

# Chapter 2

## Election by Majority Judgment: Experimental Evidence

Michel Balinski and Rida Laraki

### 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 have been proposed and 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 method \(1785\)](#), [Borda’s method \(1784\)](#) (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 (1820) was the first to propose two centuries ago they asked 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<sup>1</sup> (sometimes after dropping highest and lowest scores). They have been emulated in various schemes proposed for voting with ranges taken to be integers in  $[0, 100]$ ,  $[0, 5]$ ,  $[0, 2]$ , or  $[0, 1]$  (the last approval voting).

It is fair to ask whether any one 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 important property that a

---

<sup>1</sup>Laplace only used this model to deduce Borda’s method via probabilistic arguments. He then rejected Borda’s method because of its evident manipulability.

M. Balinski (✉)  
École Polytechnique and C.N.R.S., Paris, France  
e-mail: [michel.balinski@polytechnique.edu](mailto:michel.balinski@polytechnique.edu)

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 methods of voting 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 unacceptable *paradoxes* that occur in practice – Condorcet’s, Kenneth Arrow’s and others – and *impossibility theorems* – due to [Arrow \(1963\)](#), to [Gibbard \(1973\)](#) and [Satterthwaite \(1973\)](#). Moreover, as [Young \(1988, 1986\)](#) has shown, 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 ranking wished by society! In fact, the traditional model harbors a fundamental *incompatibility* between winning and ranking [Balinski and Laraki \(2007, 2010\)](#). 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, leads to a new theory as free of the defects as can be.

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, [McLean \(1990\)](#); [Hägele and Pukelsheim \(2001, 2008\)](#)). The impossibility and incompatibility theorems are one good reason to discard the traditional model. The 2007 experiment with the majority judgment 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. Put simply, it is an inadequate model.

The majority judgment is a new mechanism based on a different model of the problem of voting (inspired by practice in ranking wines, figure skaters, divers, and others). It asks voters to evaluate every candidate in a common language of grades – thus to *judge* each one on a meaningful scale – rather than to compare them. This scale is absolute in the sense that the merit of any one candidate in a voter’s view – whether the candidate be “excellent,” “good,” or merely “acceptable” – depends only on the candidate (so remains the same when candidates withdraw or enter). Assigning a value or grade permits comparisons of candidates, do not permit evaluations (or any expression of intensity). In this paradigm, the majority judgment 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 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. Sen (1970) proposed a model whose inputs are the voters' utilities: but satisfaction is a complex, relative notion. The satisfaction of seeing, say, Jacques Chirac (the incumbent candidate of the traditional right) elected in 2002 depends on who opposed him: many socialist voters (or others of the left) who detested Chirac were delighted to see him crush Jean-Marie Le Pen (the ever present candidate of the extreme right). So satisfaction is not independent of irrelevant alternatives and leads to Arrow's paradox. But with a common language of grades, such voters could decide to evaluate Chirac's merit as *Acceptable* or *Good* opposed to Le Pen and/or Lionel Jospin (the incumbent Prime Minister and candidate of the Socialist Party) while awarding a grade of *Poor* or *to Reject* to Le Pen.

In the real world, satisfaction of a voter depends on a host of factors that include the winner, the order of finish, the margin of victory, how socio-economic groups have voted, the method of election, etc. Utilities, we believe, *cannot* be inputs to practical decision mechanisms. Grades of a common language have an absolute meaning that permit interpersonal comparisons. Common languages exist. They are defined by rules and regulations and acquire absolute meanings in the course of being used (e.g., the points given to Olympic figure skaters, divers and gymnasts, the medals given to wines, the grades given to students, the stars given to hotels, etc.). The principal experiment of this paper shows that a common language may be defined for voters in a large electorate as well.

The majority judgment avoids the unacceptable paradoxes and impossibilities of the traditional model. The *theory* that shows why the majority judgment is a satisfactory answer to the basic challenge is described and developed elsewhere (see Balinski and Laraki (2007, 2010)). In this theory, Arrow's theorem plays a central role as well: it says that without a common language, no meaningful final grades exist. Theorems show – and experiments confirm – that while there is no method that avoids strategic voting altogether, the majority judgment best resists manipulation.

