

MATHEMATICAL GAMES

*From counting votes to making votes
count: the mathematics of elections*

by Martin Gardner

With the presidential election coming up next month in the U.S. it seems a good time to discuss the mathematics of voting, not least because there are a variety of mathematically determined paradoxes and anomalies that are particularly pertinent to this presidential contest. The way these contradictions touch on plurality voting and the proceedings of the electoral college will be dealt with here, as will the relatively new system known as approval voting. A procedure of increasing interest to political scientists, approval voting manages to avoid many of the logical inconsistencies inherent in other voting schemes.

This month's discussion of the mathematics of voting is written not by me but by Lynn Arthur Steen, professor of mathematics at St. Olaf College in Northfield, Minn. The editor of *Mathematics Magazine*, Steen writes frequently on mathematical subjects and has twice been awarded the Mathematical Association of America's Lester R. Ford award for excellence in writing. He has also edited a variety of books, including *Mathematics Today: Twelve Informal Essays* (Springer-Verlag, 1978) and (with Matthew P. Gaffney) *Annotated Bibliography of Expository Writing in the Mathematical Sciences* (Mathematical Association of America, 1976).

All that follows (up to my concluding comments on previous columns) was written by Steen, who titles his discussion of voting paradoxes and anomalies "Election Mathematics: Do All Those Numbers Mean What They Say?"

Over the past year the prospect of a three-way race for president of the United States has focused public attention on the importance of strategies for voting and on the special vagaries of the electoral college. Although the complications imposed by the electoral college are unique to presidential elections, other uncertainties imposed by three-way contests for public office are not. When the public must choose among more than two alternatives, the task of making the choice is frustratingly difficult. The

source of both the difficulties and the possible solutions is to be found in the little-known mathematical theory of elections.

The social contract of a democracy depends in an obvious and fundamental way on a simple mathematical concept, namely the concept of a majority. Barring the unlikely event of a tie, in any dichotomous ballot one side or the other must receive more than half of the votes. When there are three or more choices of approximately equal strength, however, it is unlikely that such a ballot will yield a majority decision. It is primarily for this reason that many people believe the two-party system is essential to the stability of democracy in the U.S., even though that system is neither mandated nor recognized by the Constitution.

Mathematical theory and political idealism notwithstanding, quite often the public does face a choice among three or more significant alternatives. The same problem that this year appeared as Carter v. Reagan v. Anderson has developed in other years. Such multiple-candidate contests are difficult to resolve fairly if there is no clear-cut majority, but they can easily arise in any free election. Indeed, it follows from some simple mathematics that there are practically no positions candidates in a two-way contest can take that are invulnerable to attack by a third or a fourth candidate.

If each issue in a two-candidate election is represented by a rating of voter preference on a one-dimensional scale, then regardless of the distribution of attitudes among the voters the optimal position for each candidate is the median: the point that divides the electorate into two camps of equal size. The same is true whether public opinion is distributed normally (so that the graph of position v. number of supporters has a single, centered hump), is split bimodally (so that the graph has two approximately equal humps), is skewed sharply to one side or is divided in a highly irregular way. An example of each of these distributions, with the median marked,

is given in the accompanying illustration [*upper illustration on page 18*].

Consider a two-candidate contest in which one candidate adopts a position a little to the left of the median and the other candidate begins with a position at about the middle of the right half of the population. This would be typical of a centrist candidate *C* running against a moderate right-wing candidate *R*. In this case it is reasonable to assume that as far as this particular issue is concerned the voters whose preference lies to the left of the position held by the centrist candidate *C* will favor *C*, the voters whose preference lies to the right of candidate *R* will favor *R*, and the voters whose preference lies in between will be divided about evenly between the two candidates. Under these circumstances, in a preelection poll the centrist candidate would receive a majority of the votes.

The only way for candidate *R* to improve his standing in the poll (on this single issue) is to shift his position toward the middle of the distribution, to ensure that more voters will be to his right. Moving toward the center, or to the left, will always be advantageous for the right-wing candidate. Similarly, a left-leaning candidate can improve his standing with the voters by moving toward the center, or to the right. The median position is the only one that cannot be improved on by further shifting on the part of either candidate.

