How competitive should a fair single member districting plan be?

John F. Nagle

Physics Department, Carnegie Mellon University

Election Law Journal 16, 196-209 (2017). DOI: 10.1089/elj.2016.0386

Abstract

Partisan unfairness is easily detected when the statewide vote is equally divided between two parties. But when the vote is not evenly divided, even the determination of which party is disfavored becomes controversial. This paper examines the ideal fair outcome in a two party single member district system when the statewide vote is not equally divided. It is shown that equal voter empowerment, implied by readings of the first amendment (*Shapiro v. McManus and Whitford v. Nichol)*, requires that the fraction of seats be proportional to the fraction of the statewide vote. However, strict proportionality conflicts with the single member district system, so alternative approaches are explored. Generalized party inefficiency and voter effectiveness are defined and shown to encompass many possibilities for an ideal fair seats-votes function. The best choice is fundamentally determined by the degree of geographical heterogeneity of voters of like mind. Based upon historical election results, it appears that a good approximation to a normative seats-votes function of the American system of single member districts should have competitiveness (aka responsiveness) roughly twice as large as proportionality. This is consistent with the method employed by the plaintiffs in *Whitford v. Nichol*. This method is also basically consistent with the claim of the plaintiffs in *Shapiro v. McManus*, although in this case gerrymandering is better exposed by examining symmetry.

1. Introduction

When each party receives half the total vote in a two party election for a state's congressional delegation or legislative body, everyone agrees that the fair outcome is for each party to win half the seats. While normal statistical variations occur, no one can cogently argue that an equal outcome should not be the ideal. However, precisely splitting the vote seldom happens and then one asks - what is the fair fraction of seats for a party that wins a fraction ½+x of the vote?[1] It is very often assumed that the fair outcome is a fraction ½+x of the seats; this is usually called proportionality.[2] However, proportionality has been challenged in several ways. It has been argued that a more competitive[3] system provides a more stable government by giving the winning party a more comfortable margin of seats than simple proportionality (Hirsh and Ortiz 2005). This argues that ½ + x fraction of the vote should result in ½ + Rx fraction of the seats where the responsiveness/ competiveness factor R is greater than one. A second challenge to proportionality is that actual election results strongly indicate a value of R greater than one (Goedert 2014,Wang 2016). The so-called cube 'law' promulgated early on (Kendall and Stuart 1950) proposed a factor R of three (when x is small), so 51% of the vote would result in 53% of the seats. Recently, a third challenge to proportionality has been advanced based on an

[1] It is convenient to work with fractions, so we designate S as the fraction of seats and V as the fraction of votes for one of the parties with 1 - S and 1 - V being the seats and votes for the other party. If, as usual, no other parties win seats, then V can be taken to be the votes for one party divided by the sum of the votes for the two major parties.

[2] Sometimes the ambiguous term representation is used.

[3] Competitiveness is synonymous with responsiveness because when responsiveness R is large, then a greater fraction of the districts have an expected vote within a competitive range, often considered to be 50±5%. Many reformers believe that greater competitiveness is desirable in a districting plan, although a contrarian view has also been argued (Buchler 2011).

"efficiency gap" that postulates a value R = 2 for the responsiveness (Stephanopoulos and McGhee 2015). This latter method undergirds the Wisconsin plaintiffs' case in *Whitford v. Nichol*.[4]

That responsiveness actually matters when assessing harm by putative gerrymanders, consider Maryland's congressional composition. Based on statewide election returns, Maryland is about 64% Democratic.[5] Of the 8 congressional seats, proportionality proposes that 5.1 Democratic seats would be the fair result, significantly fewer than the actual 7 Democratic seats. However, the cube 'law' (R=3) proposes that the fair number of Democratic seats would be 7.4, which would suggest that the actual outcome even slightly favors Republicans.[6] If the kind of method used by the Wisconsin plaintiffs in *Whitford v. Nichol* were to be used in the Maryland case of *Shapiro v. McManus*[7], then the appropriate value of R is clearly important.

Generally[8], the issue addressed in this paper is: What should the ideal seats-votes function S(V) be, based on fundamental principles. In addition to the agreed ideal that the fraction of seats

[4] *Whitford et al. v. Nichol et al.* IN THE UNITED STATES DISTRICT COURT FOR THE WESTERN DISTRICT OF WISCONSIN, Case: 3:15-cv-00421-bbc Document #: 43 Filed: 12/17/15

[5] Also, there are over twice as many registered Democrats as Republicans which would suggest the state is 68% Democratic. The 2012 congressional elections had 66% Democratic statewide vote and the 2014 congressional elections had 58%.

[6] Although it couldn't be any fairer given the necessary rounding to an integer and the assumed ideal R=3.

[7] CLC % Common Cause Amici Curiae Br. *Shapiro v. McManus* (14-990) SCOTUS 8-13-15.

[8] The previous paragraphs slightly over simplify the issue by focusing only on the responsiveness R when the deviation x of the fractional vote from ½ was small. A general seats-votes function S(V) is not necessarily linear in V over the entire range of V from 0 to 1. One can then define the responsiveness for any value V by the derivative R(V)=dS(V)/dV. However, we will usually just write R to be the derivative at the midpoint of the S(V) function when V is half the vote.

won should be ½ when the fraction of the vote is ½ (*i.e.*,S(½) = ½), it is also obvious that both parties should be treated equally (Hirsh and Ortiz 2005,Grofman and King 2007). If one party wins a fraction S = ½ + Rx of the seats when it receives V = ½ + x of the vote, then the other party should also win the same fraction of seats when it receives the same fraction ½ + x of votes. This means that the ideal S(V) function should be symmetric about the V = ½ , S = ½ midpoint in an S(V) graph, i.e., S(½+x) = 1 – S(½-x) for all values of x between 0 and ½. The symmetry criterion provides a very important fundamental constraint on the ideal S(V) function, but it does not fully determine S(V). In particular, it does not even determine the responsiveness R. This matters when the statewide vote V is different from ½ because a strong majority in control of redistricting could redistrict an S(V) function that would be unbiased in the sense of being symmetric, but by building in a large responsiveness, it would give the redistricting party essentially all the seats. For example, when all seats are equally competitive, the responsiveness factor R is infinite, corresponding to winner-take-all, so a majority party with V greater than ½ would win every seat. Therefore, one should not rely only on symmetry to assess harm to parties or to voters when V is substantially different from ½.[9]

In order to determine S(V) and responsiveness R generally, additional principles to symmetry must be found and applied. The analysis in this paper begins in Section 2 with a reminder that proportionality follows from the fundamental principle that all voters should be empowered equally. This equal empowerment principle is equivalent to the amici first amendment argument given in *Shapiro v. McManus* that inveighs against viewpoint discrimination. Voters with one

[9] As was noted in *LULAC* 548 U. S. at 419-420 (opinion of Kennedy, J.), " … I would conclude that asymmetry alone is not a reliable measure of unconstitutional partisanship."

viewpoint are discriminated against when their votes lead to less representational power than voters with a different viewpoint. Therefore, voters should be equally empowered at the polls. This is similar to the now well accepted argument that each district should have the same number of voters; otherwise voters in a district with more voters would be less empowered representationally than voters in a district with fewer voters.

