THE NEW YORKER

BOOKS

WIN OR LOSE

No voting system is flawless. But some are less democratic than others.

by Anthony Gottlieb

JULY 26, 2010

Can theorists engineer a better way to elect candidates?

Whenever the time came to elect a new doge of Venice, an official went to pray in St. Mark's Basilica, grabbed the first boy he could find in the piazza, and took him back to the ducal palace. The boy's job was to draw lots to choose an electoral college from the members of Venice's grand families, which was the first step in a performance that has been called tortuous, ridiculous, and profound. Here is how it went, more or less unchanged, for five hundred years, from 1268 until the end of the Venetian Republic.

Thirty electors were chosen by lot, and then a second lottery reduced them to nine, who nominated forty candidates in all, each of whom had to be approved by at least seven electors in order to pass to the next stage. The forty were

pruned by lot to twelve, who nominated a total of twenty-five, who needed at least nine nominations each. The twenty-five were culled to nine, who picked an electoral college of forty-five, each with at least seven nominations. The forty-five became eleven, who chose a final college of forty-one. Each member proposed one candidate, all of whom were discussed and, if necessary, examined in person, whereupon each elector cast a vote for every candidate of whom he approved. The candidate with the most approvals was the winner, provided he had been endorsed by at least twenty-five of the forty-one.

Don't worry if you blinked: bewildering complexity was part of the point. The election aimed to reassure Venetians that their new ruler could not have been eased into place by backroom deals. Venetians had been coming up with inventive ways to make political decisions for a couple of hundred years before they concocted this rigmarole. Earlier elections in Venice, and in other Italian communes, required a winner to be endorsed by two-thirds, or sometimes three-quarters, of the voters. The hallmark of the Venetian approach has come to be known as "approval voting," in which electors do not need to pick a favorite but may vote for several candidates they like.

In 1179, two years after a stay in Venice, Pope Alexander III reformed papal elections, perhaps because he liked some of what he saw there. Among other things, he abolished a tradition of requiring unanimity among the cardinals, and settled for a two-thirds majority instead. You would expect a two-thirds consensus to be easier to reach than unanimity, but papal conclaves in the thirteenth century seemed to go on forever. On six occasions, it took several months to choose a Pope. In 1241, by some accounts, the head of the civil administration in Rome threatened to exhume the corpse of the defunct Pope and parade it through the city in full regalia if the cardinals didn't settle on a new one. Eventually, the cardinals got the hang of it. After some tinkering over the years, the two-thirds rule was reconfirmed by the present Pope, in 2007.

"What is done by two-thirds of the Sacred College, that is surely of the Holy Ghost," Pius II said of his own election, in 1458. He did not explain why divine approval kicks in only at the two-thirds mark. Since then, mathematicians, economists, and political theorists have made their own attempts to elucidate the math of voting, and figure out better electoral systems. The story of these efforts is told in "Numbers Rule: The Vexing Mathematics of Democracy, from Plato to the Present" (Princeton; $26.95), by George Szpiro, a journalist and mathematician.

The book happens to be timely. The voting method that is used in Britain—and that has been kept on in some of its former colonies, including the United States—may finally be replaced. When neither of Britain's two main parties won a majority in parliamentary elections this May, the Conservative Party formed a coalition government with Britain's third party, the Liberal Democrats, who have long campaigned for electoral reform. One price of the coalition deal is a national referendum on changing the country's main voting system, which will be held in 2011. If Britain does switch, as most other developed countries already have, the cause of voting reform in holdouts like the United States, India, and Canada could get a boost.

In the standard British style of voting, each elector casts one vote, and the candidate with more votes than any other is elected. This is known as "first past the post"; the winner just needs to get more votes than anyone else, not achieve any threshold (such as a two-thirds majority). It is also called "winner takes all," to distinguish it from some other methods, which elect the top two or more candidates.

