Title
Designing electoral systems

Permalink
https://escholarship.org/uc/item/28t10044

Journal
Electoral Studies, 8(1)

ISSN
0261-3794

Authors
Taagepera, R
Shugart, MS

Publication Date
1989

DOI
10.1016/0261-3794(89)90021-8

Peer reviewed
Designing Electoral Systems

REIN TAAGEPERA AND MATTHEW S. SHUGART

University of California, Irvine CA 92717, USA

The insights gained while carrying out a book-length study of electoral systems are applied to evaluate existing electoral systems and to suggest guidelines for changes, if necessary.

The insights achieved through a thorough analysis of electoral systems in general enable us to evaluate the performance of specific systems. New systems to satisfy specific goals can be devised, and modifications to existing systems can sometimes be suggested.

This does not mean that we recommend far-reaching changes or 'reforms' in numerous countries. A major purpose of elections is to supply a stable institutional framework for the expression of various viewpoints. Even if imperfect, a long-established existing electoral system may satisfy this purpose better than could a new and unfamiliar system, even if it be inherently more advantageous. Those political forces which are disadvantaged by the existing rules learn to live with them, gradually devising strategies that minimize their disadvantages. What disadvantages remain do not come unexpected, and hence the level of frustration is reduced. Familiarity breeds stability. In contrast, introduction of a new electoral system inevitably involves a temporary reduction in stability. Parties, candidates, and voters have to learn new strategies while passing through a period of enhanced surprise, disappointment, and frustration. This may be worthwhile, if the existing system has serious shortcomings and the new one has clear advantages. Our study suggests that this is rarely the case, because polities with different electoral systems learn to achieve the same goals by different means. In particular, over the long run, most plurality systems are not as unrepresentative as their detractors say, and most PR systems are not as unstable as their detractors say. If there is instability or lack of representativeness, the roots often are elsewhere, so that no electoral system could do much about it. Major electoral reforms should not be undertaken lightly.

Some Examples of Unwarranted Changes in Electoral Rules

The 1958 change in French electoral rules (from PR to two-round majority) came in the context of a wider political crisis to which the previous electoral system had contributed. Hence some change in electoral rules probably was inevitable. Maybe a much smaller change than the actual one might have sufficed to alleviate the negative side effects. Maybe a still different system might have been even better. But broadly speaking, a change in 1958 was not unreasonable, and the durability of the electoral system adopted speaks in its favour. The same cannot be said about the French switch to PR for the elections of 1986.

The French Socialists had long promised to adopt PR and, prior to the 1986
elections, they implemented it. A functioning electoral system which contributed to long-term political stability was replaced. The ruling party expected to lose the next elections and wished to minimize its losses of seats. This short-term outcome was achieved, but at the cost of considerable risks to regime stability. The conservative winners of the 1986 elections planned to return to the previous electoral rules. It is hard to say whether such a move will restore the previous stability or contribute to further instability by creating a habit of changing the electoral system at the whim of the moment's majority. Once stability of an electoral system has been disturbed for partisan reasons, it may become difficult to restore stability or even specify which course of action would enhance stability.

Argentina around 1960 offered a spectacular example of an almost spasmodic alternation between two electoral systems. In 1957 it was d'Hondt, but in 1958 and 1960 it was 'list plurality with limited vote' (Nohlen, 1978: 200). In 1963 and 1965 it was back to d'Hondt, followed by 8 years of no elections. Dissatisfaction with each electoral system seemed to build up before the public had time to become familiar with it and start making full use of its inherent opportunities. Nor was there any long-term learning in the sense of experimenting with new rules; the country returned to the rules that had been felt unsatisfactory just a few years earlier.

Over a longer period (1848 to 1945) France underwent similar oscillations between basically two systems. Three periods using two-round elections in single-seat districts alternated with three periods using multi-seat districts which offered the voter a multiple vote. (A few further complications have been omitted. See Campbell, 1958.) Dissatisfaction with the existing system never led to experimentation with single-seat plurality or (until 1945) with single-vote list PR, systems which had been tested in many other countries. Greek electoral rules since 1920 have been similarly spasmodic within a limited number of alternatives.

