Fair Majority Voting
(or How to Eliminate Gerrymandering)

Michel Balinski

1. THE PROBLEM. Something is rotten in the electoral state of the United States. Mathematics is involved. Advances in computer technology—hardware and software— have permitted a great leap "forward" in the fine art of *political gerrymandering*—"the practice of dividing a geographical area into electoral districts, often of highly irregular shape, to give one political party an unfair advantage by diluting the opposition's voting strength" (according to *Black's Law Dictionary*).

It is generally acknowledged that some four hundred of the 435 seats in the House of Representatives are "safe," and many claim that districting determines elections, not votes. Recent congressional elections (especially those of 2002 and 2004)— summarized in Table 1—show the shocking impact of gerrymandering. Incumbent candidates, in tailored districts, are almost certain of reelection (over 98% in 2002 and 2004, over 94% in 2006). If an election is deemed "competitive" when the spread in votes between the winner and the runner-up is 6% or less, then 5.5% of the elections were competitive in 2002, 2.3% in 2004 and 9.0% in 2006. Many candidates ran unopposed by a candidate from one of the two major parties in all three elections. In Michigan, the Democratic candidates together out-polled the Republican candidates by some 35,000 votes in 2002, yet elected only six representatives to the Republican's nine. In the 2002 Maryland elections, Republican representatives needed an average of 376,455 votes to be elected, the Democratic representatives only 150,708. In the 2004 Connecticut elections, the Democratic candidates as a group out-polled the Republican candidates by over 156,000 votes; nevertheless, only two were elected to the Republican's three. In all three elections Massachusetts elected only Democrats: in 2002 six of the ten were elected without Republican opposition, in 2004 five and in 2006 seven. Ohio elected eleven Republican and seven Democratic representatives in 2006, and yet the Democratic candidates received 211,347 more votes than did the

Table 1. Results of 2002, 2004, and 2006 congressional elections.

	2002	2004	2006
Incumbent candidates	386	392	394
Incumbent candidates reelected	380	389	371
Incumbent candidates who lost to outsiders	4	3	23
Elected candidates ahead by $\geq 20\%$ of votes	356	361	318
Elected candidates ahead by $\geq 16\%$ of votes	375	384	348
Elected candidates ahead by $\leq 10\%$ of votes	36	22	56
Elected candidates ahead by $\leq 6\%$ of votes	24	10	39
Candidates elected without opposition	81	66	59
Republicans elected	228	232	202
Democrats elected	207	203	233

"Without opposition" means without the opposition of a Democrat or a Republican. "Democrats elected" includes one independent in 2002 and 2004 who usually voted as a Democrat.

Republican candidates. California's last redistricting is particularly comfortable: every one of its fifty-three districts has returned a candidate of the same party since 2002 (fifty were elected by a margin of at least 20% in 2002, fifty-one by at least that margin in 2004 and forty-nine in 2006, and only one candidate by less than a margin of 6% in any of those elections). Gerrymandering is widespread and decidedly ecumenical: both parties indulge.

The lack of competitiveness makes it very difficult to change the composition of the House. Compare, for example, the 2002 and 2004 election outcomes. In forty-five states exactly the same numbers of Republicans and Democrats were elected in both elections, and in four states there was a difference of exactly one. The one significant change took place in Texas: the Republicans won six more seats in 2004 than 2002. Why? Like every other state, Texas redistricted before the 2002 election. But in that election the Republicans took total control of the state government and redistricted once again for blatant and avowed partisan interests. Redistricting a second time on the basis of the same census was challenged in the courts and struck down by the Supreme Court just before the 2004 elections, too late to revert to the previous districts. In 2006, a change in the political mood of the nation shifted a mere thirty seats, 6.9% of the size of the House. Of these thirty, twenty-three were won by margins of less than 10% of the vote.

How has this situation come about? That is a long and fascinating story culminating in the Supreme Court's five-to-four decision (April 28, 2004) that upheld Pennsylvania's actual districting plan [**1**]. Everyone involved—the attorneys against, the attorneys for, and the Justices—acknowledged that the plan was a blatant political gerrymander! In view of the confused and often contradictory precedents of some forty years, four justices, led by Antonin Scalia, wished to rule the question nonjusticiable[1] because of the lack of established criteria for deciding whether a plan is fair or not...except for one, clearly stated in 1969:

> Since "equal representation for equal numbers of people [is] the fundamental goal for the House of Representatives," the "as nearly as practicable" standard requires that the State make a good-faith effort to achieve precise mathematical equality. Unless population variances among congressional districts are shown to have resulted despite such effort, the State must justify each variance, no matter how small (*Kirkpartrick v. Preisler*, 394 U.S. 526 (1969), legal references omitted).

Every one of Pennsylvania's nineteen districts has a population of either 646,371 or 646,372: by the mathematics—the one criterion accepted by the Court—the plan is perfect! Indeed, *every one* of Texas's thirty-two districts has a population of 651,619 or 651,620.

How is it possible to determine such "perfect" plans? The answer is simple: first, a fundamental advance in gerrymandering technology has been made; second, a municipality, township, or village is no longer necessarily within one district. The smallest "atom" that is never split is a census tract: the average number of inhabitants of a census tract in Pennsylvania is thirty-eight. Map-makers simply transfer census tracts from one district to another until they find equality. Pennsylvania's district plan splits twenty-nine counties and eighty-one municipalities. Computer programs newly developed for the redistricting season following the census of 2000 make it easy to create maps on a screen and to modify them, by transferring a census tract (or other geographic area) from one district to another, with a simple click of the mouse. With

---

[1] *Vieth v. Jubelirer* 541 U.S. 267 (2004).

each new map a host of information appears concerning the districts: numbers of inhabitants, numbers of votes for Bush and for Gore in the 2000 elections, numbers of African-, Polish-, or Hispanic-Americans, numbers of Catholics and Protestants, distributions of income levels, . . . , and much, much more is available. Districts in red are Republican, in blue Democratic. To facilitate "kidnapping"—placing two incumbents of the opposition party in the same district—small elephants indicate the residency of Republican incumbents, small donkeys of Democratic incumbents. The programs have brought about a fundamental change: gerrymandering has become a science instead of an art. Justice John Harlan was unusually prescient when in a 1969 dissenting opinion he called for a new system:

> The fact of the matter is that the rule of absolute equality is perfectly compatible with "gerrymandering" of the worst sort. A computer may grind out district lines which can totally frustrate the popular will on an overwhelming number of critical issues. The legislature must do more than satisfy one man, one vote; *it must create a structure which will in fact as well as theory be responsive to the sentiments of the community*... Even more than in the past, district lines are likely to be drawn to maximize the political advantage of the party temporarily dominant in public affairs (*Wells v. Rockefeller* 394 U.S. 542 (1969), my emphasis).

