We modeled each decennial redistricting period from 1970-2020 separately in *rstan* using four chains, each running 1,000 iterations. The average imputed vote share, π_{it}, in uncontested races won by Democrats (72%) and Republicans (31%) closely resembles estimates from Stephanopoulos and Warshaw (2020)(71% and 31%) and Stephanopoulos and McGhee (2015) (70% and 32%). Our analysis utilizes the imputed vote share for uncontested elections and the observed two-party vote share in contested races.

Figure 2: Efficiency Gap and Dilution Asymmetry for Selected Congressional Delegations, 1976-2020



Solid black squares denote state-year observations that would not survive the two-seat constitutional standard proposed in Stephanopoulos and McGhee (2015). Dashed lines denote average ± 1.5 standard deviations (see text).

statewide competitiveness of both parties. Second, in approximately 20% of state-year observations, the signs of Efficiency Gap and Dilution Asymmetry measures are opposite to one another. The occurrence of some sign reversals is clearly problematic in terms of the inferences one hopes to draw, but it is, again, to be expected. To see why, suppose we plotted seatshare against voteshare for each state-year observation. Any observation that falls between the single- and double-proportionality lines depicted in either panel of Figure 1 will exhibit a sign reversal.[10]

Of course, we may be less interested in the set of *all* cases than we are the set of cases that an Efficiency Gap analysis would conclude are constitutionally suspect. Those cases are depicted as black squares in the figure (using the proposed two-seat threshold in Stephanopoulos and McGhee (2015)). There are 74 cases that would be deemed problematic (57 suspiciously pro-Republican, and 17 suspiciously pro-Democratic). Of these, 14 (19%) – all of which were suspected to be pro-Republican – appear to be pro-Democratic once one adjusts for the base rate.[11] Each of these cases corresponds to a large state with a lopsided Democratic margin.

While the foregoing is analytically useful, the legal relevance of these cases is questionable because a case would only materialize on EG grounds if a Democratic majority brought an action against its own plan. A more interesting question arises if an analysis based on EG countenanced a plan that a DA-based approach rejected. In light of the fact that there is no DA equivalent to the two-seat standard suggested for EG, we proceed by adopting an approach that labels outlier plans as impermissible – specifically, plans that generate an Efficiency Gap or Dilution Asymmetry measure 1.5 standard deviations below or above zero.[12] The dashed lines in the scatterplot depict these critical thresholds: for example, an outcome with 2× Efficiency Gap of -0.4 and a dilution asymmetry of -0.25 would be impermissible using the former measure but permissible using the latter.

Our analysis reveals that based on the outlier standard, the lion's share (around 80%) of cases pass muster using either standard. Hearteningly, there are no cases where one standard would

[10] We thank an anonymous referee for making this enlightening point.

[11] The observations are California in 2000, 2006, and 2008; Illinois in 2008; and New York in 1976, 1978, 1986, 1988, 1990, 1996, 2000, 2004, 2016, and 2020.

[12] The 1.5 standard deviation threshold is chosen to assure a comparable fraction of suspect cases to that obtained using the two seat threshold.

conclude a plan is overly pro-Republican while the other would deem it overly pro-Democratic, or vice versa. However, there are numerous cases in which one standard would accept a plan while the other would condemn it: 81 in all. In 36 instances, EG flags the case while DA does not.[13] But in 45 instances, DA^P flags cases as problematic that EG does not.[14] This latter set of cases would be those in which a minority party might have an interest in arguing for the plan's unfairness.

If we were to assume, arguendo, that the DA^P standard represents the *true* level of partisan unfairness. the figures represented in the table would imply that the Efficiency Gap has a false positive rate of 6.4%, and a false negative rate of 45.4%.[15]

4.2 Severity of Disparities in Simulated Redistricting Plans

The previous section discusses the identification of potentially problematic districting plans in the set of enacted plans; however, it could be the case that maps that reduce either measure of disparity are infeasible. In this section, we consider differences in the magnitudes and signs of EG and DA^P across the range of feasible maps (as determined by simulation) in a sample of sixteen states with five or more congressional seats that vary in statewide partisan competitiveness.

For each state, we collected 2020 Presidential elections data and 5,000 simulated redistricting plans from the Algorithm-Assisted Redistricting Methodology (ALARM) Project (McCartan et al., 2022). We compare these simulated plans to the new congressional districts created during the 2021-2022 redistricting process for the coming decade. ALARM's simulated plans are designed to satisfy an equal population requirement and additional redistricting criteria that may vary from state to state. The histogram running down the left side of Figure 3 depicts, for each state, the frequency of simulations that gave rise to a particular seat share for Democrats. The median seat allocation in each state is depicted by a hashmark on the vertical axis. (States are sorted by the absolute statewide vote margin.) So, for example, the median and modal plans for Arizona yielded

[13] AL (1978), AZ (2004), CO (1976, 1980), CT (2004), FL (2006), GA (1996), IA (2004, 2010), IL (2006, 2010), IN (2018), MD (2000), MI (2006, 2012, 2014), NC (1988, 2012, 2016), NJ (2012), OH (2006, 2018), OR (1996, 2000, 2014), PA (2012, 2014, 2016), SC (1980, 2008), TX (1992, 1994), VA (2006, 2012), and WA (1980, 1996)).

[14] AL (2010, 2012, 2016), GA (1976, 1980, 1988), KY (2000, 2004, 2012), MA (1980, 1982, 1984), MD (1980, 2012, 2016), MS (1990, 1992), NJ (2020), NY (2018), OK (2012, 2014, 2016), SC (1976, 2014), TX (1976, 1978), and VA (1980).