However, proportionality is contrary to the undergirding theory invoked in *Whitford v. Nichol*. This would appear to vitiate the use of that theory in legal briefs because its non-proportional result contradicts the proportional result that is obtained from the first amendment right for equal empowerment. Nevertheless, that is not the end of the story. As discussed in Section 2, first amendment equal empowerment of all voters conflicts, both fundamentally and practically, with single member district systems, which is also part of the American legal framework.

We then turn in Section 3 to the alternative approach (McGhee 2014) being used in *Whitford v. Nichol*. Requiring the "efficiency gap" (Stephanopoulos and McGhee 2015) between parties to vanish gives a value of R=2 when a particular choice is made for "wasted" votes.[10] However, other choices lead to other plausible values of R as is shown in Section 3, and this would appear to undermine the particular R=2 result of the efficiency gap (EG) approach. Section 4 describes a new voter-centric approach based on the concept of voter effectiveness. Like the party-centric efficiency gap approach, this approach also involves definitional choice. The ensuing S(V) and their R values are mathematically derived for all choices. It is then shown that all voter-centric choices of voter effectiveness share an inherent deficiency except for the one that gives

[10] The term "wasted votes" is unfortunate in that it can be construed as disparaging some voters.

proportionality (R=1). This conflicts with the resulting responsiveness (R=2) of the party-centric approach, but it does support the definition of wasted votes used by that approach.

Section 5 discusses and compares the complex and somewhat dissonant findings from the earlier sections and offers tentative conclusions for appropriate S(V) functions to use for measuring gerrymandering when the vote share V differs from ½. The paper ends with a comparative analysis of bias in the Maryland congressional district that could inform the legal argument in *Shapiro v. McManus.*

2. Tension between a fundamental principle and single member district systems

The simplest principle upon which democracy is based is that each voter should be equally empowered, at least in terms of casting a vote. In a representative democracy, this means that each voter should be represented to the same degree. If there is a fraction V of voters of like mind in that they vote for the same representatives who then are elected to a fraction S of the seats, then the empowerment of each of those voters is proportional to the fraction S/V. Likewise, if there is a fraction 1-V of voters of opposite mind in that they vote for opposing representatives, who then are elected to a fraction 1-S of the seats, then the empowerment of each of those voters is the fraction (1-S)/(1-V). For both groups of voters to be equally empowered requires that S/V = (1-S)/(1-V). Multiplying both sides by V(1-V) and adding VS to both sides gives the simple relation S = V. This S(V) = V function is just proportionality. It affirms what most people intuitively believe is the ideal, fair outcome for an electoral system.

As is well known (Murakami 1968, Rogowski 1981) equal empowerment of all voters is obviously impossible in a single member district (SMD) system because some fraction of voters will have voted for the losing candidate and will therefore be completely un-empowered with respect to a particular legislative body for the duration of that body's term. This is, of course, an argument for a list system (Amy 2000) in which each voter in the fraction of V voters of like mind actually votes for and is represented by a fraction S of seated representatives. However, SMD is the system in the US. Thus, the closest approximation to equal voter empowerment is to empower the average voter of like mind rather than the individual voter. This broadening of equal voter empowerment to accommodate the SMD system is equivalent to avoiding "viewpoint-based discrimination" in the words of the amici brief for *Shapiro v. McManus* which emphasizes that viewpoint-based discrimination should be afforded relief under the first amendment. Only by empowering voters of like mind equally can viewpoint-based discrimination be eliminated. The same simple math in the preceding paragraph again leads to proportionality S = V as the ideal, fair outcome for an SMD electoral system when voters of like-mind are equally empowered.

Unfortunately, the SMD system not only excludes equal individual voter empowerment, it subverts proportionality and it therefore even subverts average voter empowerment, as the following examples illustrate. Suppose that the geographical distribution of voters of opposite mind is homogeneous. A simple extreme example is if voters in every household in the state are divided nearly equally. Then, no matter how the district lines are drawn, a small swing in preference would lead to all the districts being won by one party; this is winner-take-all, which has a responsiveness R of infinity. For a second example, suppose that the geographic distribution of voters is rather heterogeneous. A simple extreme example is if all the voters in a geographically distinct part of a state vote one way and all the voters in the remaining geographically distinct part of a state[11] vote the other way. Then, respecting contiguity and compactness would lead to all seats not changing for typical shifts in overall voter preference. This is described as no responsiveness, *i.e.*, R = 0.

The aforementioned extreme examples emphasize that the appropriate SMD responsiveness R depends upon the geographical distribution of voters of like minds. Unfortunately, R also depends upon how the redistricting is done. Packing voters of both like minds reduces R even while preserving symmetry.[12] Nevertheless, while it has been possible to pack Democratic

[11] With respect to geographical distinctiveness, the two peninsulas of Michigan come to mind.

[12] Bipartisan gerrymandering could achieve minimal competitiveness by drawing lines that result in a vote of 0.6 in half the districts and 0.4 in the other half. Although this example has a value of R =5 when

voters in cities to the 90% level, it seems generally more difficult to pack Republican voters in most states to the same high degree.[13] Symmetry would require the same maximum packing for both parties. Then, it would not be possible to achieve S = V proportionality because there would be no districts at the extreme ends of the S(V) function.[14] Instead, the district votes would have to move towards the middle of the S(V) graph, centered around V= ½. As a simple example, suppose that the districts' votes were distributed roughly equally over a central V region. Then the S(V) graph would be nearly[15] a straight line, as in proportionality, but with a slope (responsiveness R) greater than 1. Of course, many other symmetric distributions of district votes are possible, but symmetry and realistic geographical distribution of voters of like mind together will tend to push R to a value greater than one. However, refining this consideration quantitatively to obtain precise values of R is rather complex. Alternative approaches are therefore considered in the next two sections.

---

responsiveness is defined as ΔS/ΔV and one uses ΔV=0.1, swings in V are typically of order 0.05, so this example would have a small effective value of R.

[13] A thorough discussion of this point is given in chapter 4 of (McGann, et al. 2016). The first congressional district of Maryland is disproportionately packed as will be shown later, but even so, the Republican vote in 2012 was only 63.4%.