The new technology and the lack of criteria by which to evaluate a districting plan—defined by law or recognized by courts—together pose the problem foreseen by Justice Harlan: to find a new "structure," a new method of election.

2. A SOLUTION: FAIR MAJORITY VOTING. The aim of this paper is to provide an answer to Justice Harlan's quest: it sets forth a method of election that makes political gerrymandering impossible. If the approach is to be considered at all as a practical method of election, it must be amenable to a simple, informal, relatively nontechnical description that may be read and understood by justices, lawyers, historians, or just plain interested citizens. That is what this section seeks to provide. The formal mathematical definition of fair majority voting is given in Theorem 2, where the set of elected candidates is characterized.

By tradition and by law, every member of the House of Representatives represents a district. But in the view of the electors and of the elected, a member of Congress represents his or her state as well. Each behaves and votes in the interests of his or her state as much as in the interests of his or her district. Often the entire delegation of a state will vote identically (for example, when the issue involves the state's rights to Federal funding for one purpose or another). In actual fact, representatives represent their districts *and* their states *and* their parties. From this perspective, many electors are very badly represented, as has been observed. Gerrymandering has seriously accentuated what amounts to a disenfranchisement of voters.

A new method of election—fair majority voting—is responsive to the partisan sentiments of the state as a whole *and*, at the same time, gives to each district its own representative [**1**]. It reconciles the two dominant approaches to representative government: political parties are allotted representatives in proportion to total votes ("proportional representation"), *and* each district has one representative.

Fair majority voting (FMV) is defined as follows. Voters cast ballots in single-member districts, just as they do today in the United States. However, in voting for a candidate, each gives a vote to the candidate's party. Two rules decide which candidates are to be elected. (1) The *requisite number* of representatives each party is to have is calculated by Jefferson's method of apportionment on the basis of the total

party votes (section 4 defines it and argues why it should be chosen); (2) the candidates elected—exactly one in each district and the requisite number from each party—are determined through a procedure most easily explained by example (it is described in general in section 4 and justified in section 5).

Table 2. 2004 Connecticut congressional elections: votes.

District	1st	2nd	3d	4th	5th	Total
Republican	73,273	165,558	68,810	149,891	165,440	622,972
Democratic	197,964	139,987	199,652	136,481	105,505	779,589

The electoral system in use today elects in each district the candidate with the most votes. If these "district-winners" give to each party its requisite number of elected representatives, then FMV elects them. In the Connecticut 2004 elections (Table 2) this was not the case, because there were three Republican district-winners and two Democratic district-winners, whereas the Democrats had 156,617 more votes, so the Republicans should have elected only two representatives and the Democrats three (by the method of Jefferson).[2]

Since each district deserves one representative, the Republicans two and the Democrats three, in the FMV approach the five Republican candidates compete for their two seats and the five Democrats for their three seats just as each pair of opposed candidates compete for one seat in a district: the problem is *symmetric*. Among the Republicans the two with the most votes have the strongest claims to seats. Similarly, among the Democrats the three with the most votes have the strongest claims. If these five "party-winners" were all in different districts, FMV would elect them. But in the 2004 Connecticut election they were not.

Who, then, should be elected?

FMV can be given two symmetric explanations. The first focuses on districts. It begins by asking if the candidates with the most votes in each district—the *district-winners*—give the correct total number of seats to each party: for the 2004 Connecticut elections, the answer is no. Why? Because the distribution of votes for the various candidates is in some sense "unbalanced": the Democratic votes do not count as much as they should relative to those of their Republican opponents.[3] They should be adjusted. But the *relative* votes among the Democrats (and among the Republicans) must remain the same, because they are competing among themselves for three seats (and the Republicans among themselves for two seats). Therefore, all the Democratic votes should be scaled up (or all the Republican votes scaled down) until one more of the Democrats' *justified-votes* exceeds that of his or her Republican opponent: this happens when the scaling factor is $149,892/136,481 \approx 1.0983$ (see Table 3).[4] The district-winners relative to the justified-votes give to the parties their requisite number of seats: FMV elects them.

The second explanation takes the dual approach. Instead of choosing a district-winner in each column and scaling the votes in the rows (i.e., of the parties) so that

---

[2]In 2006, four Democrats and one Republican were elected, but the Democratic candidates' 525,673 votes and the Republican candidates' 419,895 votes entitled the parties to three and two representatives respectively.

[3]In 2006 (see footnote 2) the Republican votes did not count as much as they should have with the same districts, a beautiful example of the perniciousness of the current system.

[4]In this and the subsequent tables the justified-votes are rational numbers: they are systematically rounded to the nearest integers.

Table 3. 2004 Connecticut congressional elections: justified-votes (Democratic candidates' votes all scaled up, district-winners in bold).

District	multiplier	1st	2nd	3d	4th	5th
Republican	1	73,273	**165,558**	68,810	149,891	**165,440**
Democratic	1.0983	**217,416**	153,743	**219,270**	**149,892**	115,872

each row (party) has the number of district-winners it deserves—or beginning with the columns and justifying the votes in the rows—it begins with the rows and justifies the votes in the columns. In each row (party) choose the number of candidates the party deserves, taking those who have the most votes: the *party-winners*. If every column (district) has exactly one party-winner, they are elected. Again, this is not the outcome in Connecticut: the second district has two party-winners, the fourth none (see Table 2). Why? For the same reason as before. Here the votes in districts with no winners should be increased, and/or those in districts with more than one winner decreased. But the relative votes between the candidates in each district must remain the same. Therefore, the district votes should be scaled so that the two highest justified-vote getters among the Republicans and the three highest among the Democrats are all in different columns or districts. For Connecticut (see Table 4) it suffices to multiply the votes of the second district by $136,480/139,987 \approx 0.9749$.

