Election by Majority Judgement: Experimental Evidence

Michel Balinski, Rida Laraki

ECOLE POLYTECHNIQUE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Election by Majority Judgement: Experimental Evidence

Michel Balinski
Rida Laraki

17 décembre 2007

Cahier n° 2007-28

LABORATOIRE D'ECONOMETRIE
1rue Descartes F-75005 Paris
(33) 1 55558215
http://ceco.polytechnique.fr/
mailto:lyza.racon@shs.poly.polytechnique.fr

Election by Majority Judgement: Experimental Evidence

Michel Balinski[1]
Rida Laraki[2]

17 décembre 2007

Cahier n° 2007-28

Résumé: Le jugement majoritaire est une méthode d'élection. Cette méthode est l'aboutissement d'une nouvelle théorie du choix social où les électeurs jugent les candidats au lieu de les ranger. La théorie est développée dans d'autres publications ([2, 4]). Cet article décrit et analyse des expériences électorales conduites pendant les deux dernières élections présidentielles françaises dans plusieurs buts: (1) démontrer que le jugement majoritaire est une méthode pratique, (2) la décrire et établir ses principales propriétés, (3) démontrer qu'elle échappe aux paradoxes classiques, et (4) illustrer comment dans la pratique tous les mécanismes de vote connus violent certains critères importants. Les démonstrations utilisent des concepts et méthodes nouveaux.

Abstract: The majority judgement is a method of election. It is the consequence of a new theory of social choice where voters judge candidates instead of ranking them. The theory is explained elsewhere [2, 4]. This article describes and analyzes electoral experiments conducted in parallel with the last two French presidential elections to: (1) show that the majority judgement is a practical method, (2) describe it and its salient properties, (3) establish that it escapes the classical paradoxes, (4) illustrate how in practice the well known electoral mechanisms all fail to meet important criteria. The demonstrations introduce new concepts and methods.

Mots clés: choix social, mécanismes de vote, expérience électorale, jugement majoritaire, le paradoxe d'Arrow

Key Words: social choice, voting mechanism, electoral experiment, majority judgement, Arrow's paradox

Classification: AMS: 91B14, 91B12, 91A80.
JEL: D71, D72, C72.

[1] CNRS et Ecole Polytechnique.
[2] CNRS et Ecole Polytechnique.

Election by Majority Judgement: Experimental Evidence

Michel Balinski and Rida Laraki

École Polytechnique and C.N.R.S., Paris, France

December 17, 2007

Abstract

The majority judgement is a method of election. It is the consequence of a new theory of social choice where voters judge candidates instead of ranking them. The theory is developed elsewhere [2, 4]. This article describes and analyzes electoral experiments conducted in parallel with the last two French presidential elections to: (1) show that the majority judgement is a practical method, (2) describe it and establish its salient properties, (3) establish that it escapes the classical paradoxes, and (4) illustrate how in practice the well known electoral mechanisms all fail to meet important criteria. These demonstrations introduce new concepts and methods.

Introduction

Throughout the world the choice of one from among a set of candidates is accomplished by elections. Elections are mechanisms for amalgamating the wishes of individuals into a decision of society. Many different mechanisms have been proposed and/or used.

Most rely on the idea that voters compare candidates—one is better than another—so have lists of "preferences" in their minds. These include first-past-the-post (in at least two avatars), Condorcet's [9], Borda's [7] (and similar methods that assign scores to places in the lists of preferences and then add them), convolutions of Condorcet's and/or Borda's, the single transferable vote (also in at least two versions), and approval voting (in one interpretation).

Electoral mechanisms are also used in a host of other circumstances where winners and orders-of-finish must be determined by a jury of judges, including figure skaters, divers, gymnasts, pianists, and wines. Invariably, as the great mathematician Laplace was the first to propose two centuries ago [18], they ask voters (or judges) not to compare but to evaluate the competitors by assigning points from some range, points expressing an absolute measure of the competitors' merits. Laplace suggested the range $[0, R]$ for some arbitrary positive real number R, whereas practical systems usually fix R at some positive integer. These mechanisms rank the candidates according to the sums or the averages of their points[1] (sometimes after dropping highest and lowest scores). They have been emulated in various schemes proposed for voting with ranges taken to be integers in $[0, 100]$, or in $[0, 5]$, or the integers 0, 1, and 2, or the integers 0 and 1 (approval voting).

It is fair to ask whether any of these mechanisms—based on comparisons or sums of measures of merit—actually makes the choice that corresponds to the true wishes of society, in theory or in practice. All have their supporters, yet all have serious drawbacks: every one of them fails to meet some property that a good mechanism should satisfy. In consequence, the basic challenge remains: to find a mechanism of election, prove it satisfies the properties, and show it is practical.

The existing voting mechanisms have for the most part been viewed and analyzed in terms of the traditional model of social choice theory: individual voters have in their minds "preference" lists of the candidates, and the decision to be made is to find society's winning candidate or to find society's "preference" list from best (implicitly the winner) to worst. *All* of the mechanisms based on this model are wanting because of *paradoxes* that occur in practice—Condorcet's, Arrow's and others—and *impossibilities*—Kenneth Arrow's [1] and Gibbard-Satterthwaite's [13, 23]. Moreover, as Peyton Young has shown [24, 25], in this model finding the rank-ordering wished by a society is a very different problem than finding the winner wished by a society: said more strikingly, the winner wished by society is not necessarily the first placed candidate of the

[1] Laplace only used this model to deduce Borda's method via probabilistic arguments. He then rejected Borda's method because of its evident manipulability.

ranking wished by society! In fact, the traditional model harbors a fundamental *incompatibility* between winning and ranking [2, 4]. The mechanisms based on assigning points and summing or averaging them seem to escape the Arrow paradox (though that, it will be seen, is an illusion), but they are *all* wide open to strategic manipulation. However, evaluating merits, as Laplace had imagined, opens the door to a new theory free of these defects.

"During the Middle Ages," Richard Feynman once 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 organized, of course, into science." The idea that voting depends on comparisons between pairs of candidates—the basic paradigm of the theory of social choice—dates to medieval times: Ramon Llull proposed a refinement of Condorcet's criterion in 1299 and Nicolaus Cusanus proposed Borda's method in 1433 (see, [20, 14, 15]). The impossibility and incompatibility theorems are one good reason to discard the traditional model. The 2007 experiment with the majority judgement described in this article provides another: fully one third of the voters declined to designate one "favorite" candidate, and on average voters rejected over one third of the candidates. These evaluations cannot be expressed with "preference" lists. Thus, on the one hand the traditional model harbors internal inconsistencies, and on the other hand voters do not in fact have in their minds the inputs the traditional model imagines, rank orders of the candidates. The model doesn't work, so must be eliminated.

The majority judgement is a new mechanism based on a different model of the problem of voting (inspired by practices in ranking wines, figure skaters, divers, and others). It encompasses the traditional approach and the scoring systems. It asks voters to evaluate every candidate in a common language of grades—thus to *judge* each one on an absolute scale—rather than to compare them. Assigning a value or grade permits comparisons of candidates, comparisons of candidates does not permit evaluations (or any expression of intensity). In this paradigm the majority judgement emerges as the unique acceptable mechanism for amalgamating individuals' wishes into society's wishes. Given the grades assigned by voters to the candidates, it determines the final-grades of each candidate and orders them according to their final-grades. The final-grades are *not* sums or averages. The majority judgement avoids the paradoxes and impossibilities of the traditional model.

The *theory* that shows why the majority judgement is a satisfactory answer to the basic challenge is described and developed elsewhere (see [2, 4]). The aim of this article is to describe electoral experiments that show it provides a *practical* answer to the basic challenge and that it satisfies the important properties of social choice—or comes as close to satisfying them as possible. The demonstration invokes new methods of validation and new concepts. The experiments, and the elections in which they were conducted, are also used to show how the well known mechanisms fail to satisfy important properties.

1 Background of the experiments

The experiments were conducted in the context of the French presidential elections of 2002 and 2007. To begin, their salient features are described.

The French constitution (Article 7) states:

> The president of the Republic is elected by an absolute majority of the votes. If it is not obtained in the first round of the election, a second round is held two Sundays later. The only two candidates who may present themselves, after the eventual withdrawal of more favored candidates, are those who have the largest number of votes in the first round.

The precise mechanism used in each of the rounds[2] is implicitly the "first-past-the-post" system: it gives to each voter the possibility of casting one vote for at most one candidate and the "order of finish" is determined by the total obtained by each candidate. Except for the provision of a "run-off" between the top two finishers, this is exactly the mechanism used in the U.S. presidential elections and primaries in each state: an elector has no way of expressing her or his opinions concerning candidates except to designate exactly one "favorite." In consequence—imagine for the moment a field of at least three candidates—his or her vote counts for nothing in designating the winner unless it was cast for the "winner," for no expression concerning the remaining two or more candidates is possible.

Moreover, the first-past-the-post system is subject to *Arrow's paradox*—the winner may change because of the presence or absence of "irrelevant" candidates—as is practically every system that is used to elect a candidate throughout the world. The U.S. presidential election of 2000 is a good example (see table 1). Ralph Nader had no chance whatever to be elected, but his candidacy for Florida's 26 electoral votes alone was enough to change the outcome of the election.[3]

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

Table 1. Votes: United States presidential election of 2000.

French presidential election of 2002

The French presidential election of 2002 with its sixteen candidates is a veritable story-book example of the inanity of the first-past-the-post mechanism (see

[2]There have always been two rounds. The first direct popular election of the President in the fifth Republic (instituted in 1958) was in 1965: in the first round Charles de Gaulle had 44.64% of the vote, François Mitterrand 31.72%. Together they received 76.36%. In every subsequent election the top two together received a lower percentage. In 2002 the top seven together received 76.04%.

[3]This, of course, assumes that the vast majority of Nader's votes would have gone to Gore.

table 2). 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 François Bayrou that of the moderate Union pour la Démocratie Française (UDF, the ex-President Valéry Giscard d'Estaing's party). Arlette Laguiller was the perennial candidate of a party of the extreme left, the Lutte Ouvrière. The extreme right had two candidates, Le Pen and Bruno Mégret; the moderate right five, Chirac, Bayrou, Alain Madelin, Christine Boutin, and Corinne Lepage; the left and greens four, Jospin, Jean-Pierre Chévènement, Christiane Taubira, and Noël Mamère; and the extreme left four, Laguiller, Olivier Besancenot, Robert Hue, and Daniel Gluckstein. One group managed to present only one candidate, Jean Saint-Josse: the hunters.

J. Chirac	J.-M. Le Pen	L. Jospin	F. Bayrou
19.88%	16.86%	16.18%	6.84%
A. Laguiller	J.-P. Chévènement	N. Mamère	O. Besancenot
5.72%	5.33%	5.25%	4.25%
J. Saint-Josse	A. Madelin	R. Hue	B. Mégret
4.23%	3.91%	3.37%	2.34%
C. Taubira	C. Lepage	C. Boutin	D. Gluckstein
2.32%	1.88%	1.19%	0.47%

Table 2. Votes: French presidential election, first-round, April 21, 2002.

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.[4] Had either Chévenèment, 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.

[4] In their last 11 predictions (late February to the election), the Sofres polls showed Jospin winning 7 times, Chirac 2 times, a tie 2 times.

Anything can happen when the "first-past-the-post" (or the "first-two-past-the-post") mechanism is used! This—and the Nader Florida phenomenon—is 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 is also a clear proof that these 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 .

French presidential election of 2007

French voting behavior in the presidential election of 2007 was very much influenced by the experience of 2002. There were twelve candidates. Nicolas Sarkozy was the candidate of the UMP (Union pour un Mouvement Populaire, founded in 2002 by Chirac), its president and the incumbent minister of the interior; Ségolène Royal that of the PS; Bayrou again that of the UDF (though he announced immediately after the first round that he would create a new party, the MoDem or Mouvement démocrate); and Le Pen again that of the FN. The extreme left had five candidates—Besancenot (again), Marie-George Buffet, Laguiller (again), José Bové, and Gérard Schivardi—, the extreme right had two—Le Pen (of course) and Philippe de Villiers—and the hunters one, Frédéric Nihous.

N. Sarkozy	S. Royal	F. Bayrou	J.-M. Le Pen
31.18%	25.87%	18.57%	10.44%
O. Besancenot	P. de Villiers	M.-G. Buffet	D. Voynet
4.08%	2.23%	1.93%	1.57%
A. Laguiller	J. Bové	F. Nihous	G. Schivardi
1.33%	1.32%	1.15%	0.34%

Table 3. Votes: French presidential election, first round, April 22, 2007.

The distribution of the votes among the twelve candidates in the first round is given in table 3. In the second round Nicolas Sarkozy defeated Ségolène Royal by 18,983,138 votes (or 53.06%) to 16,790,440 (or 46.94%).

