MMP-elections and the assembly size

BY Eivind Stensholt

DISCUSSION PAPER

NHH

Institutt for foretaksøkonomi
Department of Business and Management Science

FOR 15/2019

ISSN: 1500-4066
October 2019

29.10.2019

MMP-elections and the assembly size

Eivind Stensholt,
eivind.stensholt@nhh.no mobile (047) 455 27 129
Norwegian School of Economics,
Helleveien 30
5045 Bergen
Norway

Abstract

MMP (Mixed Member Proportional) elections for legislatures have ballots with one vote in a local single seat tally and one vote for a party list in a multi-seat tally.

In Germany, the multi-seat tally occasionally violated a Participation axiom. The federal Constitutional Court declared this unconstitutional in 2008. Rules were changed. In 2017, the result was a Bundestag with 709 members, 111 of them in extra-ordinary party seats.

The paper considers two remedies against excessive assembly size. One is "faithful accounting" of ballot data in each local tally, another a change from Plurality to a Majority method. For this use, we consider IRV, i.e. Instant Runoff Voting, in combination with a 3-candidate Condorcet method. The mayoral IRV election in Burlington 2009 serves as an example, here in the special context of MMP.

Violations of the Participation criterion occur also in the usual Majority methods for single seat elections. The legal adoption of a mathematical axiom from election theory have consequences seen in the context of established impossibility theorems.

JEL classification D72

Key words: Mixed Member Proportional, Instant Runoff Voting, Participation criterion, legality, legitimacy, Burlington election 2009.

Acknowledgement: Thanks to colleague Dirk Schindler for many and useful conversations!

MMP-elections and the assembly size

Introduction

Surveying elections of legislatures, "New Handbook" (Reynolds & al. 2005) classifies 72 cases of “Proportionality”: The number of seats won by a party passing a threshold criterion is roughly proportional to its number of votes nationwide. In multi-seat constituencies, each voter chooses one party list.

There are 91 cases of “Plurality/Majority” in single-seat constituencies. Majority methods pick a winner with at least 50% support, e.g. by a 2-day election or an elimination procedure like IRV (*Instant Runoff Voting*). There are costs in money, time, and voter effort, so the dominance of Plurality is not surprising. Presidential elections are important enough to justify the costs; the fact that 78 countries use a 2-day method (e.g. France) and only 22 use Plurality, indicates that Majority helps to achieve legitimacy.

New Handbook gives 30 cases of "Mixed Systems", defined on p 179 as "A system in which the choices expressed by voters are used to elect representatives through two different systems, one *proportional representation* system and one *plurality/majority* system". Some mixed systems are of the *parallel kind*, consisting of local seat tallies and a party seat tally that are not connected. A different kind is MMP (*Mixed Member Proportional*): A ballot contains two votes; the first supports one *local* candidate in a single seat constituency and the second supports one party. The goal is to make a party’s group in the assembly, *including its local winners*, proportional to the voters’ support for the party.

For its local part, MMP inherits the strategic voting possibilities of its Plurality/Majority method. In a Plurality election, the available strategy is *Compromise*; a voter casts an *“instrumental”* vote for the one deemed best in a narrow set of “electable” candidates, rather than an *“expressive”* vote for the one best liked. Therefore, many MMP voters *split the ballot*, i.e. they support a local candidate who is not from the party they support. Ballot splitting is important for local representation with majority support. A ballot may also be split in order to change the party vote, e.g. to keep a political ally above a threshold criterion. Mudambi and Navarra (2004) discuss sides of this. Linhart & al. (2019) survey, empirically, how different variations of the MMP-theme perform. Mixing rules vary, and

may have unwanted and unexpected consequences. The Bundestag MMP election 2017 was according to new rules, adapted to decisions in the German Constitutional Court. However, instead of getting 299 Plurality winners and 299 party representatives in the 598 *ordinary* seats, the assembly got 709 members, 111 of them in *extraordinary party seats*, mainly due to the treatment of split votes in the tally. If 100000 more voters, who locally supported a CSU winner, had moved their party vote from CSU to FDP (say), the assembly would have grown again, by (2.3), from 709 to 734 members. A proposed *"faithful accounting"* (1.4) avoids this effect of natural ballot splitting. For MMP without split ballots, Csato (2016) studied some vote transfer ideas in order to get closer to overall proportionality.

However, if all parties are small, excessive assembly size may still occur. A Majority method in local tallies is essential to keep the size within bounds.

Central in the development towards the present rules was the first decision of the Federal Constitutional Court (2008). The court referred to a principle of *equal voter influence*, (5.3) below, which is so general that one may be in doubt how to test a given election method with it. However, in some voting systems it happens that addition of a set of equal ballots to a tally leads to a new result that, *according to the ballots added, is worse than the old result*. A particular form is *"negatives Stimmgewicht"*, which sometimes made a small difference in the number of seats for a party. The court found this was unconstitutional.

In election theory, mathematical axioms often have a *"normative"* flavor, but *"impossibility theorems"* show that in certain axiom sets, not all can be satisfied. The succinctness of axioms makes a *legalistic* approach understandable, but a consequence may be that it makes *legitimacy* harder to obtain. This is a theme for the last parts of the paper.

IRV (Instant Runoff Voting) and Baldwin's Condorcet method are Majority methods and natural choices for local MMP tallies, but they allow a "NoShow" paradox akin to the *"negatives Stimmgewicht"*. They have similar properties that may bring them in conflict with the decision from 2008. Variations of MMP are used or have been tried or proposed in other countries. How MMP performs in Germany, where it originated, is obviously important in discussions about its adoption.

1. MMP – Mixed Member Proportional

Notation **The parties are $Z_1, \ldots, Z_t$, of which $Z_1, \ldots, Z_k$, $k \le t$, pass a threshold criterion;**

(1.1) $\Lambda(Z_j)$ is the set of voters with Zweitstimme to party Z_j (in the whole nation);

$\Gamma(A, Z_j)$ is the subset with Erststimme to candidate A (in the constituency of A).

Sometimes a party does not nominate a candidate in a constituency. Also in other cases, many voters *split their votes*: One vote, the *Erststimme* then supports a local candidate A not from the party Z_j that they support with their other vote, the *Zweitstimme*. Thus, a local seat winner W has an *electoral basis*, Bas(W), which is a disjoint union:

(1.2) $$\text{Bas}(W) = \Gamma(W, Z_1) \cup \Gamma(W, Z_2) \cup \ldots \cup \Gamma(W, Z_t)$$

This structural feature has no rôle in today's tally rules (2019), i.e. (1.3) and (1.5) below. It is essential in the alternative tally defined by (1.4) and (1.5).

REMARK (1.1) **A local winner W may be candidate for a party Z_j with $k < j \le t$. In (1.2), an independent local winner W belongs to an "imagined party" Z_j with empty $\Lambda(Z_j)$. The threshold criterion is to obtain 5% of the Zweitstimme or to win 3 local seats.**

The Zweitstimme are intended to achieve a nationwide distribution of seats "proportional" to the size $|\Lambda(Z_j)|$ of $\Lambda(Z_j)$, $1 \le j \le k$. A suitable "accounting principle" is essential.

Presently (2019), a local winner W from party Z_j is recorded on an account for $\Lambda(Z_j)$:

(1.3) ***Present accounting:***

W's local seat counts as one full seat won by $\Lambda(Z_j)$ if W comes from Z_j, $1 \le j \le t$.

However, all voters in Bas(W), (1.2), give one Erststimme to W. Thus, an alternative is:

(1.4) ***Faithful accounting:***

A fraction $|\Gamma(W, Z_j)|\cdot|\text{Bas}(W)|^{-1}$ of W's seat counts as won by $\Lambda(Z_j)$, $1 \le j \le t$.