In a contest between two people for one job, first past the post seems to be merely common sense. But, as soon as there are three or more candidates on the slate, it can quickly go awry. The least popular candidate could easily win, if the opposition to him or her splits its votes between two or more other candidates. Say sixty per cent of voters are right of center and forty per cent are to the left. In a three-way contest with two equally popular right-wing candidates and one left-winger, a first-past-the-post vote will elect the left-winger, whom only a minority want. A dramatic variety of vote-splitting happens when a "spoiler" with no chance of winning manages to affect the outcome of an election by sucking away votes that could have reversed the positions of two front-runners in a close race. If several hundred Ralph Nader supporters in Florida had voted for Al Gore, the outcome of America's 2000 Presidential election would have been different.

Add political parties to the picture, and the winner-take-all system looks even worse. A party's share of seats in a parliament or a congress can diverge wildly from its over-all share of votes. In Britain's 1983 election, the Liberal-S.D.P. Alliance won more than twenty-five per cent of the votes but fewer than four per cent of the seats in Parliament. In this year's election, the Liberal Democrat Party won more than a fifth of the votes and less than a tenth of the seats. That's because Lib-Dem supporters today, like those of the Alliance in 1983, are thinly scattered across the land.

In a country with just two political parties, winner takes all can deliver a proportionate parliament or congress. But that's only because the supporters of political parties tend to live in clusters. Say that fifty-three per cent of American voters are Democrats and forty-six per cent are Republicans (mirroring the vote in last year's Presidential election). Then imagine that the supporters of the two parties were spread evenly throughout the country's districts—like Britain's Liberal Democrats, only much more so. In that case, the Democrats would win every seat and there would be no Republicans in Congress. Something like this happened in Wales in 1906, when Conservatives got thirty-four per cent of the vote and no seats.

Szpiro, who is more interested in math than in politics, says relatively little about how voting systems have played out in the real world. (Readers who want to enrage themselves and frighten their families should turn to William Poundstone's 2008 book, "Gaming the Vote," a masterly account of the way electoral mathematics is manipulated in America.) But it's clear that no country would pick first-past-the-post voting today. Of democracies with no significant British past, only Nepal now elects its national assembly this way.

The history of voting math comes mainly in two chunks: the period of the French Revolution, when some members of France's Academy of Sciences tried to deduce a rational way of conducting elections, and the nineteen-fifties onward, when economists and game theorists set out to show that this was impossible. Perched in the middle is the Reverend Charles Dodgson, better known as Lewis Carroll, the author of "Alice's Adventures in Wonderland" and "Through the Looking-Glass."

The first mathematical account of vote-splitting was given by Jean-Charles de Borda, a French mathematician and a naval hero of the American Revolutionary War. Borda concocted examples in which one knows the order in which each voter would rank the candidates in an election, and then showed how easily the will of the majority could be frustrated in an ordinary vote. Borda's main suggestion was to require voters to rank candidates, rather than just choose one favorite, so that a winner could be calculated by counting points awarded according to the rankings. The key idea was to find a way of taking lower preferences, as well as first preferences, into account.

Unfortunately, this method may fail to elect the majority's favorite—it could, in theory, elect someone who was nobody's favorite. It is also easy to manipulate by strategic voting. If the candidate who is your second preference is a strong challenger to your first preference, you may be able to help your favorite by putting the challenger last. Borda's response was to say that his system was intended only for honest men. The French Academy adopted Borda's method of electing members, until a new member, Napoleon Bonaparte, pressed the Academy to abandon it. It's not clear what Napoleon didn't like about it, but, given that his conquest of Venice deposed its last doge, Napoleon has the distinction of having quashed two unusual voting schemes. (Versions of Borda's method are still used in some sporting competitions, in the Eurovision Song Contest, and to elect the Parliament of Nauru, a Pacific atoll.)

The Marquis de Condorcet, a prominent reformer who became a secretary of the revolutionary National Assembly, made a deeper investigation of voting. In effect, Condorcet suggested dividing elections into a series of one-on-one contests, so that every candidate is directly compared with every other. If there is a candidate who wins every such match, it is clear who should be the over-all winner of the tournament. But even with as few as three candidates there may not be such a person.