[14] Even if a district voted 100% for party B when the statewide vote is V = ½, the S(V) curve would be expected to reach S=1 when V reaches ¾, as shown in Appendix A of (Nagle 2015), so proportionality is always unachievable at the extremes of V.

[15] Curvature in S(V) would be induced by the considerations in the previous footnote. However, this would not change the value of R inferred at the midpoint of the rank/vote function when V = ½.

3. Generalized McGhee Approach

This section generalizes the method employed to measure partisan bias that is being used by the Wisconsin plaintiffs in *Whitford v. Nichol*. This approach and the one in the next section are based on well-defined quantities. One such quantity is lost votes by party; this is the number of votes cast for losing candidates summed over all the districts in which the party's candidate loses. A possible principle for a fair S(V) function is that the number of lost votes L_A for party A be required to be set equal to the number of lost votes L_B for party B. Recently, it has been shown (Nagle 2015) that this principle of equalizing lost votes leads to proportionality. Another well-defined quantity is surplus votes by party; this has been defined (McGhee 2014) as the number of winning votes in excess of half the total vote summed over all districts in which the party's candidate wins.[16] As a winning vote in a district that is won by 100% of the vote is less effective than a winning vote in a district that is barely won by 50%, surplus votes are a disadvantage to a party; they reflect the partisan strategy of packing. Both losing and surplus votes have been described by the unfortunate adjective, "wasted" (McGhee 2014). The formal definition of wasted votes, designated W, is the sum of losing L and surplus votes, here designated E for excess; therefore $W_A = L_A + E_A$ is the number of wasted votes for party A. Then, another possible principle for a fair S(V) function is that the number of wasted votes W_A for party A be equal to the number of wasted votes W_B for party B. It has been shown (McGhee 2014) that this principle of equalizing the number of wasted votes by party also determines an

[16] One might also consider defining the excess vote to be the number of winning votes minus the number of losing votes; this gives twice the excess vote defined above. This variation makes no substantial difference as will be discussed in Section 5 (footnote 24).

S(V) function that has R=2, twice as responsive as proportionality. Figure 1 shows how the two functions differ.

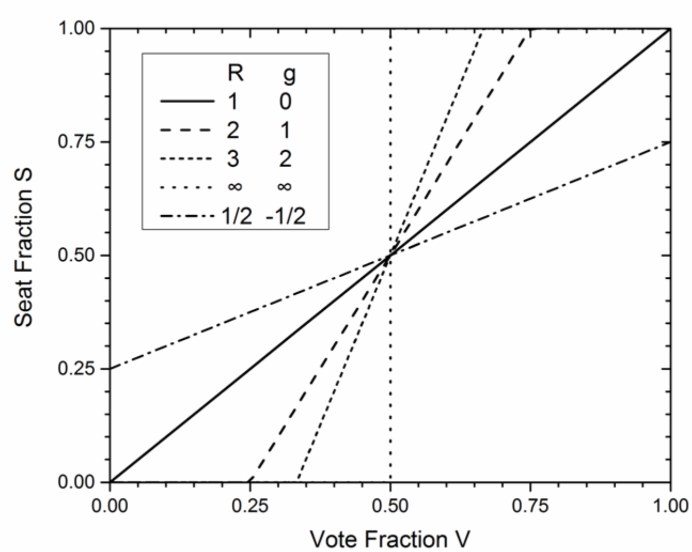


Figure 1: The solid line shows zero bias for proportional representation and the long dashed line shows zero bias when equalizing wasted votes. Other zero bias functions with other values of responsiveness R are the seat vote functions $S_g(V)$ for the party centric fairness criterion given by Eq. (1).

This paper now newly derives seats-votes functions when the quantity equalized by party is any linear combination of losing and excess votes. Specifically, L_A and L_B are designated as the losing votes of parties A and B, respectively, and E_A and E_B are designated as the excess votes of parties A and B, respectively. For a particular linear combination specified by a parameter designated as g, it is then required that

$$L_A + gE_A = L_B + gE_B. \tag{1}$$

Choosing g=0 in Eq. (1) is the same as equalizing just lost votes, whereas choosing g= 1 in Eq. (1) is the same as equalizing wasted votes.

Let us discuss possible reasons for considering various values of g. As packing is the most obvious way to obtain partisan bias, perhaps it should be the sole quantity to equalize, which is realized in Eq. (1) by allowing g to become infinitely large ($g=\infty$). Alternatively, one could argue that the ideal S(V) function should reflect voter satisfaction. Many voters are likely to feel just as happy when their candidate wins big as when their candidate just scrapes by. This argues that the penalty functions to be equalized should just include losing votes, assigning $g=0$ which does not penalize excess votes. One could even argue that some voters feel happier when their candidate wins big because it enhances their confidence in being on the right side; then, it could even be argued that g should be less than 0, thereby assigning a negative penalty for excess votes in Eq. (1). However, parties are less concerned with voter satisfaction than with the overall seats outcome, and then it is clear that excess votes are harmful to a party if it has more of them than the other party.[17] This party-centric perspective argues that g in Eq. (1) should be greater than 0 consistent with treating both losing votes and excess votes as harmful.

Let us now turn to describing, for all values of the g parameter in Eq. (1), the seat-votes functions for party-centric measures which we will henceforth designate $S_g(V)$. The mathematical derivation of $S_g(V)$ is straightforward and is given in Appendix A. A salient feature of these functions is that the value of the responsiveness R is given by $1+g$ and S is a simple linear function of V for values of V in the central region of the $S_g(V)$ plot as shown in Figure 1. Specifically,

$$S_g(V) = \tfrac{1}{2} + (1+g)(V - \tfrac{1}{2}) \qquad (2)$$

[17] Interestingly, the gap between excess votes $E_A - E_B$ for the two parties does not widen but is given just by $V - \tfrac{1}{2}$ as shown in Eq. (A6) in Appendix A. When a gerrymandering party packs the other party's voters, it wins more seats and then it also has more excess voters such that there is no change in the excess votes gap; rather, it is the gap in lost votes that widens. Nevertheless, inclusion of excess votes makes a considerable difference to the S(V) functions compared to just including lost votes.

in this central region. As V increases from ½, S_g increases becoming 1 when the vote V equals ½(2+g)/(1+g). Because S cannot exceed 1, it is impossible to satisfy Eq. (1) for larger values of V when $g > 0$. Nevertheless, the difference between the penalty functions for the two parties, namely, the two sides of Eq. (1), is least when S=1, so the fairest realizable $S_g(V)$ function assigns S=1 for V>½(2+g)/(1+g). We will designate the breakpoint in the $S_g(V)$ function as V_b = ½(2+g)/(1+g). By symmetry, there is also a break point when V is small, as shown in Figure 1 where the $S_g(V)$ function assigns S=0 for V < ½g/(1+g).