The two approaches designate the same set of winners: *they always do for two, three, or any number of parties that are apportioned seats* (see Theorem 2).

Table 4. 2004 Connecticut congressional elections: justified-votes (2nd district's candidates votes both scaled down, party-winners in bold).

District	1st	2nd	3d	4th	5th
Republican	73,273	**161,410**	68,810	149,891	**165,440**
Democratic	**197,964**	136,480	**199,652**	**136,481**	105,505
multiplier	1	0.9749	1	1	1

When there are exactly *two* parties a very simple rule yields the FMV result (see Table 5): (i) Compute the percentage of the vote for each of the two candidates in each district. (ii) Elect for each party the number of candidates it deserves, taking those with the highest percentages. Clearly, no two can be in a same district.

Table 5. 2004 Connecticut congressional elections: percentage of votes in districts (FMV winners in bold).

District	1st	2nd	3d	4th	5th
Republican	27.0%	**54.2%**	25.6%	52.3%	**61.1%**
Democratic	**73.0%**	45.8%	**74.4%**	**47.7%**	38.9%

The United States has established a strong two-party tradition. Some pretend that this is due to electing the candidate with a plurality of the votes in single-member constituencies. It is of course true that this system is extremely efficient in eliminating

candidates from small parties (in 2002 and 2004 exactly *one* representative was elected to Congress who was neither a Republican nor a Democrat, though he usually voted with the Democrats[5]). FMV can accomplish exactly the same purpose by denying seats to any party that has (say) less than 20% or 25% of the total votes in a state (this has the merit of making it clear to all that small parties are excluded, in contrast with the situation today when it is true but not stated). Although FMV has been explained in the context of exactly two parties, it can be used with any number of parties (as is made clear in sections 4 and 5).

In any case some requirements must be imposed on parties for them to be "eligible" to elect any representatives at all. Otherwise, it would be possible for a small party with relatively small numbers of votes in many districts to be apportioned one or more seats, and FMV might then elect one of that party's candidates having abnormally little support.[6]

FMV is a *practical* proposal: it is a special case of a more general voting system that was adopted by the canton and the city of Zürich (in Switzerland) and used for the first time to elect the parliament of the city on February 12, 2006 [5]. Called "biproportional" representation, it is the same as fair majority voting except that each "district" elects a number of representatives that depends upon its population and parties present lists of candidates in the districts. The system determines how many candidates from each party-list should be elected in each district, instead of designating the one candidate that is elected. (N.B. The idea and axiomatic justification for biproportional representation was developed in a series of papers [3], [4], [6]. It is discussed in the context of Mexico in [7] and [8]. The application to Zürich and the account of how a citizen's suit against the past system led to its adoption is described in [11]. FMV was first described informally in [1]; its first formal description, characterization, and proof is given in this article.)

3. THE PROS AND CONS. Fair majority voting offers many advantages and but few inconveniences.

First, it eliminates the possibility of defining electoral districts for partisan political advantage. A vote counts for a party no matter where it is cast.

Second, since parties are allotted seats on the basis of their total vote in all districts, the necessity of strict equality in the number of inhabitants per district is attenuated. This permits districting lines to be drawn that respect traditional political, administrative, and natural frontiers, and communities of common interest.

Third, the law has encouraged, and the courts have accepted, the creation of "minority-majority" districts, in which a nationally underrepresented group constitutes a voter majority sufficient to enable it to elect its own representatives. This possibility has been used for partisan purposes, for minority populations often have their own political agendas. FMV permits such districts to be defined without favoring any party.

Fourth, it is today entirely possible for a minority of the voters in the United States to elect a majority of the members of the House of Representatives just as it is possible for a minority to elect the President, as it did in 2000 and could well have done again in

---

[5]This was Bernie Sanders, an independent, who had been Vermont's sole representative since 1991. He was elected to the Senate in 2006.

[6]If no requirement were imposed, FMV would have given one seat to the Libertarians of California in 2002, its candidates having received 3.6% of the total vote. The one seat would have gone to its candidate in the tenth district who had less than 41,000 votes whereas the Democrat's candidate in the district had over 123,000 votes.

2004.[7] FMV would almost surely prevent a minority of voters from electing a majority in the House.

Fifth, FMV makes every vote count. It is inconceivable that a major party would not present a candidate in every district of a state if FMV became the electoral system: even as little as 10% or 20% of the vote against a very strong entrenched candidate would help the opposition party to elect one of its candidates in another district. The anomaly of large numbers of unopposed candidates would therefore disappear. In addition, since every vote counts, many citizens would vote who do not today (because now their votes make no difference, simply adding to huge majorities or minute minorities).

Sixth, that a state like Massachusetts has no Republican representatives at all seems ridiculous. Certainly at least 10% of the potential voters in Massachusetts have preferences for the Republican party, and should be represented by at least one of the state's ten representatives. FMV makes this possible.

Seventh, with FMV every district continues to have one representative, as required by federal law. The one major drawback is that a district's representative could have received fewer votes than his or her opponent in the district (e.g., FMV elects the Democrat with 136,481 votes in Connecticut's fourth district when the Republican candidate receives 149,891 votes; see Table 2). On the other hand, in the 2004 California election a Democrat won (in the twentieth district) with 61,005 votes, whereas a Democrat lost (in the fourth district) with 117,443 votes; also a Republican lost (in the 10th district) with 95,349 votes. This is every bit as shocking. Furthermore, there is evidence that suggests voters would accept this drawback. The results of the 2006 Zürich election were accepted without criticism, yet some party-lists were allotted more seats than other party-lists that had more votes.[8] Of course, if the candidate with the most votes in a district must always be elected, there is no escaping the present system!

Eighth, under FMV every candidate has the incentive to seek as many votes as possible. Every vote counts for a candidate *and* for his or her party, but more for the candidate than for the party, because he or she also competes for a seat among the party candidates. This is not true in traditional "proportional representation" systems, where parties present lists of candidates and an elector casts a vote for an entire list. A candidate at the top of a list of a major party is assured of election, and a candidate at the bottom of the list is assured of not being elected. Incentives are confused.

