The Majority Judgement: A New Mechanism for Electing and Ranking

PROFESSOR MICHEL BALINSKI

Introduction

Today the practice of democracy implies the election of representatives. Electoral systems define how representation is apportioned among states and provinces according to their populations, and to political parties according to the votes they receive; they define how electoral districts are determined as a function of geography and population; and – when one candidate among several is to be chosen – they define exactly how voters express their opinions and how these opinions are amalgamated to determine who is elected. Each of these definitions is a "mechanism" – "a system of parts that operate or interact like those of a machine," "an instrument or process, physical or mental, by which something is done or comes into being" (according to the *American Heritage Dictionary of the English Language*).

Across the years, different nations have adopted very different mechanisms. They are all pure inventions of the human mind. However, all of them are mathematical inventions: they transform populations or total party votes (numbers) into numbers of representatives; they partition nations, provinces or states (geo-metric figures) into collections of legislative districts; they amalgamate expressions of the opinions of each voter concerning candidates into the collective decision of the electorate. The mechanisms with which numbers are transformed, areas are partitioned, and opinions are amalgamated have properties that may be analysed mathematically: they may favour the big or favour the small, they may yield electoral districts that are "fair" or grossly unfair, they may elect the candidate preferred by the electorate or some other candidate.

Regretfully, democratic practice in almost all nations is firmly in the hands of elected politicians. They are at once the players and the referees of the electoral game. They regularly attempt to manipulate the

mechanisms in order to favour their re-election. Gross inequities are sometimes avoided via judicial rulings. But the fact is that there is today a growing body of knowledge – a theory of electoral systems – that has determined fair mechanisms to resolve most if not all of the problems that arise in the establishment of an electoral system. Although politicians have resisted recourse to this knowledge, slowly but surely it is beginning to spread, and it is inevitable that soon democratic societies will demand the use of truly fair electoral mechanisms.

This paper will address only one of these problems: How is one candidate among many to be elected or, how are several candidates to be ranked (the first designated the winner)? It is an important problem. There is a growing awareness that the mechanisms used to solve this problem – in the United States, in France, in the United Kingdom, and elsewhere – may not be electing the candidate preferred by the electorate.

Practice

First-past-the-post is probably the most used system for electing one among several candidates (United States, United Kingdom). Each voter expresses his or her opinion by naming exactly one candidate (or none). The candidates are ranked according to the number of times they are named, the winner is the candidate most often named. This is a very bad mechanism! The 2000 United States presidential election shows why (see Table 1). The mere presence of a minor candidate (Ralph Nader), who had absolutely no chance of winning, made Bush the winner instead of Gore: there is no question that the bulk of those who voted for Nader preferred Gore to Bush, but the method of election did not allow them to express it. Had Nader not been a candidate in Florida it seems certain that Gore would have had more votes than Bush, and so would have won Florida's 26 electoral votes, making Gore the winner with 291 electoral votes to Bush's 245. This is an instance of the famous *Arrow paradox*: an irrelevant candidate can change the outcome of the election.

2000 Election	Nation		Florida	
	Votes	Electoral Votes	Votes	Electoral Votes
George W Bush	50,456,002	271	2,912,790	26
Albert Gore	50,999,897	266	2,912,253	0
Ralph Nader	2,882,955	0	97,488	0

Table 1. 2000 U.S. presidential election

Two-past-the-post is used in France. Each voter expresses his or her opinion by naming exactly one candidate (or none). The candidates are ranked according to the number of times they are named, and if one candidate is named by more than 50% of the voters, he or she is elected. Otherwise, there is a run-off between the two candidates most often named to determine the order between them. This is another very bad mechanism! The 2002 French presidential election shows why (see Tables 2a, b).