The special case when the penalty function is based exclusively on excess votes, namely, $g = \infty$, warrants specific comment. As g increases, R=1+g becomes larger and the break points move closer to V = ½. In the limit $g = \infty$, the break points converge on V= ½, the responsiveness R becomes infinite, so S is 1 for all V greater than ½ and S is 0 for all V less than ½. This is just the winner-take-all extreme. From a mathematical perspective, it is just a limiting case, but one of interest because of the character of the underlying penalty function.[18]

The case of g less than 0 also warrants specific comment. This is the case when excess votes are considered to be desirable rather than harmful. This case has responsiveness R less than 1. It does not have break points to constrain S to be between 0 and 1. However, it has the unrealistic feature of requiring S to be less than 1 even when V equals 1.

[18] As mentioned in the previous footnote, $E_A - E_B = V_A - ½$. This means that it is impossible to balance excess votes when the statewide vote V_A is not ½. This is consistent with the infinite value of R when $g=\infty$.

4. New voter-centric approach

There is an important, and ultimately preferable, alternative to the party-centric measures of bias in the previous section that we will call voter-centric in this paper. Whereas party-centric measures focus on equalizing the aggregate harm done to a party, voter-centric measures focus on equalizing the average effectiveness of voters of like mind. The fundamental quantities remain those of lost and excess votes as defined at the beginning of the previous section. The crucial difference is that the quantities to be equalized are averages for voters of like mind rather than aggregate numbers of voters by party. The average effectiveness of A voters of like mind is the total number of effective A votes divided by V_A, the total number of A voters, with a similar definition for B voters. The number of effective A votes is just V_A minus the number of ineffective votes. Similar to the previous section, the number of ineffective A votes can be defined as $L_A + gE_A$ so the average A effectiveness becomes $[V_A- L_A- gE_A]/V_A$. Equalizing A and B voter effectiveness then requires

$$[V_A-L_A-gE_A]/V_A = [V_B-L_B-gE_B]/V_B \quad . \tag{3}$$

For the case g=0, the numerators in Eq. (4) are the number of winning votes (V-L) for each party, so Eq. (3) guarantees that voters of like mind are equally likely, on average, to be happy with their representative.[19] For the case g=1, the numerators in Eq. (3) are the numbers of non-wasted votes by party. A wasted vote is an ineffective vote, so the numerators are the numbers of effective votes for the parties. Dividing by the denominators then gives the average effectiveness of voters of like mind on each side of Eq. (3). The underlying principle then is that

[19] This ideal unbiased result for this particular choice of g has already been presented as a voter happiness measure (Nagle 2015). Here, we emphasize that that measure is the same, both in principle and in result, as the current voter-centric measure that is equivalent to using only lost votes.

there is no first amendment viewpoint-based discrimination (no partisan bias), when voters of opposite mind are equally effective on average.[20]

Equation (3) may be rewritten by subtracting 1 from both sides and then multiplying by -1 to give

$$[L_A+gE_A]/V_A = [L_B+gE_B]/V_B \quad . \tag{4}$$

Eq. (4) looks similar to the party-centric balance in Eq. (1) but with the major difference that the harms to the parties in the numerators is divided by the number of voters of like mind, V_A and V_B, that vote for candidates of party A and B, respectively. Because Eq. (3) is equivalent to Eq. (4), the approach in this section can be described either as equalizing the average voter effectiveness or as equalizing average voter harm.

Just as for party-centric measures, the relative proportion of lost and excess votes also gives rise to different voter-centric measures. Working out the ideal, zero bias, S(V) results for general values of g is algebraically rather complex and is deferred to Appendix B. However, the derivation of the g=1 case turns out to be quite simple. As this case is also particularly important, it is given here. We start from Eq. (3). The number of lost votes L_A is the sum of A votes for all districts won by party B candidates. The number of excess votes E_A is the sum, for all districts won by party A candidates, of A votes minus half the votes in each of those

[20] Effectiveness could equally well be called efficiency as it is the ratio of effective votes of like-minded voters divided by the total number of like-minded voters, rather like the standard definition of efficiency in the physical sciences as output/input. Our voter-centric fairness principle then requires the gap between the efficiencies of different-minded voters to be zero. However, as the term "efficiency gap" has been coined to describe the party-centric fairness principle (Stephanopoulos and McGhee 2015), the term effectiveness rather than efficiency is used here.

districts.[21] Therefore, the sum of lost and excess A votes equals the total number of A votes, namely the statewide V_A, minus half the votes in each district won by A. This V_A cancels the V_A that appears in the numerator in Eq. (4), leaving half the sum of the votes in each district won by A in that numerator. The number of districts won by A is just S_A so the numerator on the left hand side of Eq. (5) is proportional to S_A and the numerator on the right hand side of Eq. (5) is proportional to S_B, giving

$$S_A/V_A = S_B/V_B \qquad . \qquad (5)$$

Using fractional scales for S and V, both from 0 to 1, and identifying $S = S_A = 1 - S_B$ and $V = V_A = 1 - V_B$, Eq. (5) is easily solved to give

$$S = V \qquad , \qquad (6)$$

which is just proportionality.[22]

As an aside, it may be noted that, while this ideal S(V) result for g=1 is the same as for proportionality, there is a slight difference for the corresponding measure of bias if one defines that as the difference in the right and left hand sides of Eqs. (4) or (5). Then one obtains the measure of bias as

$$B' = (S-V)/2V(1-V) \qquad , \qquad (7)$$

[21] We assume that the turnout is essentially equal for all districts, thereby assuming that there is little turnout bias (McDonald 2009), typical of recent PA congressional elections (Nagle, 2015).

[22] McGann et al. (2016) have criticized the EG measure on the grounds that there are many other models for ideal fairness that would give different values of R and the voter centric model in the present paper is one mentioned. It was also implied that yet a different measure based on equating the number of wasted votes per seat won would give yet another value of R, but this second alternative is actually equivalent to their first alternative and therefore also gives the same value R=1.