Last, and most important, with FMV the House of Representatives might once again become a "mirror" or "miniature" of the electorate as a whole. Incumbents would no longer have the overwhelming advantages that they enjoy today. There would be no safe districts. The courts would be spared the trouble of having to deal with questions they are not able to adjudicate.

4. THE MATHEMATICS. Consider a state with n representatives (i.e., n districts) and m parties, where each voter casts one vote for a party-candidate in his or her district. Let $v = (v_{ij})$, where v_{ij} is the vote received by the candidate of party i in district j, and $p = (p_i)$, where $p_i = \sum_j v_{ij}$ is the total vote of party i.

FMV apportions the n seats among the m parties on the basis of the total party votes, $p = (p_1, \ldots, p_m)$. Let $a = (a_1, \ldots, a_m)$, with a_i the number of seats apportioned to party i. Exactly how should this be done? This "vector" apportionment problem has

---

[7]A switch of seventy thousand votes from Bush to Kerry (1.3% of the votes in Ohio, 0.06% of the votes of the nation) would have made Kerry president, though his vote total would have been at least three million less than Bush's.

[8]In one district, a party-list had one seat for 661 votes while another had two seats for 631 votes. A party had three seats for 1,025 votes in one district, but only two seats for 1,642 votes in another district [**5**].

been thoroughly studied (see [**9**]). The short discussion to follow draws on well-known results.

The appropriate method to apportion seats to eligible parties—those that obtain some minimum percentage of the votes—is Jefferson's. Let $\lfloor x \rfloor$ be the largest integer no larger than the real number x, and let $\lfloor\lfloor x \rfloor\rfloor = \lfloor x \rfloor$ when x is not an integer and $\lfloor\lfloor x \rfloor\rfloor = x$ or $x - 1$ when x is an integer. *Jefferson's method* (also known as *D'Hondt's*) is to take $a_i = \lfloor\lfloor \lambda p_i \rfloor\rfloor$, where λ is chosen so that $\sum a_i = n$. There are three principal reasons for choosing it [**9**]. (1) Among all acceptable methods it most favors the large parties, which tends to help the emergence of a majority party at the national level. (2) It is the unique acceptable method that guarantees each party at least its proportional share rounded down. (3) Several states have exactly two representatives. It gives two seats to the party receiving the most votes unless the party second in the running gets at least one-half the number of votes of the first party. Every other proportional method gives a seat to the runner-up party when it has less than half of the vote count of the leading party. More generally, suppose that two parties with the most votes share n seats. Then when one of them has at least $100k/(n + 1)\%$ of the total vote of the two, it is allotted at least k seats. This seems reasonable.

Let $x = (x_{ij})$, with $x_{ij} = 1$ if the candidate of party i is elected in district j and $x_{ij} = 0$ otherwise. Fair majority voting selects a $(0, 1)$-valued matrix x that satisfies the following conditions:

$$\sum_i x_{ij} = 1 \quad (j = 1, \ldots, n), \qquad \sum_j x_{ij} = a_i \quad (i = 1, \ldots, m),$$

$$v_{ij} = 0 \Rightarrow x_{ij} = 0.$$

The first equations guarantee to each of the districts exactly one representative; the second equations guarantee to each party i exactly a_i representatives; finally, the logical limitation makes it impossible for a candidate who receives no votes whatsoever to be elected. Any such x is *feasible*. The set of candidates singled out by its 1's is a *feasible delegation* (in a minor abuse of language we frequently refer to x itself as a feasible delegation).

Does a feasible delegation always exist? The example of Table 6 shows that the answer is no, so it is necessary to determine the conditions under which a feasible delegation does exist.

Table 6. Example of votes that allows no feasible delegation. (+ represents a positive vote, 0 no votes).

	1st	2nd	3d	4th	5th	6th	7th	seats
Party 1	+	+	+	+	+	+	+	2
Party 2	+	+	+	+	+	+	+	1
Party 3	+	+	+	0	0	0	0	4

There is no feasible delegation or feasible x in the example of Table 6 because four districts (the fourth through the seventh) cast all their votes for parties 1 and 2 that together deserve only three seats. Equivalently, party 3 deserves four seats but receives all of its votes from only three districts. Clearly, no feasible delegation can exist in this situation. Practically speaking, the situation is unlikely: the voters for party 3 in three districts would have to exceed in number all the voters in the other four districts. In any case, the obligation of candidates to be residents of their districts suggests that

every candidate will receive at least one vote... which is a sufficient condition for the existence of feasible delegations. The problem nevertheless begs for an answer: the fact is that there is no feasible delegation only if a situation like that illustrated in Table 6 obtains for some subset of districts and parties. To describe the general case, let K be a subset of the districts and $a(K) = \sum \{a_i : v_{ij} > 0 \text{ for some } j \in K\}$.

Theorem 1 (Feasibility conditions). *There exists a feasible delegation x if and only if $a(K) \geq |K|$ for every subset K of the districts.*

This statement is easily proved (see the end of section 6).

A *problem* (v, a) defined by an m-by-n matrix of votes v and an apportionment a satisfying $\sum a_i = n$ is said to be *feasible* if it has at least one feasible delegation x. For given row-multipliers $\lambda = (\lambda_i) > 0$ and column-multipliers $\rho = (\rho_j) > 0$ the matrices $\lambda \circ v = (\lambda_i v_{ij})$, $v \circ \rho = (v_{ij} \rho_j)$, and $\lambda \circ v \circ \rho = (\lambda_i v_{ij} \rho_j)$ are the *justified-votes* of the candidates of the different parties in the various districts.

A set of candidates elected by fair majority voting is called an *FMV-delegation*. The theorem that follows characterizes them:

Theorem 2 (FMV characterized). *Suppose that the problem (v, a) is feasible. Then:*

  (i) *There are row-multipliers λ such that electing a set of candidates with the most justified-votes $(\lambda \circ v)$ in each district j—a set of district-winners—gives every party i the number a_i of seats it deserves.*

 (ii) *There are column-multipliers ρ such that electing a set of a_i candidates with the most justified-votes $(v \circ \rho)$ of each party i—a set of party-winners—gives every district j exactly one seat.*

(iii) *There is a set of candidates that is at once a set of district-winners and a set of party-winners with respect to the justified-votes $(\lambda \circ v \circ \rho)$.*

In each case the sets of designated candidates are one and the same, though different multipliers may be used to find them. These sets are FMV-delegations.