In Anglo-Saxon countries another set of two alternatives has tended to prevail: the reformers who dislike single-seat plurality usually favour some sort of single transferable vote (STV), disregarding list PR despite its popularity and proven workability in continental Europe.2 In continental Europe, on the other hand, STV has found few adherents despite the intellectual attraction of taking into account the voters' second preferences.

Denmark and Sweden, which introduced list PR in the early 1900s, took the trouble to switch later from d'Hondt to modified Sainte-Laguë allocation formula, a shift which has little effect on seats distribution at large district magnitude (M) and becomes completely pointless when coupled with overriding stipulations regarding nationwide compensation seats and thresholds. The recent (1987) West German shift from d'Hondt to the 'Niemeyer' method (almost simple quota plus largest remainders) also will shift only a couple of seats to smaller parties (Jesse, 1985: 218).

In sum, many changes in electoral rules have not been worth the effort and disruption of stability, in retrospect. Some have been carried out for shortsighted reasons. In many cases, the range of alternatives considered has been much too narrow, and the outcome has at times represented a return to a system with proven disadvantages. The lack of understanding of the workings of electoral systems has at times led to neglect of major factors (magnitude, adjustment seats) and fascination with secondary factors (PR allocation formulas), resulting in reforms which changed almost nothing. On the other hand, some other electoral systems have changed too much, for example, by flipping from \( M = 1 \) to large-magnitude PR or vice versa, when a moderate change in magnitude might have sufficed to do away with the observed shortcomings.
Justified Changes in Electoral Rules

The observations above do not imply that changes in electoral systems are always unwarranted. Electoral reform might be desirable, from a democratic viewpoint, when the existing rules seriously dampen or overamplify or distort changes in popular opinion, or if they are confusingly complex. Amplification here refers to the phenomenon of large parties tending to receive seat shares exceeding their vote shares.

Overamplification means that small shifts in party votes result in huge shifts in their seat shares; it can result from use of plurality rule in excessively few districts. The reverse phenomenon of dampening could result from an allocation formula using a divisor so large that effectively all parties obtain almost the same number of seats regardless of their votes shares. Apart from outright dampening, a system could also underamplify the major parties' votes to the point where coalition formation becomes difficult. Both amplification and dampening are systematic features of the given electoral rules. In contrast, by distortions we mean unsystematic and unpredictable outcomes which contravene the notion that more votes should mean more seats; this is the case when a party with fewer votes obtains more seats or a party increasing its votes share from one election to the next sees its seats share reduced. As for excessively complex rules, they may produce an adequate votes-to-seats conversion pattern but be beyond the comprehension of most voters.

We will start by discussing overamplification, because numerous real cases of obvious overamplification exist. We will not discuss electoral pathologies which do not depend on the explicit electoral rules: blocking the nomination of candidates, electoral fraud, malapportionment, and gerrymander. They may distort the votes cast or the conversion into seats, and reforms might be highly desirable, but these would be largely independent of the rules to convert votes into seats.

Changes to Reduce Overamplification

Overamplification of narrow pluralities in party votes into landslide victories in terms of seat shares often occurs in small ex-British island states in the Antilles, where the phenomenon has been studied by Lijphart (1988), and elsewhere. In these 'mininations' a votes distribution of 55 per cent-45 per cent can result in a seat distribution of 45 to 5 in a 50-seat assembly, which largely does away with any British-type parliamentary opposition functions. The next elections may reverse the votes and hence the seat ratios so that, at most, only 10 previous deputies survive. The loss of continuity in parliamentary style and collective memory is appreciable, affecting political stability.