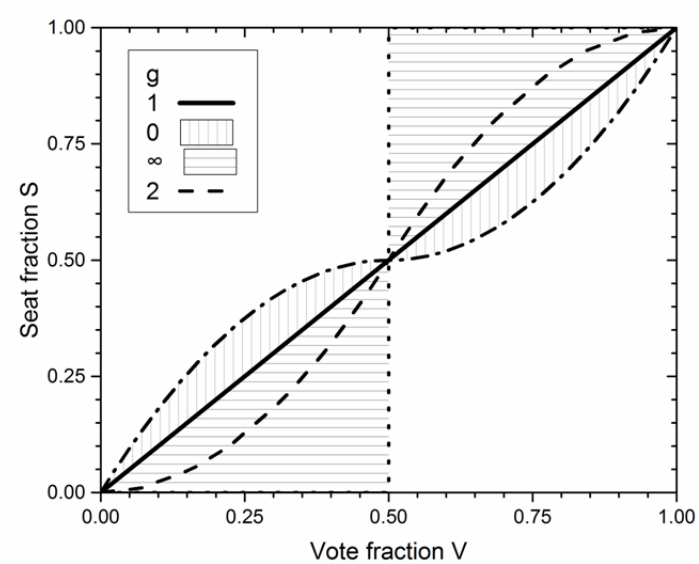
in contrast to B = S – V for party-centric proportionality. For V= ½, B′ in Eq. (7) is twice as large as B; this is a trivial change in scale like going from feet to yards that can be reconciled in Eq. (7) by replacing the 2 by a 4. The slight, but real, difference between B′ and B is then a factor of 4V(1-V) which increases B′/B as V deviates from ½, but only by 20% even when V is as large as 0.7.

Figure 2 shows results for the ideal voter-centric cases for several values of g. Except for g=1, it is possible to obtain unbiased results for a range of S values for the same value of V when V is not equal to ½. This range is shown in Figure 2 by the shaded regions for g=0 and g=∞. For g=2 the range lies between the dashed curve and the line of proportionality. The quantity that is associated with these variations is the average fraction of votes v_n in those districts that are won (or lost) by either A or B, as defined in Appendix B. Proportionality is a fair possibility for all g, but for $g \neq 1$ this only occurs when each district is won or lost by all the vote (fractional district v_n = 0 or 1). As the district votes become more realistic, fairness is only achieved for sub-proportional S/V ratios (R<1) when g < 1 and for super-proportional S/V ratios (R>1) when g>1. For realistic district vote averages in the range of 0.6 for won districts, the fair outcome is very close to the functions shown for the g=0 and g=2 cases. The responsiveness R is close to 0 for g=0 and close to 2 for g=2.

One can obtain an intuitive understanding of the range of possibilities most easily by considering the extreme possibilities for the case g = ∞. Half the votes are excess when district votes are 0 or 100% and then S has to be proportional to V to satisfy Eq. (3). The other extreme is when each district is won by a vanishingly small fraction of the vote which means that there is a negligible fraction of excess votes; although the vote V must equal ½, S can be any value from

0 to 1, so this gives the vertical line that borders the region in Figure 2 indicated by faint horizontal hatching. In contrast to this easily intuited g = ∞ case, for the g=0 case, the extreme shown by the function that bounds the region in Figure 2 indicated by vertical hatching requires somewhat tedious algebra as outlined in Appendix B.

Figure 2. Ideal (zero bias) Seats-Votes possibilities for four voter-centric measures identified by the values of g.



The range of possible unbiased S(V) results that are portrayed in Figure 3 should not be interpreted as meaning that any particular S(V) result is unbiased if it happens to fall in the range of possible unbiased values. That is because the bias depends on the average district votes and these are not determined by overall S and V. Nevertheless, the actual bias is easily calculated from the difference in the two sides of Eqs. (3) or (4). However, as Eric McGhee has kindly pointed out, the possibility that different values of S for the same vote V may give the same value of bias violates a fundamental principle for bias measures (McGhee 2016), namely, gerrymandering might be able to increase S for the same V and not be detected by the measure of bias. The mechanism to do this is to draw the lines to change the average district votes. Making districts more competitive allows a gerrymandering party that has $V > ½$ to increase its S with no change in this measure of bias when $g > 1$.[23] Since making districts more competitive increases R, this is completely consistent with the discussion in Section 2 that a majority party benefits from a larger value of R.

5. Discussion

It appears that fundamental principles tend to yield proportionality as the ideal seats-votes S(V) function. This is clearly so for voter empowerment as shown in Section 2. It is not as immediately clear for the alternative methods in Sections 3 and 4. The responsiveness R for both those methods depends upon the weight g that is placed on excess votes. For those methods to be productive, it is necessary to find a value of g that is superior. One way to choose g appears at the end of the preceding section. In the voter-centric method all values of g not equal to 1 allow violation of the general principle that a measure of bias should not allow gerrymandering. This reduces the plethora of choices to the g=1 case which leads to proportionality and responsiveness R=1. Importantly, the voter-centric result that g=1 is the only acceptable value supports the assumption (McGhee 2014) that excess votes should count equally with lost votes when evaluating harm either to parties or to voters.[24]

---

[23] Similarly, making districts less competitive when $g<1$ allows a gerrymandering party that has $V > ½$ to increase its S with no change in this measure of bias. In the less likely case that a party in control of gerrymenadering has $V < ½$, switch more and less in the preceding two cases.

[24] Furthermore, this resolves the issue brought up in footnote 16 regarding how to define excess votes precisely. If the alternative definition mentioned there were taken, the only viable voter-centric case

Turning to the 'party-centric' method in Section 3, it is tempting to argue that the value g=1 established for the voter-centric method should again apply. This would give a responsiveness R=2 instead of the voter-centric R=1. This conflict in the results of the two methods raises the question of which method is more fundamental. The party-centric method equalizes the total harm to both parties in the sense that each party ideally has the same number of wasted votes. However, one may well question why a party with fewer voters should suffer having as many wasted votes as the party with more voters. Another concern is, why shouldn't one equalize non-wasted votes? That seems logically just as plausible as equalizing wasted votes. However, that leads to the absurd result that each party has to win the same number of seats no matter what the overall vote V is. This follows mathematically because the number of non-wasted votes is just equal to the number of seats times half the votes in each district, so to equalize the number of non-wasted votes requires equalizing the number of seats. In contrast, as shown in Section 4, equalizing the average effective vote gives the same results as equalizing the average ineffective vote. Therefore, the voter-centric method is more fundamental. This means that, once again, fundamentalism leads to proportionality.[25]

On the other hand, empiricism has not supported proportionality going back as far as the cube law (Kendall and Stuart 1950). A recent study (Goedert 2014) has reported that R=2 better represents American elections. A similar result can be gleaned from Figure 1 of (Wang 2016).[26] However, there is a caveat; dominant parties have the incentive to increase their state's R value

would instead have g= ½, and there would be no other significant difference, so g=1 with the original definition of excess votes suffices.

[25] A rather different formal theory by (Hout and McGann 2009) also leads to proportionality.