The aim of this article is to describe electoral field experiments (as versus laboratory experiments) that show majority judgment provides a *practical* answer to the basic challenge. The demonstration invokes new methods of validation and new concepts. The experiments, and the elections in which they were conducted, show the well-known methods fail to satisfy important properties, and permit them to be compared.

## Background of the Experiments

The experiments were conducted in the context of the French presidential elections of 2002 and 2007. 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

**Table 2.1** Votes: United States presidential election of 2000

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

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.

The first-past-the-post system is, of course, 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 2.1). 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.<sup>2</sup>

### ***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 Table 2.2). 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 Socialiste (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.

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, trotskysts, etc., – had no choice but to vote for Chirac. His votes represented very different sentiments and intensities.

<sup>2</sup> This, of course, assumes that the vast majority of Nader’s votes would have gone to Gore.

**Table 2.2** Votes: French presidential election, first-round, April 21, 2002

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%

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.<sup>3</sup> Had either Chévènement, an ex-socialist, or Taubira, a socialist, withdrawn, most of his 5.3% or her 2.3% of the votes would have gone to Jospin, so the second round would have seen a Chirac-Jospin confrontation, as had been expected. In fact, Taubira had offered to withdraw if the PS was prepared to cover her expenses, but that offer was refused. It has also been whispered that the RPR helped to finance Taubira's campaign (a credible strategic gambit backed by no specific evidence). Moreover, if Charles Pasqua, an aging past ally of Chirac, had been a candidate – as he had announced he would be – then he could well have drawn a sufficient number of votes from Chirac to produce a second round between Jospin and Le Pen, which would have resulted in a lopsided win for Jospin. Anything can happen when the “first-past-the-post” (or the “two-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 also shows that the mechanisms invite “strategic” candidacies: candidates who cannot hope to win (or survive a first round) but can cause another to win (or to reach the second round) by drawing votes away from an opposing candidate.

### ***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

<sup>3</sup> In their last 11 predictions (late February to the election), the Sofres polls showed Jospin winning seven times, Chirac two times, a tie two times.

**Table 2.3** Votes: French presidential election, first round, April 22, 2007

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%

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.

The distribution of the votes among the twelve candidates in the first round is given in Table 2.3. 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 mammoth: 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<sup>4</sup> asked electors 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 500 signatures. They are drawn from a pool of about 47 thousand elected officials who represent the 100 departments, must include signatures coming from at least 30 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.

<sup>4</sup> By Tns – Sofres – Unilog Groupe Logica CMG, April 22, 2007.

**Table 2.4** Polls, March 28 and April 19, 2007, potential second round (IFOP)

	Bayrou		Sarkozy		Royal		Le Pen	
Bayrou	–	–	54%	55%	57%	58%	84%	80%
Sarkozy	46%	45%	–	–	54%	51%	84%	84%
Royal	43%	42%	46%	49%	–	–	75%	73%
Le Pen	16%	20%	16%	16%	25%	27%	–	–

**Table 2.5** Projected second round results, from vote in Faches-Thumesnil experiment [Farvaque et al. \(2007\)](#) (e.g., Sarkozy has 48% of the votes against Bayrou)

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

Polling results (Table 2.4) 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 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 2.4 suffices to determine the “Borda scores”<sup>5</sup> 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 [Farvaque et al. \(2007\)](#) was conducted in Faches-Thumesnil (a small town in France’s northern-most department, Nord) on election day, where the official 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 2.5): they yield the same unambiguous order of finish among the four significant candidates.

The Majority Judgment

2007 Experiment

The experiment took place in three of Orsay’s 12 voting precincts (the 1st, 6th, and 12th). Orsay is a suburban town some 22 km from the center of Paris. In 2002 it was the site of the first large electoral experiment conducted in parallel with a

<sup>5</sup> 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.

presidential election (Balinski et al. 2003, discussed below). The three precincts were chosen among the five of the 2002 experiment as the most representative of the town and its various socioeconomic 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, 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. Thus, this was a *field* experiment.