In response to the debacle of 2002, the number of registered voters increased sharply (from 41.2 million in 2002 to 44.5 million in 2007), and voter participation was mammouth: 84% of registered voters participated in both rounds. Voting is, of course, a strategic act. In 2007 voters were acutely aware of the importance of who would survive the first round. Many who believed that voting for their preferred candidate could again lead to a catastrophic second round, voted differently. Some, in the belief that their preferred candidate was sure to reach the second round, may have voted for that candidate's easiest-to-defeat opponent. Such behavior—a deliberate strategic vote for a candidate who is not the elector's favorite ("le vote utile")—was much debated by the candidates and the media, and was practiced. A poll conducted on election day[5] asked electors

[5]by Tns - Sofres - Unilog Groupe Logica CMG, April 22, 2007.

what most determined their votes. One of the seven possible answers was a deliberate strategic vote: this answer was given by 22% of those (who said they voted) for Bayrou, 10% of those for Le Pen, 31% of those for Royal and 25% of those for Sarkozy. Comparing the first rounds in 2002 and 2007 also suggests deliberate strategic votes were important in 2007: in 2002 the seven minor candidates of the left and the greens (Laguiller, Chévènement, Mamère, Besancenot, Hue, Taubira, Gluckstein) had 26.71% of the vote whereas in 2007 six obtained only 10.57% (Besancenot, Buffet, Voynet, Laguiller, Bové, Schivardi); in 2002 the five minor candidates of the right and the hunters (Saint-Josse, Madelin, Mégret, Lepage, Boutin) had 13.55% of the vote whereas in 2007 two obtained only 3.38% (Villiers, Nihous).

The very fact of being a candidate is a strategic act. To become an official candidate requires five hundred signatures. They are drawn from a pool of about forty-seven thousand elected officials who represent the one hundred departments, must include signatures coming from at least thirty departments, but no more than 10% from any one department. Both Besancenot and Le Pen appeared to have difficulty in obtaining them. Sarkozy publicly announced he would help them obtain the necessary signatures, as a service to democracy.

In the period leading up to the first round of voting, the major candidates of the right and the left—Sarkozy of the UMP and Royal of the PS—both argued strenuously against Bayrou, the centrist. Both most feared him in a one-to-one confrontation. The polls show why: as of February 2007 they consistently suggested that Bayrou would defeat either one of them in the second round. Immediately after the first round, Royal and Sarkozy both sought Bayrou's support[6] and tried to incorporate some of his ideas along with theirs. Once elected, Sarkozy, in naming many political personalities of the left to responsible political positions (ministries, commissions, a coveted international position, ...), put into effect one of Bayrou's principal promises, the appointment of persons from the left *and* the right ("l'ouverture").

Polling results (table 4) suggest that François Bayrou was the *Condorcet-winner*: he would have defeated *any* candidate in a head-to-head confrontation. Moreover, the pair by pair confrontations (of March 28 and April 19) determine an unambiguous order of finish (there is no "Condorcet cycle"): Bayrou is first, Sarkozy second, Royal third and Le Pen last. The information in table 4 (of March 28 and April 19) suffices to determine the "Borda scores"[7] among the four candidates. On March 28 the Borda-scores were: Bayrou 195, Sarkozy 184, Royal 164, and Le Pen 57. On April 19 they were: Bayrou 193, Sarkozy 180, Royal 164, and Le Pen 63. Condorcet and Borda agree on the order of finish.

Another experiment [11] was conducted in Faches-Thumesnil (a small town in France's northern-most department, Nord) on election day, where the official

[6] Royal subsequently revealed that she had offered Bayrou the position of prime-minister at that time.

[7] A candidate's *Borda-score* is the sum of the votes he or she receives in all pair by pair votes. Equivalently, with n candidates, a voter gives $n-1$ *Borda-points* to the first candidate on his/her list, $n-2$ to the second, down to 0 to the last. The sum of a candidate's Borda-points is the candidate's Borda-score.

results of the first round were close to the national percentages. Voters were asked to rank-order the candidates, permitting the face-by-face confrontations to be computed (see table 5): they yield the same unambiguous order of finish among the four significant candidates. It may be observed that once again the order of finish agrees with the Borda-ranking.

	Dec. 15	Jan. 20	Feb. 15	Mar. 15	Mar. 28	Apr. 16	Apr. 19
Bayrou	45%	49%	52%	54%	54%		55%
Sarkozy	55%	51%	48%	46%	46%		45%
Bayrou	43%	50%	54%	60%	57%		58%
Royal	57%	50%	46%	40%	43%		42%
Bayrou					84%		80%
Le Pen					16%		20%
Sarkozy	49%	51%	53%	54%	54%	53%	51%
Royal	51%	49%	47%	46%	46%	47%	49%
Sarkozy					84%		84%
Le Pen					16%		16%
Royal					75%		73%
Le Pen					25%		27%

Table 4. Polls from December 2006 to April 2007 on potential second round results (by IFOP, except on March 15 by Tns/Sofres).[8]

	Bayrou	Sarkozy	Royal	Le Pen
Bayrou	–	52%	60%	80%
Sarkozy	48%	–	54%	83%
Royal	40%	46%	–	73%
Le Pen	20%	17%	27%	–

Table 5. Projected second round results, from vote in Faches-Thumesnil experiment [11]. (*E.g.*, Sarkozy has 48% of the votes against Bayrou.)

2 The Majority Judgement

2007 experiment

The experiment took place in three of Orsay's twelve voting precincts (the 1^{st}, 6^{th} and 12^{th}). Orsay is a suburban town some 22 kilometers from the center of Paris. In 2002 it was the site of the first large electoral experiment conducted in parallel with a presidential election ([5], discussed below). The three precincts were chosen among the five of the 2002 experiment as the most representative of the town and its various socio-economic groups. Potential participants were informed about the experiment well before the day of the first round by letter, an article in the town's quarterly magazine, an evening presentation open to all,

[8] A blank indicates no figure is available. Many more Sarkozy *vs.* Royal polls were conducted.

and posters (as had been done in 2002). The various communications explained how the votes would be tallied and the candidates listed in order of finish, and showed the ballot they would be asked to use.

It is important to realize that the three precincts of Orsay were *not* representative of all of France: the order between Royal and Sarkozy was reversed, Bayrou did much better than nationally and Le Pen much worse (see table 6).

	N. Sarkozy	S. Royal	F. Bayrou	J.-M. Le Pen
National	31.18%	25.87%	18.57%	10.44%
Orsay precincts	28.98%	29.92%	25.51%	5.89%
	O. Besancenot	P. de Villiers	M.-G. Buffet	D. Voynet
National	4.08%	2.23%	1.93%	1.57%
Orsay precincts	2.54%	1.91%	1.40%	1.69%
	A. Laguiller	J. Bové	F. Nihous	G. Schivardi
National	1.33%	1.32%	1.15%	0.34%
Orsay precincts	0.76%	0.93%	0.30%	0.17%

Table 6. French presidential election, first round, April 22, 2007: national vote *vs.* vote in the three precincts of Orsay.

Ballot: Election of the President of France 2007

To be president of France,
having taken into account all considerations,
I judge, in conscience, that this candidate would be:[9]

	Excellent	*Very Good*	*Good*	*Acceptable*	*Poor*	*to Reject*
Olivier Besancenot						
Marie-George Buffet						
Gérard Schivardi						
François Bayrou						
José Bové						
Dominique Voynet						
Philippe de Villiers						
Ségolène Royal						
Frédéric Nihous						
Jean-Marie Le Pen						
Arlette Laguiller						
Nicolas Sarkozy						

Check one single grade in the line of each candidate.
No grade checked in the line of a candidate means to Reject the candidate.

Table 7. The majority judgement ballot (English translation).

[9]The question in French: "Pour présider la France, ayant pris tous les éléments en compte, je juge en conscience que ce candidat serait:" The grades in French: "Très bien, Bien, Assez bien, Passable, Insuffisant, à Rejeter." The names of the candidates are given in the official order, the result of a random draw.

On April 22, the day of the first round, after voting officially in these three precincts, voters were invited to participate in the experiment using the majority judgement. A team of three to four knowledgeable persons were in constant attendance to encourage participation and to answer questions. Voting *à la* majority judgement was carried out exactly as is usual in France: ballots were filled in the privacy of voting booths, inserted into envelopes, and then deposited in large transparent urns. A facsimile of the ballot (in translation) is given in table 7.

Several comments concerning the ballot are in order. First, the voter is confronted with a specific question which he or she is asked to answer. Second, the answers, or evaluations, are given in a language of grades that is *common* to all French citizens: with the exception of *to Reject*, they are the grades given to school children. These evaluations are *not* numbers: they are *not* abstract values or weights that a voter almost surely assumes will be added together to assign a total score to each candidate (and so may encourage him or her to exaggerate up or down), but mean the same thing (or close to the same thing) to everyone.

Contrary to the predictions of several elected officials and many Parisian "intellectuals," the voters had no problem in filling out the ballots. For the most part, one minute sufficed. The queues to vote by the majority judgement were no longer than those to vote officially (though of course the experimental vote did not require electors to sign registers or present their papers of identity). Moreover, 1,752 of the 2,360 who voted officially (or 74%) participated in the experiment: the waiting times could not have been long! In fact, the rate of participation was slightly higher because in France a voter can assign to another person a proxy to vote for him or her, and the experiment did not allow anyone to vote more than once. 19 of the 1,752 ballots were indecipherable or deliberately subverted, leaving a total of 1,733 valid ballots. Most of the voters who did not participate in the experiment said they could not take the time, some seemed to be hostile, several did not understand.

Each member of the team that conducted the experiment had the impression that the participants were very glad to have the means to express their opinions concerning *all* the candidates, and liked the idea that candidates would be assigned grades.[10] An effective argument to persuade reluctant voters to participate was that the majority judgement allows a much fuller expression of a voter's opinions. The actual system offered voters only 13 possible *messages*: to vote for one of the twelve candidates, or to vote for none. The majority judgement offered voters more than 2 billion possible messages.[11] Several participants actually stated that the experiment had induced them to vote for the first time: finally a method that permitted them to express themselves!

[10] A collection of television interviews of participants prepared by Raphaël Hitier, a journalist of *I-Télé*, attests to these facts.

[11] With twelve candidates and six grades, there are $6^{12} = 2,176,782,336$ possible messages.

The results

	Excllnt	*Very Gd*	*Good*	*Accptbl*	*Poor*	*to Rejct*	Sum
Avg./ballot	0.69	1.25	1.50	1.74	2.27	4.55	12

Table 8. Average number of grades per majority judgement ballot (of the 4.55 *to Reject*, 0.5 corresponded to no grade).

Voters were particularly happy with the grade *to Reject*, and used it the most: there was an average of 4.1 of *to Reject* per ballot and an average of 0.5 of no grade (which, in conformity with the stated rules, was counted as a *to Reject*). Voters were parsimonious with high grades and generous with low ones (see table 8). Only 52% of voters used a grade of *Excellent*; 37% used *Very Good* but no *Excellent*; 9% used *Good* but no *Excellent* and no *Very Good*; 2% gave none of the three highest grades.

Six possible grades assigned to twelve candidates implies that a voter was unable to express a preference between every pair of candidates. The number of different grades actually used by voters shows that in any case they did not wish to distinguish between every pair (see table 9) since only 14% used all six grades. This suggests that six grades was quite sufficient. A scant 3% of the voters used at most two grades, 13% at most three, suggesting that more than three grades is necessary.

1 grade	2 grades	3 grades	4 grades	5 grades	6 grades
1%	2%	10%	31%	42%	14%

Table 9. Percentages of voters using k grades ($k = 1, \ldots, 6$).

The highest grades were often multiple. 11% of the ballots had at least two grades of *Excellent*; 16% had at least two grades of *Very Good* and no grade of *Excellent*; almost 6% had at least two grades of *Good*, no *Excellent*, no *Very Good*. In all, more than 33% of the ballots gave the highest grade to at least two candidates. Thus one of every three voters did not designate a single "best" candidate! This seems to indicate that voters conscientiously answered the question that was posed. It also shows that many voters either saw nothing (or very little) to prefer among several candidates or, at the least, were very hesitant in making a choice among two, three or more candidates. Moreover, many voters did not distinguish between the leading candidates: 17.9% gave the same grade to Bayrou and Sarkozy (10.6% their highest grade to both), 23.3% the same grade to Bayrou and Royal (11.7% their highest grade to both), and 14.3% the same grade to Sarkozy and Royal (4.1% their highest to both). Indeed, 4.8% gave the same grade to all three (4.1% their highest to all three: all who gave their highest grade to Sarkozy and Royal also gave it to Bayrou!). These are significant percentages: many elections are decided by smaller margins.

This finding is reinforced by two facts observed elsewhere. First, a poll conducted on election day[12] asked at what moment voters had decided to vote for a particular candidate. Their hesitancy in making a choice is reflected in

[12] by TNS Sofres - Unilog Groupe Logica CMG, April 22, 2007, the same poll cited earlier.

the answers: 33% decided in the last week, a third of whom (11%) decided on election day itself. For Bayrou voters 43% decided in the last week and 12% on election day; for Sarkozy voters the numbers were 20% and 6%; for Royal voters, 28% and 9%; for Le Pen voters, 43% and 18%. But the "first-past-the-post" system *forced* them to make a choice (or to vote for no one)! Second, the Faches-Thumesnil experimenters [11] asked voters to rank-order all twelve candidates. They were testing "single-transferable-vote" mechanisms.[13] Rank-ordering fewer than twelve meant that those not ranked were all considered to be placed at the bottom of the list (so the mechanisms could not "transfer" votes to such candidates). 960 voters participated, only 60% of those who voted officially, and 67 ballots were invalid. Only 41% of the valid ballots actually rank-ordered all twelve candidates. 53% rank-ordered six or fewer candidates, 29% of them rank-ordered three or fewer. All of this bespeaks of a reluctance to rank-order many candidates: it *is* a difficult, time-consuming task.