The reasons for such overamplification are several. The mininations may think that they emulate the United Kingdom by using plurality in single-member districts, but they unwittingly diverge from the model. Even if a minination has an assembly size leading to the usual hinge ratio of \( n = \log V / \log S = 3 \), the outcome is not quite the same as in the United Kingdom, because the random fluctuations around the norm set by the seat-vote equations must not be forgotten. Such fluctuations are much more severe for an assembly of 50 seats than of 500. In the same minination, one election may result in near-proportionality while the next one may yield a disproportionality much more extreme than would be predicted on the basis of \( n = \log V / \log S = 3 \). To make matters worse, very small countries tend to pick assembly sizes below the cube root norm, while the United Kingdom exceeds that norm. Such mininations will have a hinge
ratio appreciably higher than the traditional $n=3$ in their seat–vote equation, resulting in a systematic built-in overamplification. Some mininations exacerbate the effect of low $S$ by using multi-seat districts but still sticking to the plurality rule, which further increases the hinge ratio $n$. Thus Mauritius shifted in 1967 from plurality in 40 single-seat districts to plurality in 20 three-seat districts (Nohlen, 1978: 151). In some cases the dominant party may play such games on purpose to whittle down the opposition, although the stakes are high: if the ruling party ever should lose, it would lose badly. In Mauritius no such purpose seemed to exist, since at least 4 adjustment seats were foreseen to compensate for ethnic disproportionality, plus another 4 adjustment seats for under-represented parties. In practice a total of 10 adjustment seats were added to the 60 regular seats both in the 1967 and the 1977 elections. The outcome was a clumsy two-stage process, whereby the second stage by-passed the election results. The same degree of proportionality could be achieved directly by using a suitable number and magnitude of districts, with a non-plurality seat allocation rule. An even more drastic example of plurality in very few multi-seat districts is supplied by the Seychelles.

Cases of moderate amplification. The simple case of Seychelles glaringly brings out and separates the effects of excessively few districts in absolute terms and compared to the cube root norm, when the plurality rule is applied. The random fluctuation effect becomes negligible if the number of districts is several hundred, as long as there are only two major parties. With three major parties (as in the United Kingdom and New Zealand in the 1920s), the unpredictability of seats on the basis of votes appears even in large nations. The predictable amplification of electoral plurality expressed by the seat–vote equations has many positive features. Cabinet durability is enhanced, while the main opposition party is not severely eliminated from the assembly, provided that the power index $n=\log V/\log S$ is not much larger than 3. If there is a semi-regular alternation of parties in power (as is the case in the United States, the United Kingdom, and New Zealand, among others), there may be even a ‘proportional tenure’ in the following sense: from 1945 to 1979, the UK Conservatives governed for 197 months and Labour did so for 190 months, in a proportion not much different from that of the total votes accumulated during that period, which were 122 million for Conservatives and 124 million for Labour (Taylor, 1984).

This ‘proportional tenure’ does not apply to third parties like British Liberals or the New Zealand Social Credit. Proportional tenure completely fails in India, where semi-regular alternation of ruling parties has not taken place. The representatives of more than one half of the electorate have been almost continuously excluded from cabinet participation. However, the same is true in Japan, which does not use plurality and has no extreme deviation from PR. Even Sweden, with its list PR in fairly large districts, has less ‘proportional tenure’ than the United Kingdom, since the socialist half of the electorate has determined the cabinet much more often than the non-socialist half. The degree of ‘proportional tenure’ seems to be quite independent of the electoral rules used.

The conclusion is that there is no pressing moral or practical reason to shift away from plurality rule in large and homogeneous countries which have practised it for a long time. However, some of them might wish to do so, not only to give ‘fairer’ representation to small parties by reduced amplification but also to avoid excessive unpredictability in the presence of many small parties. In such a case, they would be well advised not to go overboard and adopt extreme PR through the use of large-magnitude districts or nationwide adjustment seats. Two-seat districts (with
STV or list PR) would represent the minimum possible shift from one-seat districts, and yet such a reform would go a long way toward ensuring near-PR representation to third parties such as British Liberals or Alliance. Possibility of gerrymander would be severely reduced. Very small non-local parties would still be excluded, so that the need for a threshold clause is avoided. However, two-seat districts may have an effect akin to ‘bipartisan gerrymander’: both major parties may be assured one of the two seats in the district so that the competition between them is reduced. To guard against such developments, $M=3$ might be preferable to $M=2$. However, a mix of $M=3$ and $M=2$ or 4 must be avoided so as to prevent ‘magnitude gerrymander’. Any larger district magnitudes would represent a needlessly large change. They would bring few further advantages while possibly introducing new disadvantages.

