

ECONOMIC RESEARCH REPORTS

CONSTRAINED APPROVAL VOTING:
A CUSTOM-DESIGNED ELECTION SYSTEM

by

Steven J. Brams

R.R. #89-11

May 1989

**C. V. STARR CENTER
FOR APPLIED ECONOMICS**



**NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003**

CONSTRAINED APPROVAL VOTING: A CUSTOM-DESIGNED ELECTION SYSTEM

Steven J. Brams
Department of Politics
New York University
New York, NY 10003

ABSTRACT

A voting system is described that was designed for a professional association to ensure the equitable representation of different interests on its governing board. Approval voting, whereby voters can vote for as many candidates as they approve of, or find acceptable, was combined with constraints on the numbers that can be elected from different categories of members. These categories were defined by region and specialty and are illustrated by a 2×3 matrix.

The representation problem is how to assign representatives, each with one vote, to the different categories so as to approximate as closely as possible target election figures (TEFs), which give the precise numbers of seats to which each category is entitled. Allocations that are consistent (larger TEFs receive at least as many seats as smaller TEFs), or--more stringently--are based on the Hamilton method of apportionment are shown not always to produce "controlled roundings," which always exist but are not in general unique. Constrained approval voting (CAV) is a method for choosing one from the set of controlled roundings--or, possibly, from a larger set of outcomes based on looser criteria--that is most approved of by all voters, subject to the constraints.

CAV received serious consideration but was not adopted because it violated the unitary philosophy of the association--that its members should view it as a single entity. Nevertheless, its development illustrates how the professional knowledge of political scientists can be applied to problems in the private sector, wherein new questions requiring further analysis may be raised. This analysis, in the case of CAV, not only provided an understanding of its possible practical effects but also cast new light, in a specific context, on fundamental issues in the theory of representation.

CONSTRAINED APPROVAL VOTING: A CUSTOM-DESIGNED ELECTION SYSTEM

Steven J. Brams
New York University

1. Background

Unlike our colleagues in economics and some other disciplines, we are rarely asked to apply our political science expertise to the solution of practical problems in the private sector. Thus, I was surprised when a professional association approached me about advising them on the election of their governing board.

In this report on my consulting experience, I shall not reveal particulars of the association in order to disguise its identity. However, I have altered no essential facts of the study that I made, which involved designing an election system that would give equitable representation to different kinds of interests on the board.

This association was undertaking a review of its election procedures in response to pressures for reform. In particular, it was thought by some members of the association that there was underrepresentation of certain types of members on the board--and overrepresentation of other types--creating certain biases that affected the association's policies. Although a different election system was not viewed as a panacea, it was considered a possible way to ameliorate the perceived misrepresentation of certain elements on the governing board.

After much discussion, which included a consideration of proportional-representation systems (e.g., the Borda count and the Hare system of single transferable vote), I recommended an election reform called "constrained approval voting" (CAV). Under CAV, the basic feature of

approval voting (Brams and Fishburn, 1983)--that voters can vote for as many candidates as they like--was wedded to constraints placed on the numbers that could be elected in different categories.

This hybrid system significantly modified the purpose for which approval voting was originally designed. It also raised a number of questions about the properties of constraints, and their likely effects, on the representation of different interests on the board.

After describing and illustrating possible constraints and discussing the main features of CAV, I shall indicate how my recommendation was received. First, though, it is useful to give some background on the election system that the association was using and also indicate the scope of the study.

In previous elections, the members voted for a slate of candidates prepared by a nominating committee. About twice as many candidates as there were seats to be filled were nominated; members could vote only for as many candidates as there were seats to be filled--no more and no less. Those candidates with the most votes were elected to fill the open seats over a multiyear cycle. If the cycle were two years, for example, board members would serve two years, with half the seats being filled in each annual election.

No change was contemplated in the size of the board or in the terms of office of its members. The issue for the association was whether to elect board members by constituencies representing different interests of the association. What made the problem unusual was that a constituency was defined by both region and specialty

2. The Representation Problem

Assume that there are two regional divisions of the firm (A and B) and three specialty divisions (X, Y, and Z). As an illustration of the percentages of members that fall into each category on the two dimensions, consider the following 2 x 3 matrix, where the rows indicate the regional divisions and the columns the specialty divisions:

Region	Specialty			Row Total
	X	Y	Z	
A	27	16	17	60
B	21	9	10	40
Column Total	48	25	27	100

These percentages may be interpreted as targets, which the composition of the board should reflect as closely as possible.