Chirac	Le Pen	Jospin	Bayrou	Laguiller	Chévènement
19,88%	16,86%	16,18%	6,84%	5,72%	5,33%

Mamère	Besancenot	Saint-Josse	Madelin	Hue	Mégret
5,25%	4,25%	4,23%	3,91%	3,37%	2,34%

(Pasqua)	Taubira	Lepage	Boutin	Gluckstein
(0%)	2,32%	1,88%	1,19%	0,47%

Table 2a. 2002 French presidential election, first-round results (16 candidates, 72% participation), Pasqua not a candidate

Actual		No Taubira		Pasqua no Chévènement	
Chirac	Le Pen	Chirac	Jospin	Jospin	Le Pen
82,21%	7,79%	< 50%?	> 50%?	> 75%	< 25%

Table 2b. French presidential election, second-round results: actual (80% participation, left), Taubira not a candidate (centre), Pasqua a candidate and Chévènement not a candidate (right).

Jacques Chirac, the incumbent President, was the candidate of the Rassemblement pour la République (RPR), the big party of the "legitimate" right; Lionel Jospin, the incumbent Prime-Minister, that of the Parti Socialist (PS); Jean-Marie Le Pen that of the extreme right, Front National party (FN); and Francois Bayrou that of the moderate Union pour la Démocratie Francaise (UDF, the ex-President Valéry Giscard d'Estaing's party). The others were candidates of the extreme right or the extreme left. France fully expected a run-off between Chirac and Jospin, and was profoundly shocked to be faced with a choice between Chirac and Le Pen. Chirac crushed Le Pen, obtaining 82.2% of the votes in the second round, but the vast majority of Chirac's votes were against Le Pen rather than for him. The left – socialists, communists, trotskyists, . . . – had no choice but to vote for Chirac! His votes represented very different sentiments and intensities.

Most polls predicted that Jospin would have won against Chirac with a narrow majority; Sofres predicted a 50%-50% tie on the eve of the first round. Had either Chévènement, an ex-socialist, or Taubira, a socialist, withdrawn, most of his 5.3% or her 2.3% of the votes would have gone to Jospin, so the second round would have seen a Chirac-Jospin confrontation, as had been expected. In fact, Taubira had offered to withdraw if the PS was prepared to cover her expenses, but that offer was refused. It has also been whispered that the RPR helped to finance Taubira's campaign (a credible strategic gambit backed by no specific evidence). Moreover, if Charles Pasqua, an aging past ally of Chirac, had been a candidate – as he had announced he would be – then he could well have drawn a sufficient number of votes from Chirac to produce a second round between Jospin and Le Pen, which would have resulted in a lopsided win for Jospin.

The moral of the story is this: Anything can happen when the "first-past-the-post" or the "two-two-past-the-post" mechanism is used! This is again nothing but Arrow's paradox: the winner depends on the presence or absence of candidates including those who have absolutely no chance of winning. It also shows that the mechanisms invite "strategic" candidacies: candidates who cannot hope to win (or survive a first round) but can cause another to win (or to reach the second round) by drawing votes away from an opposing candidate.

Traditional Theory

Traditionally – in the common imagination and in the theory of voting – a voter is assumed to have in his or her mind a list of preferences over the candidates that is ordered from best to worst. Indeed, some countries (notably Australia and Ireland) ask voters to give their lists of preferences from best to worst. This reveals more information about voters' opinions but also may falsify their intents. A voter who believes that there is only one decent candidate is unable to express this, and his second-ranked candidate will weigh as importantly as the second-ranked candidate of a voter who believes that candidate is excellent.

The great hope (since 1299) has been to choose a *Condorcet-winner*: a candidate who beats every possible opponent face-to-face. But there may be no Condorcet-winner, as is shown by the following example (where, for example, 30% of the voters prefer A to B and B to C):

30%	32%	38%
A	B	C
B	C	A
C	A	B

	A	B	C
A	_	68%	30%
B	32%	_	62%
C	70%	38%	_

Table 3. Condorcet paradox

The table on the right gives the scores in the head-to-head confrontations between pairs of candidates. This is an instance of the famous Condorcet paradox: A defeats B with 68% of the votes, B defeats C with 62% of the votes, and C defeats A with 70% of the votes.

Borda's method uses the voters' preference lists. It assigns each candidate the sum of his votes against all other opponents, and ranks the candidates accordingly. In the above example, A's "Borda-score" is 98, B's 94 and C's 108.