Assertions (i) and (ii) of the theorem have already been illustrated; a simultaneous application of the row- and column-multipliers obtained there yields (iii) (see Table 7).

Table 7. 2004 Connecticut congressional elections: justified-votes (district-winners *and* party-winners in bold).

District	multiplier	1st	2nd	3d	4th	5th
Republican	1	73,273	**161,410**	68,810	149,891	**165,440**
Democrat	1.0983	**217,416**	149,891	**219,270**	**149,892**	115,872
multiplier		1	0.9749	1	1	1

Multiple solutions are extremely rare, but when they do exist, the same multipliers yield *all* solutions (as will become apparent in the proof of Theorem 2, which is given in section 6).

5. A JUSTIFICATION. Theorem 2 completely defines FMV. Another characterization justifies it.

$$\begin{pmatrix} 1 & & 1 & \\ & 1 & & 1 \\ & 1 & & 1 \end{pmatrix} - \begin{pmatrix} & 1 & 1 & \\ & 1 & & 1 \\ 1 & & 1 & \end{pmatrix} = \begin{pmatrix} 1 & -1 & & \\ & 1 & -1 & & 1 & -1 \\ -1 & & 1 & & -1 & 1 \end{pmatrix}.$$

Figure 1. Feasible delegations differing in two simple cycles, $m = 3$, $n = 6$, $a = (2, 2, 2)$.

Observe that the difference between any two feasible matrices x and y of a problem is a matrix of 0's, 1's, and -1's, every row and column of which sums to 0. In the example of Figure 1, the nonzero entries break down into two simple cycles, each cycle consisting of a $+1$ and a -1 in each of its rows and columns. One cycle is in the three rows and first three columns, the other is in the last two rows and last two columns. Two feasible delegations may differ by many such cycles.

Consider two feasible matrices x and y that differ in a *single* cycle of k rows and columns, as in Figure 2. The i-indices are all different and the j-indices are all different. An x-entry means its x-value is 1 and its y-value is 0, a y-entry that its x-value is 0 and its y-value 1.

$$\begin{pmatrix} x_{i(1)j(1)} & & \leftarrow & & & & y_{i(1)j(k)} \\ \downarrow & & & & & & \\ y_{i(2)j(1)} & \rightarrow & x_{i(2)j(2)} & & & & \\ & & \downarrow & & & & \\ & & y_{i(3)j(2)} & \rightarrow & & & \uparrow \\ & & & \ddots & & & \\ & & & & x_{i(k-1)j(k-1)} & & \\ & & & & \downarrow & & \\ & & & & y_{i(k)j(k-1)} & \rightarrow & x_{i(k)j(k)} \end{pmatrix}$$
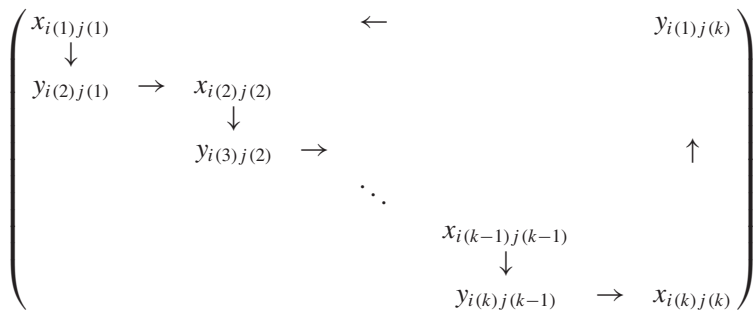
Figure 2. Feasible delegations x and y that differ in a single simple cycle.

Suppose that for a problem $(\bar{v}, a) = (\lambda \circ v \circ \rho, a)$, where $\lambda > 0$ and $\rho > 0$, the following holds in the cycle (taking $l - 1$ and $l + 1$ modulo k):

$$\bar{v}_{i(l)j(l)} \geq \bar{v}_{i(l+1)j(l)}, \quad \bar{v}_{i(l)j(l)} \geq \bar{v}_{i(l)j(l-1)}. \tag{1}$$

That is, in every row *and* every column of the cycle the candidate designated by the x-entry equal to 1 has at least as many $\bar{v}$-votes as the $\bar{v}$-votes of the candidate designated by the y-entry equal to 1.

In this case I claim that x should clearly be considered at least as good as y. For when x designates a candidate different from y that candidate has either more or the same number of "votes" $\bar{v}$ than the candidate designated by y in the same district *and also* in the same party wherever x and y differ. But each district must be assigned one representative and each party i must be given a_i representatives. So multiplying *all* the votes of either a district or a party should change nothing, since all it does is rescale the votes of a set of competing candidates in a district or in a party. This simply says that the two problems (v, a) and $(\lambda \circ v \circ \rho, a)$, for $\lambda > 0$ and $\rho > 0$, are equivalent, so if x is at least as good as y relative to votes $\bar{v}$ then the same is true relative to votes v. More particularly, if at least one of the inequalities in (1) is strict, then x should be considered strictly better than y; and if they are all equations, then x and y should be considered equally good.

This defines a binary relation $\succ$ between feasible delegations x and y that differ in a single cycle:

$x \succeq y$ when (1) holds, $x \succ y$ when (1) holds with at least one strict inequality, and $x \approx y$ when all the inequalities of (1) are equations.

Let $\pi(x) = \Pi_{x_{ij}=1} v_{ij}$ for x a feasible delegation.