Alternative (1.4) records the reality when voters from different $\Lambda(Z_j)$ support a local winner W. With Erststimme and Zweitstimme in the same ballot, it requires no change in the voting procedure. In each constituency, the tally tabulates the combinations of Erststimme and Zweitstimme. Example (1.1) shows the local seats $D(Z_j)$ according to (1.3).

Accounting rules (1.3) and (1.4) give two different definitions of $D(Z_j)$, the number of (shares of) local seats which the Erststimme tally records on $\Lambda(Z_j)$'s account, $1 \le j \le t$. $P(Z_j)$ is the non-negative number of party seats that the Zweitstimme tally awards to $\Lambda(Z_j)$, $1 \le j \le k$. The central requirement in MMP is (1.5):

(1.5) ***Proportionality:***

The ratio ρ = Zweitstimme per seat (local- and party-) for parties passing the threshold is:

$$\rho = |\Lambda(Z_j)| \cdot [D(Z_j) + P(Z_j)]^{-1},\ 1 \le j \le k; \quad P(Z_j) = 0 \ \text{ for } k < j \le t.$$

Round-offs to integers and geographical allocation of party seats are not discussed here; solutions (ρ, $P(Z_1)$, …, $P(Z_k)$) of (1.5) are in non-negative reals. If ρ decreases, properly increased $P(Z_j)$, $1 \le j \le k$, give another solution. Since $P(Z_j) \ge 0$, an increase of ρ must stop:

(1.6) **maximal ρ occurs when $P(Z_j) = 0$ for some party Z_j, $1 \le j \le k$**

More explicitly, choose Z_m, $1 \le m \le k$ so that

$$D(Z_m) \cdot |\Lambda(Z_m)|^{-1} = \max_j D(Z_j) \cdot |\Lambda(Z_j)|^{-1},\ 1 \le j \le k$$

In Example (1.1), Z_m = CSU. Rewrite (1.5) as

$$\rho^{-1} = P(Z_m)\cdot|\Lambda(Z_m)|^{-1} + D(Z_m)\cdot|\Lambda(Z_m)|^{-1} = P(Z_j)\cdot|\Lambda(Z_j)|^{-1} + D(Z_j)\cdot|\Lambda(Z_j)|^{-1},\quad 1 \le j \le k$$

Here $P(Z_m)\cdot|\Lambda(Z_m)|^{-1} \le P(Z_j)\cdot|\Lambda(Z_j)|^{-1}$. Thus, if $P(Z_m) > 0$, then $P(Z_j) > 0$, $1 \le j \le k$. Therefore, the number of party seats may be reduced until $P(Z_m) = 0$:

(1.7) **At $P(Z_m) = 0$, ρ obtains its maximal value $\rho^* = |\Lambda(Z_m)|\cdot D(Z_m)^{-1}$**

"Critical" assembly size **There are $S = \Sigma_j [D(Z_j) + P(Z_j)]$ seats, $1 \le j \le t$, and $\rho = \rho^*$ gives the *"critical"* assembly size S^*, i.e. the smallest possible size that allows proportionality (1.5):**

(1.8) $$S^* = \Sigma_j [D(Z_j) + P(Z_j)],\ \text{with } P(Z_m) = 0$$

A seat distribution is above critical size when all $P(Z_j) > 0$ for $1 \le j \le k$. S is chosen as small as possible, provided that S_o "ordinary seats" are occupied. Thus,

(1.9) **$S = \max\{S_o, S^*\}$, and if $S^* > S_o$, then $S^* - S_o$ "extra-ordinary" party seats are created.**
If $S^* < S_o$, then $S_o - S^*$ more ordinary party seats are distributed,
while the ratios in (1.5) are reduced but kept equal.

EXAMPLE (1.1) **Germany uses accounting rule (1.3) and the Bundestag has $S_o = 2D = 598$ ordinary seats. In 2017, k = 7 Zweitstimme-groups $\Lambda(Z_j)$ won seats. The data relevant for our discussion are in the table below.**

Z_j	ERSTSTIMME		SEATS			ZWEITSTIMME		Zwst/tot
	%	#	LOCAL	TOTAL	PARTY	#	%	
CDU	30.2	14030751	185	200	15	12447656	26.8	62238
SPD	24.6	11429231	59	153	94	9539381	20.5	62349
AfD	11.5	5317499	3	94	91	5878115	12.6	62533
FDP	7.0	3249238	0	80	80	4999449	10.7	62493
Linke	8.6	3966637	5	69	64	4297270	9.2	62279
Grüne	8.0	3717922	1	67	66	4158400	8.9	62066
CSU	7.0	3255487	46	46	0	2869688	6.2	62385
	97-	44966765	299	709	410	44190959	95-	

REMARK (1.2) In elections CSU operates only in Bavaria and CDU only in the 15 other states; in the Bundestag they work as one party.

REMARK (1.3) There are many split ballots. Of them,

(1.10) *at least* 44966765 – 44190959 = 775806 give Erststimme to Z_a and Zweitstimme to a party Z_b that failed to pass the threshold, i.e. $1 \le a \le k < b \le t$.

Voters who give Zweitstimme to a very small party, are naturally motivated to split their ballot; with Erststimme to a frontrunner, they do not "waste their vote". When a party narrowly fails to pass the threshold, we must expect much larger differences than in (1.10).

With accounting (1.3), it is possible but unlikely that column 3 has sum less than D (= 299). Faithful accounting (1.4) give sums below D. A conservative, rough proportionality estimate in Example (1.1) is based on (1.10):

(1.11) *At least* 299·775806/44966765 ≈ 5.2 local seat shares are on $\Lambda(Z_b)$-accounts, $k < b$

Thus, faithful accounting (1.4) gives *at most* 299 – 5.2 = 293.8 seats in column 3.

REMARK (1.4) By (1.2), local winners get cross-party support from split ballots. One party supported by center-right (CDU/CSU) and one by center-left (SPD) dominate Plurality tallies and stabilize political activity as one of them becomes the core of a government coalition. Zweitstimme elect also representatives who promote political ideas with relatively small electoral support. However, 111 extra-ordinary party seats violate the intended balance between the two kinds of representation. Our analysis of the table shows how the present accounting rule (1.3) easily can result in many more extra-ordinary party seats.

2. *Tally under present accounting rule (1.3)*

First, a Plurality tally of the Erststimme in each constituency establishes column 3 with the number of local seats won by each $\Lambda(Z_j)$. Next, the Zweitstimme determine column 4 according to (1.5). The numbers of party seats in column 5 follow by subtraction.

The Zweitstimme tally establishes column 6 and calculates, for each party, the ratio of Zweitstimme per local seat won.

For CSU this ratio is 2869688/46 ≈ 62384.5.

For all other parties, it is larger. Thus, CSU = Z_m in the notation above. By rule (1.5), all other parties reach the same fraction as CSU:

(2.1) At critical assembly size S* all parties "pay a price" of 62384.5 Zweitstimme per seat.

With its Zweitstimme and its number of local seats, CSU = Z_m sets this price.[1] Thus, before round-offs, all Zweitstimme together pay for an assembly of size

(2.2) $$S = S^* = 44190959/62384.5 = 708.4 \text{ seats.}$$

The official tally used Sainte-Laguë's method to calculate column 4; with supplementary data, it also distributes seats among the member states. Column 8 reflects the accuracy of Sainte-Laguë (1910) in achieving proportionality with round-offs to integers.