The following toy example is informative. It admits a Condorcet-winner, C.

5%	33%	34%	28%
A	A	B	C
B	C	C	B
C	B	A	A

	A	B	C
A	_	38%	38%
B	62%	_	39%
C	62%	61%	_

Table 4. The method determines the winner.

With first-past-the-post, A is first, B second and C last, written A >B > C. With two-past-the-post the order of finish is B >A > C. With Borda the order of finish is C > B > A. In short, the method determines the winner. Moreover, it pays for voters to misrepresent their preferences. If, with first-past-the-post, the 28% vote for B instead of C, B wins which is better for them. If, with two-past-the-post, the 33% vote for C instead of A in the first round, C wins which is better for them. If, with Borda, the 28% vote B > C > A instead of C > B > A, B wins which is better for them. This short discussion has raised difficulties that arise with different methods of voting: Arrow's paradox, Condorcet's paradox and the evident possibility that it pays voters to misrepresent their preferences. The fact of the matter is that this is an unavoidable conundrum of the traditional model. A reasonable method should: permit voters to list candidates in any order they wish; declare a candidate the winner if he or she is first on every voter's list; and never change the winner when some "irrelevant" candidate enters the race or withdraws.

Arrow's famous "impossibility" theorem shows that there is no method that can meet these three conditions.

A New Theory

So what is to be done? The problem of electing and ranking is vast. Nations, societies, unions and other large institutions elect presidents, senators, representatives, treasurers, judges, sheriffs, . . . ; juries (of 5 to 12) rank figure skaters, gymnasts, divers, wines, cheeses, . . . ; juries of companies rank employees ("forced ranking"); committees rank nominees for (Nobel, literary, scientific,. . .) awards and prizes; all kinds of juries rank universities, hospitals, restaurants, hotels, movies, beauty queens, muscle men, professors, students, dogs, pianists, flutists, marching bands, ... As Arthur Miller (the dramatist and one-

time husband of Marylin Monroe) once quipped, "We're ranking everybody every minute of the day."

It is a curious fact that the theoreticians of voting - the specialists of the theory of social choice – have steadfastly analysed the problem in the same way since 1299 when Ramon Llull first proposed face-to-face votes between every pair of candidates. Since then voters and judges are assumed to rank-order the competitors from best to worst, and the problem is to find a rule or mechanism to amalgamate these into the rank-order of society or the jury. The fact is curious for two reasons. First, conceiving of the problem in this way leads to paradoxes, impossibility and incompatibility theorems (notably Arrow's, but many others as well): i.e., the model leads to an inconsistent theory. Second, voters and judges invariably do not have rank-orders in their minds: instead, they evaluate the merits of the candidates or of the performances of competitors.

"During the Middle Ages," Richard Feynman wrote, "there were all kinds of crazy ideas, such as that a piece of rhinoceros horn would increase potency. Then a method was discovered for separating the ideas – which was to try one to see if it worked, and if it didn't work, to eliminate it. This method became organised, of course, into science." The time has come to discard the traditional view, replace it with a more realistic one, and accept its logical implications. Pragmatic people – faced with judging figure skaters, gymnasts, wines, pianists, . . . , – have *all* invented their own *ad hoc* methods. Instead of rank-ordering competitors, judges assign them points and the points are used to determine the juries' rank-orders. The difficulty with their methods is that they invariably add or average the points to determine rank-orders, and this opens the door to strategic manipulation and outright cheating (by giving high points to favourites, low points to their opponents), as happened in the big scandal of the 2002 Winter Olympic Games in pairs figure skating. On the other hand, the points they use are invariably well defined and constitute what we – my colleague Rida Laraki and I – now call *common languages*: the meanings of the points are clear to everyone, though there may be disagreement on how many points a competitor merits. "Kenneth is an A+ student" is a meaningful statement (or was before the age of grade inflation). Told that "Sonja's free skating performance is worth 5.9" when the traditional "0 to 6"

scale was used – that "her skating skills component is 7.75" with the newly adopted scale – or that "Xu Sang's inward flying 1½ somersault was a 9.0," means something specific to figure skating or diving enthusiasts. Other everyday examples of common languages are the number of stars given a hotel or a restaurant. With this point of view Arrow's theorem says: if there is no common language there can be no consistent decisions. This makes sense: imagine the presidents of China and the USA trying to reach a decision with no common language!