There is, of course, nothing very novel about this analysis. It is part of our common experience in presidential politics. Candidates representing the right or the left tend to begin distinctly to the right or to the left and then move progressively closer to the center as they attempt to appeal to a greater number of voters. The appeal of the median position in a two-candidate contest, however, is precisely what makes such a contest vulnerable to assault from either side by a third or a fourth candidate. In any contest with two candidates near the center a third candidate entering on the left or the right can always gain a plurality. Indeed, for practically any distribution of the electorate there are no positions in a two-candidate contest where at least one of the candidates cannot be beaten by a third. As is shown in the accompanying illustration [*lower illustration on page 18*], there is always a place along the one-dimensional continuum where a new candidate can position himself to displace one or more nearby candidates.

A single issue rarely plays a deciding role in an election. Hence election analyses based on single issues are not very helpful, unless they can be combined to show how to design a platform that will ensure a candidate's election. Shaping a winning platform is a complex business, however, because it is possible for a platform consisting entirely of winning, or majority, planks to be defeated. The

reverse side of the coin is that a majority platform can be constructed from minority planks. Hence a majority can be formed from a coalition of minorities.

To see how this paradox can arise consider the simplest possible case: a ballot to decide two unrelated, dichotomous issues, represented by resolutions *A* and *B*. In this case the voters actually have four options:

- I. Approve *A* and *B*.
- II. Approve *A* and defeat *B*.
- III. Defeat *A* and approve *B*.
- IV. Defeat *A* and *B*.

The voters who favor both *A* and *B* would choose option I as their first choice, option IV as their fourth choice and option II as their second or third choice, depending on whether they feel more strongly about *A* or about *B*. The voters who favor *A* but object to *B* might rank the four options in the order II, I, IV and III (or II, IV, I and III). In general each voter will have a preference ranking for one of the $4 \times 3 \times 2 \times 1$, or 24,

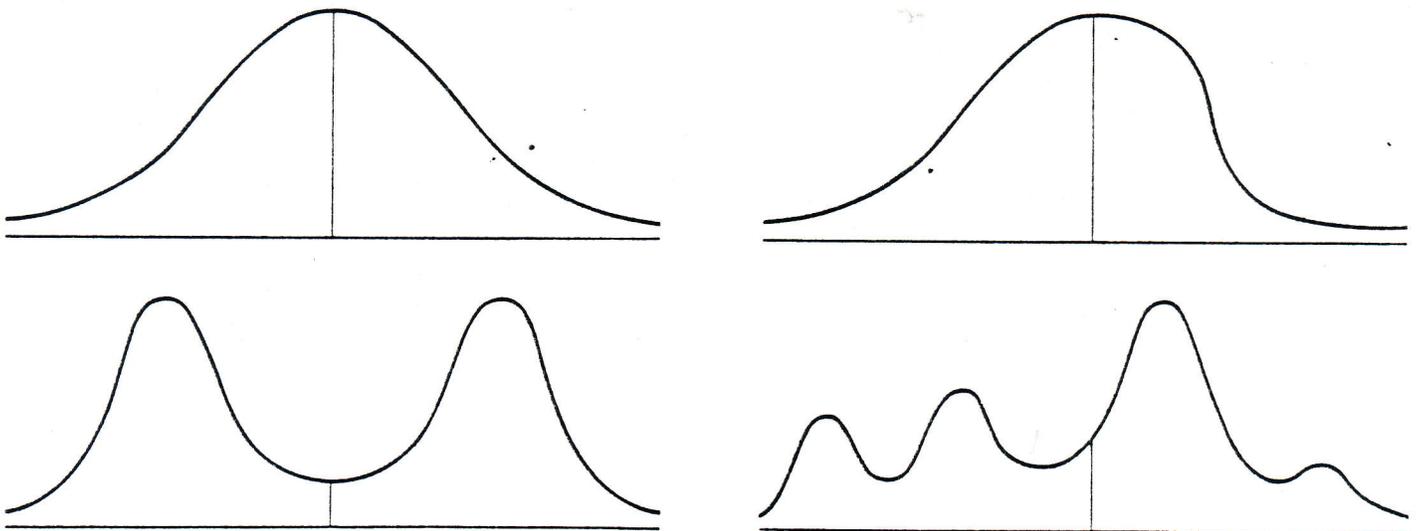
possible permutations of the four available options. (The rankings are by no means equally likely; it would be hard to imagine circumstances under which many people would rank the options in the order of preference I, IV, II, III.)