Changes to Reduce Dampening and Underamplification

No actual cases seem to exist where the largest party (or electoral alliance) is systematically disadvantaged by the election rules. Theoretically, such dampening can happen if a list PR formula with a very large divisor is used in the presence of so many parties that their number exceeds district magnitude. Under these rules a party would actually gain by splitting itself up into several ‘parties’ for the duration of elections. Dampening could also arise in two-seat districts, for any non-plurality allocation rule. The actual cases concern underamplification rather than all-out dampening. Even among those who consider near-perfect PR for major parties desirable, there is some consensus that formation of numerous splinter groups should be discouraged. These two divergent goals often make nations adopt rather large district magnitudes or even nationwide adjustment on the one hand but then compensate on the other hand by introducing representation thresholds. Thresholds represent a superfluous complication and the cutoff point chosen is arbitrary. Much the same effect could be achieved by selecting a suitable district magnitude. Of course, selecting a magnitude is also arbitrary, but this is a decision that cannot be avoided, since some magnitude has to be chosen. There is no need to add another level of arbitrariness.

Changes to Reduce Distortion of Electoral Results

Most countries have fairly well-defined proportionality profiles. Regardless of the degree of deviation from PR, the advantage ratio of any party with a given vote share tends to be pretty much the same. In the cases where different parties with the same votes share obtain disparate seat shares, gerrymander, malapportionment, or region-to-region heterogeneities usually are the cause rather than electoral rules as such, and the pattern is still predictable. Only in a few cases does the votes-to-seats conversion seem erratic from one election to the next or offer little apparent relation between vote shares and seat shares, so that the ‘wide scatter’ proportionality profile appears. In such cases the question of electoral reform arises. In general, a slight increase in district magnitude would help.

Changes to Simplify Existing Rules

We should aim at simplicity, unless complications demonstrably yield a more-than-marginally better outcome. In electoral systems as well as in income tax rules and other laws, complexity introduces an elitist inequity of its own, even if the purpose is
increased 'fairness': the more complex a system becomes the fewer people can comprehend it so as to make use of its opportunities.

Some electoral systems offer complexities that baffle even experts outside the particular country: districts, remainder distribution in super-districts in which a party can participate only if it has previously gained such-and-such combination of votes, except when it has such-and-such other redeeming features. The enabling and disabling clauses pile up on top of each other. Yet, when one works through it, most of them cancel out, or the entire elaborate structure is dominated by one last overriding clause: if you did not get your PR share on this, that, or the other level of distribution, you still get it at the nationwide adjustment. It could also work in a reverse direction: at the district level you get the minimum number of seats that any rational allocation rule would allow you, but the allocation of the remainder seats that could go one way or the other is decided by a different rule which excludes you. The latter is the case for elections in France 1986, which often have been described as following a quota allocation rule, because most seats are allocated by quota in districts. However, the final outcome is a pure d'Hondt distribution, because for the crucial though relatively few remainder seats the system reverts to d'Hondt.

Some other electoral systems are so complex that they cannot be shown by theoretical means to be equivalent to a simpler system. However, if one graphs their proportionality profiles, the result is indistinguishable from that of a simple allocation rule applied in districts of a suitable magnitude.

In conclusion, most games with remainder distribution, adjustment seats, and thresholds are not worthwhile. They may have a specific effect on the representation or non-representation of a given small party in a given region at the time the rules are introduced. But demographic and political shifts are likely to alter the picture within ten years, and electoral systems should be designed with more than ten years in mind, if they are to be a factor of stability.

A major result of our systematic quantitative study of outputs such as advantage ratios, proportionality profiles, and deviation from PR is a negative one: many inputs which have been thought to matter do not have much importance. This applies to the precise form of non-plurality allocation procedures in particular, including not only list PR but also STV and even SNTV. Though negative, this is an important finding. It tells us when not to carry out a proposed electoral reform, because the marginal improvement, if any, will be undetectably small in terms of nationwide output. The effect on district-level and intraparty politics may exist. However, the very direction of such effects, if they exist, remains as yet obscure, and hence they cannot supply an informed basis for electoral reform. We are not the first ones to suggest that allocation rules (apart from multi-seat plurality) have little importance, but in view of the attention still paid to them by some other scholars and even more by practising politicians, the point needs to be made once more in the light of new analysis.

On the positive side, a major outcome of our systematic study is confirmation that magnitude matters. In the case of multilayered systems, one has to rely on an 'effective' magnitude rather than the mechanical average of magnitudes of the lowest-level districts. The importance of magnitude has been recognized, of course, by many scholars since at least the 1950s, but actual electoral reforms have often failed to recognize its overriding effect.