Majority Judgement

The *majority judgement*[1] is a method of voting and judging that emerges as the optimal method by the criteria of the traditional theory of voting – when the voters evaluate candidates (or judges evaluate competitors).

The majority judgement itself is easy enough to explain; indeed, its explanation has persuaded some pragmatists to accept it already. The basic point of view is that voters and judges do not vote: they evaluate candidates in a common language of grades. So, to begin, there must be a common language.

Large scale voting experiments were conducted in parallel with two elections, the 2007 French and 2008 American presidential elections. In one experiment members of a scientific society were invited to participate in an experiment conducted on the web. They were given the ballot of Table 5 (Dem. means Democratic, Rep. means Republican and Ind. means Independent). The common language of grades is: *Excellent, Very Good, Good, Acceptable, Poor,* or *to Reject*. Extensive statistical analyses of the results of the experiment in the French presidential election – where the common language was *Très bien, Bien, Assez Bien, Passable, Insuffisant, à Rejeter* and over 1,700 voters participated - shows that the language was indeed common in that it was used in the same way by the voters.

The ballot deliberately poses a clear and solemn question inviting voters to evaluate the candidates. The experiments show that voters have no difficulty in giving their evaluations: in the French election

[1] The idea of the majority judgement was developed in research pursued with Rida Laraki. Its theory and practice is fully explained in our forthcoming book.

there were twelve candidates, yet most voters filled out the ballots within a minute or a minute and a half.

Ballot: Election of the President of the United States of America 2008
To be the President of the United States of America, having taken into account all relevant considerations, I judge, in conscience, that this candidate would be:

	E	*VG*	*G*	*A*	*P*	*TR*	*NO*
Michael R. Bloomberg, Ind.							
Hillary R. Clinton, Dem.							
John R. Edwards, Dem.							
Michael D. Huckabee, Rep.							
John S. McCain, Rep.							
Barack H. Obama, Dem.							
Collin L. Powell, Ind.							
W. Mitt Romney, Rep.							

You must check one single grade or "No opinion" in the line of each candidate. "No opinion" is counted as *To Reject.*
E=Excellent; VG=Very Good; G=Good; A=Acceptable; P=Poor; TR=To Reject; NO=No Opinion
Table 5. Ballot, web experiment, U.S.A., October 2008.

The usual method (first-past-the-post) offers voters nine possibilities to express their opinions: to indicate one of eight candidates, or none. The majority judgement offers 6^8 = 1, 679, 616 possibilities to express their opinions. Voters interviewed during the French 2007 experiment repeatedly voiced their appreciation for being able to adequately express their opinions with the majority judgement. They also distinctly liked the idea that each candidate, including the winner, receives a "final-grade" (the majority-grade, described below).

The results were:

	E	*VG*	*G*	*A*	*P*	*TR*
Barack H. Obama, Dem.	35.9%	32.1%	12.2%	8.4%	7.6%	3.8%

Hillary R. Clinton, Dem.	16.0%	29.0%	21.4%	16.8%	11.5%	5.3%
Colin L. Powell, Indep.	10.7%	22.1%	26.0%	26.7%	9.2%	22.1%
Michael R. Bloomberg, Indep.	3.1%	14.5%	24.4%	26.7%	9.2%	22.1%
John R. Edwards, Dem.	1.5%	13.0%	22.1%	30.5%	18.3%	14.5%
John S. McCain, Rep.	3.1%	7.6%	23.7%	21.4%	30.5%	13.7%
W. Milt Romney, Rep.	0.8%	7.6%	10.7%	27.5%	30.5%	22.9%
Michael D. Huckabee, Rep.	3.8%	3.8%	6.1%	19.8%	19.1%	47.3%

Table 6. Results, web experiment, U.S.A., October 2008

The *majority-grade* of a candidate is the middlemost (or median) of his or her grades. When there are many voters, a majority of voters assign a candidate at least his or her majority-grade, and also a majority of voters assign the candidate at most his or her majority-grade. For example, Clinton's majority-grade is *Good*: 16.0% + 29.0% + 21.4% = 66.4% assign her at least *Good* and 21.4% + 16.8% + 11.5% + 5.3% = 55.0% assign her at most *Good*. It is *Good*+ because the percentage above *Good* is greater than the percentage below *Good* (otherwise it would have been a *Good* -).