While scrutinizing such contests, Condorcet noticed that a loop may arise resembling the children's playground game rock-paper-scissors. Scissors cut paper, paper wraps rock, but rock blunts scissors—so none of them is the strongest. The same could happen in a run of head-to-head votes. If more people prefer Alphonse to Benoît than prefer Benoît to Alphonse, and more prefer Benoît to Claude than Claude to Benoît, it could also be the case that more prefer Claude to Alphonse than Alphonse to Claude. In this sort of loop, which was later called a "cycle" by Charles Dodgson, the notion of an over-all winner breaks down. Condorcet's own fortunes broke down when he fell out with the Jacobins over the execution of Louis XVI, was branded a traitor, and died mysteriously in prison.

After the Academy dropped Borda's method, it plumped for a simple suggestion by the astronomer and mathematician Pierre-Simon Laplace, who was an important contributor to the theory of probability. Laplace's rule insisted on an over-all majority: at least half the votes plus one. If no candidate achieved this, nobody was elected to the Academy. For government elections, in which winners must be found somehow, Laplace was confident that a victor would emerge fairly quickly. Elections to the French National Assembly, and to the French Presidency, currently borrow from both Laplace and Condorcet. They use a two-stage method that guarantees a winning candidate with an over-all majority, by conducting a runoff round between the top two candidates from the first round. This practice may

not last. In France, altering the voting rules to entrench your power is an acceptable weapon in politics, and the parliamentary voting system has been changed a dozen times since 1870.

National politics weren't on Dodgson's mind, it appears, when he first became interested in the theory of voting, in the early eighteen-seventies. Ostensibly, he was pondering the best way for the governing body of Christ Church, Oxford, where he was a tutor in mathematics, to decide on the design for a controversial belfry, and to pick new members of the college. As to what explained his sudden interest in college politics, some people—notably the late economist and Dodgson scholar Duncan Black—have suggested that Alice Liddell, who inspired the Wonderland tale, in 1862, was at the bottom of it. Alice's father, the head of Christ Church, had forbidden Dodgson further contact with his daughters, and meddling in college politics may have been Dodgson's way of getting back at him.

For college elections, Dodgson first proposed a version of Borda's method, and also a version of Condorcet's (though he appears not to have known about Borda's and Condorcet's work). Later, he developed an interest in politics beyond the walls of Christ Church, and, in the eighteen-eighties, he tried to find ways to secure equitable representation in Parliament for minorities. Another early advocate of proportional representation was John Stuart Mill, who, in 1861, wrote about the critical distinction between "government of the whole people by the whole people, equally represented," which was the ideal, and "government of the whole people by a mere majority of the people exclusively represented," which is what winner-takes-all elections produce. (The minority that Mill was most concerned to protect was the "superior intellects and characters," who he feared would be swamped as more citizens got the vote.)

The key to proportional representation is to enlarge constituencies so that more than one winner is elected in each, and then try to align the share of seats won by a party with the share of votes it receives. These days, a few small countries, including Israel and the Netherlands, treat their entire populations as single constituencies, and thereby get almost perfectly proportional representation. Some places require a party to cross a certain threshold of votes before it gets any seats, in order to filter out extremists.

The main criticisms of proportional representation are that it can lead to unstable coalition governments, because more parties are successful in elections, and that it can weaken the local ties between electors and their representatives. Conveniently for its critics, and for its defenders, there are so many flavors of proportional representation around the globe that you can usually find an example of whatever point you want to make. Still, more than three-quarters of the world's rich countries seem to manage with such schemes. (Britons who adamantly oppose all forms of proportional representation appear to draw strength from the thought that these places are, after all, run by foreigners.)