Destructive strategic vote splitting under rule (1.3)

The table hints that many split ballots give Erststimme to CDU/CSU and Zweitstimme to FDP. Supporters of CDU/CSU expect that they cannot win many party seats. Intending to help FDP, a potential ally that did not pass the threshold in the previous election of 2013, they move to Λ(FDP). Imagine that x more with Erststimme to CSU (in the table) do so. The Erststimme tally will not change. CSU keeps 46 local seats, but the price that all parties pay, drops from 62384.5 to

$$\rho = \frac{2869688 - x}{46} = 62384.5 - \frac{x}{46}$$

The critical assembly size increases from 708.4 to

(2.3) $$S^* = \frac{44190959}{62384.5 - x/46}$$

---

[1] In a study of consequences of imagined changes, e.g. in vote-splitting, one must check whether they cause another party to get the Z_m-rôle; the table shows that switches between CSU (without party seats) and CDU (with only 15 party seats) may occur.

If the vote splitting is popular and $x = 1434844 = 2869688/2$, then the price ρ is cut in half, S^* doubles to 1416.8, and the relative strength of CSU in the Bundestag is cut in half.

One must assume that vote-splitting in order to help FDP with more Zweitstimme started long before election, e.g. from $|\Lambda(\text{CSU})| = 2869688 - x$ with $x < 0$:

If $x = -225442$, $\rho^* = 67285$ and critical size $S^* = 44190959/67285 = 657$ seats.

REMARK (2.1) At this spot, i.e. $\rho^* = 67285$, the rôle of Z_m switches between CSU and CDU, since also $67285 = 12447656/185$; see the table. Further back, some voters with Erststimme to CDU and some voters with Erststimme to CSU may already have split their votes and moved into Λ(FDP). At what ρ-value it really started, is anybody's guess.

REMARK (2.2) Vote-splitting where voters in Λ(FDP) change Erststimme and give it to CDU/CSU may have helped CDU/CSU to get some of their local victories, also resulting in a decreased ρ and an increased critical size S^*; see the table.

REMARK (2.3) Counted as one party, CDU/CSU = Z_m, with $12447656 + 2869688 = 15317344$ Zweitstimme and $185 + 46 = 231$ local winners, gets ratio $15317344/231 = 66309$. Critical size is $S^* = 44190959/66309 = 666$ seats. S^* is reduced by $709 - 666 = 43$ party seats, 15 from CDU and 28 from others. CDU is stronger in some states than in others and got party support even where CDU's local candidate did not win. A strong single-state CSU got the Z_m-rôle.

Legitimacy v Legality

Candidacy for an office Properties of election methods tend to raise questions about both legality and legitimacy, perhaps at the same time. In this paper, legitimacy includes the voters' general acceptance of methods and results as being fair. Hettlage (2018) is concerned about the following aspect: *Some seats go to a politician who was not candidate for it*. One may well see "degrees of candidacy": A local candidate runs for a unique office as the representative of the constituency. Party lists name candidates for the 299 ordinary party seats, although voters cannot express preferences inside a party list. However, the 111 extra-ordinary seats did not even exist before the election. Nobody could register as candidate for any of them. We should elect our representatives for specific offices. In Hettlage's words: *"Without candidate, no mandate."* The more citizens think that extra-

ordinary party seats raise a question of legitimacy, the more they do so. Whether they also raise legal questions is perhaps too much of a "technical" issue for voting laypersons.

The electoral basis **Who "owns" a local winner W? Disregarding both conceptuality and essential realities, the present accounting (1.3) declares, de facto, that Λ(Z), or even Z, is the sole "owner" when W is nominated by party Z. Conceptually, every Erststimme for W may come in split ballots, e.g. if W is independent (no vote from Λ(Z), Remark (1.1)). Realistically, a popular local candidate from a small party Z in the political center may win (with Λ(Z) giving just a small contribution). With accounting rule (1.4), W's knowledge of Bas(W), defined in (1.2), should influence W's relations with the electorate and with Z.**

A consequence of the ownership notion in accounting (1.3) is that normal voter behavior, intended to help a small political ally, explains most of the 111 extra-ordinary seats in 2017. We must assume that most voters involved, supporting a local CSU-candidate but moving from Λ(CSU) to Λ(FDP), say, were not aware that the tally would react by blowing up the assembly size and reduce CSU's relative size. When voters receive and digest this information, what will it mean for the legitimacy of the tally rules?

Neglected ballot information **The present accounting rule (1.3) makes a very weak connection between Erststimme and Zweitstimme. One may as well cut each ballot in half, and collect the Erststimme and the Zweitstimme in separate boxes: The tally is the same. A vast amount of ballot information now neglected, can be used to help keeping S* within bounds.**

3. *Tally under faithful accounting rule (1.4)*

Faithful accounting (1.4) makes use of more ballot information than the present rule (1.3) does. It records the composition of Bas(W) in (1.2) and produces a different column 3 for local seats in the table. This requires a much more thorough handling of the ballot data; see also Remark (1.3).

Fortunately, special features in the table allow a reasonable estimate of the local entry for Λ(CSU) under faithful accounting (1.4): CSU runs *only* in Bavaria, and CSU-candidates win *all* 46 local seats, with *precisely* 3255487 Erststimme, and there are no other local winners that Λ(CSU) may have shares in. Let CSU get u Zweitstimme in split ballots:

(3.1) CSU receives both Erststimme and Zweitstimme from 2869688 – u voters.

Most likely, few voters find reason to support CSU with Zweitstimme but not with Erststimme; we may expect u to be small. Assume there are roughly the same number of voters in all constituencies. Λ(CSU)'s own local seat fractions add up to approximately

(3.2) $$\frac{2869688 - \mathbf{u}}{3255487} \cdot 46 = 40.5487 - \mathbf{u} \cdot 0.00001413$$

This number replaces 46 in the table's column 3.

Λ(CSU)'s ratio of Zweitstimme to this new entry is

(3.3) $$\rho = \frac{2869688}{2869688 - \mathbf{u}} \cdot \frac{3255487}{46} \approx \frac{2869688}{2869688 - \mathbf{u}} \cdot 70771$$

It seems safe to assume that CSU has the Z_m-rôle also under rule (1.4) and sets the price (slightly) above 70771.[2] With the new column 3, rules (1.4) and (1.5) give a critical size S*:

S* ≤ 44190959/70771 ≈ 624.4 seats

However, here the proportionality rule (1.5) concerns only those (estimated) 293.8 shares of local winners that are won and "paid for" by the $\Lambda(Z_a)$ with a $\leq$ k; see (1.11). Adding the 5.2 local seat shares won by the $\Lambda(Z_b)$ with k < b, the estimate is

(3.4) S* ≈ 629.6 seats

---

[2] **CDU's 15 party seats is a measure of CSU's local advantage. CDU keeps its own shares of 185 local winners, but its gained shares are in a much smaller number of constituencies; CDU may still need party seats. .**

Vote splitting that turns from destructive to beneficial under rule (1.4)

If x more voters who give Erststimme to CSU now split their vote by moving from Λ(CSU) to Λ(FDP), they also move their shares in local winners to the account of Λ(FDP). They reduce $|\Lambda$(CSU)$|$ to 2869688 – x; under rule (1.4) the new ratio for CSU is a modification of (3.3):

(3.5)
$$\rho = \frac{\mathbf{2869688 - x}}{\mathbf{2869688 - x - u}} \cdot \mathbf{70771}$$

Thus, if CSU = Z_m also after the action, the new price increases (slowly) with x, and S* decreases. At some stage, CSU may lose the Z_m-rôle, most likely to CDU. In the table, however, CDU does not have the same special features as CSU; e.g. what shares in the local seats won by SPD-candidates will faithful accounting (1.4) record on Λ(CDU)'s account?