The targets in any election will depend not only on the numbers of members in the association that fall into each category but also on the numbers of continuing board members in each category. For example, if elections are held over a two-year cycle, and cell AY already contains 24 percent of the continuing board members, the 16 percent shown in the table should be reduced to 8 percent as a target--and underrepresented cells increased accordingly--to ensure that members from AY do not continue to be overrepresented but instead are properly represented at the 16-percent level (the average of 24 and 8) on the next board.

To be sure, this and the other percentages shown in the table are an ideal: given that each board member has one vote, no allocation of board members (and therefore votes) to each category will perfectly mirror the

percentages. Although a system of weighted voting of members could lead to a better fit of the votes to the percentages in each category, no consideration was given to endowing board members with different numbers of votes.

To illustrate the next step in the construction of an election system, assume that the percentages in the above matrix are indeed the targets, and six new members are to be elected to the board. (It is purely coincidental that this number exactly matches the number of cells: the latter number could be fewer or more than the number of candidates to be elected, and the analysis would be the same as that which I shall describe next.) Multiplying the target percentages by 6, one obtains the following target election figures, or TEFs (the column and row sums are separated somewhat from the six cell entries for clarity):

1.62	0.96	1.02	3.60
1.26	0.54	0.60	2.40
2.88	1.50	1.62	6.00

The remainders of each of the TEFs--except for 6.00, the total number of candidates to be elected--obviously preclude a perfect matching of (whole) representatives to the cells. But this fact does not imply that one cannot narrow down possibilities to those that are, in some sense, best-fitting. To do so, consider the following set of constraints:

1. Row and column minima. Rounded down, the TEF column sums are 2, 1, and 1; the row sums are 3 and 2. Assuming these as minima for the totals of regional and specialty representatives, respectively, there are 65

distinct cases that satisfy these constraints, as exhaustively enumerated in Figure 1.¹

Figure 1 about here

2. Row and column maxima. Rounded up, the TEF column sums are 3, 2, and 2; the row sums are 4 and 3. Assuming these as maxima for the totals of regional and specialty representatives, respectively, 30 of the 65 cases satisfying constraint 1 are excluded. Specifically, the row maxima exclude none of the 65 cases, but the first, second, and third column maxima exclude 8, 11, and 11 cases, respectively, as shown in Figure 2. (It is worth

Figure 2 about here

noting that these column-maxima constraints are mutually exclusive: the cases that each excludes are not excluded by either of the other two maxima constraints.) The 35 cases that remain as admissible are also shown in Figure 2.

3. Cell minima and maxima. Rounding down and up the TEFs of all cells--as opposed to the column and row sums--gives a minimum and maximum for each cell. Satisfying these minimal and maximal cell constraints reduces the 35 admissible cases meeting constraints 1 and 2 to just nine:

<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>9</u>
211	211	112	112	201	111	111	111	102
110	101	200	110	111	210	201	111	210

Note that the sum of the first row in the first four cases is 4, whereas the sum of this row in the last five cases is 3.

FIGURE 1. 65 POSSIBLE ALLOCATIONS OF CANDIDATES THAT SATISFY
CRITERION 1

Sum of First Row = 4 (30 cases)

400	310	310	310	301	301	301	220	220
011	101	011	002	110	020	011	101	011
220	211	211	211	211	211	211	202	202
002	200	110	101	020	011	002	110	020
202	130	121	121	121	112	112	112	103
011	101	200	110	101	200	110	101	110
031	022	013						
200	200	200						

Sum of First Row = 3 (35 cases)

300	300	300	210	210	210	210	210	210
111	021	012	201	111	102	021	012	003
201	201	201	201	201	201	120	120	120
210	120	111	030	021	012	201	111	102
111	111	111	111	111	111	102	102	102
300	210	201	120	111	102	210	120	111
030	021	021	021	012	012	012	003	
201	300	210	201	300	210	201	210	