Lemma 1 (Order between "neighbors"). *Suppose x and y are feasible delegations of a problem (v, a) that differ in a single cycle. Then $x \succ y$ if and only if $\pi(x) > \pi(y)$.*

To prove it, suppose $x \succ y$. Then inequalities (1) hold with at least one strict, for some $\lambda > 0$ and $\rho > 0$, so

$$\prod_{x_{ij}=1, y_{ij}=0} \lambda_i v_{ij} \rho_j > \prod_{x_{ij}=0, y_{ij}=1} \lambda_i v_{ij} \rho_j,$$

or

$$\prod_{x_{ij}=1, y_{ij}=0} v_{ij} > \prod_{x_{ij}=0, y_{ij}=1} v_{ij},$$

implying, since x and y only differ in that single cycle,

$$\pi(x) = \prod_{x_{ij}=1} v_{ij} > \prod_{y_{ij}=1} v_{ij} = \pi(y).$$

Now suppose that $\pi(x) > \pi(y)$. The proof of the lemma is completed by showing that for any positive real N there are multipliers λ and ρ so that the $\bar{v}$-values of the equivalent problem satisfy

$$\bar{v}_{i(l)j(l)} = \delta, \quad \bar{v}_{i(l+1)j(l)} = N - \delta$$

for all l, where $0 < \delta < N$ and $l + 1$ is taken modulo k. A tedious but straightforward calculation reveals that such multipliers exist and that $\delta = N/(1 + r^k)$, where

$$r = \frac{v_{i(1)j(1)} v_{i(2)j(2)} \cdots v_{i(k)j(k)}}{v_{i(2)j(1)} v_{i(3)j(2)} \cdots v_{i(1)j(k)}} = \frac{\pi(x)}{\pi(y)}.$$

The number r is positive. $\pi(x) > \pi(y)$ implies $r > 1$, so $\delta > N - \delta$ and $x \succ y$.

Notice that when $x \approx y$ *every* candidate in the cycle has exactly the same (rescaled) vote, meaning that the delegations corresponding to x and y are really equally good relative to $\lambda \circ v \circ \rho$, hence with respect to v. Notice also that r depends only on the cycle and not on x and y (e.g., different pairs of delegations can differ in the same cycle). Thus when x and y differ by a single cycle, $\pi(x) = r\pi(y)$ for some *factor r of the cycle*.

To illustrate what has just been said, consider again the example of Connecticut ($m = 2$, $n = 5$, $a = (2, 3)$, v the votes of Table 2) and feasible matrices x and y that differ by a single simple cycle (underlined),

$$x = \begin{pmatrix} 1 & \underline{1} & 0 & \underline{0} & 0 \\ 0 & \underline{0} & 1 & \underline{1} & 1 \end{pmatrix}, \quad y = \begin{pmatrix} 1 & \underline{0} & 0 & \underline{1} & 0 \\ 0 & \underline{1} & 1 & \underline{0} & 1 \end{pmatrix}.$$

Since $165{,}558 \times 136{,}481 > 149{,}891 \times 139{,}987$, $\pi(x) > \pi(y)$ or $x \succ y$. Taking $N = 300{,}000$, the calculation[9] gives $\delta = 152{,}777$ and $N - \delta = 147{,}223$, so $x \succ y$. Multipliers that give the result are $\lambda = (1,\ 1.139680)$ and $\rho = (1,\ 0.922798,\ 1,\ 0.982204,\ 1)$, so

$$\bar{v} = \gamma \circ v \circ \rho = \begin{pmatrix} 73{,}273 & \underline{152{,}777} & 68{,}810 & \underline{147{,}223} & 165{,}440 \\ 225{,}616 & \underline{147{,}223} & 227{,}539 & \underline{152{,}777} & 120{,}242 \end{pmatrix}.$$

Two definitions are necessary.

- A feasible delegation x is *best with respect to* $\prec$ if there exists no feasible delegation y for which $x \prec y$.
- A feasible delegation x *maximizes* $\pi(x)$ if $\pi(x) \geq \pi(y)$ for every feasible delegation y.

Lemma 2. *A feasible delegation x maximizes $\pi(x)$ if and only if x is best with respect to $\succ$.*

To see the truth of this lemma, suppose x maximizes $\pi(x)$ but is not best with respect to $\succ$. Then there is a feasible y satisfying $x \prec y$, and by Lemma 1 $\pi(x) < \pi(y)$, a contradiction.

For the converse, suppose x is best with respect to $\succ$ but $\pi(x)$ is not maximized. Then there exists a delegation y with $\pi(x) < \pi(y)$. x and y may differ in several cycles. Letting $r_1, r_2, \ldots, r_m$ be their respective factors, $\pi(y) = r_1 r_2 \cdots r_m \pi(x)$. By hypothesis, $r_1 r_2 \cdots r_m > 1$, so $r_i > 1$ for some i. Let z be the feasible delegation that differs from x only the ith cycle. Then $\pi(z) = r_i \pi(x)$ with $r_i > 1$, contradicting the fact that x is best with respect to $\succ$.

Theorem 3 (Characterization). *A feasible delegation is an FMV-delegation if and only if it is best with respect to $\prec$.*

The idea of building a partial order on feasible delegations from comparisons of "smallest" possible changes is closely linked to the concept of "coherence" or "consistency" [2].

6. THE PROOFS. The truth of the assertions made about fair majority voting may be established via (at least) two arguments. One is by appealing to more general results concerning biproportionality ([3], [4], [6]). The other, which is new, is more direct and is pursued here.

The natural computational idea that stems from Lemmas 1 and 2 is to pick some feasible delegation, then ask whether a "neighboring" one—differing by a single cycle— is better with respect to the relation $\succ$: if yes, then take it, and repeat; if no, then try to show that a best feasible delegation has been found. The proof shows that linear programming can be used to implement this idea.

The geometric mean of a set of n numbers is the nth root of the product of the numbers, so maximizing the geometric mean of the respective votes of a delegation is equivalent to maximizing $\pi(x)$. But maximizing $\pi(x)$ over the feasible delegations is equivalent to finding an x that solves

$$\max_{x} \sigma(x) = \sum_{i,j} x_{ij} \log v_{ij}, \qquad (2)$$

---

[9]Here the numbers are rounded to the nearest integer.

when

$$\sum_j x_{ij} = a_i, \quad \sum_i x_{ij} = 1, \quad x_{ij} \geq 0, \ (i = 1, \ldots, m; \ j = 1, \ldots, n) \qquad (3)$$

and $v_{ij} = 0 \Rightarrow x_{ij} = 0$. This is a linear program whose feasible solutions are bounded—more specifically, it is a transportation problem—so it always has a solution at an extreme point of the polytope defined by the equation and inequality constraints (when nonempty), and the extreme points are precisely the feasible $(0, 1)$-valued matrices x, that is, the feasible delegations. Thus if there exists a feasible delegation the linear program must have a solution. An optimal solution to the linear program is shown to be an FMV-delegation (as defined in Theorem 2).