The alternative voting method that will be put to a referendum in Britain is not proportional representation: it would elect a single winner in each constituency, and thus steer clear of what foreigners put up with. Known in the United States as instant-runoff voting, the method was developed around 1870 by William Ware, an American architect, whose other monuments are some of Harvard University's best-known buildings. Instant-runoff voting has been used for elections to Australia's House of Representatives since 1918, and is currently making patchy progress in the United States. This year, it was used to select the best picture in the Academy Awards—and was dropped in Burlington, Vermont, after use in two mayoral races. Still, it's the only alternative voting method that's making any headway at all in American public elections.

In instant-runoff elections, voters rank all or some of the candidates in order of preference, and votes may be transferred between candidates. The idea is that your vote may count even if your favorite loses. If any candidate gets more than half of all the first-preference votes, he or she wins, and the game is over. But, if there is no majority winner, the candidate with the fewest first-preference votes is eliminated. Then the second-preference votes of his or her supporters are distributed to the other candidates. If there is still nobody with more than half the votes, another candidate is eliminated, and the process is repeated until either someone has a majority or there are only two candidates left, in which case the one with the most votes wins. Third, fourth, and lower preferences will be redistributed if a voter's higher preferences have already been transferred to candidates who were eliminated earlier.

At first glance, this is an appealing approach: it is guaranteed to produce a clear winner, and more voters will have a say in the election's outcome. Look more closely, though, and you start to see how peculiar the logic behind it is. Although more people's votes contribute to the result, they do so in strange ways. Some people's second, third, or even lower preferences count for as much as other people's first preferences. If you back the loser of the first tally, then in the subsequent tallies your second (and maybe lower) preferences will be added to that candidate's first preferences. The winner's pile of votes may well be a jumble of first, second, and third preferences.

Such transferrable-vote elections can behave in topsy-turvy ways: they are what mathematicians call "non-monotonic," which means that something can go up when it should go down, or vice versa. Whether a candidate who gets through the first round of counting will ultimately be elected may depend on which of his rivals he has to face in subsequent rounds, and some votes for a weaker challenger may do a candidate more good than a vote for that candidate himself. In short, a candidate may lose if certain voters back him, and would have won if they hadn't.

Supporters of instant-runoff voting say that the problem is much too rare to worry about in real elections, but recent work by Robert Norman, a mathematician at Dartmouth, suggests otherwise. By Norman's calculations, it would happen in one in five close contests among three candidates who each have between twenty-five and forty per cent of first-preference votes. With larger numbers of candidates, it would happen even more often. It's rarely possible to tell whether past instant-runoff elections have gone topsy-turvy in this way, because full ballot data aren't usually published. But, in Burlington's 2006 and 2009 mayoral elections, the data were published, and the 2009 election did go topsy-turvy.

Although the quirks of instant-runoff voting are an extreme case, the misbehavior of voting schemes in general has been known to social scientists since the mid-twentieth century. That was when Kenneth Arrow, an economist at Stanford, examined a set of requirements that you'd think any reasonable voting system could satisfy, and proved that nothing can meet them all when there are more than two candidates. So designing elections is always a matter of choosing a lesser evil. When the Royal Swedish Academy of Sciences awarded Arrow a Nobel Prize, in 1972, it called his result "a rather discouraging one, as regards the dream of a perfect democracy." Szpiro goes so far as to write that "the democratic world would never be the same again," as if Arrow's mathematics had given a boost to the likes of Kim Jong Il. (The spread of democracy, as it happens, increased after the publication of Arrow's proof.)

There is something of a loophole in Arrow's demonstration. His proof applies only when voters rank candidates; it would not apply if, instead, they rated candidates by giving them grades. First-past-the-post voting is, in effect, a crude ranking method in which voters put one candidate in first place and everyone else last. Similarly, in the standard forms of proportional representation voters rank one party or group of candidates first, and all other parties and candidates last. With rating methods, on the other hand, voters would give all or some candidates a score, to say how much they like them. They would not have to say which is their favorite—though they could in effect do so, by giving only him or her their highest score—and they would not have to decide on an order of preference for the other candidates.