REMARK (3.1) **Under faithful accounting (1.4), the strategic vote splitting considered above increases $|\Lambda$(FDP)$|$. This may well help the recipient to pass a threshold. However, in the calculation of FDP's number of party seats, an increased $|\Lambda$(FDP)$|$ is counteracted by an increased account for Λ(FDP) in column 3.**

4. *Faithful accounting with a 50% requirement*

Splitting ballots is natural and acceptable in MMP, e.g. when the purpose is to help a political ally to pass the Zweitstimme threshold. When voters from several Zweitstimme-groups $\Lambda(Z_j)$ join force and support one of two local front-runners in the Plurality part, many will even see it as laudable (Dowding and VanHees 2008).

The reduction of S* from 708.4 to 629.6 in the election of Example (1.1) indicates that faithful accounting (1.4) is a useful remedy against an intolerable assembly size which, under the present rule (1.3), may be due to vote-splitting.

However, with faithful accounting (1.4), voters who support party Z, but split their vote and support a local winner W not from Z, will increase the account for $\Lambda(Z)$ in column 3; thereby they counteract their own Zweitstimme. A voter may perceive this as a dilemma, but with faithful accounting, it is still a rational choice for many to give Erststimme to the most "electable" local candidate A who in the Bundestag will cooperate with Z. Helping A, a small number of Z-supporters may decisively increase the size of a wider political camp. If A does not win, their Zweitstimme is a fallback security with real support for Z.

Faithful accounting (1.4) harnesses split ballots for a well-functioning MMP. However, intolerably high S* may still be due to an unfortunate development of the party structure, where vote-splitting and faithful accounting become insufficient remedies:

EXAMPLE (4.1) **Assume that eight parties have 10% each of the Zweitstimme, and one party has 20%, all uniformly distributed. Unless massive splitting of ballots helps a smaller party, the 20% party takes all D local seats and the rôle of Z_m. By proportionality, each 10% party gets D/2 seats. Critical Bundestag size becomes S* = 5D = 1495.**
Generally, local winners come from center-right or center-left; that is one purpose of MMP. This gives supporters of small parties an incentive to cast their Erststimme for center-right or center-left, Remark (1.4). However, if less than 50% of the Zweitstimme are payment for local seats, the remaining purchasing power may create more than D party seats. Even with faithful accounting (1.4), one cannot expect the incentive to be strong enough to avoid this: The 10%-parties easily pass the threshold, and it may seem pointless [hopeless] to split ballots in order to support [challenge] a local candidate from the 20%-party.

Majority methods

MMP is an ambitious and noteworthy attempt to include a large single-seat component in a proportional representation of those $\Lambda(Z_j)$ that pass the threshold. For legitimacy, local representatives ought to get a majority; to this MMP adds the importance of a tolerable S*.

With single-seats only, as in most Plurality elections, many voters avoid "wasting their vote", disregard accusations of "favorite betrayal", and in all sincerity cast an "instrumental vote" for a frontrunner. One expects the winner will receive substantial support.

With faithful accounting (1.4), MMP still gives an incentive to vote instrumentally with the Erststimme, but the Zweitstimme is a fallback security, which weakens the incentive.

A Majority method is likely to give a high price ρ, and $S^* \le S_o = 2D$ should become normal. To achieve wide distribution of the bills with the present kind of ballots, a very simple Erststimme tally method may be considered:

(4.1) ***A Majority tally for Plurality ballots*** **Let W and R be winner and runner-up in a local Erststimme Plurality tally. A ballot without Erststimme to W or R then counts as half a ballot with Erststimme to W and half a ballot with Erststimme to R.**

In a post-election study of Example (1.1) with its local Plurality data, method (4.1) seems particularly suitable. Perhaps it may also be useful in real elections: It gives voters an incentive to choose one of (usually) two frontrunners. A supporter of W has joined with more or less likeminded voters and strengthened their side, while supporters of R keep full real weight of their Zweitstimme. De facto, a "know-not" in $\Lambda(Z)$ skips the Erststimme part of MMP; recording half a voter's share in W on the account of $\Lambda(Z)$ is a "taxation" to spread the bills for local winners and to keep S* within bounds.

(4.2) ***"Conventional" Majority tallies*** **Ranked ballot methods obtain a majority, but in many cases, voters are not obliged to choose one of the n! strict rankings of all n candidates. Often they are only required to rank c candidates as 1, 2, …, c, while n–c are omitted.** ***Symmetrizing*** **an incomplete ballot is equivalent to replacing it by (n-c)! ballots with all possible extensions to n candidates, each of weight 1/(n-c)!. Method (4.1) is IRV with symmetrizing.**

A real case is the IRV-election in Example (4.2); the *"pictogram"* is a spatial model of the election. It is also convenient to reshape an n-candidate ballot from voter v as an nxn-matrix Con(v): 1 in entry (i, j) indicates that v ranks candidate i ahead of candidate j; other entries are 0. With n=3, we may have, e.g.

$$\mathrm{Plu}(v) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}; \quad \mathrm{Con}(v) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}; \quad \mathrm{Bor}(v) = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

Plu(v) and Bor(v) are auxiliary columns derived from Con(v); they show the top-ranked candidate and the *"Borda-points"* awarded by v (in general n–1, n–2, …, 0).

In the penultimate round, we symmetrize all 8833 remaining ballots in Example (4.2), Burlington 2009. Candidates are in order IRV-winner (K), runner-up (W), and eliminated (M):

$$\text{(4.3)} \quad \begin{matrix} K \\ W \\ M \end{matrix} \quad \mathrm{Plu} = \begin{bmatrix} 2982 \\ 3297 \\ 2554 \end{bmatrix}; \quad \mathrm{Con} = \begin{bmatrix} 0 & 4541.5 & 4121.5 \\ 4291.5 & 0 & 3952 \\ 4711.5 & 4881 & 0 \end{bmatrix}; \quad \mathrm{Bor} = \begin{bmatrix} 8663.0 \\ 8243.5 \\ 9592.5 \end{bmatrix}$$

Bor is the sum of columns in Con. Although Plu(v) is trivially obtained from Con(v), Plu cannot be obtained from Con. Information is lost in data aggregation.

Both IRV (Instant Runoff Voting) and Baldwin's Condorcet variation (Baldwin 1926) are elimination methods that tally ranked ballots.

- **IRV eliminates the candidate with the smallest number of top-ranks in Plu.**
- **Baldwin eliminates the candidate with the smallest Borda-sum in Bor.**

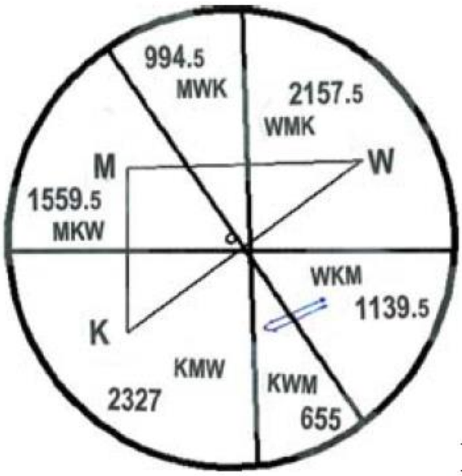
The matrix Con illustrates that a Condorcet winner scores above 50% in each pairwise comparison, and therefore above average Borda score; thus, Baldwin never eliminates a Condorcet winner. Both methods lead to a final tally round with two frontrunners, where the winner obtains a majority.

A common claim is that such methods require significant extra expense or effort. However, Australia has long experience with IRV. One Australian remedy to ease the voter's burden is the practice that parties issue "how-to-vote cards" suggesting a continued preference ranking to their own supporters. If such cards, with a standard format, become mandatory and perhaps even a default choice in the Erststimme part of the MMP ballot, it is hard to see these claims as important objections.