FIGURE 2. CASES THAT SATISFY CRITERION 1 AND ARE EITHER EXCLUDED OR INCLUDED BY CRITERION 2

30 Cases Excluded

(a) 8 cases in which the first column sums to more than 3:

400 310 301 211 300 210 201 111
011 101 110 200 111 201 210 300

(b) 11 cases in which the second column sums to more than 2:

220 211 130 121 031 210 201 120 111
011 020 101 110 200 021 030 111 120

030 021
201 210

(c) 11 cases in which the third column sums to more than 2:

211 202 112 103 013 210 201 111 102
002 011 101 110 200 003 012 102 111

012 003
201 210

35 Cases Included

Sum of First Row = 4 (16 cases)

310 310 301 301 220 220 211 211 211
011 002 020 011 101 002 110 101 010

202 202 121 121 112 112 022
110 020 200 101 200 110 200

Sum of First Row = 3 (19 cases)

300 300 210 210 210 201 201 201 120
021 012 111 102 012 120 111 021 201

120 111 111 111 102 102 021 021 012
102 210 201 111 210 120 300 201 300

012
210

The three constraints, progressively applied, have reduced the number of admissible cases from 65 (constraint 1) to 35 (constraints 1 and 2) to 9 (constraints 1, 2, and 3). The satisfaction of these three constraints results in what is called a controlled rounding, which can always be found for any matrix (Cox and Ernst, 1982).²

Put more directly, a controlled rounding is one in which, for every column and row, the sum of its cell TEFs, rounded down or up, equals the column or row (total) TEF, rounded down or up. Constraints 1 and 2 give all possible cases that are roundings of the column and row TEFs; constraint 3 limits these to exactly those that are also roundings of the cell TEFs.

3. Further Narrowing: The Search May Be Futile

There are various criteria by which one could reduce the number of cases still further. For example, define an integer representation to be cell-consistent if the TEF of a cell that is assigned a larger integer is at least as great as one that is assigned a smaller integer. Thus, in controlled-rounding case #1 above, the TEF of cell BZ is 0.60 and that of BY is 0.54; yet BZ is assigned a 0 and BY a 1, which makes this representation cell-inconsistent.

In fact, the only two cases that are cell-consistent are #2 and #8. Allocation #2 is cell-consistent because cell AX is the largest TEF (1.62) and receives the only 2 seats that are assigned to a cell; BY is the smallest TEF (0.54) and receives the only 0. Allocation #8 is cell-consistent because only 1's are assigned to all cells; since all assignments are the same, no smaller TEF receives a larger assignment than a larger TEF. When the integer assignments to the column and row sums are also consistent (i.e.,

larger TEFs are not assigned smaller integers), then an allocation is said to be consistent.

Another criterion for reducing the number of integer representations--in this case, to exactly one--is the Hamilton method of rounding (Balinski and Young, 1982), which proceeds in two steps:

1. Allocate to each category--both the six cells and the column and row sums--the integer portion of its TEF (i.e., its number to the left of the decimal point);

2. Of those seats remaining (out of the 6 to be allocated in our example), allocate them to the TEFs with the largest remainders--starting with the TEF with the biggest remainder--until the 6 seats are exhausted.

To illustrate the Hamilton method for the TEFs given earlier, the integer allocations according to step 1 are as follows (note that the column sums total 4 and the row sums total 5):

1	0	1	3
1	0	0	2
2	1	1	4/5

The remaining seats are now allocated, according to step 2, on the basis of the TEFs having the largest remainders:

- 3 to cells in the 2 x 3 matrix (to which 3 seats have already been allocated);
- 2 to the column sums (to which 4 seats have already been allocated);
- 1 to the row sums (to which 5 seats have already been allocated).

These assignments give as a final allocation

2	1	1	4
1	0	1	2
3	1	2	6

Observe that this allocation is the same as (consistent) allocation #2, given in section 2. Allocation #8, on the other hand, is consistent but not Hamilton, which illustrates

Proposition 1. Hamilton allocations are always consistent, but consistent allocations are not always Hamilton.

The second part of this proposition is proved by allocation #8. The first part follows from the fact that, by step 1 of the Hamilton method, TEFs with larger integer portions never receive fewer seats than TEFs with smaller integer portions; by step 2, TEFs with larger remainders never receive fewer seats than TEFs with smaller remainders. Hence, larger TEFs can never be assigned fewer seats than smaller TEFs.