Now, for the sake of simplicity suppose 500 voters (say at a party convention) are divided into three caucuses as follows: caucus *X*, with 150 votes, ranks the four options in the order I, II, III, IV; caucus *Y*, with 150 votes, ranks them II, IV, I, III, and caucus *Z*, with 200 votes, ranks them III, IV, I, II. In this case caucuses *X* and *Y*, with 300 votes, favor the approval of resolution *A*, whereas caucuses *X* and *Z*, with 350 votes, favor the approval of resolution *B*. Because there are different voters making up these majorities, however, the platform consisting of the planks "Approve *A*" and "Approve *B*" will be defeated by the 350-vote block of caucuses *Y* and *Z*!

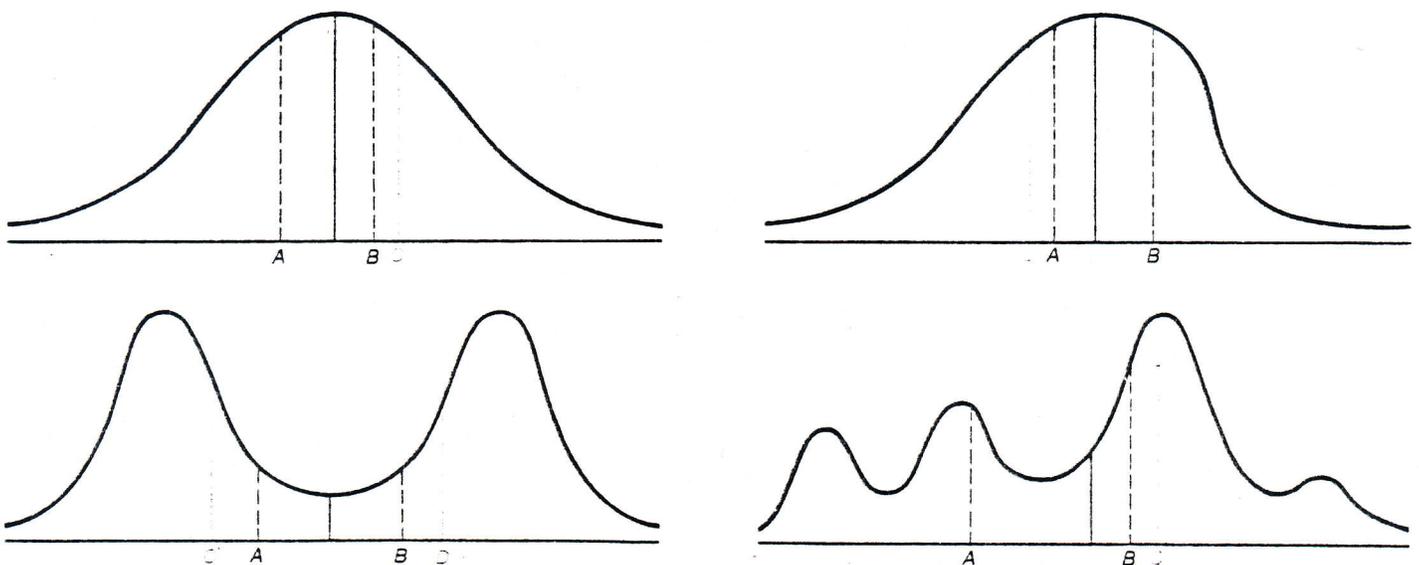
This surprising phenomenon is a special case of the well-known anomaly of cyclic majorities: If three voters respectively prefer *A* to *B* to *C*, *B* to *C* to *A*, and

C to *A* to *B*, then any candidate can be defeated by some other candidate by a vote of two to one in a two-candidate contest. When the issues in an election create cyclic majorities, no set of positions on the issues is invulnerable to assault by a new coalition of minorities another factor that encourages third and fourth-party candidates.