[15] The false positive rate is equal to false positives over (false positives plus true negatives), or $36/563 = 6.4\%$. The false negative rate is equal to false negatives over (false negatives plus true positives), or $45/99 = 45.4\%$.

5 of 9 districts with Democratic majorities; only a minuscule fraction generated a map with 3 of 9 Democratic majority districts. (Seat allocations never generated by simulation are assumed infeasible and not shown.)

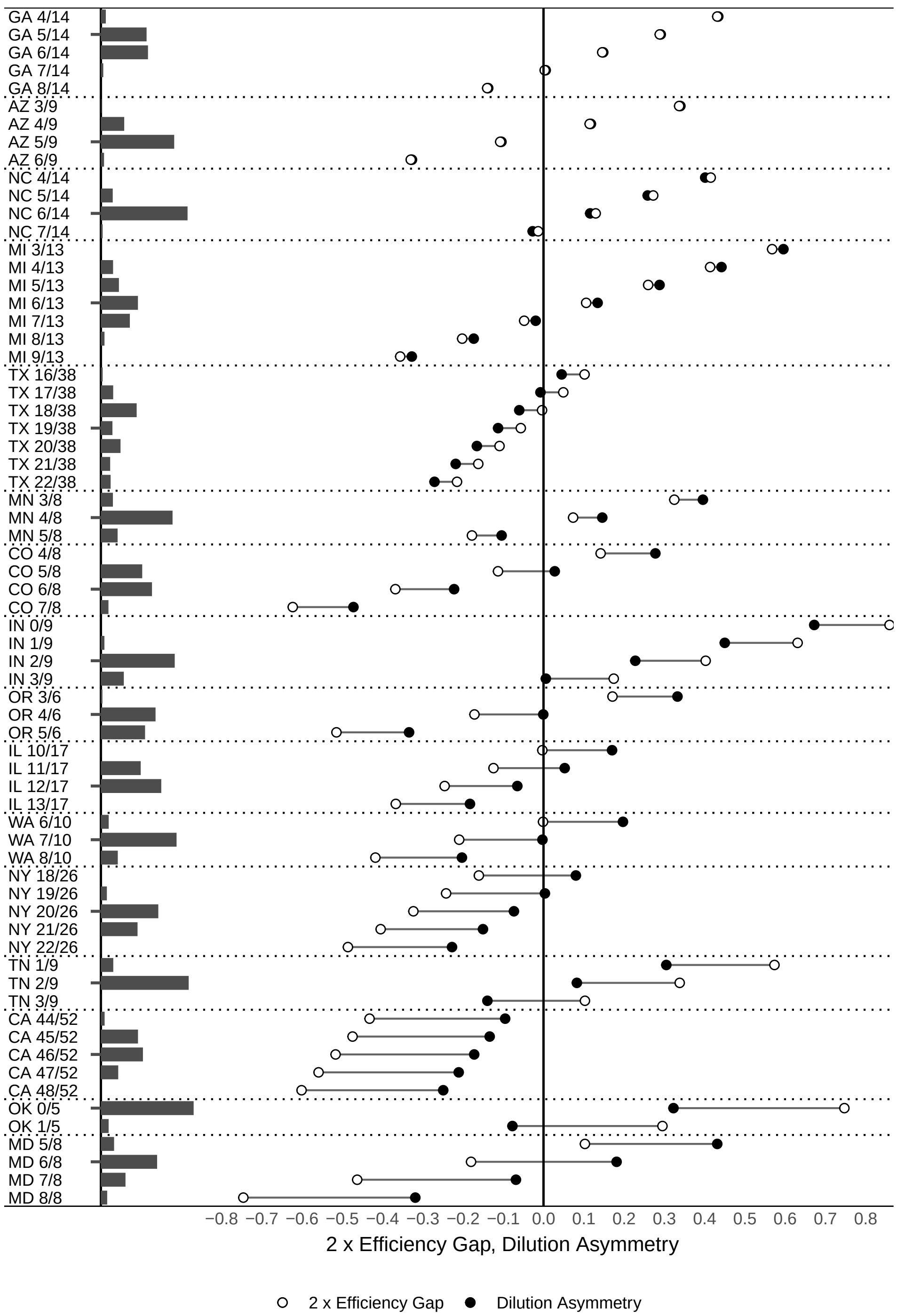
The black circles in Figure 3 depict partisan Dilution Asymmetry calculated using the equal turnout formula in Equation (4), while white circles depict the 2 × the Efficiency Gap using the formula in Equation (2).[16] (For reasons discussed above, EG is multiplied by two to put it on the same scale as DA^P.) First, consider the estimates for Georgia and Arizona. The 2020 presidential winner in these states was determined by less than half a percentage point, so the base rate issue should be negligible no matter what the seat allocation. This is precisely what we see: $2 \times EG$ and DA^P are nearly identical for those states. By comparison, the difference between the two measures for lopsided states like Maryland (33% margin for Biden), California (29% margin) and Oklahoma (33% margin for Trump) is quite substantial, averaging -0.37, -0.34, and 0.39, respectively.