The intent was to find out whether real, uncontrollable voters of widely differing opinions and incentives could intelligently evaluate many candidates using the ballots of the majority judgment. The outcome was unknown and risky: perhaps few would cooperate or the evaluations would prove too difficult, perhaps a minor candidate would emerge victorious or the winner would receive a very low grade, perhaps indeed the results would simply be chaotic. The analysis of voters' behavior shows that the results make sense and that they evaluated honestly; in any case, they had no incentive to evaluate strategically. This permits a comparison of different methods of voting based on a real "preference profile" of voters in a real election; had the experiment itself been real and binding, some voters would have voted strategically, which would have precluded a valid comparison of methods.

It is important to appreciate 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 2.6).

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 judgment. A team of three to four knowledgeable persons were in constant attendance to encourage participation and to answer questions. Voting *à la* majority judgment 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 2.7.

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.

**Table 2.6** French presidential election, first round, April 22, 2007: national vote vs. vote in the three precincts of Orsay

	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 2.7** The majority judgment ballot (English translation)

<b>Ballot: Election of the President of France 2007</b>						
<i>To be president of France, having taken into account all considerations, I judge, in conscience, that this candidate would be:</i> <sup>6</sup>						
	<i>Excellent</i>	<i>Very Good</i>	<i>Good</i>	<i>Acceptable</i>	<i>Poor</i>	<i>to Reject</i>
Olivier Besancenot						
Marie-George Buffet						
G�rard Schivardi						
Fran�ois Bayrou						
Jos� Bov�						
Dominique Voynet						
Philippe de Villiers						
S�gol�ne Royal						
Fr�d�ric 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.

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 judgment 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. Nineteen of the 1,752 ballots were indecipherable or deliberately subverted, leaving a total of 1,733 valid ballots.

<sup>6</sup> The question in French: “Pour pr sider la France, ayant pris tous les  l ments en compte, je juge en conscience que ce candidat serait:” The grades in French: “Tr s bien, Bien, Assez bien, Passable, Insuffisant,   Rejeter.” The names of the candidates are given in the official order, the result of a random draw.

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.<sup>7</sup> An effective argument to persuade reluctant voters to participate was that the majority judgment 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 judgment offered voters more than 2 billion possible messages.<sup>8</sup> 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.

**The Results**

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 2.8). 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 2.9) 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.

The highest grades were often multiple. Almost 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.

**Table 2.8** Average number of grades per majority judgment ballot

	<i>Excellent</i>	<i>Very Good</i>	<i>Good</i>	<i>Acceptable</i>	<i>Poor</i>	<i>to Reject</i>	<i>Sum</i>
Avg./ballot	0.69	1.25	1.50	1.74	2.27	4.55	12

**Table 2.9** Percentages of voters using  $k$  grades ( $k = 1, \dots, 6$ )

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

<sup>7</sup> A collection of television interviews of participants prepared by Raphaël Hitier, a journalist of *I-Télé*, attests to these facts.

<sup>8</sup> With twelve candidates and six grades, there are  $6^{12} = 2,176,782,336$  possible messages.

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<sup>9</sup> asked at what moment voters had decided to vote for a particular candidate. Their hesitancy in making a choice is reflected in 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 [Farvaque et al. \(2007\)](#) asked voters to rank-order all twelve candidates. They were testing “single-transferable-vote” mechanisms.<sup>10</sup> 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). Nine hundred and sixty 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. Fifty-three percent 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.

Of the 1,733 valid majority judgment ballots,<sup>11</sup> 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 judgment ballots, less than 1.7% of the possible messages would have been used. Those that were the same among the 1,733 valid ballots 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 judgment in the three precincts is given in Table 2.10. Since every candidate was necessarily assigned a grade – assigning no grade meant assigning a *to Reject* – each candidate had exactly the same number of

<sup>9</sup> by TNS Sofres – Unilog Groupe Logica CMG, April 22, 2007, the same poll cited earlier.

<sup>10</sup> 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.

<sup>11</sup> 559 in the 1st precinct, 601 in the 2nd, 573 in the 3rd.

**Table 2.10** Majority judgment results, three precincts of Orsay, April 22, 2007

	<i>Excellent</i>	<i>Very Good</i>	<i>Good</i>	<i>Acceptable</i>	<i>Poor</i>	<i>to Reject</i>
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%

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.<sup>12</sup> *Everyone* with some knowledge of French politics who was shown the results with the names of Sarkozy, Royal, Bayrou and Le Pen hidden invariably identified them: the grades contain meaningful information.