EXAMPLE (4.2) **In the Mayoral IRV-election of Burlington, Vermont, 2009, nine possible ballot types were in the tally when k = 3 candidates remained[3]. Let X** indicate ballots with top-rank to X but no ranking of the other two. Arrows illustrate the symmetrizing:**

MWK	M**	MKW	KMW	K**	KWM	WKM	W**	WMK
767	↙455↘	1332	2043	↙568↘	371	495	↙1289↘	1513
994.5		1559.5	2327		655	1139.5		2157.5

A unique "circle pictogram" (Stensholt 1996) shows the political landscape:



Burlington 2009. **Three chords define a small triangle T covering 0.000098 of the circle area. T is usually so small that, like here, a *candidate triangle* △MWK may be adapted with perpendicular bisectors close to the chords. Six areas are proportional to the sizes of the new categories, with full voters and "half-voters". The pictogram is a *"spatial model"*: Each voter ranks the candidates according to their distance from the voter; here the model includes half-voters. Sometimes it is essential to consider full voters and half-voters separately. In particular, only 495 voters put themselves in the WKM-area, where half-voters count for 1289/2. With top-ranks (|M|, |K|, |W|) = (2554, 2982, 3297), M was eliminated; finally K defeated W.**

REMARK (4.1) **Symmetrizing gives the same result as the official tally, which discarded 455 ballots that were emptied after elimination of M. A pictogram illustrates the political landscape, and half-voters are included for verisimilitude. By showing both top-ranks and pairwise comparisons, a pictogram shows the 3-candidate IRV tally and the Condorcet relation of pairwise comparison.**

[3] **Data from Rangevoting.com. After elimination of other candidates, the three major candidates remained: M = Montroll (democrat); K = Kiss (progressive); W = Wright (republican).**

REMARK (4.2) **The same IRV-tally is obtained with all vote vectors of the form**

$$(|MWK|, |MKW|, |KMW|, |KWM|, |WKM|, |WMK|) = (994.5,\ 1559.5,\ r,\ 2982 - r,\ 3297 - s,\ s)$$

Thus, the IRV tally ignores subsidiary rankings from supporters of the two finalists, here K and W. In Australia, such a vector with unpublished (r, s), is also equivalent to the standard tally report, but with integers when X-ballots, as usual, are not allowed.**

Condorcet cycles "K beats W beats M beats K" occur for $r < 1862.5 < s$. In the pictogram of a cyclic election, the voter-free triangle T covers the circle center. Since T, empirically, is very small, cycles are very rare in ordinary elections, but overrepresented in "Monte-Carlo" simulations of IRV using two popular probability distributions, IC (*Impartial Culture*) and IAC (*Impartial Anonymous Culture*), e.g. (Smith 2010 – 2016).

In *"Perfect Pie-sharing"* T has shrunk to a point; the locus is a curve in the (r, s)-plane. Along the curve, Condorcet's relation of pairwise comparisons must be transitive, since it ranks candidates by proximity to the center. Perfect Pie-sharing generalizes Duncan Black's *"Single-Peak"* condition (Black 1948), which is satisfied with (r, s) = (2982, 3297); then T is a point on the periphery of the pictogram.

REMARK (4.3) The Burlington election illustrates "*monotonicity failures*" and is close to a "*NoShow paradox*". Adducing these as flaws, some critics attack the use of IRV. In 2008, the Federal German Constitutional Court declared an analogue in MMP to be unconstitutional.

Monotonicity failures

Being closest to the pictogram center, M is Condorcet winner, i.e. M beats both W and K in pairwise comparisons. In IRV-elections, a Condorcet winner will, *if not eliminated*, also become IRV winner. Obviously, IRV never eliminates a candidate with above 1/3 of the top-ranks. The main ingredient of a 3-candidate *"monotonicity failure"* in a non-cyclic election is a Condorcet-winner M with less than 1/3 of the top-ranks. In general, *non-monotonic events* come from anti-M voters, in Burlington moving between KWM and WKM.

The vote vectors obtainable by such action, with $-655 \le x \le 1139.5$, are:

$$(|MWK|, |MKW|, |KMW|, |KWM|, |WKM|, |WMK|) = (994.5,\ 1559.5,\ 2327,\ 655 + x,\ 1139.5 - x,\ 2157.5) \quad (4.4)$$

Tiebreaks in the IRV-tally happen, by (4.3), at x-values

$-428 = 2554 - 2982$; $-125 = (4291.5 - 4541.5)/2$; $743 = 3297 - 2554$

(4.5) Anti-M voters succeed in eliminating M when they keep $x \in (-428, 743)$:

$x < -428$: K is eliminated, finally M wins over W;
$-428 < x < -125$: M is eliminated, finally W wins over K;
$-125 < x < 743$: M is eliminated, finally K wins over W;
$743 < x$: W is eliminated, finally M wins over K.

At the midpoint of $[-428, 743]$, $x = 157.5$, W and K are joint Plurality winners. When x passes an end-point, but not –125, there is a *"non-monotonic effect"*. There are two kinds, *tricks* and *traps*. Vote vectors on each side are *"monotonicity failures".*

***A "trick":* If x passes one endpoint of $[-428, 743]$ and ends inside, M loses the IRV-victory by being eliminated. Either: tricky WKM-voters switch to KWM, when x passes – 428, W wins. Or: tricky KWM-voters switch to WKM, when x passes 743, K wins.**

***A "trap":* This is a trick in reverse; x passes one endpoint and ends outside $[-428, 743]$. Either: some KWM-voters give away top-ranks to winner W, and M gets into the final. Or: some WKM-voters give away top-ranks to winner K, and M gets into the final (indicated in the pictogram).**

REMARK (4.4) ***Lack of motivation among anti-M voters*** **WKM-voters who feared M and expected W to lose in the final, might get the unfortunate idea to obtain or consolidate K's victory by switching to KWM. However, there was no danger that they would become "trap victims" and pass the "brink" at x = 743; see the arrow in the pictogram. Those who count as half-voters in the anti-M areas could hardly be motivated to change their top-ranks:**

> **Only 495 W-supporters stated a subsidiary preference for K over M. Thus, in order to reach x = 744, *at least* 249 (= 744 – 495) W**-voters, *who did not care to distinguish between the leftists,* K *and* M, would still have to follow the arrow in the pictogram and *rank* K *ahead of* W *in their ballots.***

The brink at x = 743 is therefore too far away to be passed with x = 0 as start point. Similarly, the other brink at x = – 428, with switched rôles for K and W, is too far away.

REMARK (4.5) **Even if an election is much closer to a brink than Burlington 2009 was, a non-monotonic effect is very unlikely. The Condorcet-winner and another candidate, (here M and K) may have a close fight to become challenger of the Plurality winner (here W), and the sign of $|M|-|K|$ is decisive. Obviously,**

[|M|+|K|+|W|]/3 > |MWK|+|MKW| > |WKM|+|KWM|

Thus, relatively few anti-M voters influence |M|−|K| in a non-monotonic effect, and their motivation to join a Pushover action is naturally weak, Remark (4.4). Many more anti-W voters influence |M|−|K| in the fight for a place in the final. Moreover,

one anti-W voter, v, moving from KMW to MKW,
neutralizes two voters **moving from WKM to KWM.**

If v consider M as likely Condorcet winner and therefore best challenger to W, v's move shows the idea behind the *"Compromise strategy"*. It is the most common form of strategic voting, widely practiced in Plurality elections. Moreover, v comes from a large "reservoir" of naturally motivated full KMW-voters (2043 of them in the Burlington pictogram).

Purely random events always occur. The difference, |K|−|M|, is more exposed to random events changing the number of ballots with top-ranks for K or for M than to random events concerning the few anti-M ballots. (In Burlington 2009, the two last tally rounds were so clear that random last-day changes of the vote vector could not have mattered.)