Second, while we can sign the difference between the two measures, with Democratic-leaning states producing $EG - DA^P < 0$ and Republican ones $EG - DA^P > 0$, the substantive import of any difference may depend critically on the seat allocation. This is perhaps best illustrated using the state of Washington, which split approximately 60-40 for Biden in 2020. A feasible strictly proportional delegation would anticipate 6 Democratic seats out of 10, and indeed DA^P falls extremely close to zero for maps generating that anticipated seat allocation. The estimate of $2 \times EG$, by contrast, is approximately 0.2, suggesting an anti-Democratic gerrymander and a base rate induced bias of EG away from zero. By contrast, a map with 7 Democratic seats minimizes the Efficiency Gap at close to zero, but the pro-Democratic shift in seats suggests a negative Dilution Asymmetry – in other words, the base rate bias is toward zero.

Third, conditions do arise in the simulations where adjusting for the base rate flips the sign of the estimate of unfairness. This can be seen most dramatically for a subset of feasible maps in Colorado, Illinois, Maryland, New York, Oklahoma, and Tennessee. For example, nearly 40% of simulated redistricting plans for Illinois suggest a Republican advantage using EG, and a Demo-

[16] We employ the equal turnout formulas for clarity of exposition: different simulations give rise to subtly different turnouts across districts, yielding minute perturbations of both measures using the alternative versions of the measures that simply add noise to the figure.

Figure 3: Efficiency Gap and Partisan Dilution Asymmetry
in a Sample of Simulated Districting Plans



The black circles denote the partisan Dilution Asymmetry; white circles denote the Efficiency Gap. The histograms on the left depict, for each state, the frequency of simulations that allocate a particular seat share for Democrats.

cratic advantage using DA^P. For Maryland, 56% of simulated plans appear to favor the Democrats using EG but Republicans using DA^P. Across all states, the mean absolute difference between $2 \times EG$ and DA^P is 0.1875 in simulated plans that feature a sign reversal.

Note, finally, that situations may arise in which (a) the signs of EG and DA^P are opposite, but for one seat allocation, EG is large in absolute value and DA^P small; while for another seat allocation, EG is small in absolute value and DA^P large. A case in point is the comparison between Illinois with maps that give 11 seats to the Democrats and (much rarer) maps that assign 10.

4.3 The Historical Trajectory of Partisan Disparities

A common perception is that partisan polarization, coupled with sophisticated software for drawing district boundaries, has led to a situation in which partisan gerrymandering has reached levels heretofore unheard of, and that these gerrymanders have tended to favor Republicans. While the motive to gerrymander is indisputable, it is valuable to examine whether the historical data evinces strong trends in the prevalence and/or partisan direction of gerrymandering, and whether different measures might cause us to draw erroneous conclusions.

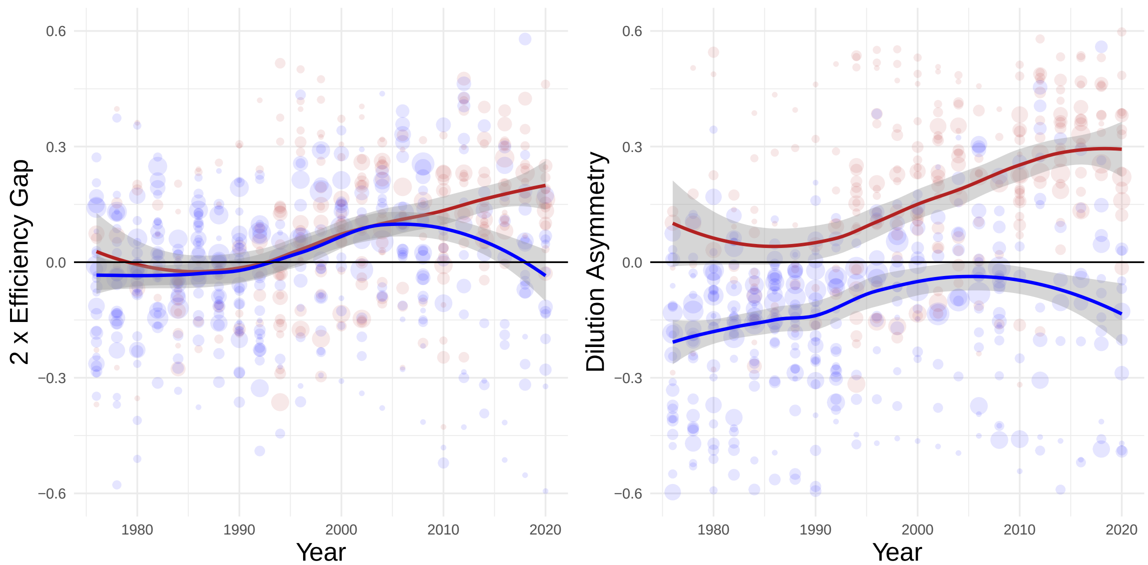
With this in mind, the left panel of Figure 4 plots 2× the Efficiency Gap over time. We partition the data into state-year observations where statewide voteshare in the state leaned Democratic (in blue) and Republican (in red). Tokens correspond to individual states (scaled by the size of their House delegations), with a locally smoothed regression line indicating the overall trend.

Using the Efficiency Gap as our diagnostic, we would conclude that prior to the mid-1990s, the average absolute level of gerrymandering was low for both blue and red states. More recently, there has been an increase in the measure, evincing a clear pro-Republican trend driven primarily by red states, but also (from the mid-2000s to mid-2010s) blue states.[17]

Turning to the second panel, which displays Dilution Asymmetry over time, a somewhat different picture emerges: red states tend to display significant pro-Republican dilution asymmetry over the entire time series. But blue states – which, recall, showed either low efficiency gap or moderately pro-Republican efficiency gap, now demonstrate significant pro-Democratic Dilution Asymmetry

[17] Figure 7 of Jackman (2017), which uses data from 1972-2016 and does not distinguish between red and blue states, displays a similar pattern for the Efficiency Gap.