The evidence conclusively demonstrates two facts. (1) The "first-past-the-post" system forces voting for one candidate when in fact many voters do not wish to make a single choice. (2) The age-old view of voting (and the basic assumption of the traditional model of social choice theory)—that voters have rank-orders of the candidates in their minds—is not a reasonable model of reality.

Of the 1,733 valid majority judgement ballots,[14] 1,705 were different. It is surprising they were not all different! Had all those who voted in France in 2007 (some 36 million) cast different majority judgement ballots, less than 1.7% of the possible messages would have been used. Those that were the same among the 1,733 valid ballots (or messages) of the experiment contained only *to Reject*'s or were of the type an *Excellent* for Sarkozy and *to Reject* for all the other candidates. The opinions of voters are richer, more varied and complex by many orders of magnitude than those they are allowed to express by all current systems!

The outcome of voting by majority judgement in the three precincts is given in table 10. Since every candidate was necessarily assigned a grade—assigning no grade meant assigning a *to Reject*—each candidate had exactly the same number of grades. Accordingly, the results may be given as percentages of the grades received by each candidate. In fact, there were relatively few ballots that assigned no grade to a candidate[15] (even though no grade explicitly meant *to Reject*). Most close observers of French politics who were shown the results with the names of the candidates hidden were able to correctly identify Sarkozy, Royal, Bayrou and Le Pen.

[13] These elect the candidate who is ranked first by a majority. If there is no such candidate, then candidates are eliminated, one by one, their votes "transferred" to the next on the lists, until a candidate *is* ranked first by a majority. The choice of who to eliminate may differ. One mechanism eliminates the candidate ranked first least often; another eliminates the candidate ranked last most often. In the experiment the first elected Sarkozy, the second elected Bayrou.

[14] 559 in the 1^{st} precinct, 601 in the 2^{nd}, 573 in the 3^{rd}.

[15] No grade was assigned to each of the candidates in the following percentages: Nihous 7.2%, Schrivardi 5.8%, Laguiller 5.3%, Villiers 4.3%, Buffet 4.3%, Voynet 4.3%, Bové 4.2% Besancenot 3.2%, Bayrou 2.9%, Le Pen 2.7%, Royal 1.8%, Sarkozy 1.7%.

	Excellent	*Very Good*	*Good*	*Acceptable*	*Poor*	*to Reject*
Besancenot	4.1%	9.9%	16.3%	16.0%	**22.6%**	31.1%
Buffet	2.5%	7.6%	12.5%	20.6%	**26.4%**	30.4%
Schivardi	0.5%	1.0%	3.9%	9.5%	24.9%	**60.4%**
Bayrou	13.6%	30.7%	**25.1%**	14.8%	8.4%	7.4%
Bové	1.5%	6.0%	11.4%	16.0%	**25.7%**	39.5%
Voynet	2.9%	9.3%	17.5%	**23.7%**	26.1%	20.5%
Villiers	2.4%	6.4%	8.7%	11.3%	15.8%	**55.5%**
Royal	16.7%	22.7%	**19.1%**	16.8%	12.2%	12.6%
Nihous	0.3%	1.8%	5.3%	11.0%	26.7%	**55.0%**
Le Pen	3.0%	4.6%	6.2%	6.5%	5.4%	**74.4%**
Laguiller	2.1%	5.3%	10.2%	16.6%	**25.9%**	40.1%
Sarkozy	19.1%	19.8%	**14.3%**	11.5%	7.1%	28.2%

Table 10. Majority judgement results, three precincts of Orsay, April 22, 2007. The percentage in bold in the row of a candidate indicates the column of her/his majority-grade.

The *majority-grade* of a candidate is his or her median grade. It is simultaneously the highest grade approved by a majority and the lowest grade approved by a majority. For example, Dominique Voynet's majority-grade (see table 10) is *Acceptable* because a majority of $2.9\%+9.3\%+17.5\%+23.7\% = 53.4\%$ believe she merits at least that grade and a majority of $23.7\%+26.1\%+20.5\% = 70.3\%$ believe she merits at most that grade.

The *majority-ranking* orders the candidates according to their majority-grades. However, with twelve candidates and six grades some candidates will necessarily have the same majority-grade. The general theory [2, 4] shows that two candidates are never tied for a place in the majority-ranking unless the two have *precisely* the same set of grades. But when there are many voters, as is typical in most elections, the general rule for determining the majority-ranking may be simplified. Three values attached to a candidate—called the candidate's *majority-value*—are sufficient to determine the candidate's place in the majority-ranking:

$$(p, \alpha, q) \text{ where } \begin{cases} p = & \% \text{ of grades above majority-grade,} \\ \alpha = & \text{majority-grade, and} \\ q = & \% \text{ of grades below majority-grade.} \end{cases}$$

A mnemonic helps to make the definition of this order clear: supplement a majority-grade (other than *Excellent* or *to Reject*) by a "mention" of $\pm$ or 0 that depends on the relative sizes of p and q and call it the *majority-grade**:

$$\alpha^* = \begin{cases} \alpha^+ & \text{if } p > q, \\ \alpha^0 & \text{if } p = q \\ \alpha^- & \text{if } p < q, \end{cases}$$

(the possibility that $p = q$ is slim). Thus, for example, Sarkozy's majority-value

is $(38.9\%, \mathit{Good}, 46.9\%)$ and his majority-grade* is Good^-. Naturally, α^+ is better than α^0, and both are better than α^-.

Consider two candidates A and B with majority-values (p_A, α_A, q_A) and (p_B, α_B, q_B). A ranks ahead of B, and (p_A, α_A, q_A) ahead of (p_B, α_B, q_B), when

- A's majority-grade* is better than B's (or $\alpha_A^* \succ \alpha_B^*$), or
- their majority-grade*'s are both α^+ and $p_A > p_B$, or
- their majority-grade*'s are both α^- and $q_A < q_B$.

To illustrate,

- Bayrou with (44.3%,Good^+,30.6%) ranks ahead of Royal with (39.4%,Good^-,41.5%) because Good^+ is better than Good^-,
- Besancenot with (46.3%,Poor^+,31.2%) ranks ahead of Buffet with (43.2%,Poor^+,30.5%) because 46.3% > 43.2%, and
- Royal with (39.4%,Good^-,41.5%) ranks ahead of Sarkozy with (38.9%,Good^-,46.9%) because 41.5% < 46.9%.

It is practically certain that this rule for deciding the order among the majority-values suffices to give an unambiguous order of finish in any election with many voters. A more detailed discussion of tie-breaking rules is given in the appendix.

	Majority-ranking	$p =$ Above maj.-grade	$\alpha^* =$ The majority-grade*	$q =$ Below maj.-grade	Natl. rank.	Orsay rank.
1^{st}	Bayrou	**44.3%**	Good^+	30.6%	3^{rd}	3^{rd}
2^{nd}	Royal	39.4%	Good^-	**41.5%**	2^{nd}	1^{st}
3^{rd}	Sarkozy	38.9%	Good^-	**46.9%**	1^{st}	2^{nd}
4^{th}	Voynet	29.8%	$\mathit{Acceptable}^-$	**46.6%**	8^{th}	7^{th}
5^{th}	Besancenot	**46.3%**	Poor^+	31.2%	5^{th}	5^{th}
6^{th}	Buffet	**43.2%**	Poor^+	30.5%	7^{th}	8^{th}
7^{th}	Bové	34.9%	Poor^-	**39.4%**	10^{th}	9^{th}
8^{th}	Laguiller	34.2%	Poor^-	**40.0%**	9^{th}	10^{th}
9^{th}	Nihous	45.0%	*to Reject*	–	11^{th}	11^{th}
10^{th}	Villiers	44.5%	*to Reject*	–	6^{th}	6^{th}
11^{th}	Schivardi	39.7%	*to Reject*	–	12^{th}	12^{th}
12^{th}	Le Pen	25.7%	*to Reject*	–	4^{th}	4^{th}

Table 11. The majority-values (p, α, q) and the majority-ranking, three precincts of Orsay, April 22, 2007.
(The column headed "Natl. rank." is the national rank-order according to the current system. The column headed "Orsay rank." is the rank-order in the three precincts of Orsay according to the current system.)

The majority-grades and the majority-values for the experiment are given in the order of the majority-ranking in table 11. The majority-ranking is very

different from the rank-ordering obtained in the three precincts of Orsay with the current system. Sarkozy had the highest number of *Excellents*, but also the highest number of *to Rejects* among the three serious candidates. *Every* grade of the candidates counts in determining their majority-grades and the majority-ranking. This explains why Bayrou—third according to the official vote in Orsay's three precincts—is first according to the majority judgement. Le Pen—fourth according to the official vote—is last according to the majority judgement because 74.4% of the voters graded him *to Reject*. Whenever a majority of the voters assigns a same grade to a candidate, that is necessarily the candidate's majority-grade. Another marked difference with the current system is the green candidate Voynet's fourth-placed finish (instead of seventh-placed): the electorate was able to express the importance it attaches to problems of the environment while giving higher grades to candidates it judged better able to preside the nation. Once elected, Sarkozy recognized this importance: his new government has one "super-ministry," the Ministry of Ecology and Sustainable Development.

Notice that the "raw" majority judgement results make a very strong case for ranking Bayrou first, Royal second and Sarkozy third for the following reason. Except for the *Excellents*, whose percentages taken alone give the opposite rank-ordering, the percentages of at least *Very Good*, at least *Good*, ..., at least *Poor*, **all** agree with that order (see table 12). Practically any reasonable election mechanism will agree with this ranking of the three important candidates.

	At least					
	Excellent	*Very Good*	*Good*	*Acceptable*	*Poor*	*to Reject*
Bayrou	13.6%	43.3%	69.4%	84.2%	92.6%	100%
Royal	16.7%	39.4%	58.5%	75.3%	87.5%	100%
Sarkozy	19.1%	38.9%	53.2%	64.7%	71.8%	100%

Table 12. Cumulative majority judgement grades, three precincts of Orsay, April 22, 2007.

Validation

The result of the second round on May 6, 2007, in the three voting precincts of Orsay was

Ségolène Royal: 51.3% Nicolas Sarkozy: 48.7%

The results of the face-to-face confrontations between every pair of candidates may be estimated from the majority judgement ballots[16] by comparing their respective grades (see table 13). In particular, Royal defeats Sarkozy with 52.3% of the vote, a "prediction" of the outcome of the second round within 1%. The participants seem to have expressed themselves in the majority judgement ballots in conformity with the manner in which they actually voted. The 1% difference is easily explained. 26% of the voters did not participate in the experiment; and the last two weeks of the campaign may have changed perceptions.

[16] The information in table 10 does not suffice.

The closeness of the estimate to the outcome shows the majority judgement ballots are consistent with the observed facts.

The estimates of table 13 show Bayrou to be the Condorcet-winner, which is consistent with all polls. Moreover, the estimates of the face-to-face races determine an unambiguous order of finish—it is the order given in the table—so there is no Condorcet-cycle. This order is almost the majority-ranking: only Nihous and Villiers are permuted (their respective majority-values are (45.0%,*to Reject*,–) and (44.5%,*to Reject*,–), very close). The information in the table suffices to compute the Borda-scores of each candidate (the sum of the percentages in each of their rows): Bayrou 867, Royal 812, Sarkozy 711, Voynet 634, Besancenot 581, Buffet 552, Bové 486, Laguiller 481, Villiers 430, Nihous 369, Schivardi 335, Le Pen 342. The Condorcet and Borda orders of finish are the same except for the last two candidates.

	Bay	Roy	Sar	Voy	Bes	Buf	Bov	Lag	Vil	Nih	Sch	LP
Bayrou	–	56	60	77	77	81	83	83	84	90	90	86
Royal	44	–	52	73	74	78	81	80	77	85	87	81
Sarkozy	40	48	–	59	61	64	66	66	77	75	75	80
Voynet	23	27	41	–	56	59	67	67	66	75	79	74
Besancenot	23	26	39	44	–	53	60	61	62	69	74	70
Buffet	19	22	36	41	47	–	57	59	61	68	73	69
Bové	17	19	34	33	40	43	–	51	56	62	66	65
Laguiller	17	20	34	33	39	41	49	–	56	62	66	64
Villiers	16	23	23	34	38	39	44	44	–	54	56	59
Nihous	10	15	25	25	31	32	38	38	46	–	53	56
Schivardi	10	13	25	21	26	27	34	34	44	47	–	54
Le Pen	14	19	20	26	30	31	35	36	41	44	46	–

Table 13. Face-to-face elections, percentages of votes estimated from majority judgement ballots, three precincts of Orsay, April 22, 2007.
It shows, for example, Royal winning 52% of the vote against Sarkozy and, symmetrically, Sarkozy winning 48% of the vote against Royal. The percentage of ballots that give to both candidates of a pair the same grade is split evenly between them.

	Major			Leftist						Rightist		
	Bay	Roy	Sar	Voy	Bes	Buf	Bov	Lag	Sch	Vil	Nih	LP
Estimate 1	25.6	25.6	28.4	3.5	4.9	2.6	1.6	1.6	0.4	2.3	0.5	2.9
Actual	25.5	29.9	29.0	1.7	2.5	1.4	0.9	0.8	0.2	1.9	0.3	5.9
Estimate 2	25.3	25.4	27.4	3.4	4.6	2.5	1.5	1.5	0.3	1.9	0.4	5.8

Table 14. First round vote, percentages of votes estimated from majority judgement ballots, three precincts of Orsay, April 22, 2007.
(In estimate 1, the percentage of ballots that give to several candidates the same highest grade is split evenly among them. In estimate 2, the same assumption is made except when Le Pen is one of the several candidates, in which case he is accorded the entire percentage.)