[26] One can also extract a value of R ≅ 2.7 from Figures 2 and 3 that Wang obtained from simulations of "fantasy delegations" using nation-wide data.

and this tends to increase the empirical R above what would be the ideal unbiased value.[27] Nevertheless, a value of R greater than 1 is consistent with the qualitative argument in Section 2 that is based on likely geographical distributions of voters of like mind. Therefore, one must be reconciled to the realization that the ideal value of S when V differs from ½ is not determined by a simple fundamental principle but by geographical voter heterogeneity, and that may well be different for different states.[28] It would be interesting to measure such geographical distribution as well as to model the ideal responsiveness R based upon it, and to apply such information to each state, but that project is beyond the scope of this paper and perhaps it is even impossible. At this time it appears that the best one can do is to appeal to the empirical result that R is approximately 2 when averaged over many states (Goedert 2014,Wang 2016).[29]

Having tentatively decided on a value of $R \cong 2$, it may be of interest to consider the detailed S(V) function. So far in this paper, the only viable function that conforms to the empirical responsiveness R=2 is McGhee's party-centric wasted votes (g=1) function shown in Figure 1 and again in Fig 3. This function has the artificial break points at V= ¼ and V= ¾ which means that a deeply minority party should receive no seats at all, an unpalatable feature. Figure 3 compares two other functions with R=2. Unlike the party-centric g=1 function, the voter-centric

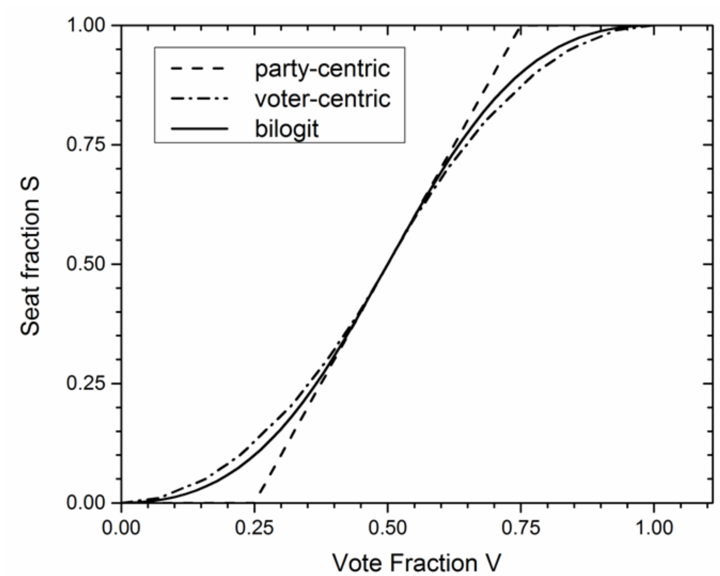
[27] Statistical samples should therefore not include results when the dominant party was in control of the redistricting process, which, unfortunately, increases the uncertainty in determining R empirically.

[28] Unfortunately, different values of R for different states can theoretically lead to anti-majoritarian results nationally (McGann, et al. 2015; McGhee 2016), which motivates deciding on a common value of R. Contrary to what (McGann, et al. 2016, p. 218) wrote, proportionality is not required for the common value of R to avoid national anti-majoritarian results (McGhee 2016).

[29] Note, however, that (McGann, et al. 2016, p. 72) obtain R=1.52; the smaller value could be related to their methodological addition of a random variable to obtain an S(V) curve.

g=2 function for realistic distributions of voters of like mind[30] provides a smooth curve and is considerably more favorable to deeply minority parties. Figure 3 also shows that the R=2 bilogit taken from a historically valuable family of functions (King and Browning 1987) is numerically quite similar.[31] Of course, if different values of R were to be preferred, corresponding functions from these two families could then be chosen.

Figure 3. Comparison of three S(V) functions with the same level of responsiveness R=2.



At this point, it is appropriate to comment on the perspective that it may be too much to attempt to prescribe ideal responsiveness and the ideal S(V) function (Nagle 2015;McGann, et al. 2016). The implication then is that one only needs to assure symmetry to avoid political bias. However, as elaborated in the introduction, there is a weakness in relying only on symmetry. A

[30] It was assumed that the winning district vote average fraction was 0.6.

[31] The bilogit is the much easier one to calculate.

strong majority in control of redistricting could redistrict an S(V) function that would be unbiased in the sense of being symmetric, but with a large responsiveness which would give the redistricting party essentially all the seats. Therefore, one should not rely only on symmetry when V is significantly different from ½. Fortunately, for states with V close to ½ it doesn't matter much whether one uses symmetry alone or a symmetric R=2 function. However, the goal in this and recent papers was to provide a measure of bias for all states and there seems to be a developing consensus on what the ideal competitiveness/responsiveness should look like even though the supporting arguments differ.[32] However, it should be remembered that symmetry is an absolute fundamental principle (McGann, et al. 2016). It should therefore be applied first in the analysis of partisan bias. Being less fundamental, responsiveness should be applied second to those cases in which the vote is persistently different from V = ½.

To see how the generalities in the preceding paragraph play out in a particular case, let us return to the case of Maryland mentioned in the introduction. The value of R=2 in the efficiency gap method used by the plaintiffs in *Whitford v. Nichol* would suggest that a fair number of Democratic congressional seats in Maryland would be 6.2, but is that enough different to support the plaintiff's case in *Shapiro v. McManus*? One of the suggestions of (Grofman and King 2007) is that a threshold for justiciability might consist only of deviations exceeding one seat, so it might appear, based solely upon the efficiency gap, that Maryland gerrymandering is not egregious enough to warrant action. A closer look suggests otherwise.

[32] In particular, the choice of R=2 is the same as obtained using the efficiency gap (Stephanopoulos and McGhee 2015) but the theoretical underpinning is different.

Figure 4 shows the seats-votes curves for the two parties derived from the Maryland 2012 congressional results.[33] If there were no bias, then there would be no difference between the curve for Democrats and the curve for Republicans, but Figure 4 shows substantial differences.[34] It is also of interest to compare the different values obtained from different measures of bias.[35] The simplest bias measure is the percentage difference in seats evaluated from the seats-votes curve when the statewide party vote is 50%; the value of this measure is labelled S in the legend of Figure 4. A recently proposed measure (Mm in the legend) is the percentage difference in vote for the median seat minus the mean seat (McDonald and Best 2015,Wang 2016). The Mm value is much smaller than the S value because the effective responsiveness is so large.[36] That makes the difference gap between the S(V) curves tall and thin and it is the thin direction that the Mm method measures in contrast to the S method that measures the tall direction. A recently proposed geometric bias measure (Nagle 2015) measures the difference in area between the two curves in Figure 4 with the value shown for G in the legend to Figure 4. This G measure incorporates both the seats dimension that is focused on by the S measure and the votes dimension that is focused on by the Mm measure. Unlike the S and Mm measures which focus on the central part of the S(V) curve, the G measure fully takes into account all districts, especially those at the extreme, such as the packed 1st district. The EG value in the legend comes

[33] These seats-votes curves were obtained by the method of (Nagle 2015) that shifts the vote, but does not use the flawed uniform shift method.