The "primal" simplex method implements the natural computational idea mentioned earlier. It begins with an arbitrary feasible matrix x, then finds a better neighboring extreme point—a better feasible matrix that differs from x in a single cycle—if such exists, and repeats. If this process halts at x, then it is an optimal solution to the problem, that is, a best feasible matrix.

Proof of Theorems 2 and 3. First, suppose that x is best with respect to $\succ$. Then by Lemma 2, x maximizes $\pi(x)$ (equivalently $\sigma(x)$).

Duality theory supplies the appropriate multipliers. The dual problem is

$$\min_{u, w} \sum_i a_i u_i + \sum_j w_j$$

when

$$u_i + w_j \geq \log v_{ij} \quad (i = 1, \ldots, m; \ j = 1, \ldots, n).$$

All optimal solutions x and u, w of the respective programs satisfy $x_{ij}(u_i + w_j - \log v_{ij}) = 0$, so $x_{ij} = 1$ implies $u_i + w_j = \log v_{ij}$, and $u_i + w_j > \log v_{ij}$ implies $x_{ij} = 0$.

Define the multipliers to be $\lambda_i = e^{-u_i}$. Then $e^{w_j} \geq \lambda_i v_{ij}$ for all i and j, and $x_{ij} = 1$ only if $\lambda_i v_{ij} = \max_h \lambda_h v_{hj} = e^{w_j}$, which establishes the first part of Theorem 2. Symmetrically, taking $\rho_j = e^{-w_j}$, the same conditions imply that $e^{u_i} \geq v_{ij}\rho_j$ for all i and j, and $x_{ij} = 1$ only if $v_{ij}\rho_j = \max_h v_{ih}\rho_h = e^{u_i}$, proving the second part of Theorem 2.

Notice, however, that a stronger conclusion can be drawn: there exist column-multipliers ρ such that choosing a_i candidates with the most justified-votes for each party i gives every district exactly one representative *and* the a_i candidates all have the same number of justified-votes. It is easy to adjust the column-multipliers of Theorem 2 to obtain equality. Suppose v_i^* is the lowest justified-vote of party i's winners: decrease the column-multipliers of its other winners so that each has v_i^* justified-votes. This changes nothing between winners and losers because the losers whose justified-votes are decreased were already below all the winners from their parties (see Table 8).

Finally, if both λ and ρ are defined as indicated, the duality conditions imply $1 \geq \lambda_i v_{ij}\rho_j$ for all i and j, and $x_{ij} = 1$ only if $\lambda_i v_{ij}\rho_j = \max_{hk} \lambda_h v_{hk}\rho_k = 1$. This proves more than the third part of Theorem 2. Namely, there are multipliers so that the adjusted-votes of winners are all exactly 1 (or any other number $c > 0$ obtained by replacing λ by $c\lambda$), those of all others lower or equal to 1 (or c). For example, multiplying Connecticut's Republican row by $136{,}481/161{,}410 \approx 0.8456$ in Table 8

Table 8. 2004 Connecticut congressional elections: party-winners with the same justified-votes (compare with Table 4).

District	1st	2nd	3d	4th	5th
Republican	50,516	**161,410**	47,038	149,891	**161,410**
Democratic	**136,481**	136,480	**136,481**	**136,481**	102,935
multiplier	0.6894	0.9749	0.6836	1	0.9756

does the trick (here the winners all have $c = 136,481$ adjusted-votes). This completes the proof of Theorem 2, and that an x that maximizes $\pi(x)$ is an FMV-delegation.

For the converse of Theorem 3, suppose that x is an FMV-delegation as defined by (i) in Theorem 2. Then $x_{ij} = 1$ implies $\lambda_i v_{ij} \geq \lambda_h v_{hj}$ for every h. So if y is any feasible delegation

$$\prod_{x_{ij}=1} \lambda_i v_{ij} \geq \prod_{y_{ij}=1} \lambda_i v_{ij}$$

implying

$$\pi(x) = \prod_{x_{ij}=1} v_{ij} \geq \prod_{y_{ij}=1} v_{ij} = \pi(y),$$

so x maximizes $\pi(x)$, implying x is best with respect to $\succ$.

If x is an FMV-delegation as defined by (ii) or (iii) in Theorem 2, a similar deduction shows x is best with respect to $\succ$. This completes the proof of Theorem 3. ∎

The values $\pi(x)$ implicitly assigned by FMV to every feasible matrix x furnish a complete order. However, there may be pairs of feasible matrices x and y with $\pi(x) > \pi(y)$ for which it is impossible to find a sequence of feasible matrices, beginning with y and ending with x, such that each feasible matrix z in the sequence is succeeded by another better than or as good as z and differing from it by a single cycle. So there are reasons for questioning the "validity" of this complete order. On the other hand, there is no need for a complete order. All that is necessary is to be able to demonstrate that the solution retained is better than (or at least as good as) any other.

An "unfortunate" consequence of the proof via optimization is that it invites the idea that it might be preferable to maximize some other function of the votes—although the function that is maximized is a *consequence* of accepting the idea that rescaling the votes of candidates of a party or of a district yields an equivalent problem. For example, it has been suggested that a feasible delegation should be elected whose *total vote* (or average vote of its candidates) is a maximum rather than a feasible set for which the *product of the votes* (or geometric mean) is a maximum [**10**]. This is not a reasonable idea because of what it implies (and shows, incidentally, the importance and extreme sensitivity of the choice of function to be optimized, a fact that is often forgotten).

The proof just presented, when modified so as to maximize the total vote of a feasible delegation (instead of the product) establishes counterparts to Theorem 2. The counterpart to (i) reads:

When (v, a) is feasible, there are party-addenda (or row-addenda) λ such that electing a set of candidates with the most justified-votes ($\lambda_i + v_{ij}$) in each

district—a set of district-winners—gives every party i the number a_i of seats it deserves.

The analogous statements to (ii) and (iii) of Theorem 2 hold as well. In each case linear programming duality gives the result when the problem is to maximize $\sum\{v_{ij} : x_{ij} = 1\}$ for $x = (x_{ij})$ a feasible matrix.