The majority judgement ballots may also be used to estimate the extent of deliberate strategic voting (not in accord with voters' convictions) in the first round under the current system (see table 14). It is naturally assumed that a candidate receiving the highest grade accorded by a voter would receive his or her one vote. But since a third of the voters gave their highest grade to more

than one candidate, an assumption must be made concerning their behavior. Estimate 1 naively assumes such votes are split evenly among the candidates receiving the highest grade. Estimate 2 takes into account Le Pen's very peculiar niche in the far right of the French political spectrum: it assumes that when a voter's highest grade goes to Le Pen and others, then her or his vote goes to Le Pen only (if you vote far right it is more strategic to vote for Le Pen, but why not add the others if you can!). This second assumption explains almost perfectly what happened to the far right, and seems to be the better model. Comparing estimate 2 with the actual vote suggests that 6.3% of the 13.8% for the six candidates of the left and greens (so a little less than half of their votes according to estimate 2) went to Royal and Sarkozy, three-quarters of them for Royal, one-quarter for Sarkozy. Contrary to the stated opinions of most political observers, it seems that Bayrou voters backed him by conviction not strategy.

It happens that the majority judgement winner coincides with the Cordorcet-winner and the Borda-winner. It also happens that the majority-ranking almost coincides with the unambiguous order of the face-to-face winners: only the 9^{th} and 10^{th} placed candidates (Nihous and de Villiers) are interchanged. And the majority judgement coincides with the Borda-ranking as well except for one more interchange in the last two places (Le Pen and Schivardi). When elections are really clear cut affairs, most reasonable mechanisms may be expected to give close to the same results. Table 12 explains why the first three finish in that order according to the majority judgement, as well as by Condorcet's and Borda's methods.

	Roy	Sar	Bay	LP	Bes	Vil	Voy	Bov	Buf	Lag	Nih	Sch
12^{th} prct.	32.0	26.6	20.2	10.0	2.7	2.5	2.3	1.3	1.2	0.8	0.2	0.0
National	31.2	25.9	18.6	10.4	4.1	2.2	1.9	1.6	1.3	1.3	1.2	0.3
	Sar	Roy	Bay	LP	Bes	Vil	Buf	Voy	Bov	Lag	Nih	Sch

Table 15. Actual percentages, first round, April 22, 2007, in Orsay's 12^{th} precinct (top row of percentages with names of candidates above) and all of France (bottom row of percentages with names of candidates below).

Some persons have averred that the majority judgement necessarily favors centrist candidates. This is neither true in theory nor in practice, despite the fact that Bayrou was a centrist candidate. First, observe that Bayrou's share of the vote was considerably higher in the three precincts of Orsay than in the entire nation: winning in Orsay's three precincts implies little about what might have happened nationally. Second, consider the actual first round percentage results in the 12^{th} precinct. They were close to the result in all of France (when the percentages of Royal and Sarkozy are permuted and Buffet is moved up two slots in the order of finish, see table 15).

Bayrou was as much a centrist candidate in the 12^{th} precinct as he was in the three precincts. Yet, in the 12^{th} precinct Bayrou was *not* the majority judgement winner (see table 16 for the results of the four major candidates): Royal is first, Bayrou second, Sarkozy third, and Le Pen last. In practice and in theory a candidate receives a high majority-grade when he or she is assigned

many high grades and few low grades: this is no more reserved to a centrist candidate than to any other candidate. 42.4% of Royal's grades were above *Good*, only 40.8% of Bayrou's were above *Good*.

	Majority-ranking	$p =$ Above maj.-grade	$\alpha^* =$ The majority-grade*	$q =$ Below maj.-grade
1^{st}	Royal	**42.4%**	*Good+*	40.1%
2^{nd}	Bayrou	**40.8**%	*Good+*	31.4%
3^{rd}	Sarkozy	38.0%	*Good−*	**48.7%**
12^{th}	Le Pen	30.9%	*to Reject*	–

Table 16. The majority-values (p, α, q) and the majority-ranking, Orsay's 12^{th} precinct, April 22, 2007.[17]

	Bayrou	Royal	Sarkozy	Le Pen
Bayrou	–	53.5%	59.0%	82.8%
Royal	46.5%	–	54.3%	77.9%
Sarkozy	41.0%	45.7%	–	77.7%
Le Pen	17.2%	22.1%	22.3%	–

Table 17. Projected second round results, Orsay's 12^{th} precinct. (*E.g.*, Sarkozy has 41% of the votes against Bayrou.)

The results of the face-to-face confrontations between the pairs of major candidates deduced from the majority judgement ballots in the 12^{th} precinct are given for the four major candidates in table 17. Bayrou is again the Condorcet-winner despite Royal's majority judgement victory: Why?

		Excllnt	*Very Gd*	*Good*	*Accptbl*	*Poor*	*to Rejct*
Bayrou's grades	by Royal	7%	33%	29%	16%	9%	6%
	by Sarkozy	6%	28%	30%	19%	9%	8%
Sarkozy's grades	by Royal	3%	10%	16%	15%	11%	45%
	by Bayrou	6%	22%	24%	17%	6%	25%
Royal's grades	by Bayrou	7%	26%	26%	20%	13%	9%
	by Sarkozy	3%	13%	22%	24%	18%	21%

Table 18. Grades given to three major candidates by voters who gave their highest grade to one of the others, from majority judgement ballots, three precincts of Orsay, April 22, 2007[18] (*e.g.*, "by Royal" means by those voters who gave their highest grade to Royal).

[17] The majority-grades and the majority-ranking of the candidates after Sarkozy is the same as for the three precincts except that Besancenot obtains a *Poor−*, and de Villiers is placed 9^{th} and Nihous 10^{th}. Some may dispute ranking Royal above Bayrou. For a discussion see the appendix where several tie-breaking rules are discussed.

[18] A Tnes-Sofres poll of March 14-15, 2007 showed 72% of Royal voters (respectively, 75% of Sarkozy voters) giving their votes to Bayrou in a second round against Sarkozy (respectively, against Royal).

The reason is clear enough. Bayrou was the *second* choice of a very large number of voters, so against Royal alone in the current system he would naturally take a large number of Sarkozy's votes and against Sarkozy alone he would naturally take a large number of Royal's votes. The majority judgement ballots show that the voters who gave Sarkozy their highest grade strongly preferred Bayrou to Royal, those who gave Royal their highest grade strongly preferred Bayrou to Sarkozy, whereas those who gave their highest grade to Bayrou evaluated Royal and Sarkozy about equally (see table 18).

Face-to-face confrontations ignore how the electorate *evaluates* the respective candidates (just as the 2002 run-off ignored the respective evaluations of Chirac and Le Pen) except, of course, that one is evaluated higher than the other. Two thirds of the second highest grades are merely *Good* or worse (see table 19). This is why being second in the rankings of voters has very different meanings.

Grades:	*Excellent*	*Very Good*	*Good*	*Acceptable*	*Poor*	*to Reject*
Highest	52%	37%	9%	2%	0%	1%
Second highest	–	35%	41%	16%	5%	3%
Third highest	–	–	26%	40%	22%	13%

Table 19. Distributions highest grade, second highest grade, third highest grade, three precincts of Orsay, April 22, 2007.

First ranked candidates (such as Sarkozy and Royal) often elicit strong support and strong opposition. Second ranked candidates are often centrists (such as Bayrou). In consequence, a second ranked candidate is often favored in face-to-face confrontations, so is favored by Condorcet's method. Such centrist candidates are even more favored by Borda's method: when there are many marginal candidates of the right and the left, the second ranked candidates garner many Borda points because they are ahead of most of them. But this is not true with the majority judgement: the evaluations—the *grades* of the second ranked candidates—decide, not the place in the ranking. The Condorcet-winner—never used as a mechanism of election—and Borda's method—almost never used—are highly suspect for this among other reasons.

The closeness of the actual results in Orsay's 12^{th} precinct to the national results (when Sarkozy takes the place of Royal) suggests that Sarkozy could have been first in the majority-ranking at the national level. The voting in the 1^{st} precinct, where Bayrou did exceptionally well, reinforces this impression. The actual first round percentages were Sarkozy 32.6%, Bayrou 29.4% and Royal 25.2%, yet Sarkozy placed a very close second in the majority-ranking to Bayrou's first. The projected second round winners show Bayrou to be the Condorcet-winner in this precinct as well.

Common language

The theoretical underpinnings of the majority judgement require that voters (or judges, when the problem is to rank competitors or alternatives) evaluate the candidates in a language of grades that is common to them all. Evaluations

should be absolute, not relative. Therefore the question to be confronted by a voter must not suggest "how do you compare the candidates," but instead address "how do you evaluate each candidate." The question posed and the language of grades offered in the ballot must make this distinction clear. Polls in the 2007 French presidential elections illustrate the point (see table 20). The question on the left suggests an evaluation, the question on the right a comparison. The results show the well known fact that "yes" or "no" answers can yield strikingly varying results as a function of the question posed.

	Question: Would each of the following candidates be a good President of France?		Question: Do you personally wish each of the following candidates to win the presidential election?	
	Yes	No	Yes	No
Bayrou	60%	36%	33%	48%
Sarkozy	59%	38%	29%	56%
Royal	49%	48%	36%	49%
Le Pen	12%	84%		

Table 20. Polling results, March 20 and 22, 2007 (Bva).

What constitutes a "good" common language, how is one to test whether a language of grades or of measurement is "good," and, indeed, why can one assume that a common language exists at all?

Common languages assuredly do exist because they have been routinely invented, learned through use, and commonly understood in a host of applications, including ranking figure skaters, gymnasts, divers, pianists, wines and students (these and other practical uses of common languages of measurement are investigated in [4]). In particular, the Chopin International Piano Competition has used a number scale since its establishment in 1927 (though the range of the numbers has changed over time). Schools and universities either give number grades or letter grades together with their numerical "equivalents."

The numbers, of course, are abstract and mean nothing until they are defined. The "natural" language of words are their definitions. Using numbers suggests that the mechanism for amalgamating the grades of many judges will be to take their sum or average (as has the Chopin competition since 1927), and may well induce judges or voters (or teachers and professors) to assign the grades strategically in view of their ultimate use. For this reason it is better to choose a "natural" language, although repeated use eventually converts numbers into words that have well-defined meanings (*e.g.*, when a professional judge says a dive in an international competition is an "8.5," all of his or her peers will know exactly what that means, whether they agree or not).

Finding a language of grades that is common to all the *voters* in a society is less easy since it must be understood the first time it is used. France mainly uses a 0 to 20 grading systems in its schools and universities, but it also uses the six descriptive words of the majority judgement ballots (with the exception of *to Reject*), words familiar to all French school children. A "good" language should

contain a sufficient number of grades to enable voters to express themselves as fully as they wish, which argues in favor of a language with many grades. It should also be common to all voters—that is, be used and understood "in the same way" by all voters—which argues for a language with few grades. The choice that was made in this experiment appears to have been judicious for several reasons.

First, all of the grades were used a significant number of times (see table 8).

Second, six grades were sufficient, for only 14% of all the voters used all six grades, suggesting that more grades would have been used by very few. 73% used four or five grades, and the average was 4.5 grades per ballot (see table 9).

	Three prcts.	1^{st} prct.	6^{th} prct.	12^{th} prct.	Samples of 100 Avg. (σ)	Range	Disjoint samples of 50 Avg. (σ)	Range
Excellent	0.7	0.7	0.7	0.7	0.7 (.07)	0.6/0.8	0.7 (.12)	0.5/0.9
Very Good	1.3	1.2	1.2	1.4	1.2 (.13)	1.1/1.5	1.3 (.16)	1.1/1.5
Good	1.5	1.5	1.4	1.6	1.5 (.13)	1.4/1.7	1.5 (.27)	0.9/1.8
Acceptable	1.7	1.7	1.7	1.8	1.8 (.15)	1.7/2.1	1.7 (.27)	2.1/2.6
Poor	2.3	2.3	2.3	2.2	2.3 (.19)	2.1/2.7	2.3 (.19)	2.1/2.6
to Reject	4.6	4.8	4.6	4.3	4.5 (.29)	4.1/4.8	4.5 (.41)	4.1/5.3

Table 21. Average number of words per majority judgement ballot, 2007 Orsay experiment. (σ is the standard deviation. 10 random samples of 100 and 10 disjoint random samples of 50 were taken.)

Third, it is possible to test whether the six "words" used in this experiment constituted a "common" language or did not. The idea is to ask whether the voters used the language in the same way: Did subsets of the voters use each of the words on average about the same number of times? One approach to answering this question is to compare the use of the words in the ballots coming from the naturally defined subsets that are the voting precincts. Another is to take random samples—or random disjoint samples—from among the 1,733 ballots. Table 21 shows that each of the three voting precincts—the 1^{st} with 559 voters, the 6^{th} with 601 voters, and the 12^{th} with 573 voters—used the language in almost exactly the same way, which of course agreed with the use of the language by the entire population. It also suggests that similar results obtain when random subsets of 100 and when random disjoint subsets of 50 are chosen from the 1,733 ballots. A more detailed analysis takes many random samples of k voters and studies the distribution of the average number of words per ballot for different values of k [4]. This is quite remarkable: the outcomes in the different precincts are different—and the outcomes on different samples are different—but the use of the language is the same.