[34] The 1st congressional district is packed with Republicans and that accounts for the large difference in the two curves at 0 and 8 party seats in Figure 4.

[35] Many different measures of bias, along with some of their faults, have been described recently (Nagle 2015). Except for the EG, these measures approximate asymmetry.

[36] As discussed earlier, a large R is an effective way to gerrymander, even without violating symmetry. The R value at V=50 in Fig. 4 is 3.6, close to the value R=3.8 given by McGann, et al. 2016, page 91.

from the efficiency gap method and agrees better with the G measure than with the S and Mm measures.[37] For perspective, the G measure for the 2012 congressional election in Pennsylvania, arguably the most gerrymandered state in the nation, is 9.4% (Nagle 2015) , not that much larger than the 7.8% in Maryland.[38] This scrutiny of Maryland election results therefore supports the plaintiffs' case in *Shapiro v. McManus* that Maryland has been effectively gerrymandered.

[37] It must be noted, however, that the EG method is subject to excessive volatility when the district votes change by small amounts. For example, a small change in the vote in Maryland's 6th district would have switched that seat, resulting in the EG value jumping from +9.5% to -3%. An even more striking example supposes that there are three districts, with partisan divisions of 40%, 50% and 60%. Random events, such as a few voters in the 50% district not making it to the polls, produces EG values that jump between +18% and -18%. Such jumps, also shared by other simple measures of bias, invalidate the underlying concept that the bias in a districting plan should change only gradually with time (Nagle 2015). The G measure of bias does that generally, and specifically, the G values differ by small amounts when the Pennsylvania 2012 and 2014 congressional elections are compared and also when the Maryland 2012 election is compared to either the Maryland 2014 election or the above mentioned counterfactual. On the other hand, when there are many legislative districts, as in the Wisconsin legislature which is being contested in *Whitford v. Nichol*, such jumps become statistically small and the EG method is likely to be appropriate.

[38] The value of the simple MD bias is reported as 14 by McGann, et al. 2016). This is smaller than the S value in the legend in Fig. 4 due to introduction of random uncertainty in the S(V) curve. They also report a value 25.3 for an average, so-called, symmetric bias, which should be divided by 2 for numerical comparison to other measures. This is again smaller than their S value because it partially takes into account the thickness dimension as well as the length dimension in the S(V) plot.

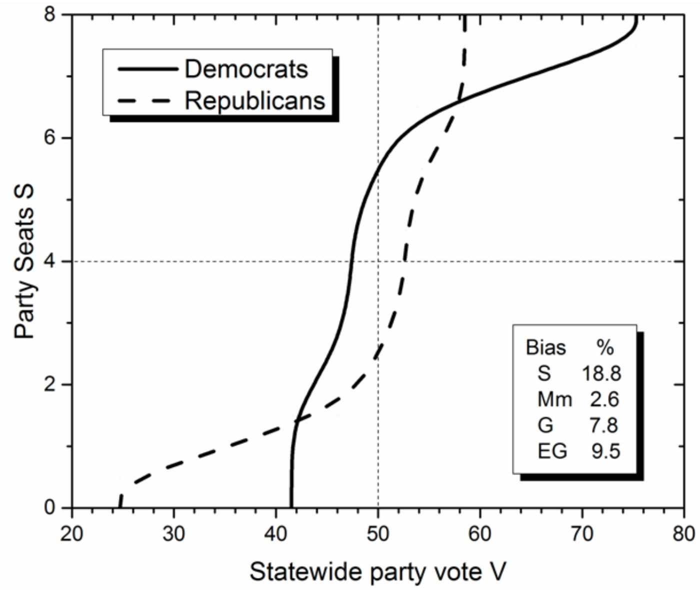


Figure 4. Seats vs. votes curves for Maryland 2012 congressional election. The lower right legend gives values of bias for four different methods discussed in the text.

6. Conclusions

1. The value of responsiveness/competitiveness R is important to evaluate the harm done by gerrymandering when the statewide vote is not evenly split. Courts typically require that an intentional gerrymander actually harm voters of like mind; results of calculations of that harm differ considerably for different values of R.

2. The first amendment protection against viewpoint discrimination at the core of *Shapiro v. McManus* requires that voters of like mind be equally empowered at the ballot box compared to voters of opposite mind. This leads to proportionality (R=1) as a first amendment principle.

3. The party-centric efficiency gap method underlying the plaintiff's case in *Whitford v. Nichol* gives R=2, but a foundationally superior voter-centric method re-affirms that R=1 is the abstract ideal. The latter method does confirm that "wasted votes" is the quantity of importance in both the party-centric and voter-centric approaches.

4. Both theoretical considerations and empirical studies lead to the conclusion that the American electoral framework of single member districts (SMD) conflicts with proportionality. Empirical results suggest that a value R=2 is an appropriate realistic ideal for the American SMD system in agreement with the plaintiff's case in *Whitford v. Nichol.*

5. Deviations from symmetry, properly measured, remain the first important measure of partisan bias, but symmetry should be supplemented by consideration of the value of R when the vote deviates from being evenly split between parties.

6. In support of *Shapiro v. McManus,* current Maryland congressional districting is significantly biased, as found primarily by lack of symmetry but also by a too high responsiveness in the appropriately constructed seats-votes curve.

Appendix A. Derivation of Eq. (2) in the text for the generalized McGhee approach

Let Σ^A sum over all districts n won by party A, normalized by the number of districts N.

Let Σ^B sum over all districts n won by party B, similarly normalized by multiplying by 1/N.

The fraction of seats won by party A is $S_A = \Sigma^A 1$. (A1a)

Similarly, $$S_B = \Sigma^B 1 \quad . \tag{A1b}$$

Let v_n be the fractional vote for party A in district n.

The statewide district weighted* fraction of votes for party A is $V_A = \Sigma^A v_n + \Sigma^B v_n$. . (A2)

The district weighted fraction of votes lost by party A is $L_A = \Sigma^B v_n$ (A3a)

Similarly, $$L_B = \Sigma^A (1-v_n) \quad . \tag{A3b}$$

The district weighted fraction of excess votes for party A is $E_A = \Sigma^A (v_n - ½)$ (A4a)

Similarly, $$E_B = \Sigma^B (½ - v_n) \tag{A4b}$$

Let us now rewrite Eq. (1) in the text as $0 = (L_A - L_B) + g(E_A - E_B)$. (A5)

Applying Eqns. (A1-4) to Eq. (A5) and combining terms yields

$$0 = (V_A - S_A) + g(V_A - ½) \quad , \tag{A6}$$

Rearranging terms in Eq. (A6) yields $S_A = (1+g)V_A - ½g$. (A7a)

Similarly, when A is replaced by B $S_B = (1+g)V_B - ½g$. (A7b)

Subtracting (A7b) from (A7a) yields $S_A - S_B = (V_A - V_B)(1+g)$. (A8)

Eq. (A8) is the same as Eq. (2) in the text when S_A is identified with S, V_A is identified with V, $S_B = 1 - S_A = 1 - S$, and $V_B = 1 - V_A = 1 - V$.