Should nations with excessively complex electoral rules simplify them? The very argument against introducing them in the first place is also an argument for keeping them: the improvement would be marginal. The advantages of simplicity would be
outweighed by the disadvantage of unfamiliarity. However, if changes were to be made for whatever external reasons, they should aim at simplifying the rules rather than adding another 'corrective' layer of complexity such as another threshold stipulation or another escape clause from thresholds. Minimal fine tuning of magnitude usually suffices to obtain the desired output.

Unless deviation ($D$) from PR regularly surpasses 10 per cent, it is hard to see why any nation with multi-seat districts and a non-plurality allocation rule should strive for a closer approximation to perfect PR. Some amplification of large party vote shares is desirable for stability, and the philosophical goal of perfect PR remains unachievable—or meaningless, if one merely strove for PR of parties rather than individual views. For those nations which should wish to reduce deviation from PR, a shift to larger district magnitudes is the simplest device. However, such a change might not reduce $D$ as much as intended, since the psychological effect may increase minor parties' votes (law of conservation of $D$).

For the sake of cabinet stability, many countries with multi-seat districts might benefit from somewhat increased amplification of the major party vote shares. If the effective number of parties is large and cabinet durability is low, an increased squeeze on the small parties may be advisable. Often talk to this effect focuses on thresholds, sometimes with special dispensation for territorial minorities. A more elegant solution again would use district magnitude. Lower district magnitudes penalize minor parties without affecting regional parties of comparable vote shares. But once again, electoral reform should not be undertaken lightly, and its effects on the party constellation should not be overestimated. Social Democrat pre-eminence in Sweden and the contrasting multiparty constellation in neighbouring Finland are not caused by differences in electoral systems, and the existing constellations might survive even major changes in the electoral system.

A Much-Discussed System: West Germany

As mentioned previously, discussion of electoral reform in several countries focuses on West Germany as an example to imitate. The attractive aspects of the West German system are the low number of parties combined with a high degree of PR for the parties which surpass the threshold on votes. A disadvantage is the relative complexity of the system. One must also ask whether the low number of parties is caused by the electoral system or by country-specific factors as in Austria.

Despite the extreme proportionality of its outcomes and voters' freedom to use their list vote on fairly small parties without 'wasting' it, the West German system has maintained a party system with effective number of parties 2.5 like typical $M=1$ systems. We might hypothesize that the psychological effect of the single-member districts carries over to the list vote as well, allowing for a near two-party system despite a high effective magnitude.

That there is such a district-level psychological effect despite the large effective magnitude is suggested by Fisher's (1973) result. In 1961, 1965 and 1969, 13 to 38 per cent of voters for the minor party lists voted for one of the larger parties in the single-seat district. The decline in the minor parties' ability to win district-level seats (none has won a district seat since 1957) suggests that the district-level psychological effect of $M=1$ plays a role in bringing about an essentially two-party system, even though ticket-splitting is allowed and some occurs.

The nature of the party system may also be in part attributed to a rather high threshold (5%), below which a party cannot win a seat unless it has obtained at least
Designing Electoral Systems

three district seats. While such a threshold would not shut out an established party of, say 10 to 20 per cent (while the psychological effect of the districts might), the threshold poses a major obstacle to starting a new party.

The degree to which the disproportionalities of the single-member district contests are offset depends on the proportion of the total seats which are compensatory (and the magnitude of the compensatory districts, in the event that the whole country is not one such district, as it is effectively in West Germany). In Mexico, where a somewhat similar electoral system was adopted in 1977, only 25 per cent of the total 400 seats are compensatory (versus 50% in West Germany) and those 100 seats are further divided into five-seat constituencies. Thus the Mexican system does not compensate to a great extent for the disproportionalities of the single-member district contests. The reformed Mexican electoral system merely grants representation (some would say token representation) to parties that could rarely have hoped to have won any seats under the old pure single-member district system, given the overwhelming strength of Mexico's official party. Since parties which obtain more than 60 direct seats are excluded from compensatory seats, potential medium-sized parties, which could seriously challenge the ruling party, do not profit at all from compensation (Nohlen, 1984).

The success of this personalized PR, or as we prefer to call it, compensatory member, system in West Germany has made it an attractive option for many countries considering electoral reform. The system is under consideration in the United Kingdom, New Zealand and Venezuela. Each of these countries has recently had a party system resembling West Germany's but with higher values of $D$, which have led to dissatisfaction with the current rules. However, there is no guarantee that emulation of West German practices would produce the same results.