Fourth, the estimates of the second round results based on the majority judgement ballots in the three precincts together and in each of them singly were close to the observed outcomes as well, as shown in table 22. They assumed: (1) when a voter gave a higher grade to one candidate over the other he or she would vote for that candidate in the second round; and (2) when voters gave the same grades to several candidates their votes were split equally among those candidates. The closeness of the estimates to the observed outcomes

suggests these assumptions were well founded, implying the language permitted the voters to correctly express their preferences and their indifferences.

	Three precincts		1^{st} precinct		6^{th} precinct		12^{th} precinct	
	Estm.	Outcm.	Estm.	Outcm.	Estm.	Outcm.	Estm.	Outcm.
Royal	52.3%	51.3%	48.2%	47.2%	54.4%	53.7%	54.3%	52.6%
Sarkozy	47.3%	48.7%	51.8%	52.8%	45.6%	46.3%	45.7%	47.4%

Table 22. Second round results, percentages of votes estimated from first round majority judgement ballots *vs.* actual outcomes, Orsay, April 22, 2007.[19]

Properties of the majority judgement

Given a common language, the majority judgement—the *majority-grade* and the *majority-ranking*—has been proven to be the *only* mechanism that is acceptable according to several different criteria (see [2, 4] for precise definitions and results). Here we only describe and illustrate the salient properties that are enjoyed by the majority judgement in the context of the experiment. *All* of the other mechanisms mentioned in this article violate several of these properties.

Ordinal. The common language is ordinal, so the mechanism used must be ordinal as well. The majority judgement is ordinal. Mechanisms based on sums or averages of points are not ordinal.

Respects majority judgement. The majority-grade (or median) is the unique mechanism which guarantees that when a majority of the electorate gives a grade g to a candidate, that candidate's majority-grade is g. Everyone of a majority can give a point score of p to a candidate, but that candidate's average will certainly not (in general) be p.

Transitive. The majority-ranking is transitive; moreover, one candidate is necessarily ranked ahead or behind another, unless both have identical sets of grades. The Condorcet-paradox shows that the Condorcet criterion is not transitive. Identifying instances where it has occurred in practice is rare because of lack of information, but it has been observed [17].

Satisfies IIA. The majority judgement satisfies independence of irrelevant alternatives. The grades are absolute not relative, so if some candidate drops out, the remaining candidates' grades remain the same. None of the mechanisms whose inputs are rank-orders satisfy IIA (including first-past-the-post, Borda's and its generalizations to scoring systems, and the single transferable vote).

Monotone. If every grade of a candidate is replaced by the same or a better grade, the candidate's place in the majority-ranking cannot be lower. If every grade of a candidate is replaced by a strictly better grade, the candidate's majority-grade must be raised. Monotonicity is not satisfied by the single transferable vote: if a winning candidate C is raised in the lists of some voters but otherwise the lists remain the same, C may no longer be the winner. Nor

[19] Royal's scores are consistently though slightly over estimated. This probably reflects changes in opinions in the two weeks that separated the two rounds of voting (due, in particular, to the televised debate between the two candidates).

is it satisfied by the French first-past-the-post with run-off system: if in 2007 Sarkozy's first round vote had increased at the expense of Royal, Bayrou could have finished second, the run-off would have been between Sarkozy and Bayrou, and Bayrou would (might) have won.

Resists strategic manipulation. Take a candidate, say Ségolène Royal, whose majority-value is

$$(39.4\%, \textit{Good}, 41.5\%).$$

Only a voter who can change Royal's majority-grade or majority-value by changing the grades they give her can have any strategic impact. Who are those voters and what are their motivations to change?

Suppose a voter believes a candidate merits a grade of g and the further the majority-grade is from g the less she or he likes it (a reasonable motivation[20]). Then the voter's optimal voting strategy is to give the candidate the grade g: the majority judgement is *strategy-proof-in-grading*[21]. Why is easy to see. 39.4% of the voters believe Royal merits above *Good*, 19.1% that she merits *Good*, 41.5% that she merits below *Good*. If a voter believes she merits better than *Good*, suppose it is a *Very Good*: giving a still higher grade can change nothing in the majority-value; giving a lower grade can only lower (a tiny bit) the majority-value, which is not the voter's intent. If a voter believes she merits worse than *Good*, suppose it is an *Acceptable*: giving a still lower grade can again change nothing; giving a higher grade can only raise (a tiny bit) the majority-value, which is not the voter's intent. If a voter believes Royal merits a *Good*, giving a higher or lower grade can only raise or lower (a tiny bit) her majority-value, which is not the voter's intent. *Q.E.D.*

Similar reasoning shows that the majority-grade mechanism is *group strategy-proof-in-grading*: a group of voters that share the same beliefs (*e.g.*, they belong to a same political party) have the same optimal strategy, namely, to give to the candidates the grades they believe they merit. For if they believed Royal merited better than *Good* and all raised the grade they gave her, her majority-value would remain the same; if all lowered the grades they gave her, her majority-value would decrease and perhaps her majority-grade as well (not their intent). If they believed Royal merited worse than *Good* and all lowered the grades they gave her, her majority-value would remain the same; if all raised the grades they gave her, her majority-value would increase and perhaps her majority-grade as well (not their intent). If, finally, they believed she merited a *Good*, and all either raised or lowered the grades they gave her, her majority-value and perhaps her majority-grade as well would either increase or decrease (not their intent).

These "strategy-proof-in-grading" properties are certainly not true of any mechanism based on sums or averages of points, nor of Borda's and its derivatives. If any voter either raises or lowers the points given a candidate—or raises or lowers a candidate's place in the voter's list—, that candidate's sum or average increases or decreases (a tiny bit)—and the candidate may be raised or lowered in the final ranking. And if many voters either raise or lower the points

[20] The voter's preferences in grading are said to be "single-peaked."

[21] In an entirely different context a related technical result is proved in [21].

given a candidate—or raise or lower a candidate's place in their lists—that candidate's sum or average increases or decreases a lot—and the candidate is very likely to be raised or lowered in the final ranking.

The strategy of a voter may, however, focus on the final ranking of the candidates rather than on the their final grades. It is impossible to completely eliminate the possibility of strategic manipulation if a voter is prepared for a candidate's final grade to be either above or below what she or he thinks the candidate merits: there is no mechanism that is "strategy-proof-in-ranking."[22] But the majority judgement best resists such manipulation. Take the example of Bayrou with a $Good^+$ and Royal with a $Good^-$, their respective majority-values being,

$$\text{Bayrou: } (44.3\%, \textit{Good}, 30.6\%) \qquad \text{Royal: } (39.4\%, \textit{Good}, 41.5\%).$$

How could a voter who wished Royal to be ranked higher than Bayrou manipulate? By changing the grades assigned to try to lower Bayrou's majority-value and to raise Royal's majority-value. But the majority judgement is *partially strategy-proof-in-ranking*: those voters who can lower Bayrou's majority-value cannot raise Royal's, those who can raise Royal's majority-value cannot lower Bayrou's. For suppose such a voter can lower Bayrou's. Then he or she must have given Bayrou a *Good* or better: but having preferred Royal to Bayrou the voter gave a grade of better than *Good* to Royal, so he or she cannot raise Royal's majority-value. Symmetrically, a voter who can raise Royal's majority-value must have given to her a *Good* or worse, so to Bayrou a worse than *Good*, so the voter cannot lower Bayrou's majority-value. Compared with mechanisms that sum or average, the majority judgement cuts in half the possibility of manipulation, however bizarre a voter's motivations (or whatever may be a voter's utility function).

As a matter of fact, 32.9% of the voters gave a higher grade to Royal than to Bayrou. Their types are summarized in table 23. The 9.2% of voters of type A—who gave an *Excellent* or *Very Good* to Royal and an *Acceptable* or worse to Bayrou—can do nothing to raise Royal's majority-value or to lower Bayrou's. On the other hand, if all of the types C, D and F lowered Bayrou's grade to *Acceptable* (it serves no purpose to lower them further) then his majority-value would go below Royal's. But that is unlikely, because most voters prefer voting in accord with their convictions (especially when they are asked to give absolute evaluations of candidates rather than relative comparisons).

A more reasonable scenario would be: one-quarter of the type B voters, who gave a mere *Acceptable* to Royal, raise her grade up to *Very Good* (more is of no use); one-third of the types C, D and E, who see only a slight difference between Royal and Bayrou, change (but more than indicated in table 23 is of no use); and one-half of the types F and G, who see a more substantial difference between the two candidates, change (again, more than indicated in table 23 is of no use). This scenario implies that 38% of the Royal voters who are able to have an impact by giving grades strategically do so (by way of comparison, a

[22] In the context of the traditional model, this is the Gibbard-Satterthwaite theorem.

poll on election day showed 31% of Royal supporters voted strategically). The result is to change the candidates' majority-values to

Bayrou: (42.2%,*Good*,36.6%) Royal: (42.0%,*Good*,40.8%),

so both have the majority-grade* $Good^+$, but Bayrou remains ahead in the majority-ranking. This shows how the majority judgement resists manipulation; it also shows that the amount of useful exaggeration is in any case limited. In contrast, mechanisms based on summing (including Borda's) or averaging points share *none* of the safeguards against manipulation discussed above.

Type	% ballots	*Excell-ent*	*Very Good*	*Good*	*Accep-table*	*Poor*	*To Reject*	Strategic change
A	9.2%	R	R		B	B	B	cannot
B	2.8%		| ←	←	←R	B	B	1/4
C	6.3%	R	B→	→	→ |			1/3
D	6.9%		R	B→	→ |			1/3
E	2.4%		| ←	←R	B			1/3
F	3.2%	R		B→	→ |			1/2
G	2.1%		| ←	←R		B	B	1/2

Table 23. Strategic voting: could Royal have won in Orsay's three precincts? (Type A voters, for example, gave an *Excellent* or *Very Good* to Royal, an *Acceptable* or worse to Bayrou. The arrows indicate increases and decreases in grades; the bar | that no purpose is served by going further.)

Voters' utilities. In theory the motivations of voters and their satisfaction are modelled by their "utilities." The utility function of a voter is at once complex and completely unknown. It is plausible to imagine that a voter would like a candidate's final grade to be as close as possible to the grade he or she believes the candidate merits, ... but it ain't necessarily so! In the "plausible" case, the candidate's utility function is absolute, otherwise it becomes relative, *i.e.*, what counts are the candidates' final rankings not their final grades. But in any case, the majority judgement mechanism makes no assumptions whatever about the voters' utilities. It depends only on what can be known in practice. It is "strategy-proof" for large classes of reasonable utility functions, and, when nothing is known about them, it best combats strategic manipulation, as the above example shows. The fact that voters share a common language of grades makes no assumptions about the voters' utilities: utilities measure the satisfactions of voters, grades measure the merits of candidates.

Grades for candidates. Voters who participated in the experiment were delighted with the idea that the majority judgement assigns grades to candidates. The majority-grade is a signal that expresses the electorate's appreciation of a candidate. Chirac's "triumph" with over 82% of the vote in 2002 would have been very different with the majority judgement. Chirac would have won, but his grade would have been modest, Le Pen's a *to Reject*. Voynet's grade in the 2007 experiment clearly expresses the electorate's concern with environmental problems, whereas the official vote completely failed to do so. Le Pen's grade in

the 2007 experiment shows the electorate's strong refusal of his ideas, whereas according to the official vote he was one of the major candidates. Even when there is exactly one candidate—which often occurs—the majority judgement may be used to disclose the electorate's evaluation of that candidate.

The majority judgement is *grade-consistent* in the following sense: if there are two separate parts of an electorate and the majority-grade of a candidate in each is a g, then the majority-grade of the candidate is a g in the whole electorate as well. This idea is suggested by the following concept invented [26] to characterize the scoring methods (that assign a fixed number of points to each place in a voter's ranking, such as Borda's, or first-past-the-post). A method is *winner-consistent* if the method used in each of two separate parts of an electorate makes candidate C the winner, then the method used in the whole electorate must make C the winner as well. The same idea may be used to characterize the point summing methods [4]. But scoring (and point summing) methods are all highly manipulable. The majority judgement is not winner-consistent, and that is a good property: winning is a relative concept that puts aside absolute evaluations and so opens the door to all the inconsistencies (the different intensities of the two parts of the electorate should count!).

Every vote counts. A husband and wife with opposite opinions sometimes skip voting since their votes "cancel each other out." There are many situations where one or a group of voters' ballots cancel each other out if a mechanism based on summing or averaging points or a scoring method is used. For example, one voter gives the same number of points to opposing candidates; or several voters give points to opposing candidates that sum to the same total; or the inputs are rank-orders, and a group of voters places every candidate in every slot of their rankings the same number of times. But this is not true of the majority judgement: every grade contributes to the determination of the majority-ranking (even when a voter gives the same grade to every candidate). Moreover, whatever may be a voter's grade or whatever may be the grades of a group of voters, there exists a situation where the voter or the group of voters is decisive, that is, counting the voter's or the group of voter's ballot(s) gives one outcome, not counting it or them gives another outcome.[23]

Freedom of expression. Some critics have averred that a voter should be forced to "make up his or her mind" by expressing a clear cut preference between any two candidates. The first-past-the-post system has this property (unless the voter abstains or hands in a blank ballot). Any mechanism in which the input is a rank-order of the candidates forbids the voter from expressing any intensity of preference: the second ranked candidate is only that, whatever the voter's evaluation. But why limit any voter's freedom of expression? Shouldn't someone who sees no discernable difference between two or more candidates be allowed to record this? Shouldn't a voter who believes his or her second ranked candidate is merely acceptable or worse be allowed to express this? The majority judgement gives voters complete freedom of expression (within the bounds of the language).