One such method is widely used on the Internet—to rate restaurants, movies, books, or other people's comments or reviews, for example. You give numbers of stars or points to mark how much you like something. To convert this into an election method, count each candidate's stars or points, and the winner is the one with the highest average score (or the highest total score, if voters are allowed to leave some candidates unrated). This is known as range voting, and it goes back to an idea considered by Laplace at the start of the nineteenth century. It also resembles ancient forms of acclamation in Sparta. The more you like something, the louder you bash your shield with your spear, and the biggest noise wins. A recent variant, developed by two mathematicians in Paris, Michel Balinski and Rida Laraki, uses familiar language rather than numbers for its rating scale. Voters are asked to grade each candidate as, for example, "Excellent," "Very Good," "Good," "Insufficient," or "Bad." Judging politicians thus becomes like judging wines, except that you can drive afterward.

Approval voting, as used in the last round of old Venetian elections, is a rating system with just two grades: approved and not approved. In the late thirteenth century, cardinals began to use it to elect Popes, too, though they stopped in 1621, perhaps because combining it with the Holy Ghost's two-thirds rule complicated matters. Approval voting was revived in the nineteen-seventies, when several American mathematicians and political scientists independently began to study it. (It is explored and defended by the N.Y.U. game theorist Steven Brams in his 2008 study "Mathematics and Democracy.") Campaigns to introduce approval voting for public elections have so far failed, but many mathematicians seem to like the way it works. The American Mathematical Society and the Mathematical Association of America, among others, use it for internal elections, though the larger Institute of Electrical and Electronics Engineers, which also adopted it, came to have second thoughts.

Range and approval voting deal neatly with the problem of vote-splitting: if a voter likes Nader best, and would rather have Gore than Bush, he or she can approve Nader and Gore but not Bush. Above all, their advocates say, both schemes give voters more options, and would elect the candidate with the most over-all support, rather than the one preferred by the largest minority. Both can be modified to deliver forms of proportional representation.

The fact that they escape Arrow's proof, though, doesn't mean that approval and range voting have no hidden kinks or paradoxes. Whether such ideas can work depends on how people use them. If enough people are carelessly generous with their approval votes, for example, there could be some nasty surprises. In an unlikely set of circumstances, the candidate who is the favorite of more than half the voters could lose. Parties in an approval election might spend less time attacking their opponents, in order to pick up positive ratings from rivals' supporters, and critics worry that it would favor bland politicians who don't stand for anything much. Defenders insist that such a strategy would backfire in subsequent elections, if not before, and the case of Ronald Reagan suggests that broad appeal and strong views aren't mutually exclusive.

Why are the effects of an unfamiliar electoral system so hard to puzzle out in advance? One reason is that political parties will change their campaign strategies, and voters the way they vote, to adapt to the new rules, and such variables put us in the realm of behavior and culture. Meanwhile, the technical debate about electoral systems generally takes place in a vacuum from which voters' capriciousness and local circumstances have been pumped out. Although almost any alternative voting scheme now on offer is likely to be better than first past the post, it's unrealistic to think that one voting method would work equally well for, say, the legislature of a young African republic, the Presidency of an island in Oceania, the school board of a New England town, and the assembly of a country still scarred by civil war. If winner takes all is a poor electoral system, one size fits all is a poor way to pick its replacements.

Mathematics can suggest what approaches are worth trying, but it can't reveal what will suit a particular place, and best deliver what we want from a democratic voting system: to create a government that feels legitimate to people—to reconcile people to being governed, and give them reason to feel that, win or lose (especially lose), the game is fair. The novelty of range and approval voting in modern politics is so great that we can't know how they'll work out without running experiments. At the very least, it would help to know how voters have dealt with such schemes in the past. Perhaps someone should start rummaging around for old ballots in the Doge's Palace. ♦

ILLUSTRATION: JOOST SWARTE