Figure 4: Evaluating the Historical Incidence of Partisan Gerrymanders



Observations in which statewide voteshare leaned Democratic (Republican) shaded in blue (red); Tokens indicate state-year observations, with larger tokens corresponding to larger House delegations.

over the course of the time series, although the disparity was somewhat muted in the early 2000s. Also notable is the fact that average pro-Democratic Dilution Asymmetry at the beginning of the time series is only slightly lower in magnitude than pro-Republican Dilution Asymmetry today.

5 Reckoning with Proportionality

The foregoing suggests that when an Efficiency Gap analysis yields unintuitive conclusions about the direction of bias in a proposed districting plan, base rate neglect may be the culprit. Our proposed measure of dilution asymmetry adjusts for base rate neglect and thus avoids this issue.

However, a naive application of Dilution Asymmetry raises a different issue: DA^P prioritizes an allocation of seatshare proportional to voteshare, but as noted above, this may be infeasible under plurality rule with single member districts. An illustrative case discussed in Duchin et al. (2019) is Massachusetts: around 35% of voters in that state favored Republican Donald Trump in the 2020 presidential election. But because Republicans aren't geographically concentrated in any specific area of the state, it is essentially impossible to draw a map that creates a district with a Republican majority. Unsurprisingly, Dilution Asymmetry is very large in that state: -0.714 in 2020, so a naive application of the measure would suggest massive anti-Republican bias.

A related issue is highlighted by the very high false negative rate of the Efficiency Gap under the assumption that Dilution Asymmetry is the correct measure. One interpretation is that an EG-based analysis is missing a lot of problematic maps, but another is that DA erroneously flags majoritarian maps that are very much the historical norm.

In this section, we take two approaches to addressing the issue of proportionality. The first is to examine how differentially weighting "cracked" and "packed" voters affects the extent to which minimizing dilution asymmetry requires hewing closely to a potentially unachievable proportionality standard. The second is simulation-based, and involves comparing seat-vote curves calculated for enacted maps and ensembles of simulated maps.

5.1 Differentially Weighting "Cracked" and "Packed" Voters

From a normative perspective it is not altogether clear whether the harm to voters associated with being cracked – that is, being placed in a district where one is a minority with no hope of electing a favored candidate – is equally severe to that of being packed – that is, being placed in a district where one is part of a partisan supermajority (e.g., Bernstein and Duchin, 2017). For this reason, Nagle (2017) proposes differentially weighting minority votes and votes in excess of a majority. He argues that because packing is the "the most obvious way" to create partisan bias, we might wish to weight packed votes more than cracked ones. The normative justification he provides for a weighting scheme that gives higher weight to cracked votes is that "some voters feel happier when their candidate wins big because it enhances their confidence in being on the right side" (p. 200).

Letting $\lambda \in \mathbb{R}^+$ denote the weight on packed votes relative to cracked ones, Tapp (2019) refers to $\lambda = 1, 2$ as "the only natural cases." $\lambda = 1$ weights packed and cracked votes equally (as above), whereas $\lambda = 2$ assigns double the weight to packed votes. Analytically, assigning double the weight to packed votes also addresses a concern expressed by Judge Griesbach in his dissenting opinion in *Gill.* He argues that

> ... in reality, all you need to win an election in a two-candidate race is one more vote than the other candidate, not 50%-plus-one of the total votes. For example, if the Indians defeat the Cubs 8 to 2, any fan might say that the Indians "wasted" 5 runs, because they only needed 3 to win yet scored 8. Under the Plaintiff's theory, however, the Indians needed 5 runs to beat the Cubs that day: 4 runs to reach 50% of the total runs, plus one to win. That, of course, is absurd.[18]

Insofar as the margin in the two-party vote is twice the excess of the majority over 50%+1, $\lambda = 2$ would seem to account for this concern.

From the perspective of the individual voter, a potentially more appealing basis for assessing the normative foundation for weighting packing and cracking differently concerns the principal-agent relationship between the representative and the voter. On the one hand, to the extent that a voter believes her interests will be better represented by a like-minded legislator than an unlike-minded one, being part of a large majority will result in less agency loss between the representative and that voter than being part of a minority. In other words, *packing may mitigate the selection problem that lies at the heart of representation.* This is the approach taken in a recent paper by McCartan and Kenny (2022), who characterize harm as befalling voters not represented by copartisans. Note also that given increasing partisan polarization, the magnitude of loss from having a mismatched legislator is also increasing – the greater the gap in preferences between a voter and her legislator, the less she may feel her interests are effectively represented. Concern with the selection issue mitigates in favor of weighting cracked votes more heavily than packed ones.

On the other hand, favoring the minimization of cracking over packing will tend to reduce the average competitiveness of districts (Tapp, 2019, p. 603). Ceteris paribus, we might anticipate that noncompetitive districts are more prone to moral hazard problems than competitive ones, as greater electoral security could encourage incumbent shirking or corruption.[19] Accordingly, this would mitigate in favor of weighting packing as more damaging than cracking.

With this said, the weighting approaches proposed in Nagle (2017) and Tapp (2019) suffer from

[18] 218 F. Supp. 3d 837 (W.D. Wis. 2016), 958.