The evidence conclusively demonstrates that the age-old view of voting – and the basic assumption of the traditional model of social choice theory – is not a reasonable model of reality.

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 2.10) 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 Balinski and Laraki (2007, 2010) 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-gauge* – are sufficient to determine the candidate’s place in the majority-ranking:

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

<sup>12</sup> No grade was assigned to each of the candidates in the following percentages: Nihous 7.2%, Schivardi 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%.

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  $\pm$  that depends on the relative sizes of  $p$  and  $q$  and call it the *majority-grade\**:

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

(the possibility that  $p = q$  is slim). Thus, for example, Sarkozy’s majority-gauge is (38.9%, *Good*, 46.9%) and his majority-grade\* is *Good*<sup>−</sup>. Naturally,  $\alpha^+$  is better than  $\alpha^-$ .

Consider two candidates  $A$  and  $B$  with majority-gauges  $(p_A, \alpha_A, q_A)$  and  $(p_B, \alpha_B, q_B)$ .  $A$  ranks ahead of  $B$ , and  $(p_A, \alpha_A, q_A)$  ahead of  $(p_B, \alpha_B, q_B)$ , when

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

To illustrate,

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

It is practically certain that this rule for deciding the order suffices to give an unambiguous order of finish in any election with many voters.

The majority-grades and the majority-gauges for the experiment are given in the order of the majority-ranking in Table 2.11. 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. Le Pen – fourth according to the official vote – is last according to the majority judgment because 74.4% of the voters graded him *to Reject*. 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 judgment 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*, etc., at least *Poor*, all agree with that order (see Table 2.12). Practically any reasonable election mechanism will agree with this ranking of the three important candidates.

**Table 2.11** The majority-gauges ( $p, \alpha, q$ ) and the majority-ranking, three precincts of Orsay, April 22, 2007

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

The columns headed “Natl. rank.” and “Orsay rank.” are the national rank-orders by the current system

**Table 2.12** Cumulative majority judgment grades, three precincts of Orsay, April 22, 2007

	At least					
	<i>Excellent</i>	<i>Very Good</i>	<i>Good</i>	<i>Acceptable</i>	<i>Poor</i>	<i>to Reject</i>
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%

**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 judgment ballots<sup>13</sup> by comparing their respective grades (see Table 2.13). 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 judgment ballots in conformity with the manner in which they actually voted. The 1% difference is easily explained. Twenty-six percent of the voters did not participate in the experiment; and the last two weeks of the campaign may have changed perceptions. The closeness of the estimate to the outcome shows the majority judgment ballots are consistent with the observed facts.

<sup>13</sup> The information in Table 2.10 does not suffice.

**Table 2.13** Face-to-face elections, percentages of votes estimated from majority judgment ballots, three precincts of Orsay, April 22, 2007

	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	–

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

**Table 2.14** First round vote, percentages of votes estimated from majority judgment ballots, three precincts of Orsay, April 22, 2007

	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

The estimates of Table 2.13 show Bayrou to be the Condorcet- and the Borda-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.

The majority judgment 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 2.14). 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

**Table 2.15** Actual percentages, first round, April 22, 2007, in Orsay’s 12th precinct (top row of percentages with names of candidates above) and all of France (bottom row of percentages with names of candidates below)

	Roy	Sar	Bay	LP	Bes	Vil	Voy	Bov	Buf	Lag	Nih	Sch
12th	32.0	26.6	20.2	10.0	2.7	2.5	2.3	1.3	1.2	0.8	0.2	0.0
Ntnl	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 2.16** The majority-gauges ( $p, \alpha, q$ ) and the majority-ranking, Orsay’s 12th precinct, April 22, 2007<sup>14</sup>

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

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.

Some persons have averred that the majority judgment 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 12th precinct. They were close to the result in all of France when the percentages of Royal and Sarkozy are permuted (see Table 2.15).

Bayrou was as much a centrist candidate in the 12th precinct as he was in the three precincts. Yet, in the 12th precinct Bayrou was *not* the majority judgment winner (see Table 2.16): Royal was first.