Controlled rounding #3, in addition to the Hamilton allocation (#2), has column and row sums identical to that of the Hamilton allocation. However, this allocation is cell-inconsistent and hence inconsistent. By Proposition 1, it cannot be Hamilton, because Hamilton allocations are a subset of consistent allocations. (Except for possible ties, in which a seat might be randomly assigned at step 1 or or step 2, Hamilton allocations are unique.)

So far, it would appear, a Hamilton allocation is the most sensible of the (consistent) controlled rounding allocations. But there is a rub: one may not exist.

As an illustration of this situation, consider the following percentages and TEFs for the allocation of 6 seats, which differ relatively little from those used in the earlier example:

<u>Percentages</u>	<u>Target Election Figures</u>
30 10 19 59	1.80 0.60 1.14 3.54
18 12 11 41	1.08 0.72 0.66 2.46
48 22 30 100	2.88 1.32 1.80 6.00

Now applying the Hamilton method to the TEFs--both the cells and the column and row sums--one obtains the following allocations for each:

2 0 1 4
1 1 1 2
3 1 2 6

Although the column allocations sum to their Hamilton allocations, the first row sums to 3 (not 4), and the second row sums to 3 (not 2). This example proves that Hamilton allocations to the cells may not agree with Hamilton allocations to the column or row sums.

The fact that the Hamilton allocations in this example are unique, but not a controlled rounding, immediately implies

Proposition 2. A controlled rounding that is Hamilton may not exist.

Recall that a requirement of a controlled rounding is that the sums of the rounded cell TEFs for each column and row equal the corresponding rounded column and row TEFs.

Finally, to settle the question of the existence of a consistent controlled rounding, consider the following percentages for a 2 x 2 matrix and, given 3 seats are to be filled, the TEFs:

Percentages

30	22		52
21	27		48
<hr/>			
51	49		100

Target Election Figures

0.90	0.66		1.56
0.63	0.81		1.44
<hr/>			
1.53	1.47		3.00

The only cell-consistent allocation is to assign 1 seat to all entries except the lowest (0.63). But the consistency of the column sums demands that the first column receive 2 seats, when in fact the sum of its cell-consistent entries (0 + 1) is 1. This example proves

Proposition 3. A controlled rounding that is consistent may not exist.

There is a final difficulty with cell-consistent controlled roundings, illustrated by the following TEFs:

Target Election Figures

0.73	0.47	0.90		2.10
0.48	0.48	0.02		0.98
0.99	0.46	0.47		1.92
<hr/>				
2.20	1.41	1.39		5.00

In this case, I have not specified the percentages on which the TEFs are based, but they can easily be reconstructed by multiplying the TEFs by 20.

The cell-consistent allocation of seats shown below is not consistent because the first two columns, and the second two rows, do not sum to values consistent with their column and row TEFs:

1	0	1		2
1	1	0		1
1	0	0		2
<hr/>				
2	2	1		5

However, there is a new difficulty in this example: the second row is entitled to only 0.98 seats, but the cell-consistent assignments for this row sum to 2.

When the discrepancy between the cell-consistent sum of a column or row and its TEF is greater than 1.0, we say that it does not satisfy quota. Put another way, the column or row TEF, rounded either up or down, is not equal to the cell-consistent sum. In our example, 0.98 rounded up is 1, but the cell-consistent sum of the second row is 2, which proves

Proposition 4. A cell-consistent rounding may not satisfy quota.

An assignment of seats that violates quota, of course, is not a controlled rounding.

I conclude that because the narrowing-down criteria discussed in this section--consistent allocations and Hamilton allocations--may be incompatible with all controlled roundings, they cannot reliably be used to distinguish either a very few or a single best allocation. In addition, we have shown that a cell-consistent allocation not only may be inconsistent but also fail to satisfy quota--and therefore be a controlled rounding. To be sure, other criteria have been proposed for filtering out the best-fitting controlled roundings, but they are not tied, in my view, to fundamental principles of fair representation.³

For the purpose of choosing an elected board that is a reasonable approximation of the TEFs, a case can be made that any of the controlled roundings is good enough: each cell receives representation within one seat of what it is entitled to, and so does each geographical region and functional category. In the absence of theoretical criteria for singling out a best controlled rounding, I suggested to the association an empirical solution that had, underlying it, a theoretical rationale tied to the notion of voter sovereignty.