[19] This effect might be mitigated in the presence of competitive primaries. However, if closed primaries yield ideologically extreme representatives who often succeed in competitive general elections, then cracked voters may be particularly disadvantaged.

a more basic mathematical issue. Letting L^p denote the total number of votes for party p in districts won by the opposing party, and E^p the total number of votes for party p in excess of 50% plus 1 in districts won by party p, the authors propose a disparity metric based on a weighted measure of dilution risk:

$$\frac{L^p + \lambda E^p}{V^p}, \tag{5}$$

where λ is permitted to take on nonnegative values on the real line. The difficulty with this approach is that it is not a generalization of the conditional probability interpretation of vote dilution. Specifically, the dilution risk for a voter from party p can be decomposed into the sum (or equivalently, twice the average) of $\Pr(packed|p) + \Pr(cracked|p)$. Examining (5), it is clear that for sufficiently high values of λ, the measure will exceed one and thus can no longer be interpreted as a conditional probability. A simple correction is to constrain $\lambda \in [0, 1]$ and define the weighted dilution risk in terms of (twice) the weighted average of the conditional packing and cracking probabilities, yielding the following Weighted Dilution Asymmetry measure:

$$DA^P_{s,t}(\lambda) = 2\left(\frac{\lambda E^D_{s,t} + (1-\lambda)L^D_{s,t}}{V^D_{s,t}} - \frac{\lambda E^R_{s,t} + (1-\lambda)L^R_{s,t}}{V^R_{s,t}}\right) \tag{6}$$

When $\lambda = \frac{1}{2}$, this collapses to the expression in equation (3).

Assuming equal turnout in districts, some tedious algebra allows (6) to be expressed analgously to (4) as follows:

$$DA^P(\lambda) = \frac{\lambda(\nu - \sigma)}{\nu(1-\nu)} + 2(2\lambda - 1)(\theta^D - \theta^R), \tag{7}$$

where θ^p denotes the proportion of party-p voters that reside in districts with party-p majorities. When $\lambda = \frac{1}{2}$, the second term goes to zero and the first collapses to the expression in equation (4).

Recall that the *un*weighted dilution asymmetry measure was minimized at proportionality (vote-share equal seatshare, or $\nu = \sigma$). Using the weighted measure, however, things become considerably more complicated owing to the fact that *how* an individual's vote is wasted now matters, as captured by the difference in θ^p parameters. Moreover, this feature of $DA^P(\lambda)$ reveals that we cannot be sure that the monotonicity implied by McGhee's efficiency principle is preserved for all values of λ: this is because we cannot generically increase σ without considering its effects on the θ^p

parameters.[20]

We will now show that the values of θ^D and θ^R will also affect whether minimizing $DA(\lambda)$ requires proportionality for $\lambda \neq \frac{1}{2}$. We proceed by deriving nonparametric bounds for (θ^D, θ^R) pairs. First, suppose $\nu \leq 0.5$. Then (θ^D, θ^R) is bounded between $(0, 1)$ (corresponding to all districts having an R majority) and $(1, \frac{1-2\nu}{1-\nu})$ (corresponding to a maximally efficient Democratic gerrymander). Next, suppose $\nu > 0.5$. The associated bounds are $(\frac{2\nu-1}{\nu}, 1)$ and $(1, 0)$. Substituting these pairs into the right side of (7), setting equal to zero, and solving for σ results in two piecewise functions that bound a range of seatshares consistent with zero dilution asymmetry for a given voteshare.

Because those functions are unwieldy and unintuitive, for expositional purposes we present the region they define graphically in Figure 5 at two values of λ: $\frac{1}{3}$ (cracking penalized more than packing) and $\frac{2}{3}$ (packing penalized more than packing). Also depicted is the seatshare that achieves $DA^P(\lambda) = 0$ given intermediate values of θ^D and θ^R. This is represented by the thick black line in each panel.

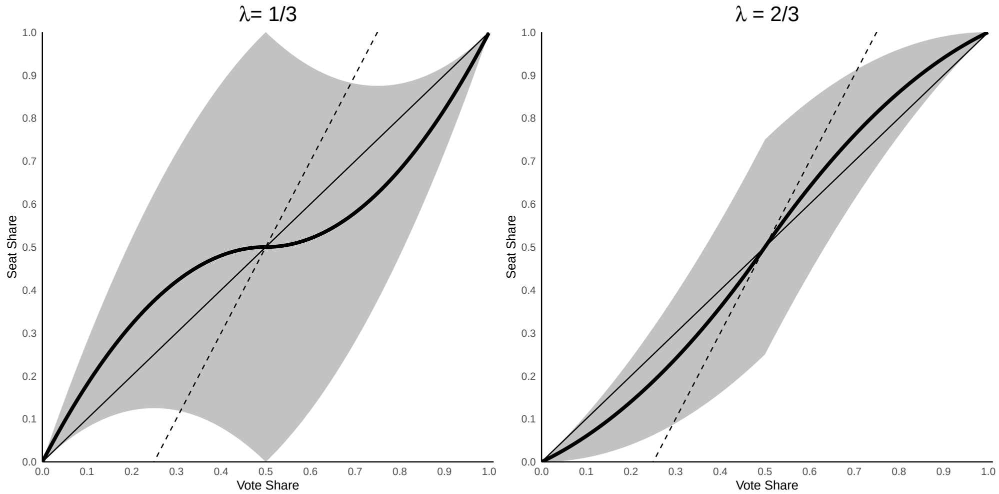
The figure shows that once we relax equal weighting of cracking and packing ($\lambda \neq \frac{1}{2}$), proportionality is neither necessary nor sufficient for minimizing dilution asymmetry. Indeed, depending on the configuration of voters, a proportional allocation may yield substantial disparate impact when one that departs from proportionality may imply none. This is captured by the intermediate curve depicted in each figure: when $\lambda = \frac{2}{3}$, this seat-to-vote curve is significantly steeper than strict proportionality for a broad range of values of voteshare.