The Supreme Court of Minnesota (2009) rejected a claim that non-monotonicity made IRV unconstitutional:

> ***"Although it is apparently undisputed that the IRV methodology has the potential for a non-monotonic effect, there is no indication, much less proof, of the extent to which it might occur, and so there is no way to know whether the alleged burden will affect any significant number of voters."***

To prove that a non-monotonic effect has occurred, one must prove that the vote vector really had been different from the one established in the tally, and that sufficiently many voters ranking a Condorcet winner last really had moved from one component of the vote vector to another. That is a hopeless task. Moreover, the court disregarded that the natural mechanism behind Compromise dominates non-monotonic effects, but presumably, the court was restricted to consider claims and arguments from the two sides. For another comment to theoretic possibilities and lack of empiricism, see (5.7) below.

Wasted votes

There are obvious circumstances in favor of the winner K in Burlington: The anti-W voters clearly chose K to challenge W, and K's victory, was an unusually didactic demonstration of IRV's dual emphasis on primary support (K beats M in top-ranks) and on subsidiary support

(K beats W in Borda-points and Condorcet's relation). However, in IRV "two silver" is *always* better than "one gold and one bronze", and "always" seems to be too often.

Electoral Reform Society (UK) compares IRV with Plurality [4]:

"Voters can vote for their favourite candidate
without worrying about wasting their vote."

Here, advocates of IRV emphasize the vote transfer after a voter's favorite is out. In reality, unfortunately, this does not include supporters of the runner-up, eliminated in the final: Of the W-supporters in Burlington, 2008 had ballots with information that *could* count, but did not under IRV (1513 WMK v 495 WKM). Most of them supported W through all tally rounds.

The tally ignored their massive subsidiary preference for* M *v* K. *Their votes were wasted.
In IRV, supporters of a potential runner-up always have reason to worry.

In a referendum 2010, Burlington repealed the IRV-method, and went back to a (modified) 2-day election, a method that, incidentally, would have given the same result. However, the issue was legitimacy. W-supporters, enraged by their grass-root experience with IRV, must have meant more for the repeal than theory for monotonicity failures.

IRV always ignores the subsidiary rankings of the supporters of the runner-up. A "NoShow paradox" would have strengthened M's and weakened K's case even more; see (5.6), and poured more fuel on the post-election political fire in Burlington.

(4.6) ***Baldwin's elimination method*** **To avoid "wasted votes", it is natural to consider a Condorcet method, perhaps after an initial sequence of IRV-eliminations, but while at least three candidates remain. Like IRV, also Baldwin is a Majority method with its one-by-one eliminations and a final with two candidates. Plurality and Borda are both "positional", i.e. point-awarding methods. Elimination according to a weighted sum of Plu and Bor in (4.3) may give a better balance between the two dual goals of IRV.**

REMARK (4.6) **Being a Condorcet method, Baldwin comes with a snag: Like the Burlington pictogram, *every* pictogram of a candidate triple {M, X, Y} with Condorcet winner M shows M as the candidate closest to the center. This illustrates a dubious incentive to campaigners:**

---

[4] https://www.electoral-reform.org.uk/voting-systems/types-of-voting-system/alternative-vote/

Leave controversial issues to others and remain neutral! An improved idea may be to use IRV eliminations until three candidates remain, and then switch to Baldwin.

MMP with faithful accounting and local Majority tallies

The 2017 Bundestag election was the first with rules (1.3) and (1.5). If complete ballot data in the 299 constituencies are available for research, one may explore the effect of changing to accounting rule (1.4) and the Majority method (4.1). "Conventional" methods (4.2) may be more attractive than (4.1), but see Remark (4.7).

REMARK (4.7) **Under faithful accounting (1.4), a voter's influence on the outcome will be mainly through the Erststimme or mainly through the Zweitstimme. By allowing incomplete rankings of local candidates, and treating an emptied ballot as in (4.1), a ballot keeps about the average influence of its Zweitstimme, without contributing to an intolerably high S*. Moreover, a ballot with emptied symmetrized Erststimme is not by force recorded as supporting an unwanted local candidate, but just as taxed for fiscal reasons.**

However, the NoShow paradox unfortunately raises new questions of legality under German law. Because of a decision on NoShow (negatives Stimmgewicht) in the Constitutional Court, one cannot expect a change to be entirely smooth sailing.

5. Constitutionality and legitimacy

"Overhang" is a mathematical concept in political and legal discussions about MMP. Assume party Z_r receives $D(Z_r)$ local seats from a total of S. A strictly proportional distribution of all S seats based on Zweitstimme only, would give it

$$Q(Z_r) \ = \ S \cdot \{|\Lambda(Z_r)|\} \,/\, \{\Sigma_j |\Lambda(Z_j)|\} \quad \text{seats}, \quad 1 \le j \le k.$$

(5.1) *Definition* $D(Z_r) - Q(Z_r)$ depends on S; if positive, it is the nationwide *overhang* for Z_r.

In Example (1.1), strict proportionality at $S = S_o = 598$ is obtained

at price 44190959/598 = 73898 Zweitstimme per seat; CSU pays for only Q(CSU) seats,

Q(CSU) = 2869688/73898 = 38.8 seats.

Local tallies and accounting rule (1.3) give entry D(CSU) = 46 in the table;

at $S = S_o = 598$, the overhang of CSU is 46 – 38.8 = 7.2 seats.

The last overhang disappears at S = S* = 709, and CSU affords to pay for all 46 seats.

Overhang seats together with the rules for allocating party seats to *party and state* sometimes gave a result where a suitable increase [decrease] of $|\Lambda(Z)|$ would have decreased [increased] the total seat number for party Z. One effect is just a reversal of the other; both of them are examples of *"negatives Stimmgewicht"*. New ballots change the result contrary to the intention expressed in them, and so do their removal.

A *decrease* of $|\Lambda(Z)|$, with no concomitant change in Erststimme or in Zweitstimme for the other parties that passed the threshold, could

be due to moves from $\Gamma(C, Z)$ to $\Gamma(C, Z')$,

where Z' did not pass the threshold even with the increased support, and C is anyone of the local candidates in state $\mathscr{A}$. Negatives Stimmgewicht could then, with *less* Zweitstimme for Z, give Z an extra seat in some other state $\mathscr{B}$. Since the move from one voter category to another paradoxically improves the result according to the originally planned ballots, this is akin to Pushover in single-seat elections. But, since the only relevant change is less Zweitstimme for Z in state $\mathscr{A}$, one may as well see it as akin to NoShow and a violation of (5.5) and the Participation criterion.

Technical background for court decisions 2008 and 2012

Local seats were "paid" with Zweitstimme *inside each single state*. The mechanism behind the "paradoxical" gain of a seat through negatives Stimmgewicht was as follows:

1) Z had its proportionality claim in $\mathcal{A}$ filled with local winners (like CSU in Bavaria 2017);

2) the (imagined) reduction of party support for Z in $\mathcal{A}$ was big enough to turn the seat of some local winner W from Z into an overhang seat, i.e. Z had no longer enough Zweitstimme in $\mathcal{A}$ to pay for it, and W was temporarily "forgotten";

3) the reduction was small enough for Z to keep up its nationwide proportionality claim;

4) by the rules, Z's claim became stronger in state $\mathcal{B}$, and Z got a new seat there;

5) Z still kept the overhang seat with W in its assembly group, thus being one seat stronger than before.

To hit a stochastic interval **x participating voters in state $\mathcal{A}$ move out from Λ(Z). Then, in step 2), if α < x for some α, a local seat won by W for Z becomes an overhang seat, but in step 3), if x < β for some β, Z keeps up the total number of seats Z claimed by proportionality, and gains a seat in state $\mathcal{B}$. Post-election analysis, occasionally found such an interval (α, β). This could hardly be predicted, while |Λ(Z)| cannot be controlled. Necessarily,**

(5.2) (α, β) depends on ballot data and is a stochastic interval.