[23] See [2], or [4] for proofs.

3 Other voting mechanisms

Approval voting

On April 21, in the first round of the French presidential election of 2002—well before we had any inkling of even working on the general problem of electing and ranking—one of us initiated an approval voting experiment,[24] conducted under the same general conditions as the experiment of 2007, in five of Orsay's twelve precincts[25] and the one precinct of Gy-les-Nonains, a small country town in Loiret. 2,597 of the 3,346 who voted officially (or 78%) participated in the experiment, 2,587 ballots were valid.

Officially, voters were confronted with having to give their one vote to one of sixteen candidates in the official vote. The ballot of the experiment consisted of a list of the candidates together with instructions saying:

> "*Rules of approval voting.* The elector votes by placing crosses [in boxes corresponding to candidates]. He may place crosses for as many candidates as he wishes, but not more than one per candidate. The winner is the candidate with the most crosses."

The instructions are deliberately neutral: no question is asked, no language is suggested, the explanation is purely relative.[26]

On average the voters cast 3.15 crosses per ballot (the distribution is given in table 24). The actual system offered voters 17 possible messages, approval voting offered more than 65 thousand.[27] Of the 2,587 valid ballots, 813 were different. Voters expressed their relief at having the possibility of casting crosses for as many candidates as they wished.

Crosses	0	1	2	3	4	5	6	7	8	9	10/16
Ballots	36	287	569	783	492	258	94	40	16	6	6
% Ballots	1.4	11.1	22.0	30.3	19.0	10.0	3.6	1.5	0.6	0.2	0.2

Table 24. Number of ballots with k crosses, $k = 0, 1, \ldots, 16$, approval voting experiment, five voting precincts of Orsay and Gy-les-Nonains, first round, April 21, 2002.

This experiment offered a rare opportunity to show that the expressed preferences of voters are far from being "single-peaked" with regard to a left/right political spectrum, *i.e.*, there exists no alignment of the candidates by which

[24] The idea to experiment approval voting on a large scale in parallel with a presidential election actually goes back to 1995, when Balinski and Laurant Mann prepared a basic plan, but were too late to realize it. For a detailed account of the 2002 experiment see [5].

[25] $1^{st}, 5^{th}, 6^{th}, 7^{th}$, and 12^{th}.

[26] This is standard practice. The 2007 ballot for the election of the officers of the Society for Social Choice and Welfare gives similarly neutral instructions: "You can vote for any number of candidates by ticking the appropriate boxes."

[27] With 16 candidates there are $2^{16} = 65,536$ possible messages. With the majority judgement there are 6^{16} or some 2.8 trillion possible messages.

a voter who most prefers any candidate C increasingly dislikes other candidates the further they are from C in the alignment. For if there were such an alignment, the total number of possible sincere messages—messages that are consistent with the voters' preferences—could be at most 137.[28]

The outcomes in the six voting precincts with approval voting and with the official voting are given in table 25. The one significant difference between them is that Le Pen is third in the official vote, eleventh in the approval vote (otherwise, Laguiller moves up three places to behind Madelin and Bescancenot moves up one place to behind Taubira). The four most important candidates—Chirac, Le Pen, Jospin and Bayrou—all lost relative support in approval voting, whereas every one of the minor candidates gained relative support. If Orsay and Gy-les-Nonains were at all representative of France, the results of the experiment showed that the indecision of the country—the lack of enthusiasm for any one candidate or party—was even more extreme than the usual method of voting indicated. No candidate received anywhere near a majority of the ballots (no "legitimacy" is added to the first-placed candidate, contrary to the claims made for approval voting [8]). Whereas we had entered into this experiment persuaded by the usual "common sense" arguments that approval voting was a good idea, the results left us with a distinct feeling that it is not a reasonable mechanism. We did not know exactly why. Now we believe we do.[29]

	% ballots with crosses	% of all crosses	Official vote first round
Jospin	40.5%	12.9%	19.5%
Chirac	36.5%	11.6%	18.9%
Bayrou	33.5%	10.7%	9.9%
Chevènement	30.3%	9.6%	8.1%
Mamère	28.9%	9.2%	7.9%
Madelin	21.3%	6.8%	5.0%
Taubira	18.9%	6.0%	3.2%
Lepage	17.9%	5.7%	2.8%
Besancenot	17.6%	5.6%	3.1%
Laguiller	15.4%	4.9%	3.7%
Le Pen	14.6%	4.6%	10.0%
Hue	11.5%	3.6%	2.7%
Saint-Josse	7.8%	2.5%	1.7%
Boutin	7.8%	2.5%	1.3%
Mégret	7.7%	2.4%	1.3%
Gluckstein	4.3%	1.4%	0.8%
Total	314.6%	100%	100%

Table 25. Approval voting results, five precincts of Orsay and Gy-les-Nonains, first round, April 21, 2002.

[28] The crosses would have to be consecutive with regard to the alignment: there are 16 such messages with one cross, 15 with two, 14 with three, ..., 1 with sixteen and 1 with none.

[29] For a different analysis of this experiment see [19].

The result of the second round on May 5, 2002 in the five precincts of Orsay and the one of Gy-les-Nonains was

Jacques Chirac: 89.3% Jean-Maire Le Pen: 10.7%

The electorate's will expressed by approval votes is not sufficient to "predict" this outcome (nor therefore the result of any other face-to-face confrontation). Crosses and no crosses do not communicate enough information. The problem is the frequency with which voters assigned crosses to two candidates or no crosses to two candidates (see table 26).[30]

	Jospin	Chirac	Bayrou	Mamère	Chévènement	Le Pen
Jospin	–	34%	44%	75%	56%	48%
Chirac	34%	–	66%	51%	54%	64%
Bayrou	44%	66%	–	55%	60%	61%
Mamère	75%	51%	55%	–	52%	61%
Chévènement	56%	54%	60%	52%	–	54%
Le Pen	48%	64%	61%	61%	54%	–

Table 26. Percentages of same votes (both crosses or both no crosses) to two candidates, approval voting results, five precincts of Orsay and Gy-les-Nonains, first round, April 21, 2002.

Three estimates of a face-to-face vote between Chirac and Le Pen were calculated. In each, if a candidate has a cross and the other does not, the first is given 1 vote, the second is given none. The first estimate gives $\frac{1}{2}$ vote to each candidate if both have crosses or neither do: giving crosses and giving no crosses to both candidates means the voter is indifferent between them. This yields the estimate

Jacques Chirac: 61% Jean-Maire Le Pen: 39%

The second estimate gives $\frac{1}{2}$ vote to each if both have crosses, otherwise 0: giving crosses to both candidates means indifference between them; zeros say nothing concerning the two. This yields the estimate

Jacques Chirac: 79% Jean-Maire Le Pen: 21%

The last estimate gives no vote to each if both have crosses or both do not: no indifference is deducible. This yields the estimate

Jacques Chirac: 80% Jean-Maire Le Pen: 20%

None of these estimates comes close to the actual result. Several crosses on a voter's approval ballot—and even more so, several no crosses—does not mean the voter is indifferent among the corresponding candidates. This shows that the approval voting mechanism does not permit the voters to correctly express their preferences or their indifferences.

[30] The analyses are confined to the more important candidates.

In this experiment approval voting was presented and appears to be a mechanism that simply adds crosses: implicitly the vote is relative, it asks voters to make pair-by-pair comparisons. As a consequence, it invites strategic voting and is for that reason subject to Arrow's paradox. For if some candidates drop out, voters may change their assignments of crosses. For example, a voter's favorite candidate drops out so the voter gives a cross to a candidate to whom he or she had not given a cross before. This may change the order-of-finish among the remaining candidates. Circumstantial evidence for such behavior is given below.

On the other hand, approval voting may be presented and viewed as a mechanism that is a special case of the majority judgement when the common language of grades consists of two words. When there are exactly two grades mathematically the approval voting ranking is the majority-ranking. But in this model, in this perception of the process, the vote is absolute, it asks voters to evaluate the candidates. In this case the voter must be posed a question and be offered a common language of words that make it clear the grades have absolute meanings. This has not been the case in any of the theoretical discussions or applications of approval voting, where the question posed, the addition of crosses and the analyses of results all suggest the point of view that what is important is comparisons. Had anyone thought about crosses and no crosses as absolute evaluations, they would (or should) have immediately pointed out that approval voting is a mechanism that excludes Arrow's paradox, so satisfies IIA!

The contrast between absolute evaluations and relative comparisons may be seen in the very different questions posed in two 2007 polls (see above, table 20): "Would each of the following candidates be a good President of France?" and "Do you personally wish each of the following candidates to win the presidential election?" The first poses an absolute question, the second a relative one. The first invites an evaluation, the second suggests a contrast. The answers are, in consequence, completely different. Significantly, the first question elicited a "yes" for the four major candidates considerably more in keeping with their *Good* or better grades in the 2007 majority judgement experience than did the second question.

If a cross is interpreted as an "approve"—so implicitly no cross is interpreted as a "disapprove"—then the winning candidate in the 2002 experiment, L. Jospin, is elected with a majority-grade of "disapprove," for that is the will of a majority of 59.5% of the electorate. It is unacceptable to elect a candidate of whom a majority disapproves! More grades are needed.

The crosses, it turns out, were used in the same way by the voters: there were on average 3.15 crosses per ballot over all six precincts, and about the same number in each. This does not, however, imply that the two "words" constituted a common language of absolute grades because usage includes strategic behavior, and perhaps what was in common was the strategic behavior. The point is this: if voters assign crosses because of absolute evaluations of the merits of candidates then the language is common; otherwise, the language is not common. If the behavior is absolute, Arrow's paradox cannot arise; if it is not absolute, the paradox can arise since the crosses assigned depend on the set of candidates.

Another experiment that was conducted in 2007 in parallel with the first round of the French presidential election provides data that allows a circumstantial analysis of this issue.

The Baujard-Igersheim experiment [6] tested two mechanisms at once[31]—approval voting (and a point-summing mechanism with points 0, 1 or 2, discussed below)—in six different voting precincts[32] with 2,836 participants (62% of those who voted officially). The approval voting ballot stated:

> *Instructions*: You indicate, among the 12 candidates, those that you support. To do so encircle the name of that or those candidates whom you support. You may encircle one name, several names or no name... The candidate elected with [this] method is the one who receives the highest number of supports.

On average the voters cast 2.33 circles per ballot. Moreover, each of the six precincts did approximately the same, so the circles were used in about the same way by all voters. The outcomes over the six precincts are given in table 27. Again, no candidate had circles in a majority of the ballots; again, the (four) major candidates all lost relative support in approval voting whereas every one of the others gained; again, as a language, the mechanism failed because the winner's grade—expressed by the majority—was "not support."

	% ballots with circles	% of all circles	Official vote first round
Bayrou	49.7%	21.4%	23.0%
Sarkozy	45.2%	19.4%	34.1%
Royal	43.7%	18.8%	23.6%
Besancenot	23.7%	10.2%	4.1%
Voynet	16.9%	7.3%	2.1%
Le Pen	11.6%	5.0%	7.6%
Bové	11.5%	4.9%	1.1%
Laguiller	9.3%	4.0%	1.0%
Villiers	9.0%	3.9%	1.7%
Buffet	7.4%	3.2%	0.8%
Nihous	3.4%	1.5%	0.6%
Schivardi	1.4%	0.6%	0.3%

Table 27. Approval voting results, Illkirch, Louvigny and Cigné, April 22, 2007 [6].

The analysis of the absolute *vs.* relative vote issue is based on the considerable information found in the majority judgement ballots. Since the language is common to random samples of 50 or 100 voters from the three precincts in

[31] One ballot contained both. This permits analyses of potential interest. On the other hand, the participants expressed themselves twice simultaneously, which may have induced inter-dependencies.

[32] Three precincts in Illkirch (Alsace), two in Louvigny (Basse-Normandie) and one in Cigné (Mayenne).

Orsay, it is reasonable to hypothesize that the distribution of grades is common to the voters anywhere in France (*nota bene*: the language is common, not the evaluations of the candidates). In the approval voting experiment there were 2.33 circles per ballot. If voting behavior was based on an absolute scale only, then voters would cast circles either for the candidates deemed *Excellent*, or those deemed *Very Good* or better, or *Good* or better, ... But (see table 8) there are on average 0.69 *Excellent*'s, 1.94 *Very Good*'s or better, and 3.44 *Good*'s or better: none of these agrees with 2.33, suggesting that the behavior is not purely absolute.

Each majority judgement ballot assigns a grade to every candidate. The highest grade is given to one or more candidates; the second highest to one or more candidates; and so on down the list. Their averages may be computed (see table 28): they are common to all three precincts as well. If voting behavior was based on a relative scale—assuming these averages are common to all of France—then 2.33 should be about equal to 1.64, or 3.83, or ... It isn't, suggesting that the behavior is not purely relative.