The accompanying diagram [top illustration on page 23] shows how the four options from which party planks in the example must be constructed create a variety of cyclic majorities, thereby explaining how a platform consisting of majority planks can represent the will of only a minority. The arrows joining various platforms depict voting dominance: the platform to which an arrow points will always lose to the platform at which the arrow originates in a dichotomous contest. The winning caucuses in each case appear beside the corresponding arrow. As this distribution demonstrates, any possible platform can be defeated by some other platform, and so a real convention whose divisions resem-



Four possible shapes of public opinion



Two candidates near median (*A* and *B*) can be defeated by a third candidate (*C*) and sometimes a fourth (*D*)

ble the ones given in this example could become mired in an unending sequence of platform motions, with each motion defeating the one before.

The phenomenon of cyclic majorities is also responsible for the most famous election paradox, Kenneth J. Arrow's 1951 proof that certain generally accepted desiderata for voting schemes are logically inconsistent. If there are only two candidates, no problems arise. If three or more candidates appear on a single ballot, however, chaos reigns.

There are diverse schemes other than plurality voting for determining the winner in an election. Many were suggested by 18th-century scholars concerned about implementing the democratic ideals of the French Revolution. Although some of these proposals are so complex as to be completely impractical, several are still in common use, in particular the method of assigning points that reflect degrees of preference to the candidates in a contest (where the candidate receiving the most points is the winner) and various methods of holding runoff elections. Yet as Arrow has shown, none of these schemes—indeed, no method other than a rational benevolent dictatorship—satisfies such commonsense rules as: If *A* is preferred to *B*, and *B* is preferred to *C*, then *A* should be preferred to *C*. Cyclic majorities reduce all voting schemes to unpredictable mystery. (For a discussion of Arrow's proof see "Mathematical Games" for October, 1974.)

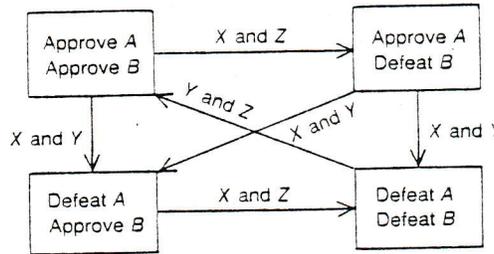
Another important problem with voting in three-option contests is that in many circumstances a vote for the candidate a person prefers most will increase the likelihood that the candidate he prefers least will be elected. (This dilemma was the one often seen in Anderson's candidacy. Many voters who preferred Anderson to Carter and Carter to Reagan believed most Anderson votes would be at Carter's expense.) The anomaly frequently leads thoughtful voters to what is called (depending on a voter's point of view) insincere or sophisticated voting.

If sophisticated voting is widely practiced, it can lead to a state of serious confusion where no one votes for his first choice, and so the public will be effectively camouflaged. An Anderson backer for whom Carter was a second choice might have voted for Carter instead of for Anderson in order to prevent the election of Reagan. If there were enough Anderson backers who reasoned this way, of course, some Reagan supporters might have begun to support Anderson to prevent Carter's reelection. The process of second-guessing the voting strategies of other segments of the electorate can quickly lead to an absurd hierarchy of insincerity in which the votes cast fail to reflect real preferences. Such a process, which it should be added is more a part of game

Options

- I. Approve resolutions A and B.
- II. Approve resolution A and defeat resolution B.
- III. Defeat resolution A and approve resolution B.
- IV. Defeat resolutions A and B.

Caucus	Policy	Votes	Order of Preference
X	Favors A strongly and favors B mildly	150	I, II, III, IV
Y	Opposes B strongly and favors A mildly	150	II, IV, I, III
Z	Opposes A strongly and favors B mildly	200	III, IV, I, II



Three caucuses voting on two platform planks create cyclic majorities

theory than of classical voting theory, rarely gives a legitimate mandate to the victor.