4. Constrained Approval Voting (CAV)

Specifically, I proposed, as a starting point, that the association use approval voting: members would be able to vote for as many candidates as they approved of, or found acceptable, rather than--as under the extant system--be restricted to voting for exactly as many candidates as there are seats to be filled. But, as before, members could still vote for any candidates, irrespective of their regional or specialty designation, which now would be explicitly indicated on the ballot.

There seemed no good reason to force voters to vote for an arbitrary number of candidates. Rather, I argued, they should be permitted the more flexible option of voting for as many candidates as they liked, the advantages of which have been discussed in detail elsewhere (Brams and Fishburn, 1983; Nurmi, 1987; Merrill, 1988).

The flexibility afforded by approval voting made sense to association members, who testified that many members did not have sufficient knowledge to make more than two or three intelligent choices. Thus, the requirement that they vote for, say, six candidates forced less knowledgeable members (usually newer) to make less-than-informed judgments, often influenced by casual advice from more senior members.

My second recommendation--to restrict the domain of possible outcomes--was designed to counteract a possible bias that approval voting might introduce. Specifically, if members of the largest categories, like AX with 27 percent of the members, tended to concentrate their votes on candidates in this category, they could unduly affect the election.

Indeed, even under the extant system whereby voters were required to vote for an entire slate, "slate engineering" by the nominating committee was considered necessary to thwart voters from electing too many members

of one type. As an example of how this tool was used, if the continuing board was overrepresented by AX members, the nominating committee might propose AX candidates who were not so well-known in order to diminish their chances of election.

This is a form of manipulation well-known to political scientists. But it is an informal device that on occasion had not worked as planned, which is one reason why the association was interested in exploring alternatives that offered more formal protection. Presumably, an election system that explicitly ensured the fair representation of different interests would also gain the confidence of voters; as a consequence, whatever outcome it produced would be considered more legitimate.

If the admissible outcomes are restricted to the set of controlled roundings and not a particular one, then the choices of voters would matter in deciding not only who is elected from each cell but also, within limits, how many. Of course, limiting voters to outcomes only in the set of controlled roundings is radically different from the usual application of approval voting--namely, to the election of single winners in multicandidate elections, with no restrictions on who can be elected.

Indeed, one might argue that the nine controlled roundings in our example are too restrictive a set. The 35 (or even 65) cases available if criterion 3 (or 2 as well) is lifted would give the voters more control in the choice of a board and hence greater sovereignty.

Thereby, the board's composition would be more responsive to their voting. Whereas the controlled roundings guarantee that an integer representation is no more than one seat from the TEFs, the less restrictive set of, say, 35 cases would permit 26 additional outcomes--each of which

leads to the election of at least one different candidate--and still guarantee the column and row sums are within one seat of their TEFs.

The acceptability of this set versus the nine controlled roundings depends upon the importance one attaches to the principle that the number of cell seats--versus the regional and specialty totals--should all be within one seat of the TEFs. To put this matter somewhat differently, the designation of what outcomes are admissible will depend on whether it is thought that more popular (i.e., approved) candidates should be permitted to win at the price of causing greater deviations than one seat from the cell TEFs.

Given a choice has been made of the set of outcomes deemed admissible, the one selected under CAV will be that with the greatest total number of approval votes. To illustrate this calculation, assume that the admissible outcomes are the nine controlled roundings in our earlier example. What is common to these cases, as can be seen from looking at their intersection, is the certain election of exactly one candidate from the three cells AX, AZ, and BX:

$$\begin{array}{ccc} 1 & 0 & 1 \\ 1 & 0 & 0 \end{array}$$

This means that the biggest vote-getters in each of these cells will be guaranteed election, whichever of the nine controlled roundings wins. The votes of these candidates can then be set aside.

The particular controlled rounding that wins will be the one in which the total approval vote of the three remaining "discretionary" choices is greatest. In case #1, for example, one would sum the votes of the runner-up in cell AX, the winner in cell AY, and the winner in cell BY to complete the six choices for this case. This total would be compared with analogous

totals in the eight other cases, given by the sums of the votes of the three best-performing candidates, in the appropriate cells, who are not certain winners.