Grades:	Three prcts.	1^{st} prct.	6^{th} prct.	12^{th} prct.
Avg. no. highest	1.64	1.51	1.62	1.80
Avg. no. second highest	2.19	2.08	2.16	2.34
Avg. no. third highest	2.76	2.73	2.78	2.76

Table 28. Average number of highest, second highest, and third highest grades, three precincts of Orsay, April 22, 2007.

Behavior in the 2007 approval voting experiment is better explained as a mixture of absolute and relative behavior:

- a voter casts circles for every candidate deemed above a *Good*; and
- if the the voter deems no candidate above a *Good*, he or she casts circles for every candidate receiving his or her highest grade.

This behavior implies an average of 2.26 circles per approval ballot in the three Orsay precincts, an average of 2.09 in the 1^{st}, of 2.27 in the 6^{th} and of 2.43 in the 12^{th}. This is in substantial agreement with the 2.33 observed in the 2007 approval voting experiment.[33]

Another observation reinforces the idea that voters express relative opinions in approval voting. The 2.33 on average approvals of 12 candidates in the 2007 Baujard-Igersheim experiment is an approval rate of 19.4%. The 3.15 on average approvals of 16 candidates in the 2002 Orsay experiment is an approval rate of 19.7%. This is incredible stability. It cannot be that a fifth of the candidates are always *Good* or above independent of who the candidates are (see, *e.g.*, table 29). Behavior that sees voters approving of some 20% of the candidates suggests

[33]Applying this behavior to the majority judgement ballots of the Orsay experiment to simulate an approval vote gives the following percentages of ballots with circles: Bayrou 51.1%, Royal 44.8%, Sarkozy 44.1%, Besancenot 16.8%, Voynet 14.5%, Buffet 11.6%, Villiers 9.9%, Bové 9.0%, Laguiller 9.0%, Le Pen 8.7%, Nihous 3.2%, Schivardi 2.6%.

they are making relative evaluations just as they are asked to do, not absolute evaluations.

We conclude that the approval voting experiments exhibited behavior that was not purely absolute. There are two implications: first, Arrow's paradox cannot be excluded; second, this realization of approval voting is not an instance of the majority judgement with two grades.

Voting by points and summing

The well nigh universally used mechanism for combining many number grades into one—in skating, diving, gymnastics, piano, wine and other competitions—is to add them or to find their average. Recently, bloggers and others in the U.S.A. and France (and surely other countries) have suggested the same idea for voting (though the scales have varied). Some have suggested that an "easier" way to realize the majority judgement would be to assign a 5 to *Excellent*, a 4 to *Very Good*, down to a 0 to *to Reject*, and then simply add the numbers. Why use the numbers 5 down to 0 instead of (say) 10, 7, 6, 3, 1 and -2 is not explained. In any case, adding or averaging numbers of some arbitrary scale is a very misguided idea.

How to construct a scale of measurement is a science in and of itself. "Measurement theory" classifies scales according to their types (see, *e.g.*, [16]). "Nominal measures" use scales that only assign categories (*e.g.*, a postal or telephone code): the only meaningful comparisons are "equal" or "not equal." "Ordinal measures" use scales that only assign an order (*e.g.*, the A, B, C, D, E, F school grades, the six word language of the Orsay experiment): the only meaningful comparisons are "equal," "greater than" and "less than." "Interval measures" use number scales that assign an order but where also equal intervals have equal significance (e.g., Celsius and Fahrenheit temperatures): the meaningful comparisons are those of ordinal measurement, but it also makes sense to add, to subtract, and to find averages. Finally, "ratio mesures" use number scales that are interval measures but where also zero has an absolute meaning (e.g., length, price, Kelvin temperatures): the meaningful comparisons are those of interval measures, but it also makes sense to multiply and divide.

Numerical languages used in practice—for evaluating students, skaters, earthquake damages, wines, divers, ...—define what is meant by the numbers. Denmark's new seven-grade number language adopted for the academic year 2006–07 (in order to conform with the new European Credit Transfer Accumulation System's ECTS grading scale[34]) is a good example: 12, 10, 7, 4, 2, 0, or -3. For sums and averages to make any sense at all this scale must be an interval measure. The language of grades is described as follows:

- 12 (A) – *outstanding*, no or few unconsiderable flaws, 10% of passing students,

[34]The previous Danish number scale had ten integers: 0 through 13 *without* 1, 2, 4 and 12. The information concerning the Danish grading systems was found in http://en.wikipedia.org/wiki/GPA, Dec. 5, 2007.

- 10 (B) – *excellent*, few considerable flaws, 25% of passing students
- 7 (C) – *good*, numerous flaws, 30% of passing students,
- 4 (D) – *fair*, numerous considerable flaws, 25% of passing students,
- 2 (E) – *adequate*, the minimum acceptable, 10% of passing students,
- 0 (Fx) – *inadequate*,
- -3 (F) – *entirely inadequate*.

To be an interval measure, the numbers must be related to the percentages of passing students. Imagine that all the real numbers from 2 ("the minimum acceptable") up to 12 are the passing grades (they could be points obtained in an examination).[35] What grade should be assigned to a 5.7? That grade whose number (2, 4, 7, 10 or 12) is closest to 5.7, namely, *good*. Any number from the interval $[5.5, 8.5]$ should be mapped into a *good*. By the same token any grade from the interval $[2, 3]$ is mapped into an *adequate*, from $[3, 5.5]$ into a *fair*, from $[8.5, 11]$ into an *excellent*, and from $[11, 12]$ into an *outstanding*. The five numbers (2, 4, 7, 10, 12) were chosen so that the intervals occupy, respectively, the percentages of the whole equal to the percentages of passing grades specified in the definition: $[2, 3]$ occupies 10% of the interval from 2 to 12, $[3, 5.5]$ occupies 25%, $[5.5, 8.5]$ occupies 30%, $[8.5, 11]$ occupies 25% and $[11, 12]$ occupies 10%.

But, is it reasonable to use numerical scales in voting? The answer is a resounding no, for several reasons.

First, the numbers mean nothing unless they are defined: proposals to use weights give them no definition. Their only real "meaning" is found in their strategic use. This induces comparisons, which immediately leads to Arrow's paradox. In the traditional model Arrow's paradox arises when a candidate drops out because that may change the order of finish among the others. Here it may arise when a candidate drops out because the strategies of voters may change, provoking a change in the order of finish among the others. Suppose a 0, 1, 2 scale is used, a voter believes several candidates are decent and the rest bad, gives a 2 to one "preferred" decent candidate, 1's to the others, 0's to the bad candidates. If the candidate with the 2 drops out, the voter may give a 2 to another "decent" candidate. Circumstantial evidence for such behavior is found in the Baujard-Igersheim 0, 1, 2 experiment [6].

The other ballot of that experiment stated:

> *Instructions*. You give a grade to each of the 12 candidates: either 0, or 1, or 2 (2 the best grade, 0 the worst). To do so, place a cross in the corresponding box. ... The candidate elected with [this] method is the one who receives the highest number of points.

The instructions are neutral: nothing is said concerning the meaning of 0, 1 or 2. The numbers induce relative, so strategic, behavior. Other numbers could

[35] This analysis results from a theoretical argument developed in [4].

have been given. For example, −1, 0, and +1: mathematically there is strictly no difference, but were these numbers used the behavior of the voters would almost surely have been different.

On average a ballot contained 1.68 "2's," 2.69 "1's," and 7.64 "0's." Behavior throughout the six precincts was very similar, so the "0's," "1's," and "2's" were used in the about same way. However, the evidence suggests that voters used the numbers in a relative sense not an absolute sense. On average the "2's" were used 1.68 times per ballot. If voters used the "2's" as an absolute indication of merit then its use should correspond to an evaluation of either *Excellent*, or at least *Very Good*, or at least *Good*,... But there are on average 0.69 *Excellent*'s, 1.94 at least *Very Good*'s, still more at least *Good*'s: none agrees with 1.68, so the behavior seems not to be purely absolute. On the other hand, 1.68 is in substantial agreement with the average number of highest grades regularly given in the Orsay experiment, 1.64 (see table 28), suggesting that the "2's" are purely relative.

Second, when numbers are used, they may well not be used in the same way at all: when a 0 to 100 scale is used, some voters may view 80 to be an excellent grade, others may see it as a merely middling grade.

Third, even if the numbers do provide a common language, they will almost certainly not be a proper interval measure, for that depends on who the candidates are and how the voters give their grades. For example, the 0 to 20 scale used in France is a common language, but an 18, 19 or 20 is unheard of in philosophy or literature, so the scale is not an interval measure. Once the distribution of the grades is known—after many elections (or many examinations)—it is possible to determine whether the scale is an interval measure and, if not, to correct it (as did the Danes). But then it is too late, since the weights must be announced ahead of time. Candidates and elections are much rarer than students and examinations, so it is not possible to "learn" and determine norms as the Danes did.

	Excellnt	*Very Gd*	*Good*	*Acceptbl*	*Poor*	*to Rjct*	Sum
Avg/ballot all	0.69	1.25	1.50	1.74	2.27	4.55	12
Avg/ballot four	1.57	2.34	1.94	1.49	0.99	3.68	12

Table 29. Average number of grades per majority judgement ballot all candidates, and four important candidates (Bayrou, Le Pen, Royal and Sarkozy, normalized to sum to 12).

Fourth, even if it turned out that the scale did approximate an interval measure, the procedure *depends on irrelevant alternatives*, it is subject to Arrow's paradox: for if one or several candidates drop out, the distribution of the remaining grades will almost certainly be different, so the scale is no longer an interval measure. The weights would then have to be changed to obtain a scale that makes it an interval measure, which could change the rank-order among the remaining candidates. When, for example, only the four important candidates are present—Bayrou, Le Pen, Royal and Sarkozy—the distribution of the grades (normalized to sum to 12) is entirely different (as may be seen in table

29). (This change in unimportant to the majority judgement because it is a purely ordinal method where no adding or averaging is done.)

Finally, there may well be situations where the numbers are at once a common language and an interval measure: possible examples are those used in evaluating wines, divers and figure skaters, where the judges are professionals who have learned the meanings of the numbers and scales. But in this case, as in all cases when numbers are used, adding (or averaging) is a bad idea because among all possible mechanisms for amalgamating the numbers it is the most manipulable, so the most open to exaggeration and outright cheating.

Traditional mechanisms

The traditional mechanisms are Condorcet's, Borda's and their derivatives, and the single-transferable vote. None offers the voters complete freedom of expression, none yield the electorate's evaluations of the candidates.

Condorcet's method has never, it seems, been used in elections. Perhaps because it is not transitive and, since it ignores evaluations, because it strongly favors "centrist" candidates (often second in voters' rank-orders, but not held in high esteem). It was, for a very short time, used to rank figure skaters, doubled—in case of an intransitivity—by Borda's rule (see [4]; in fact, the exact rule has been proposed and defended [10]).

Borda's method was adopted in about 1784 to elect members of France's Academy of Sciences until a newly elected member, Napoléon Bonaparte, insisted it be discarded in 1800, presumably because it is highly manipulable. Laplace had argued that this system would determine even the most honest voters "to rank last the most dangerous opponents to their favorites," observing, moreover, that "the institutions which adopted it has led them to abandon it." Otherwise, it never seems to have been used in large elctorates. It violates IIA, it ignores intensities, in Laplace's words, it gives "a big advantage to candidates of mediocre merit." Nevertheless, arguments for it, alone or in convolutions, continue to be made to the present day [22].

The single transferable vote is, of course, used, notably in Australia and Ireland. And yet, it violates IIA, ignores intensities, is not monotone and its rules for which candidate to drop are completely arbitrary—why drop the candidate least often first rather than the candidate most often last?—and can lead to different outcomes (in the Faches-Thumesnil experiment [11], respectively, Sarkozy and Bayrou).

4 Conclusion

Physicists rejected the ptolemaic model that affirmed the moon, planets and stars revolved about the earth because—in Feynman's pithy words—"it didn't work": it did not fit the observed facts. The majority judgement experiment proves that the model on which the theory of social choice and voting is based is simply not true: voters do not have preference lists of candidates in their

minds. Moreover, forcing voters to establish preference lists only leads to inconsistencies, impossibilities and incompatibilities. The model has led to important concepts, to criteria for testing the acceptability of voting mechanisms, and to a beautiful body of mathematical results, but it has failed to establish a *science* of social choice that deals with the actual practice of voting as well as the theory of voting because its premises are false. The time has come for social choice and voting theorists to turn to a new model.

The experiment shows that the model proposed here—that voters have evaluations of candidates in their minds and accept to express them in a common language—is much closer to the observed facts. Moreover, the model leads to a coherent theory.

The experiment shows the majority judgement is a practical mechanism. The theory shows—and the experiment illustrates—that it satisfies almost every criterion that has been advanced across the years to test whether a method of voting is acceptable. It is not completely impervious to manipulation. But there exists no method that is. The majority judgement best resists manipulation by several criteria, as the experimental evidence has illustrated and mathematical arguments have proven [4]. It offers voters the greatest freedom of expression and yields evaluations of all candidates (even when there is only one). Science is of course not static: more experiments will reveal more about the behavior of voters and their strategies, so perhaps other means will be found to express their opinions and to amalgamate them into society's opinion.