The implication of choosing a feasible delegation that wins the most votes is that what is significant in the votes between two candidates is their *absolute difference* rather than their *relative difference*. Thus, in particular, when candidate A has 100,100 votes and candidate B has 100,050 votes the "margin of victory" viewpoint equates this result to A receiving a hundred votes and B receiving fifty votes. That seems ridiculous, because the first case is a very narrow victory, whereas the second is an overwhelming victory. Representation is a proportional idea. The appearance of the geometric mean is a result, not a cause.

On the other hand, the two approaches frequently give identical results (as they do for Connecticut). In particular, when there are exactly *two* parties and the total number of votes in each district is the same, they always give identical results. This is easy to see, for when there are two parties a very simple rule (analogous to that for FMV) yields the result: Assign to each candidate the vote margin over his or her opponent (negative if he or she has fewer votes). For each party choose the number of candidates it deserves, taking those with the highest margins. No two can be in a same district. Clearly, when the vote total does not vary from district to district, the two simple rules agree.

It remains only to establish Theorem 1. It can be proved in many ways, for example by using the "max-flow, min-cut" theorem of network flows. Perhaps the easiest here is to rely once again on duality in linear programming: solve

$$\max_x \sigma(x) = \sum_{(i,j):v_{ij}>0} x_{ij}$$

when

$$\sum_j x_{ij} \leq a_i, \quad \sum_i x_{ij} \leq 1, \quad x_{ij} \geq 0.$$

If $\sigma(x) = n$ a feasible delegation exists; if $\sigma(x) < n$ none exists.

The dual linear program is

$$\min_{u,w} \tau(u, w) = \sum_i a_i u_i + \sum_j w_j$$

when

$$u_i \geq 0, \quad w_j \geq 0, \quad u_i + w_j \geq \begin{cases} 0 & \text{when } v_{ij} = 0, \\ 1 & \text{when } v_{ij} > 0. \end{cases}$$

Optimal solutions exist to both programs with all of the variables x, u, and w taking values equal to 0 or 1, and $\max \sigma(x) = \min \tau(u, w)$.

Suppose that no feasible delegation exists. Then $\sigma(x) = \tau(u, w) < n$. The situation is pictured in Figure 3 (a "$\oplus$" inside the figure means that the corresponding value of v_{ij} is positive or 0, and a "0" that the corresponding value of v_{ij} is 0). Let $K = \{j : w_j = 0\}$ and $I = \{i : u_i = 0\}$, implying that $v_{ij} = 0$ for i in I and j in K, so

Figure 3. Schematic representation of the matrix of votes v.

there are $n - |K|$ of the w_j with value 1. Since $\tau(u, w) = n - |K| + \sum_i a_i u_i < n$, it must be the case that $|K| > \sum_i a_i u_i$. But $v_{ij} > 0$ and j in K implies that $u_i = 1$, so $|K| > \sum\{a_i : v_{ij} > 0 \text{ for some } j \text{ in } K\}$.

For the converse, suppose a feasible delegation does exist. Then $\sigma(x) = \tau(u, v) = n$. Let K be any subset of the districts. Take $w'_j = 0$ if $j \in K$, and $w'_j = 1$ otherwise; take $u'_i = 1$ if $v_{ij} > 0$ for some $j \in K$, and $u'_i = 0$ otherwise. Then

$$\tau(u', w') = \sum_i a_i u'_i + \sum_j w'_j \geq \tau(u, w) = n.$$

But this simply says that $a(K) + (n - |K|) \geq n$, so $a(K) \geq |K|$, and completes the proof.

REFERENCES

1. M. Balinski, *Le suffrage universel inachevé*, Éditions Belin, Paris, 2004.
2. ———, What is just? this MONTHLY **112** (2005) 502–511.
3. M. Balinski and G. Demange, Algorithms for proportional matrices in reals and integers, *Math. Prog.* **45** (1989) 193–210.
4. ———, An axiomatic approach to proportionality between matrices, *Math. of Op. Res.* **14** (1989) 700–719.
5. M. Balinski and F. Pukelsheim, Matrices and politics, in *Festschrift for Tarmo Pukkila*, E. Liski, J. Isotalo, J. Niemelä, S. Puntanen, and G. Styan, eds., University of Tampere, Finland, 2006, 233–242.
6. M. Balinski and S. Rachev, Rounding proportions: Methods of rounding, *Math. Sci.* **22** (1997) 1–26.
7. M. Balinski and V. Ramírez, Mexican electoral law: 1996 version, *Electoral Studies* **16** (1997) 329–340.
8. ———, Mexico's 1997 apportionment defies its electoral law, *Electoral Studies* **18** (1999) 117–124.
9. M. Balinski and H. P. Young, *Fair Representation: Meeting the Ideal of One Man, One Vote*, Yale University Press, New Haven, CT, 1982; 2nd ed., Brookings Institution Press, Washington, DC, 2001.
10. D. Gale, private communications (he raised questions leading to these answers).
11. F. Pukelsheim and C. Schuhmacher, Das neue Zürcher Zuteilungsverfahren für Parlamentswahlen, *Aktuelle Juristische Praxis* **5** (2004) 505–522.

MICHEL BALINSKI, a Williams graduate, studied economics at MIT and mathematics at Princeton. He has taught at Princeton, University of Pennsylvania, the CUNY Graduate Center, Yale, and SUNY Stony Brook. Since 1982 he has been Directeur de Recherche de classe exceptionnelle of the CNRS at the École Polytechnique, Paris. He has enjoyed short-term appointments elsewhere, including Santa Monica, Ann Arbor, Paris, Lausanne, Grenoble, Vienna, Mexico City, and Santiago. He is the founding editor of *Mathematical Programming*, and a past President of the Mathematical Programming Society. His principal current interests are the theory and applications of ranking and the design of fair electoral systems.
Laboratoire d'Économétrie, École Polytechnique, 1 rue Descartes, 75005 Paris, France
michel.balinski@polytechnique.edu

A Polar Coordinates Graph for Valentine's Day

$$r = |\tan\theta|^{1/|\tan\theta|}, 0 \leq \theta \leq \pi$$

—Submitted by Dwight Boddorf, Brockway, PA