If one admits as admissible the 35 cases that meet criteria 1 and 2, what is common to these cases, as can be seen from comparing them in Figure 2, is the certain election of one candidate from cell AX. Thus, this set of cases makes five rather than three choices discretionary, although the constraints that these cases--as well as the nine controlled roundings--must satisfy still impose restrictions on the possible outcomes (e.g., the five discretionary choices cannot all be chosen from one cell).

At the least stringent level, criterion 1 alone does not necessitate the certain election of any candidates. As shown in Figure 1, the election of four candidates from cell AX is even permitted in one case, which is more than a 2-seat (122-percent) deviation from its TEF of 1.80.

In my opinion, this deviation is inordinately large, and I therefore recommended tighter restrictions. Both the nine controlled roundings, and the 35 cases that ensure the row and column totals will be within one seat of the TEFs, seemed to me acceptable, with the choice depending on how much leeway it is felt the voters should be permitted.

Clearly, approval voting, by allowing voters to vote for as many candidates as they like, gives voters greater sovereignty than does restricting their votes to a fixed number. But where one draws the line on voter sovereignty by precluding the election of candidates whose choice deviates "too far" from giving a representative board--based on the different categories--is obviously a value judgment. I believe that the analysis I have described clarifies trade-offs that such a judgment entails.

5. Aftermath

A majority of the members of the association that I worked with felt that its formal breakdown into regional and specialty categories would violate its unitary philosophy--that it should be viewed as a single entity by its members. Consequently, the association opted to continue to use slate engineering as a device to ensure, insofar as possible, a representative board. And once the categorization of candidates was rejected, approval voting was seen as a secondary issue and not deemed desirable without the constraints.

My recommendations, nevertheless, stimulated a good deal of interest and received considerable support. Although these recommendations were not adopted, I believe my experience illustrates how our professional knowledge can be applied to problems in the private sector, wherein new questions requiring further analysis may be raised. This analysis, in the case of CAV, not only provided an understanding of its possible practical effects but also cast new light, in a specific context, on fundamental issues in the theory of representation. This marriage of theory and practice, in my opinion, is salutary for the discipline.

NOTES

1. There seems to be no efficient algorithm for generating these cases, but for a 2×3 matrix a hand calculation is feasible. Computer spreadsheets were used to find integer allocations for larger matrices, reflecting finer breakdowns of the association.

2. For larger arrays (i.e., three dimensions or more), the general existence of a controlled rounding seems not to have been established.

3. For an excellent analysis of technical criteria for finding better-fitting biproportional allocations (i.e., those proportional to the TEFs in two dimensions, as here), see Balinski and Demange (1986) and Gassner (1988). In a political context, biproportionality might be based on geographical constituencies and political parties rather than regions and specialities. Because controlled roundings may not satisfy a requirement as weak as consistency--not to mention give allocations compatible with a specific apportionment method like Hamilton--I believe it proper that the voters themselves be able to express themselves in the choice of a particular final outcome, as discussed next.

REFERENCES

- Balinski, M. L., and Gabrielle Demange (1986). "Generalizing Proportionality; or How to Simultaneously Apportion Seats to Regions and to Parties." Laboratoire d'Econométrie de l'Ecole Polytechnique, Paris, mimeo.
- Balinski, Michel L., and H. Peyton Young (1982). Fair Representation: Meeting the Ideal of One Man, One Vote. New Haven, CT: Yale University Press.
- Brams, Steven J., and Peter C. Fishburn (1983). Approval Voting. Cambridge, MA: Birkhäuser Boston.
- Cox, Lawrence H., and Lawrence R. Ernst (1982). "Controlled Rounding." INFOR 20, no. 4 (November): 423-432.
- Gassner, Marjorie (1988). "Two-Dimensional Rounding for a Quasi-Proportional Representation." Université Libre de Bruxelles, mimeo.
- Merrill, Samuel, III (1988). Making Multicandidate Elections More Democratic. Princeton, NJ: Princeton University Press.
- Nurmi, Hannu (1987). Comparing Voting Systems. Dordrecht, Holland: D. Reidel.