Changes in methods of election inevitably provoke changes in the behavior of candidates and voters. Today's voting methods—and in particular, the first-past-the-post systems—incite candidates to obtain the support of a majority of the voters and to forget the others. Voters are urged to give their allegiance to one party and oppose the others. Voters are unable to express their appreciations of the candidates (even when there are but two candidates, let alone more). Political strategy focuses on one important point: to gather 51% of the vote. Minorities may be ignored, even offended. The majority judgement incites candidates to seek the highest possible evaluation of every voter. Minorities cannot be ignored. Voters are confronted with a much more serious question—how do you evaluate the candidates?—and are given the means to express themselves. In consequence, instead of focusing on 51% of the electorate up to election day, then once pronounced the winner claim to represent 100% the next day, a candidate is motivated to address his appeal to the entire nation before as well as after the election. The strategies of the political campaigns with today's voting methods cannot be imagined as those with the majority judgement.

Ecclesiastes poses the question:

> "Is there *any* thing whereof it may said, See, this *is* new?"

Indeed, one century ago, Sir Francis Galton [12] had the germ of the idea of the majority judgement. He proposed the median as the solution to the *budget problem*:

> A certain class of problems do not as yet appear to be solved according to scientific rules, though they are of much importance and

> of frequent recurrence. Two examples will suffice. (1) A jury has to assess damages. (2) The council of a society has to fix on a sum of money, suitable for some purpose. Each voter, whether of the jury or the council, has equal authority with each of his colleagues. How can the right conclusion be reached, considering that there may be as many different estimates as there are members? That conclusion is clearly *not* the *average* of all the estimates, which would give a voting power to "cranks" in proportion to their crankiness. One absurdly large or small estimate would leave a greater impress on the result than one of reasonable amount, and the more an estimate diverges from the bulk of the rest, the more influence would it exert. I wish to point out that the estimate to which least objection can be raised is the *middlemost* estimate, the number of votes that it is too high being exactly balanced by the number of votes that it is too low. *Every other estimate is condemned by a majority of voters as being either too high or too low, the middlemost alone escaping this condemnation.*[36]

Acknowledgements

The 2007 Orsay experience could not have been realized without the generous support of Orsay's Mayor, Mrs. Marie-Hélène Aubry, the staff of the Mayor's office, and our friends and colleagues who sacrificed their Sunday (a beautiful spring day) to urging voters to participate and explaining the idea: Pierre Brochot, Stéphanie Brochot, David Chavalrias, Sophie Chemarin, Clémence Christin, Maximilien Laye, Jean-Philippe Nicolai, Matias Nuñez, Vianney Perchet, Jérôme Renault, Claudia Saavedra, Gilles Stoltz, Tristan Tomala, Marie-Anne Valfort, and Guillaume Vigeral. Thanks to them, the experiment was successful and its expense limited to the costs of ballots, envelopes, and posters.

References

[1] Kenneth J. Arrow, *Social Choice and Individual Values*, Yale University Press, 1951, 2nd ed., 1963.

[2] Michel Balinski and Rida Laraki, "A theory of measuring, electing and ranking," *Proceedings of the National Academy of Sciences, U.S.A.*, vol. 104 (2007) pp. 8720-8725.

[3] Michel Balinski and Rida Laraki, "Le jugement majoritaire: l'expérience d'Orsay," *Commentaire*, No. 118, summer 2007, pp. 413-419.

[4] Michel Balinski and Rida Laraki, *One-Value, One-Vote: Measuring, Electing and Ranking* (tentative title), forthcoming.

[36] Our emphasis.

[5] M. Balinski, R. Laraki, J.-F. Laslier and K. Van Der Straeten, "Le vote par assentiment: une expérience," Cahier du Laboratoire d'Économétrie, École Polytechnique, N^o 2003-013.

[6] Antoinette Baujard and Herrade Igersheim, "Expérimentation du vote par note de du vote par approbation lors de l'élection présidentielle française du 22 avril 2007 – Premiers résultats," Report prepared for the Centre d'analyse stratégique, 2007.

[7] Le Chevalier Jean-Charles de Borda,"Mémoire sur les lections au scrutin," *Histoire de l'Académie royale des sciences* (1784), pp. 657-665.

[8] Steven J. Brams and Peter C. Fishburn *Approval Voting*, Birkhäuser, Boston, 1983.

[9] Le Marquis Jean Antoine Caritat de Condorcet, *Essai sur l'application de l'analyse la probabilité des décisions rendues à la pluralité des voix*, Paris, 1785, l'Imprimerie royale.

[10] Partha Dasgupta and Eric Maskin, "The fairest vote of all," *Scientific American*, March 2004, pp. 97.

[11] E. Farvaque, H. Jayet, and L. Ragot, "Quel mode de scrutin pour quel 'vainqueur'?: une expérience sur le vote préférentiel transférable," working paper, Laboratoire Equippe, Universités de Lille, May 2007.

[12] Francis Galton, "One vote, one value," Letter to the editor, *Nature* vol. 75, Feb. 28, 1907, p. 414.

[13] Allan Gibbard, "Manipulation of voting schemes: a general result," *Econometrica* vol. 41 (1973), pp. 587-601.

[14] G. Hägele and F. Pukelsheim, "Llull's writings on electoral systems," *Studia Lulliana* vol. 41 (2001), pp. 3-38.

[15] G. Hägele and F. Pukelsheim, "The electoral systems of Nicolaus Cusanus," to appear in Hg. G. Christianson, T.M. Izbicki, and C.M. Bellitto, eds., *The Church, the Councils and Reform: Lessons from the Fifteenth Century*, Catholic University of America Press, Washington, DC.

[16] D. H. Krantz, R. D. Luce, P. Suppes and A. Tversky, *Foundations of Measurement, Vol. 1*, Academic Press, New York, 1971.

[17] P. Kurrild-Klitgaard, "An empirical example of the Condorcet paradox of voting in a large electorate," *Public Choice* vol. 107 (1999), pp. 1231-1244.

[18] Pierre-Simon, Marquis de Laplace, *Théorie analytique des probabilités*, 3rd ed., in *Œuvres Complètes de Laplace*, t. 7, pp. v and clii-cliii.

[19] J.-F. Laslier and K. Van Der Straeten, "Vote par assentiment pendant la présidentielle 2002 : analyse d'une expérience," *Revue Française de Science Politique* vol. 54, pp.99-130.

[20] Iain McLean, "The Borda and Condorcet principles: three medieval applications," *Social Choice and Welfare* vol. 7 (1990), pp. 99-108.

[21] Hervé Moulin, "On strategy-proofness and single peakedness," *Public Choice* vol. 35 (1980), pp. 437-455.

[22] Donald D. Saari, *Chaotic elections! A Mathematician Looks at Voting*, American Mathematical Society, 2001.

[23] Mark A. Satterthwaite, "Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions," *Journal of Economic Theory* vol. 10 (1973), pp. 187-217.

[24] H. P. Young, "Condorcet's theory of voting," *American Political Science Review* vol. 82 (1988), pp. 1231-1244.

[25] H. P. Young, "Optimal ranking and choice from pairwise comparisons," in B. Grofman and G. Owen, eds., *Information Pooling and Group Decision Making*, JAI Press, Greenwich, CT, 1986, pp. 113-122.

[26] H. P. Young, "Social choice scoring functions," *SIAM Journal of Applied Mathematics* vol. 28 (1975), pp. 824-838.

Appendix

Majority judgement tie-breaking rule

The majority-values of the candidates are used to calculate the majority-ranking. If the majority-grades[37] with mention (the "majority-grades*") are different, then they determine the order between two majority-values:

$$(p, \alpha^*, q) \succ (s, \beta^*, t) \text{ if } \alpha^* \succ \beta^*,$$

where $\alpha^* \succ \beta^*$ if $\alpha \succ \beta$ and $\alpha^+ \succ \alpha^0 \succ \alpha^-$. So suppose the majority-grades with mention are the same. Then

$$(p, \alpha^+, q) \succ (s, \alpha^+, t) \text{ if } \begin{cases} p > s \\ p = s \text{ and } q < t \end{cases} \qquad (\text{recall } p > q, s > t),$$

$$(p, \alpha^0, q) \succ (s, \alpha^0, t) \quad \text{if} \quad 100 - p - q > 100 - s - t \quad (\text{recall } p = q, s = t),$$

[37] When the number of voters is odd, the majority-grade is the median, or the one middle grade. In the extremely unlikely case that there is an even number of voters and a candidate's two middle grades are different, the general theory shows that the lower of the two middle grades must be the majority-grade.

$$(p, \alpha^-, q) \succ (s, \alpha^-, t) \text{ if } \begin{cases} q < t \\ q = t \text{ and } p > s \end{cases} \qquad (\text{recall } p < q, s < t).$$

It is highly unlikely that $p = s$ or $q = t$.

The rule is balanced with regard to higher and lower grades. Call it the *coherent-rule* since it is the logical consequence of the general rule (see [2, 4]) that determines an order between any pair of candidates unless both have *exactly* the same sets of grades. It is practically certain that it will give an unambiguous order of finish in any election with many voters.[38] The order may be shown to be monotone and transitive.[39]

The coherent-rule ranks Royal ahead of Bayrou in Orsay's 12^{th} precinct (see table 16) when their majority-values are

$$\text{Royal } (42.4\%, \textit{Good}, 40.1\%) \qquad \text{Bayrou: } (40.8\%, \textit{Good}, 31.4\%),$$

for both have the majority-grade* $\textit{Good}^+$, and $42.4\% > 40.8\%$.

It could be argued that Bayrou should rank ahead of Royal because $40.8\% - 31.4\% = 9.4\%$ is weightier than $42.4\% - 40.1\% = 2.3\%$. More generally, imagine the grades of the candidates ordered from best to worst on sea-saws that are balanced at the median *Good*: that which weighs more heavily towards the better-than as versus the worse-than grades should correspond to the candidate who is ranked higher. This intuitive idea suggests the simple *difference-rule*:

$$(p, \alpha, q) \succ (s, \beta, t) \text{ if } \begin{cases} \alpha \succ \beta \text{ or} \\ \alpha = \beta \text{ and } p - q > s - t, \end{cases}$$

(where no mentions $\pm$ or 0 are added) that puts Bayrou ahead of Royal.

The difference-rule was suggested by David Gale who at first believed it was a specialization of the general rule. The coherent-rule, deduced from underlying principles, is more subtle: if the weight of one candidate leans toward the better-than side, the weight of the the other leans toward the worse-than side then the better-than side is of course ahead. If both candidates lean toward the better-than side, then the one whose better-than side is the weightier is ahead of the other—so Royal with 42.4% is ahead of Bayrou with 40.8%—because there are more voters who really care about Royal than there are who really care about Bayrou. Symmetrically, if both candidates lean toward the worse-than side, then the one whose worse-than side is the weightier is behind the other.

The difference-rule favors centrist candidates in comparison with the coherent-rule because it only contrasts better-than with worse-than, ignoring the absolute weight of better-than or the absolute weight of worse-than.

The difference-rule is also less resistant to manipulation than the coherent-rule. It may be shown that if a group of voters is able to change the order

[38] If not, the general rule may be used.

[39] To show that it is transitive, consider any majority-value (p, α, q) and associate with it the vector $(\alpha, \sigma, \beta, \gamma)$ where: $\sigma = \text{sign}(p - q)$, ($p = q$ means $\sigma = 0$ and $+ > 0 > -$); $\beta = p(\text{if } \sigma = +) - q(\text{if } \sigma = -)$ and $\gamma = -q(\text{if } \sigma = +) + p(\text{if } \sigma = -) + (100\% - p - q)(\text{if } \sigma = 0)$. The given rule defines a lexicographic order on $(\alpha, \sigma, \beta, \gamma)$.

between two candidates with the coherent-rule then necessarily that group can also change the order with the difference-rule [4]. On the other hand, there are instances when a group of voters can change the outcome with the difference-rule but not with the coherent-rule.

To better see what is going on, imagine two candidates R and B with majority-values

$$R\text{: } (40\%, \textit{Good}, 20\%) \qquad B\text{: } (30\%, \textit{Good}, 20\%),$$

R ranks ahead of B by both rules. But if $5 + \epsilon\%$ (where $\epsilon > 0$) of the voters who gave B grades worse than *Good* increase them to better than *Good* they can change the order with the difference-rule but not with the coherent-rule, since that would give the majority-values

$$R\text{: } (40\%, \textit{Good}, 20\%) \qquad B\text{: } (35 + \epsilon\%, \textit{Good}, 15 - \epsilon\%).$$

To change the order with the coherent-rule would require $(10 + \epsilon\%)$ to increase the grades they give B from worse to better than *Good*—twice as many voters—for that yields the majority-values

$$R\text{: } (40\%, \textit{Good}, 20\%) \qquad B\text{: } (40 + \epsilon\%, \textit{Good}, 10 - \epsilon\%).$$

Now consider the real problem of table 23. The majority-values were

$$\text{Bayrou: } (44.3\%, \textit{Good}, 30.6\%) \qquad \text{Royal: } (39.4\%, \textit{Good}, 41.5\%).$$

so Bayrou ranked well ahead of Royal by both rules. Now suppose the voters made the strategic changes indicated in table 23 except that all of the type D voters changed as indicated (not just 1/3 of them). This means 42.6% of the voters able to have a strategic impact try to do so. Then their respective majority-values become

$$\text{Bayrou: } (42.2\%, \textit{Good}, 41.2\%) \qquad \text{Royal: } (42.0\%, \textit{Good}, 40.8\%).$$

With the coherent-rule Bayrou remains ahead, with the difference-rule Royal takes the lead. Thus the coherent-rule better resists manipulation.