When there are eight candidates and six grades some candidates must be assigned the same majority-grade. How are they to be ranked? Suppose a is the majority-grade, p the % of grades higher than α, q the % of grades lower than a. Then the *majority-gauge* is (p; a±; q), where $p > q$ implies a is endowed with a+, and otherwise it is endowed with a- .

Thus Clinton's majority-gauge is (45.0%, *Good*+, 33.6%). The majority-gauges determine the *majority-ranking*: the rank-order of the candidates determined by the majority judgement. Naturally, *a*+ ranks higher than *a*-, which suffices to rank-order all the candidates except Bloomberg and Edwards who both have the majority-grade *Acceptable*+. If two candidates have an *a*+, then the one with the larger *p* ranks higher; and if two candidates have an *a*-, then the one with the higher *q* ranks lower. So Bloomberg with *p* = 42.0% ranks higher than Edwards with *p* = 36.6%.

	p	*a* ±	*q*
1st Barack H. Obama	35.9%	*Very Good*+	32.0%
2nd Hillary R. Clinton	45.0%	*Good*+	33.6%
3rd Colin L. Powell	32.8%	*Good_*	41.2%
4th Michael R. Bloomberg	2.0%	*Acceptable*+	31.3%
5th John R. Edwards	36.6%	*Acceptable*+	32.8%
6th John S. McCain	33.4%	*Acceptable_*	44.2%
7th W. Milt Romney	46.6%	*Poor*+	22.9%
8th Michael D. Huckabee	33.5%	*Poor_*	47.3%

Table 7. Majority-ranking, web experiment, U.S.A., October, 2008

It is easy to see that the majority judgement possesses many desirable properties. The Condorcet paradox is impossible for there can be no cycle of the type: candidate A leads candidate B, B leads C, and C leads A. The Arrow paradox is impossible as well: when a candidate enters the race or withdraws the majority-gauges of the other candidates remain the same, so the majority-ranking among them remains the same as well. The majority judgement also resists strategic manipulation. One or *all* of those who gave Clinton the grade *Very Good* (that is, above her majority-grade) cannot change her majority-gauge *except* to lower it (presumably not their intention since they gave her a higher grade than her majority-grade). Similarly, one or all of those who gave her *Acceptable, Poor* or *to Reject* (that is, below her majority-grade) cannot change her majority-gauge *except* to raise it (presumably not their intention since they gave her a lower grade than her majority-grade). And the same holds for all candidates. Although all strategic

manipulation cannot be eliminated, the majority judgement best resists it according to several different criteria.

The majority judgement is not an *ad hoc* invention based purely on intuition. It is the logical outcome of the search for an optimal method of election given that the merits of candidates (or competitors) are to be evaluated. The fact that it is widely and easily accepted intuitively attests to its robustness.

References

Balinski, Michel and Laraki, Rida (2007). "Le jugement majoritaire: l'expérience d'Orsay," *Commentaire* no. 118, été 2007, pp. 413-419.

_____ (2007). "A Theory of Measuring, Electing and Ranking", *Proceeding of the National Academy of Sciences USA*, May 22, 2007, vol. 104, no. 21, pp. 8720-8725.

_____ (2007). "Election by Majority Judgement: Experimental Evidence." *Cahier du Laboratoire d'Économétrie de l'École Polytechnique*, December 2007, no. 2007-28.

_____ . *Majority Judgement: Measuring, Ranking and Electing*, to appear in 2010, M.I.T. Press.

web-site:http://ceco.polytechnique.fr/jugement majoritaire.html

U.S. Patent pending for "The Qualified-Majority Judgement" (including the multi-criteria case).