Electoral Reform in the United States

In the United States there are frequent attempts at representation of minority groups through 'constructive gerrymander', making some districts ethnically non-random. What this means is that some sort of PR is sought within the framework of single-seat districts, but this is a contradiction in terms. Multi-seat districts are unavoidable, if one desires some resemblance to proportionality. Small district magnitude (2- or 3-seat districts) would represent the smallest change in status quo.

This is not a plea that the United States shift to multi-seat PR, but if some form of socio-political PR is desired, multi-seat districts are required. Opposing multi-seat districts because of alleged instability of parliamentary PR regimes is irrelevant to countries with presidential cabinets. Recently, some shift of attitudes regarding PR in the United States can be detected, with non-plurality systems being adopted in some cities and in presidential primaries of both political parties.

Conclusions

We have no emotional attachment to any electoral system, and our systematic quantitative study has made us see disadvantages in all electoral systems. Yet most of the longstanding electoral systems do the job. Keeping the ills we know may be better than jumping into the unknown. On the other hand, our studies have yielded a clearer idea about what features can be independently manipulated and which are locked into reciprocal relationships. We cannot recommend or discommend electoral reform for
any given polity without making value judgements. But we can tell how the outcomes would change or fail to change, if a certain rule were altered.

For new polities which have to choose some electoral rules, we would recommend low-magnitude multi-seat districts (perhaps \( M=3 \) or \( 5 \)—but not \( 4 \)) with some rule other than plurality or limited vote, encumbered with no complexities such as adjustment seats, thresholds, multistage elections, or multitiered seat allocations. Practice has shown that all district magnitudes from \( M=1 \) to \( M=150 \) can work but the extreme values tend to be more conducive to problems. Hence our mild preference for intermediate values. However, if there are many small social segments whose exclusion might threaten regime stability, PR in larger districts may be the only alternative, in spite of the slightly increased cabinet instability which could result.

Notes

1. This article is excerpted from Rein Taagepera and Matthew Soberg Shugart, Seats and Votes: The Effects and Determinants of Electoral Systems, (New Haven: Yale University Press, forthcoming). All subsequent notes containing chapter or appendix citations refer to this book.
2. As a third alternative, interest in West German-style rules has been growing recently in Canada, the United Kingdom and New Zealand (Lijphart, 1987).
3. See Appendix C4.
4. The seat–vote equations, in which \( n \) is an exponent, \( V=\text{total votes} \), and \( S=\text{assembly seats} \), is explained in chapter 14.
5. The tendency of legislatures to be of a size that is the cube root of population is explained in chapter 15.
6. The discussion of Seychelles is omitted in this excerpt.
7. See Appendix C4.
8. Any increase in magnitude beyond \( M=1 \) reduces the probability of ‘manufactured’ parliamentary majorities (based on electoral plurality short of majority) and correspondingly increases the incidence of coalition cabinets. However, the aforementioned examples of Japan and Sweden indicate that single-party cabinets can occur with magnitudes of \( M=4 \) or even much larger. It depends on the pre-existing party constellations and traditions. In surprisingly many cases, lack of parliamentary majorities in stable democracies leads to minority cabinets rather than coalitions; the minority cabinet incidence is as high as 68 per cent in Denmark and Sweden (Lijphart, 1984:61). The same has been the case in the United Kingdom and Canada when neither major party achieved parliamentary majority, and a shift to slightly larger district magnitudes is not likely to change the traditional attitudes.
10. Legal thresholds still differ in two ways from empirical thresholds imposed by low district magnitude: (1) parties which reach the threshold immediately reach proportional representation; (2) the assembly can contain no parties smaller than a specified minimum size.
11. Proportionality profiles of electoral systems and advantage ratios for parties are explained in chapter 7.
12. Effective magnitude, as a modification of district magnitude, is derived in chapter 12.
13. For the index of deviation from proportionality, see chapter 10.
14. See chapter 11.
15. See chapters 8 and 9 for effective number of parties.
16. Of course, the stability of Latin American democracy (PR with presidentialism) has not been great, either. Yet to attribute this to PR would be a misreading of the Latin American experience. If institutional factors can be blamed at all, the finger should be pointed at the timing of elections or presidentialism itself, not at PR. See Shugart, 1988.
References