If x voters moved and hit the target, i.e. x ϵ (α, β), it cannot be proved. It is a kind of skeet shooting – blindfolded and unintended. In reverse, with increased |Λ(Z)|, Z loses a seat. Moreover, not just x moving voters are involved, but all voters whose ballots define α or β.

Rulings in the Constitutional Court

In its ruling 2008, the German Federal Constitutional Court states, in para 92, that

(5.3) ***all voters should have the same influence on the outcome of the election.*[5]**

Given the unavoidable stochastic component in all vote vectors, one might perhaps interpret this as the same *expected* influence – in some sense. However, when a specific post-election analysis concludes that a voter group would have been better off by not participating, there is reason to look into the matter. Negatives Stimmgewicht had consequences for a small number of seats, but published examples gave precise values, e.g.

(5.4) ***1987, CDU in Baden-Württemberg would, with 18705 votes less have gained one seat.***

---

[5] **Among voters who supported a local CSU-winner in Example (1.1), one in Λ(FDP) influenced the result more than one who stayed in Λ(CSU); faithful accounting (1.4) in column 3 limits a voter's influence accordingly.**

One would like to see the interval (α, β) of (5.2). Anyway, there is some legal/political/public pressure in such statements. The court followed its interpretation of principle (5.3) and declared the MMP-rules then used, to be *unconstitutional*.

After another decision in 2012, came the present tally rules. With rules (1.3) and (1.5) together with (1.9), i.e. enough party seats to avoid overhangs, negatives Stimmgewicht disappeared, which is good for transparency. However, another consequence is an uncontrollable number $S^* - S_o$ of extra-ordinary party seats. The critical size S^* forces its way into public attention. How large can an unpredictable S^* become, and be tolerable?

Local Majority tallies and the principle (5.3)

Keeping S^* within bounds calls for a Majority method in the local single-seat elections, and IRV is a natural choice. However, it has properties that again raise a question of compatibility with the principle (5.3).

The monotonicity failure in Burlington, Example (4.2), has received wide publicity, but IRV also allows the NoShow paradox, more akin to negatives Stimmgewicht in MMP.

It seems like an awkward loop. Old MMP-rules allowed negatives Stimmgewicht; a court found this an illegal violation of (5.3); rules (1.3) and (1.5) avoided it by abandoning overhangs; this in turn leads to excessive assembly size. IRV is an important and natural remedy, but it allows NoShow, which violates the *Participation criterion*. A consequence of the principle (5.3), (5.5) is its definition for preferential single seat elections:

**(5.5*) "… the addition of a ballot, where candidate A is strictly preferred to candidate B, to an existing tally of votes should not change the winner from candidate A to candidate B."*
(Wikipedia: *Participation criterion*).**

The "Abstention strategy" is the reversal: By removing a ballot with A strictly preferred to B (equivalently, an A-supporter stays home), A wins instead of B. It reveals the same property of the method as the NoShow accident does.

A NoShow is *"strong"* if A is *top-ranked in the new ballot* that causes B to win. Obviously, this cannot happen in IRV if the tally, before the extra ballot, was without tiebreaks. Weak NoShow in IRV appeared as an entertaining story in (Fishburn and Brams 1983).

An almost-NoShow in Burlington 2009

Burlington 2009 was close to weak NoShow; see the pictogram and (5.6): Assume that h W-supporters up-rank M from WKM to WMK, but x WMK-voters do not show up.

The new vote vector is:

$$(|MWK|, |MKW|, |KMW|, |KWM|, |WKM|, |WMK|) =$$
$$(994.5,\quad 1559.5,\quad 2327,\quad 655,\quad 1139.5 - h,\quad 2157.5 + h - x)$$

- W is eliminated if $3297 - x < 2554$, i.e. $x > 743$; see (4.3).
- Next, M beats K if $4711.5 + h - x > 4121.5 - h$, i.e. $x < 590 + 2h$.

Thus, a weak NoShow has occurred, and

(5.6) M becomes IRV winner instead of K *when* x WMK-*ballots are removed*, if $x \in (743, 590 + 2h)$; this requires $h \geq 78$

Obviously, (5.6) violates criterion (5.5), but there is nothing anomalous in this: Removing WMK-ballots reduces both W's advantage over M in top-ranks and M's advantage over K in pairwise comparison. Weak NoShow just demonstrates that W's advantage over M disappears first with the removed ballots. Passing x = 743, enough WMK-ballots remain to keep M ahead of K: Anti-K voters apply the *"Abstention strategy"* to avoid K.

Of course, in (5.6) anti-K voters have a simpler way to avoid election of K: 429 WMK-voters could move to MWK, eliminate K (see the Burlington pictogram), and M wins with the *"Compromise strategy"*. However, this is also quite unrealistic.

A weak NoShow case is valuable as a *very strong signal* that some important feature in the preference distribution should get attention. With h ≥ 78, Abstention *could* have helped Condorcet-winner M to win the IRV-election ahead of K, but it is the tally's neglect of the W-supporters' subsidiary rankings (1515 WMK, 495 WKM) that harms IRV's legitimacy.

NoShow, wasted votes, and legality v legitimacy

If the appellants in Minnesota, see Remark (4.5), had been better prepared, could they then have obtained a court decision on NoShow in IRV, one year after the German Constitutional Court had ruled against negatives Stimmgewicht? With the adjusted vote vector (h ≥ 78), it would have been easy to count and *prove* an "almost-anomaly" where 744 WMK-voters by staying home, would have got M as winner instead of K. See (5.4) for a similar case in MMP.

Baldwin's method (4.6) helps, but has its own shortcomings, like all election methods. Like IRV, all Condorcet methods allow NoShow, except some with n < 4 candidates (Moulin 1988). Baldwin allows NoShow also with n = 3, but the increase of a voting category must either start or end in a Condorcet cycle; this is easy to check. As a practical problem, NoShow is therefore smaller than it is in IRV. To design with some wisdom remedies that cope with unwanted assembly size, will cause some new grief. That is nothing new under the sun, but raises a question of cost v benefit. A legal ban on *all* election methods that allow violations of the Participation criterion may turn out to be counterproductive.

Improving legitimacy is a different approach. The Participation criterion and other axioms organize a theory, but tally rules should promote legitimacy. It is not good for legitimacy

- **that M, as Condorcet winner with two *massive* pairwise victories at x = 0, h ≥ 78, is eliminated, and**
- **that W's supporter group (here the largest), experience that the tally ignores their subsidiary rankings and wastes their votes.**

Burlington 2009 was not even an almost-anomaly, just close to a weak NoShow. It illustrates the incompatibility of two axiomatic principles adduced by IRV advocates:

Principle 1: ***Avoid the "Burying strategy".***

The Borda Count is the prime example of an election method destroyed by Burying: If P and Q are frontrunners in a field of seven candidates, one voter who switches from PQRSTUV to PRSTUVQ, i.e. "buries Q", compensate for six voters who vote QPRSTUV. IRV avoids Burying completely, since the tally ignores all subsidiary rankings in a ballot with P on top while P is still in the race. Without exceptions, and easily understandable, this "legalistic" principle certainly has impact.

Principle 2: ***No votes are to be wasted*.**

In Burlington 2009, Example (4.2), this implies that the subsidiary rankings of the W-supporters must influence the tally. However, the Baldwin tally (4.6) with three candidates also opens for Burying:

In Example (4.2), Condorcet-winner M is clear Baldwin-winner, but vulnerable to Burying: If 930 K-supporters move from KMW to KWM, M turns to be Borda loser; see (4.3); Baldwin/Borda-elimination hits M, and finally K wins over W.

This necessarily creates a Condorcet cycle, "M beats K beats W beats M". Burying is hard to use in Condorcet elections and is at most a minor nuisance.