QED.

Appendix B. General voter-centric ideal results

We begin by collecting the districts according to which party won. In Figure 5, those districts won by party A are collected to the left of the S value on the horizontal axis and those won by party B are on the right side of the S value, so S is the number of seats won by party A. The vertical axis in Figure 5 is for the district vote for party A. All that we need for this analysis is the average A vote for those districts won by A and the average A vote for those districts won by B. These are shown by two solid horizontal lines in Figure 5. It is convenient to define x to be the average A vote for districts won by B and 1-y to be the average A vote for districts won by A. The quantities in Eqns. (A5) and (A6) depend on x and y as follows:

$L_A = x(1\text{-}S)$, $L_B = yS$, $E_A = (½ \text{-}y)S$, $E_B = (½ \text{-}x)(1\text{-}S)$, $V_A = (1\text{-}y)S + x(1\text{-}S)$, and $V_B = (1\text{-}x)(1\text{-}S) + yS$. (B1)

The lost and excess votes are shown by rectangles in Figure 5. Given S and g, Eq. (3) (and its equivalent Eq. (4)) in the main text requires that x and y are not independent; y can be determined from x through a quadratic formula

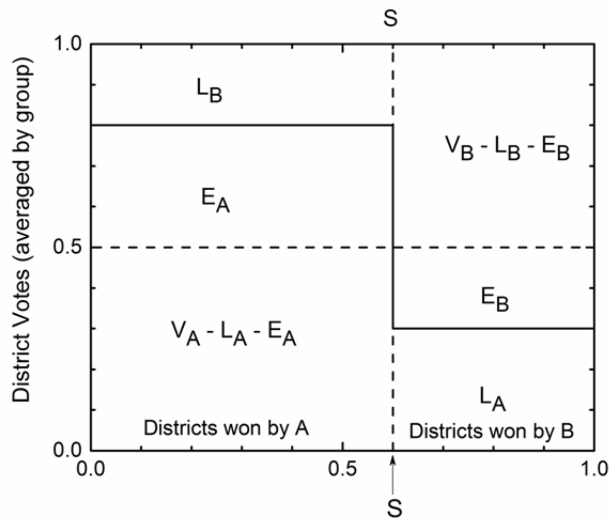
$$y^2(1-g) - y[2-g+g(1-S)/S]/2 + C(x,g,S) = 0 \qquad \text{(B2)}$$

where $$C(x,g,S) = ½\ \{g(1-S)x/S + [(1-g)x(1-2x)+x](1-S)^2/S^2\}, \qquad \text{(B3)}$$

which is straightforward to solve numerically. Then, the value of V_A is calculated in Eq. (B1). For S> ½, the range of possible x is from 0 to 0.5, but except for g=1, y ranges from 0 to a value less than ½ that depends upon S and g.

This mode of calculation gives a range of V for a given value of S. Calculation for many values of S then allows one to determine the range of S for a given value of V. The two shaded regions in Figure 3 shows the S(V) ranges for g = ∞ and g=0. For all g these regions are bounded by the proportionality line and by a function that is obtained for x=0.5 when S> ½ and by y= ½ when S< ½. The latter bounding functions are also shown for g=2 in Figure 3.

Figure 5. Illustration of basic average quantities. The district votes have been averaged within each group of districts won by voters of like mind.



Acknowledgement: The author thanks Eric McGhee for his seminal analysis of the theory of ideal responsiveness and for valuable discussion during the course of this work.

Bibliography

Amy, Douglas J. 2000. *Behind the Ballot Box: A Citizen's Guide to Voting Systems*. Westport, CT: Praeger Publishers.

Buchler, Justin. 2011. *Hiring and Firing Public Officials: Rethinking the Purpose of Elections*. New York, NY: Oxford University Press.

Goedert, N. 2014. "Gerrymandering or Geography? How Democrats Won the Popular Vote but Lost the Congress in 2012." *Research and Politics* 1: 1-8.

Grofman, B. , and G. King. 2007. "The Future of Partisan Symmetry as a Judicial Test for Partisan Gerrymandering after Lulac V. Perry." *Election Law Journal* 6: 2-35.

Hirsh, S., and D. Ortiz. 2005. "Beyond Party Lines: Principles for Redistricting Reform." http://www2.illinois.gov/gov/reformnow/Documents/press%20releases/Principles%20for%20redistricting%20reform.pdf. (2005).

Hout, E. van der, and A. J. McGann. 2009. "Proportional Representation within the Limits of Liberalism Alone." *Brit. J. Political Science* 39: 735-54.

Kendall, M. G., and A. Stuart. 1950. "The Law of Cubic Proportion in Election Results." *British Journal of Sociology* 1: 183-97.

King, G., and R. X. Browning. 1987. "Democratic Representation and Partisan Bias in Congressional Elections." *American Political Science Review* 81: 1251-73.

McDonald, Michael D. 2009. "The Arithmetic of Electoral Bias, with Applications to U.S. House Elections." *APSA 2009 Toronto Meeting Paper. Available at SSRN*.

McDonald, Michael D., and Robin E. Best. 2015. "Unfair Partisan Gerrymanders in Politics and Law:
A Diagnostic Applied to Six Cases." *Election Law Journal* 14: 312-30.

McGann, A. J., A. S. Smith, M. Latner, and J. A. Keena. 2015. "A Discernable and Manageable Standard for Partisan Gerrymandering ". *Election Law Journal* 14: 295-311.

———. 2016. *Gerrymandering in America: The House of Representatives, the Supreme Court, and the Future of Popular Sovereignty*. 1st ed. New York, NY: Cambridge University Press

McGhee, E. 2016. "Measuring Efficient Partisan Bias." *Presentation at the Annual Conference of the Midwest PoliticalScience Association, Chicago, IL.* x: x.

———. 2014. "Measuring Partisan Bias in Single-Member District Electoral Systems." *Legislative Studies Quarterly* 39: 55-85.

Murakami, Y. 1968. *Logic and Social Choice*. London: Routledge & Keegan Paul Ltd.

Nagle, J. F. 2015. "Measures of Partisan Bias for Legislating Fair Elections." *Election Law Journal* 14: 346-60.

Rogowski, R. 1981. "Representation in Political Theory and in Law." *Ethics* 91: 395-430.

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

Wang, S. S.-H. 2016. "Three Tests for Practical Evaluation of Partisan Gerrymandering." *Stan. L Rev.* 68: 1263-321.