The results of the face-to-face confrontations between the pairs of major candidates deduced from the majority judgment ballots in the 12th precinct are given for the four major candidates in Table 2.17. Bayrou is again the Condorcet-winner despite Royal’s majority judgment victory: Why?

The reason is clear. 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 judgment ballots show that the voters who gave Sarkozy their highest grade strongly preferred Bayrou to Royal, those who

<sup>14</sup> 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 9th and Nihous 10th.



**Table 2.17** Projected second round results, Orsay’s 12th precinct (e.g., Sarkozy has 41% of the votes against Bayrou)

	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 2.18** Grades given to three major candidates by voters who gave their highest grade to one of the others, three precincts of Orsay, April 22, 2007<sup>15</sup>

		<i>Excellent</i>	<i>Very Good</i>	<i>Good</i>	<i>Acceptable</i>	<i>Poor</i>	<i>to Reject</i>
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 2.19** Distributions highest grades, three precincts of Orsay, April 22, 2007

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

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 2.18).

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 2.19). This is why being second in the rankings of voters has very different senses and aggregating them as does Borda is not meaningful.

First ranked candidates often elicit strong support and strong opposition. Second ranked candidates are often centrists. 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

<sup>15</sup> 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).

garner many Borda points because they are ahead of most of them. But this is not true with the majority judgment: the evaluations – the *grades* of the second ranked candidates – decide, not the place in the ranking.

The closeness of the actual results in Orsay’s 12th 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.

*Common Language*

The theoretical underpinnings of the majority judgment 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 2.20). The question on the left suggests an absolute evaluation, the question on the right a relative 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.

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 Balinski and Laraki (2010)). 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.”

**Table 2.20** Polling results, March 22, 2007 (Bva)

	Question:		Question:	
	Would each of the following candidates be a good President of France?		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%		

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 does 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 0–20 grading systems in its schools and universities, but it also uses the six descriptive words of the majority judgment 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 2.8).

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. About 73% used four or five grades, and the average was 4.5 grades per ballot (see Table 2.9).

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, i.e., are the distributions of the grades used similar? Different approaches may be used to answer this question, but several, very simple direct tests show convincingly that the grades did constitute a common language in the experiment.<sup>16</sup> One 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 2.21 shows that each of the three voting precincts – the 1st with 559 voters, the 6th with 601 voters, and the 12th 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. The outcomes in the different precincts are different – and the outcomes on different samples are different – but the use of the language is practically the same.

---

<sup>16</sup> An extensive investigation, Balinski and Laraki (2010), uses many of the standard statistical tests to confirm this finding.

**Table 2.21** Average number of words per majority judgment ballot, 2007 Orsay experiment ( $\sigma$  is the standard deviation; 10 random samples of 100 and 10 disjoint random samples of 50 were taken)

	Three	1st	6th	12th	Samples of 100		Disjoint samples of 50	
	prcts.	prct.	prct.	prct.	Avg. ( $\sigma$ )	Range	Avg. ( $\sigma$ )	Range
<i>Excellent</i>	0.7	0.7	0.7	0.7	0.7 (0.07)	0.6/0.8	0.7 (0.12)	0.5/0.9
<i>Very good</i>	1.3	1.2	1.2	1.4	1.2 (0.13)	1.1/1.5	1.3 (0.16)	1.1/1.5
<i>Good</i>	1.5	1.5	1.4	1.6	1.5 (0.13)	1.4/1.7	1.5 (0.27)	0.9/1.8
<i>Acceptable</i>	1.7	1.7	1.7	1.8	1.8 (0.15)	1.7/2.1	1.7 (0.27)	2.1/2.6
<i>Poor</i>	2.3	2.3	2.3	2.2	2.3 (0.19)	2.1/2.7	2.3 (0.19)	2.1/2.6
<i>to Reject</i>	4.6	4.8	4.6	4.3	4.5 (0.29)	4.1/4.8	4.5 (0.41)	4.1/5.3