Arrow's theorem shows there is no "perfect" voting scheme for multicandidate elections. The procedure known as approval voting, however, manages to reflect a popular will without inducing anyone to vote insincerely. In approval voting each voter marks on the ballot every candidate who meets with his approval, and the candidate who receives the most votes of approval is the winner.

With this system it is never to a voter's advantage to withhold a vote for his first choice while voting for a less preferred candidate. Indeed, if most candidates seem to have an equal chance of winning, a rational voter should vote for all the candidates he believes are above the average of those running. To vote for more candidates would give unnecessary support to individuals the voter does not endorse, whereas to vote for fewer candidates (say to vote only for one's first choice) is to withhold support from an acceptable compromise candi-

date and to risk victory by an unacceptable candidate.

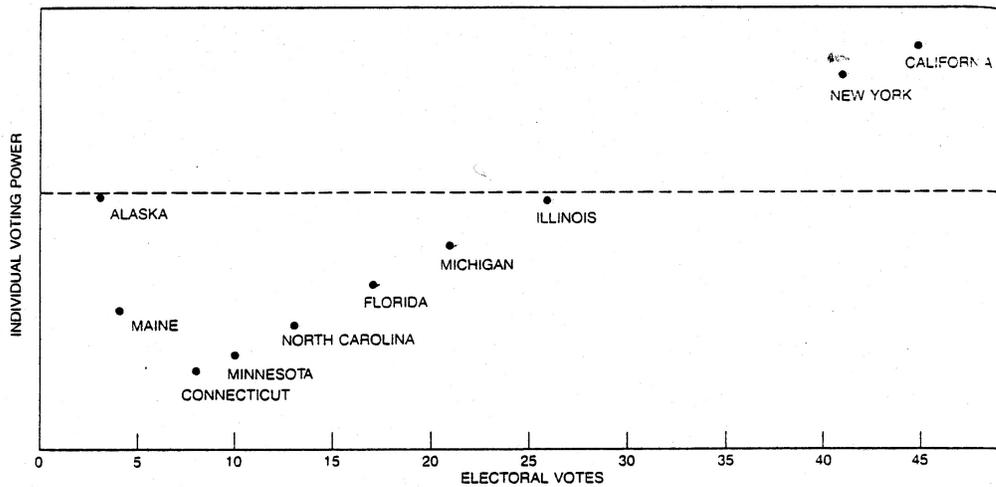
Steven J. Brams, professor of politics at New York University, has described approval voting with the phrase "One man, *n* votes." It is an apt description because approval voting is merely a way of letting a person cast as many votes as he wishes, one for each acceptable candidate. It is easy to count votes that have been cast under this system, and no runoff elections are needed. For both theoretical and practical reasons approval voting is a good compromise between the single-vote ballot that encourages insincerity and the complete preference ordering whose complexity renders it useless in any practical situation.

The accompanying illustration [below] shows how approval voting might compare with plurality voting, runoff voting and point voting in an entirely hypothetical three-way contest. The number of voters supporting each of the six possible rankings of candidates are listed in the column "Total votes," and since C would receive the largest block of first-

Order of preference	Total votes	Approval votes	
		First choice	First and second choices
A, B, C	30	20	10
A, C, B	5	5	0
B, A, C	20	10	10
B, C, A	5	5	5
C, A, B	10	5	5
C, B, A	30	20	10
Total	100	65	35

Plurality voting	Runoff voting	Point voting	Approval voting
A 35	A 35 + 20 = 55	A 35 (3) - 30 (2) - 35 = 200	A 25 + 10 + 15 = 50
B 25	C 40 - 5 = 45	B 25 (3) + 60 (2) - 15 = 210	B 15 - 10 - 20 = 45
C 40		C 40 (3) - 10 (2) - 50 = 190	C 25 - 15 - 0 = 40

Comparison of voting methods for a three-way race



Graph of individual's voting power in various states

choice votes, he would win in a plurality contest. In a runoff election *B* would be eliminated, and *A* would pick up enough second-choice votes (from those who had first voted for *B*) to defeat *C* by 55 votes to 45. In the simplest system of point voting first choices are assigned three points, second choices two points and third choices one point. Because of the large number of voters (60) for whom *B* is the second choice, with this voting scheme *B*, who was eliminated in the runoff, would be the winner.