More generally, assessing the conditions under which deviations from proportionality are appropriate for a given weighting of harms from packing and cracking requires more information than that contained in statewide seat- and voteshare alone: it also requires accounting for the fraction of voters from each party that may be expected to be represented by a copartisan.[21] Neither efficiency gap nor unweighted dilution asymmetry takes these quantities into account.

[20] That being said, whether or not the efficiency principle is a desirable normative property may itself be contestable: if we differentially weight packing and cracking, situations may emerge in which we can improve the lot of voters from one party while lowering that party's seat share.

[21] Equivalently, assessing appropriate deviations from proportionality requires evaluating average district competitiveness, which is inversely correlated to the fraction of voters from each party represented by a co-partisan.

Figure 5: Nonparametric bounds on Dilution Asymmetry-Minimizing Seatshares



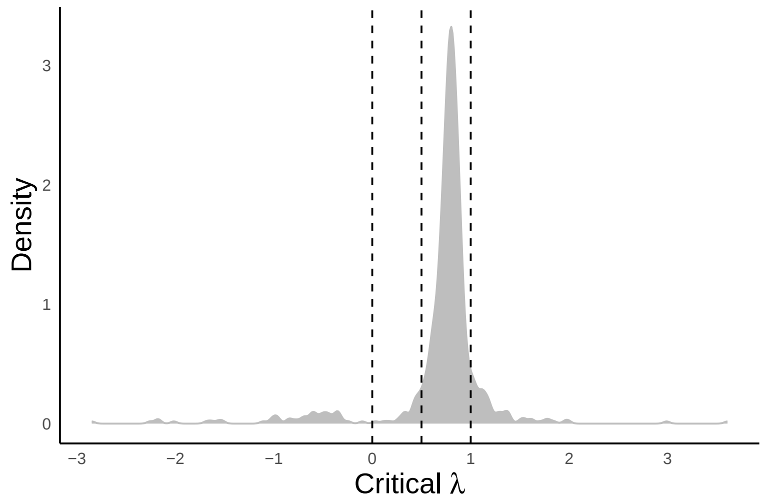
Gray regions represent seatshares that achieve $DA^P(\lambda) = 0$ given voteshare and some feasible (θ^D, θ^R) pair. The thick black curve denotes the seatshare that achieves $DA^P(\lambda) = 0$ for an intermediate feasible (θ^D, θ^R) pair. Dashed lines denote proportionality and double-proportionality. See text for discussion.

Another way of approaching the necessity of proportionality flips the foregoing on its head: rather than stipulating weights a priori and then diagnosing dilution asymmetry, we can ask a different question: for a given (nonproportional) allocation of voters, is there a feasible value of λ between zero and one that would drive $DA^P(\lambda)$ to zero? This is equivalent to asking whether it is *possible* to construct a welfarist justification for a particular map.

To calculate this critical value for a given plan, we set the right side of equation (7) to zero, solve for λ, and input the realized value of ν,σ, θ^D, and θ^R. We do this for the 662 state-year congressional election observations discussed above. Figure 6 displays a kernel density plot of the critical values of λ.

For a large majority of cases (550 out of 662, or 83%), there exists some feasible critical value of λ that can drive the weighted dilution measure to zero. Importantly, in 95% of *those* cases, the critical λ lies between $\frac{1}{2}$ and 1. In other words, most of the time, if we assign sufficient weight to packed votes relative to cracked ones in our disparity measure then we can reach a conclusion that there is no disparity at all.

Figure 6: Distribution of Critical λs that Drive Weighted Dilution Asymmetry to Zero for Selected Congressional Delegations, 1976-2020



5.2 Is Proportionality Truly Infeasible?

The unweighted variant of Dilution Asymmetry prioritizes proportionality. However, as noted above, achieving proportionality may be infeasible given a state's political geography and the impartial implementation of redistricting principles such as compactness and contiguity. Of particular concern is a potential scenario wherein adherence to a Dilution Asymmetry-derived criterion could conceivably result in the dismissal of ostensibly neutral redistricting schemes in states, like Massachusetts, where proportionality is either improbable or impossible.