Principle 2 when three candidates remain improves legitimacy. An IRV elimination prologue is an ordeal defined by the voters, and exposed to random ballot changes. Principle 1 still prevents Burying to influence what candidates that reach a Baldwin final.

The ubiquity of strategic voting

Even small sets of very reasonable desiderata may be incompatible. Arrows theorem (1950) for preferential elections is the prime example. Wilson (1972) pruned the axiom system by removing a "Pareto condition", and strengthened Arrow's impact on Social choice theory.

The G-S theorem (Gibbard 1973, Satterthwaite 1975), concerns preferential election methods for n ≥ 3 candidates that just pick a unique winner. By G-S, two properties are incompatible, i.e. being non-dictatorial, and never allowing any strategic voting. (The definition is broad: With winner W, strategic voting allows a ballot B to be replaced by ballot B' and give a new winner W' that B ranks higher than W.) It changes *at least two entries in the vote vector*. Thus, a tally cannot detect it. A NoShow accident is different; one entry in the vote vector has grown past a critical value, getting a worse winner according to the entry. This is easy to detect. IRV allows a No-Show accident every time it allows Pushover.

These theorems should remind legislatures and courts: A declaration that some property in an election method is mandatory or prohibited, may cause worse problems than it solves. If they find that some flaw makes a certain voting method *illegal*, it may be difficult to find useful methods without equally annoying flaws, or worse.

Depending on the election method, a claimed "anomaly", e.g. a NoShow paradox or a type of strategic voting, may be anywhere on a scale from non-existing to disastrous. Anomalies may be particularly frequent in Condorcet cycles, but normal preference distributions have a structure seen in the pictograms: closeness to Perfect Pie-sharing make cycles very rare.

The *possibility* of a non-monotonic effect in IRV is not rare, but is a minor nuisance on level with Burying in 3-candidate Baldwin, although it is in vogue to attack IRV for it (Ornstein and Norman 2014; Miller 2017).

Majority methods in the local tallies of MMP

In a Majority single-seat method used to tally Erststimme in an MMP context with faithful accounting (1.4), a flaw seems less serious since a voter, who has supported the local runner-up, keeps full voting power in the Zweitstimme tally. One task for the Zweitstimme tally is to achieve proportionality; this will also give compensation to "victims" of any flaw in the Erststimme tally. A less zealous implementation of the principle (5.3) should be acceptable in the local elections of MMP, especially when it improves legitimacy.

In the tradition of Black (1948), we should see voter behavior as an essential background, including the empirical rareness of Condorcet cycles when 3 candidates remain in IRV:

(5.7) ***"In deciding which system is better or worse in a particular context, judgments should be comparative, holistic, pragmatic, and guided by evidence about behavior as well as by theory."*** **(Nagel 2013)**

Concluding remarks

A volatile number of extra-ordinary party seats is unfortunate; there may well be many more than the 111 in the 2017 election. Faithful accounting (1.4) and a Majority method in the Erststimme tally, considered above, are strong remedies against oversized assemblies. They should avoid complications, improve transparency and promote legitimacy. How often will they achieve $S^* \leq S_o$ for given S_o in real elections? This is a natural topic for continued investigation.

In a way, Zweitstimme overrule Erststimme: They already establish proportionality – in some sense – but the present accounting rule (1.3) and the Plurality method are the main reasons for the large $S^* - S_o$. By normal civil case rules, the Constitutional Court will only treat them if somebody competent demand it, e.g. in a new constitutional complaint. With rule (1.4), *voter groups* $\Lambda(Z_a)$ get proportional influence on local and party seats, $1 \leq a \leq k$. The two rulings by the Constitutional Court (2008, 2012) offer 83 pages of legal background. Voting laypersons can hardly relate to the legal and political reasons given for the present reality. Transparency is important for legitimacy, and therefore a goal both in the operation of an election method and the reasons given for it.

The Erststimme tally comes first, and the name (erst = first) makes some voters believe that Zweitstimme are less important. Even if established wisdom says the opposite, there are real reasons to emphasize the Erststimme. *"Personalized voting"* in single-seat constituencies gives the district representation its special legitimacy. Political fragmentation under strict proportionality in the Weimar republic was an experience still remembered when the idea of MMP first took form after WW2.

References

Arrow, K (1950), *A difficulty in the concept of social welfare*, Journal of Political Economy, pp.328-346.

Baldwin, J. M. (1926), *The technique of the Nanson preferential majority system of election*, Proc. Roy. Soc. Victoria n. s. 39, pp. 42-52.

Black, D. (1948), *On the rationale of group decision making*, Journal of Political Economy, pp. 23-34.

Bundesverfassungsgericht (2008), Decision of July 3^{d}, 38 pages.

Bundesverfassungsgericht (2012), Decision of July 25th, 45 pages.

Csato, L (2016), *A mathematical evaluation of vote transfer systems*, https://www.researchgate.net/publication/305186342

Dowding, K and VanHees, M (2008), *In praise of manipulation*, British Journal of Political Science, Vol. 38, pp. 1 – 15.

Fishburn, P. C. and Brams, S. J. (1983), *Paradoxes of preferential voting*, Math. Mag., 56, pp. 207-214.

Gibbard, A (1973), *Manipulations of voting schemes: A general result*, Econometrica, 41, pp.587-601.

Hettlage, M. C. (2018) *Die Figur des Nicht-gewählten Abgeordneten im deutschen Wahlrecht*; Europolis-Online.org.

Linhart, E, Raabe, J, and Statsch, P (2019) Mixed-member proportional electoral systems – the best of both worlds? Journal of Elections, Public Opinion, and Parties vol. 29, pp. 21-40.

Miller, N.R. (2017) *Closeness matters: monotonicity failure in IRV elections with three candidates*, Public Choice 173, pp. 91-108.

Minnesota Supreme Court (2009), *Minnesota Voters Alliance, et.al. (appellants) vs The City of Minneapolis, et.al. (respondents).*

Moulin, H (1988), *Condorcet's principle implies the No-Show paradox*, Journal of Economic Theory 45, pp.53-64.

Mudambi, R. and Navarra. P. (2004), *Electoral Strategy in Mixed Systems of Representation,* European Journal of Political Economy, Vol. 20, pp. 227 – 253.

Nagel, J (2013), *Comments on Ornstein and Norman, Frequency of monotonicity failure under Instant Runoff Voting, estimates based on a spatial model of elections.* http://www-personal.umich.edu/~ornstein/documents/Ornstein-Norman-comments.pdf

Ornstein, J.T. and Norman, R.Z. (2014), *Frequency of monotonicity failure under instant runoff voting: estimates based on a spatial model of elections*, Public Choice 161, pp. 1-9.

Reynolds, A., Reilly, B., Ellis, A. (2005), *Electoral System Design. The New International IDEA Handbook*, International Institute for Democracy and Electoral Assistance.

Sainte-Laguë, A. (1910), *La representation proportionelle et la méthode des moindre carrées,* Comptes rendus des séances de l'Academie des sciences 151, pp 377-378.

Satterthwaite, M. A. (1975), *Strategy proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions*, J. Econ. Theory, 10, pp.187-217.

Smith, W. D. (2010-2016), *3-candidate instant runoff voting: Master list of paradoxes and their probabilities,* RangeVoting.org

Stensholt, E. (1996), Circle pictograms for vote vectors, SIAM Review 38, pp. 96-119.

Wilson, R. (1972), *Social choice theory without the Pareto principle*, Journal of Economic Theory, pp. 478-486.

NHH

NORGES HANDELSHØYSKOLE
Norwegian School of Economics

Helleveien 30
NO-5045 Bergen
Norway

T +47 55 95 90 00
E nhh.postmottak@nhh.no
W www.nhh.no