The results of approval voting depend on whether voters find only their top choice acceptable or whether they could accept some other choices as well. (Because there are only three candidates in this example it is assumed that no one votes for all three; such a vote is legal, but it would be wasted since it would raise each candidate's total by the same amount.) In this case, with 65 voters approving only their first choice, *A* would receive 50 votes of approval and win the election. If some voters choose to approve two of the three candidates, however, *B* stands to gain most because of the large number of people who rank him as their second choice. With approval voting a shift in the number of candidates meeting the approval of even a small number of voters can easily change the outcome of the election. Hence the implementation of this voting scheme would necessitate a transformation of campaign strategies, from trying to convince voters that a candidate is the best choice to trying to convince them that he is acceptable.

In the U.S., of course, presidential elections are held by the totally different rules of the electoral college. Through most of U.S. history the electoral college has served mainly to impose a unit rule

on individual states so that the winner of the popular vote in each state receives that state's entire electoral vote. According to the Constitution of the U.S., there are other significant consequences of this system that affect the outcome of three-candidate contests (in particular provisions for transferring the responsibility of deciding a presidential election from the electoral college to Congress), but here we shall examine only the consequences of the unit rule.

The most widely held view of the electoral college's unit rule has been that it favors smaller, or less populated, states, because the number of votes accorded to each state in the college is two more than its number of representatives. In relative terms these two extra votes, which represent the two senators from each state, do increase the voting strength of smaller states and diminish that of larger ones.

Paradoxically, however, the effective strength of a state in a presidential election is actually proportional to the population of the state raised to the 3/2 power. And as a result individual votes cast in the largest states are as much as three times as important as those cast in the smallest ones. This surprising conclusion is a direct consequence of elementary probability theory, and it is consistent with the spending record of candidates in recent presidential elections: candidates do devote disproportionate resources to the larger states at the expense of the smaller ones.

The "3/2 rule" is based on the assumption that candidates will generally match one another's campaign efforts in the various states. (Comparison of candidates' allocations of time and money in recent election campaigns shows that the assumption is entirely realistic.) The

reasoning begins with the obvious: Each candidate seeks to maximize his expected electoral vote, which is the sum over all 50 states of the product of each state's electoral vote and the probability that the candidate will win a majority in that state. By expressing this relation in the form of an equation and taking into account candidates' tendencies to match one another's campaign efforts from state to state it can be shown that the optimal way to maximize the expected electoral vote is to allocate campaign resources approximately in proportion to the 3/2 power of the electoral vote of each state. Thus although California has about four times the electoral vote of Wisconsin (45 compared with 11), the 3/2 rule would suggest that candidates should devote $4^{3/2}$, or 8, times more resources to California than to Wisconsin.

Another way to understand why larger states gain power rather than lose it in electoral-college politics is to examine the likelihood that any particular vote may be decisive in swinging the state for or against a particular candidate. This measure of decisiveness is the traditional way of gauging the power of an individual voter. What is needed is a measure of the average number of votes necessary to reverse the result of an election in each state.

Calculations show that the decision power of an individual in a state with v electoral votes (to be cast as a unit in the electoral college) is proportional to \sqrt{v} . Since the power of a state in the electoral college is magnified by the number of electoral votes cast by the state, the contribution of each state to the presidential decision is approximately proportional to v times \sqrt{v} , or $v^{3/2}$.

In order to gauge the relative voting power of individuals in different states the large-state bias created by the 3/2 rule must be weighed against the small-state bias of the two-senator electoral-college bonus. The significance of an individual's vote, instead of being equal for all voters, is determined by the individual's share of his state's power, and as is shown in the accompanying illustration [page 26] the different states' powers are decidedly unequal. (The broken line on the graph marks the hypothetical even distribution of power.)

Elections will always remain a matter of passion more than of logic, based on belief more than on reason. As these examples demonstrate, however, the mathematics of elections can have subtle and unexpected consequences. As in many other realms of human experience, naive expectations can be shattered by simple mathematical structures disguised as paradoxes and anomalies.