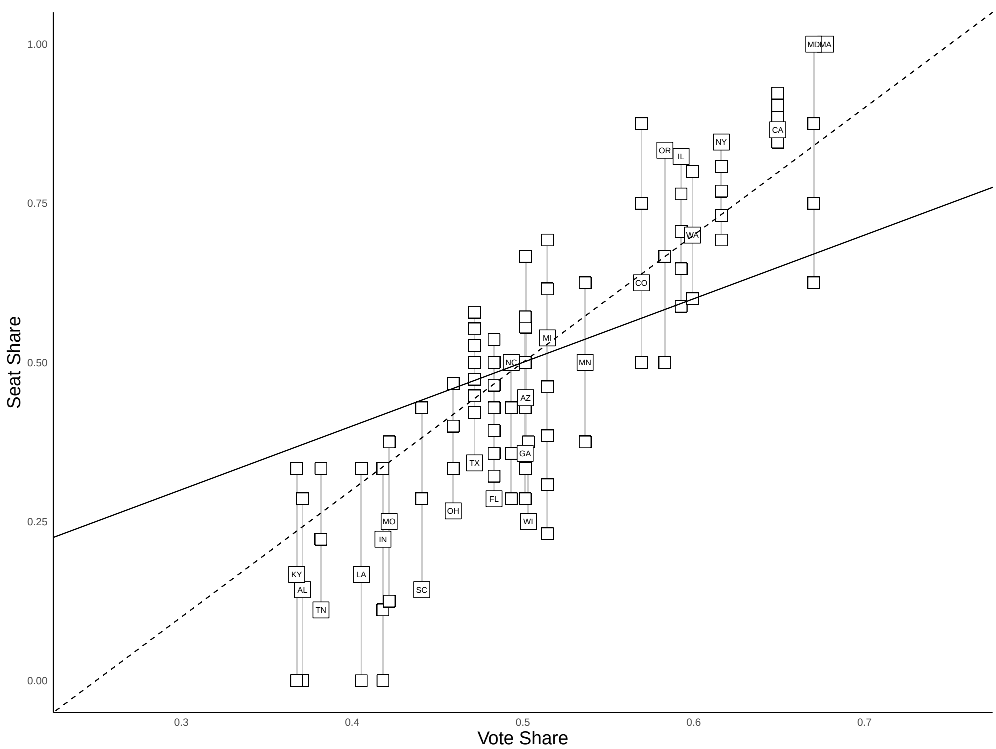
In this section, we employ non-partisan simulated congressional redistricting plans to establish viable state-specific baselines (Chen, Rodden et al., 2013; Fifield et al., 2020; DeFord, Duchin, and Solomon, 2021; Chikina, Frieze, and Pegden, 2017)[22] for the unweighted Dilution Asymmetry. In particular, we again relied on 2020 Presidential election data and 5,000 simulated redistricting plans from the ALARM Project.

The seats-votes share curve in Figure 7 depicts, for each state, the feasible range of seat shares across simulated plans, with the labeled squares corresponding to enacted plans. (States are sorted

[22] An alternative seat benchmark, put forth by (Eguia, 2022), is the proportion of the statewide population garnered within the various jurisdictions (counties and municipalities) clinched by each party.

by the Democratic voteshare.) So, for example, in the top right, Maryland and Massachusetts both adopted plans that would award Democrats a clean sweep even though they received approximately 65% of the Presidential vote. Turning to the simulated plans, none of the 5,000 simulated redistricting plans for Massachusetts could produce a more proportionate outcome. By contrast, in Maryland, a large majority of simulations (69%) anticipated a map with 5-6 districts with Democratic majorities; only a minuscule fraction generated a map with 8 of 8 Democratic majority districts (6.34%).

Figure 7: Feasibility of Proportionate Seat Allocations in an Ensemble of Simulated Plans



Each square indicates a feasible seat share for a given state obtained via simulation. Labeled squares correspond to the enacted plan. The solid line corresponds to perfect proportionality of votes to seats; the dashed line corresponds to double proportionality.

The solid line in the Figure depicts proportionality, while the dashed line indicates double proportionality. (To maximize legibility, we zoom in to votes in the realistic 25-75% range.) Just five states – Arizona, Colorado,[23] Michigan, Minnesota, and North Carolina – enacted maps that award proportionate or nearly proportionate seat allocations. In most *adopted* plans (71%), the partisan majority in the legislature is larger than the partisan majority within the population. However, the presence of a winner's bonus within enacted plans should not be misconstrued as implying that more proportional plans are infeasible (for example, in the event that a state court demanded one). In fact, 42% of simulated redistricting plans across all states achieve proportionality, and at least one proportional plan can be drawn in 20 of 24 states. Among those states where proportionality appears infeasible – California, Massachusetts, New York, and Wisconsin – only Wisconsin is competitive.

6 Conclusion

The Efficiency Gap, a popular diagnostic for partisan gerrymandering, captures disparities in the number of wasted votes by party. We identify potential issues with this approach and trace their origins to base rate neglect: a failure to account for the underlying partisan distribution in a given state. We then show that a measure of vote dilution asymmetry that adjusts for the base rate avoids these problems. In practice, a failure to account for the base rate may lead to erroneous conclusions about both the magnitude and direction of bias against voters from one or another party in districting plans. Also, whereas a longitudinal analysis based on actual vote returns in congressional races suggests that extreme partisan gerrymandering is a recent phenomenon, an analysis using our preferred measure shows that within-state partisan disparities have been ever-present.

At first blush, a criterion rooted in the Efficiency Gap, which requires a majority party bonus by construction, might seem more congruent with historical elections than the proportionality standard implied by the Dilution Asymmetry. Our analysis provides a guide for the practical application of Dilution Asymmetry based on two commonly overlooked perspectives. First, we establish that nearly all historical congressional elections exhibit *zero weighted* Dilution Asymmetry once one accords significantly more weight to packed votes than cracked votes. Second, we assess the viability

[23] Colorado's adopted plan appears to favor Republicans under EG but Democrats under the DA^P

of an unweighted DA^P standard via simulation. Our analysis of thousands of impartial simulated redistricting plans conforming to state criteria and political geography, reveals that proportional seat allocations are feasible in 84% of the states we consider.

While the recent *Rucho* decision suggests that federal courts consider partisan bias in gerrymandering nonjusticiable, state courts that take this kind of bias into account when approving or rejecting districting proposals must still rely on diagnostic tools. Our analysis suggests that Efficiency Gap should be down-weighted in comprehensive evaluations of partisan disparity relative to a measure that adjusts for base rates, and that simulation methods be employed to assess whether maps favored by such a measure are feasible given geographic and other constraints.

While we explicate a partisan measure of dilution asymmetry in this paper, there is no reason that the methodology can and should not be applied along other dimensions, be they racial or geographic (e.g., urban/rural). One next step in this research agenda is to use the insights developed here to develop more robust measures of vote dilution in these areas.

Another avenue for future exploration concerns the appropriate balancing of the harms from packing and cracking. Normative theories of representation prioritize citizens' ability to select like-minded representatives (Mansbridge, 2009), suggesting that the harm of excessive cracking may be worse than the harm of excessive packing. Yet, our analysis indicates that many plurality rule elections appear to down-weight the harm resulting from votes cast for the losing candidate relative to the harm of surplus votes for the winning candidate. Whether wasted vote claims in the context of recent U.S. elections can be supported by theoretical accounts of the harm from redistricting thus remains unclear.

References

Barton, Jeffrey T. 2022. "Fairness in plurality systems with implications for detecting partisan gerrymandering." *Mathematical Social Sciences* 117: 69–90.

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

Caughey, Devin, Chris Tausanovitch, and Christopher Warshaw. 2017. "Partisan Gerrymandering and the Political Process: Effects on Roll-Call Voting and State Policies." *Election Law Journal: Rules, Politics, and Policy* 16 (4): 453–469.

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

Chikina, Maria, Alan Frieze, and Wesley Pegden. 2017. "Assessing significance in a Markov chain without mixing." *Proceedings of the National Academy of Sciences* 114 (11): 2860–2864.

Cover, Benjamin Plener. 2018. "Quantifying Partisan Gerrymandering: An Evaluation of the Efficiency Gap Proposal." *Stanford Law Review* 70: 1131.

DeFord, Daryl, Moon Duchin, and Justin Solomon. 2021. "Recombination: A family of Markov chains for redistricting." *Harvard Data Science Review* 3 (1): 3.

Duchin, Moon. 2018. "Gerrymandering metrics: How to measure? What's the baseline?" *arXiv e-prints* p. arXiv:1801.02064.

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

Eguia, Jon X. 2022. "A measure of partisan advantage in redistricting." *Election Law Journal: Rules, Politics, and Policy* 21 (1): 84–103.

Fifield, Benjamin, Michael Higgins, Kosuke Imai, and Alexander Tarr. 2020. "Automated redistricting simulation using Markov chain Monte Carlo." *Journal of Computational and Graphical Statistics* 29 (4): 715–728.

Hodges, Zachary. 2016. "2016 Congressional Candidates.".
URL: *https://doi.org/10.7910/DVN/IARAAF*

Jackman, Simon. 2017. "Assessing the Current North Carolina Congressional Districting Plan: Amended Report.".

Jacobson, Gary C. 2015. "It's nothing personal: The decline of the incumbency advantage in US House elections." *The Journal of Politics* 77 (3): 861–873.

Jiang, Wenxin, Gary King, Allen Schmaltz, and Martin A Tanner. 2020. "Ecological Regression with Partial Identification." *Political Analysis* 28 (1): 65–86.

Mansbridge, Jane. 2009. "A "selection model" of political representation." *Journal of Political Philosophy* 17 (4): 369–398.

McCartan, Cory, Christopher Kenny, Tyler Simko, Shiro Kuriwaki, George Garcia III, Kevin Wang, Melissa Wu, and Kosuke Imai. 2022. "ALARM Project: 50-State Redistricting Simulations.".
URL: *https://doi.org/10.7910/DVN/SLCD3E*

McCartan, Cory, and Christopher T. Kenny. 2022. "Individual and Differential Harm in Redistricting." Working Paper, Harvard University.

McGann, Anthony J., Charles Anthony Smith, Michael Latner, and J. Alex Keena. 2015. "A Discernable and Manageable Standard for Partisan Gerrymandering." *Election Law Journal* 14 (4): 295–311.

McGhee, Eric. 2017. "Measuring Efficiency in Redistricting." *Election Law Journal: Rules, Politics, and Policy* 16 (4): 417–442.

McGhee, Eric M. 2014. "Measuring Partisan Bias in Single-Member District Electoral Systems." *Legislative Studies Quarterly* 39 (1): 55–85.

MIT Election Data & Science Lab. 2017. "U.S. House 1976–2020.".
URL: *https://doi.org/10.7910/DVN/IG0UN2*

Nagle, John F. 2017. "How Competitive Should a Fair Single Member Districting Plan Be?" *Election Law Journal* 16 (1): 196–209.

Stephanopoulos, Nicholas O., and Christopher Warshaw. 2020. "The Impact of Partisan Gerrymandering on Political Parties." *Legislative Studies Quarterly* 45 (4): 609–643.

Stephanopoulos, Nicholas O., and Eric M. McGhee. 2015. "Partisan Gerrymandering and the Efficiency Gap." *University of Chicago Law Review* 82: 831–900.

Stephanopoulos, Nicholas O., and Eric M. McGhee. 2018. "The Measure of a Metric: The Debate over Quantifying Partisan Gerrymandering." *Stanford Law Review* 70 (5): 1503–1568.

Tapp, Kristopher. 2019. "Measuring Political Gerrymandering." *The American Mathematical Monthly* 126 (7): 593–609.

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

Warrington, Gregory S. 2019. "A comparison of partisan-gerrymandering measures." *Election Law Journal: Rules, Politics, and Policy* 18 (3): 262–281.