**Table 2.22** Counts of usage of grades by ballot, 2007 Orsay experiment

		Number of times Grades used in a ballot								
	Prct.	0	1	2	3	4	5	6	7	8–12
<i>Excellent</i>	1st	47.0%	43.1%	7.7%	1.6%	0.2%	0.2%	0.0%	0.0%	0.2%
	6th	46.6%	41.8%	8.7%	2.0%	0.7%	0.0%	0.2%	0.0%	0.2%
	12th	51.1%	37.3%	7.9%	2.3%	0.9%	0.2%	0.0%	0.0%	0.3%
<i>Very good</i>	1st	30.2%	40.3%	19.7%	6.8%	1.1%	1.3%	0.5%	0.2%	0.0%
	6th	28.8%	37.9%	22.0%	7.2%	2.7%	0.8%	0.3%	0.3%	0.0%
	12th	26.0%	37.9%	20.4%	8.2%	4.4%	2.1%	0.7%	0.3%	0.0%
<i>Good</i>	1st	24.3%	35.1%	22.2%	11.4%	4.7%	1.4%	0.7%	0.2%	0.0%
	6th	26.3%	35.1%	20.5%	10.1%	5.3%	2.2%	0.3%	0.2%	0.0%
	12th	21.8%	30.4%	25.5%	12.0%	7.2%	2.3%	0.3%	0.3%	0.2%
<i>Acceptable</i>	1st	23.3%	29.3%	20.0%	16.8%	6.4%	3.6%	0.2%	0.0%	0.4%
	6th	22.6%	28.8%	24.1%	13.0%	6.5%	3.7%	0.3%	0.5%	0.5%
	12th	22.5%	23.0%	24.6%	17.1%	7.3%	3.8%	0.5%	0.9%	0.2%
<i>Poor</i>	1st	16.5%	20.0%	22.9%	15.9%	14.0%	5.5%	2.9%	1.4%	0.9%
	6th	16.3%	24.0%	19.5%	17.0%	9.5%	5.7%	5.8%	1.0%	1.3%
	12th	23.2%	20.8%	18.5%	15.2%	10.6%	6.1%	3.1%	1.4%	1.0%
<i>to Reject</i>	1st	3.0%	6.1%	10.7%	12.0%	16.3%	17.2%	10.4%	9.3%	15.0%
	6th	4.7%	4.7%	9.2%	17.0%	18.1%	14.5%	11.0%	7.3%	13.6%
	12th	7.0%	7.3%	14.5%	14.0%	14.5%	13.8%	7.3%	7.0%	14.7%

Table 2.22 simply gives the number of times each of the grades was used in each of the voting precincts. For example, the three percentages in bold type say that in the 1st precinct the grade *Very Good* was used twice in 19.7% of the ballots, in the 6th precinct it was used twice in 22.0% of the ballots, and in the 12th precinct it was used twice in 20.4% of the ballots. They are remarkably the same in all three precincts.

Fourth, the estimates of the second round results based on the majority judgment ballots in the three precincts together and in each of them singly were close to the observed outcomes as well, as shown in Table 2.23. They assumed: (1) when a voter gave a higher grade to one candidate than the other he or she would obtain 1 vote in the second round; and (2) when voters gave the same grades to both candidates

**Table 2.23** Second round results, percentages of votes estimated from first round majority judgment ballots vs. actual outcomes, Orsay, April 22, 2007<sup>17</sup>

	Three precincts		1st precinct		6th precinct		12th precinct	
	Estimated	Outcome	Estimated	Outcome	Estimated	Outcome	Estimated	Outcome
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%

each would obtain  $\frac{1}{2}$  vote in the second round. 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.

*Properties of the Majority Judgment*

Given a common language, the majority judgment – the *majority-grade* and the *majority-ranking* – has been proven to be the *only* mechanism that is acceptable according to several different criteria (see Balinski and Laraki (2007, 2010) for precise definitions and results). Here, we only describe and illustrate the salient properties that are enjoyed by the majority judgment 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 – no measure of intensity between grades is implied – so the mechanism used must be ordinal as well. The majority judgment is ordinal: the majority-ranking is independent of any parametrization of the language. Mechanisms based on sums or averages of points are not ordinal.

**Respects the majority.** 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. 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 Kurrild-Klitgaard (1999).

**Satisfies IIA.** The majority judgment 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

<sup>17</sup> Royal’s scores are consistently though slightly overestimated. This probably reflects changes in opinions in the 2 weeks that separated the two rounds of voting (due, in particular, to the televised debate between the two candidates).

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 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-gauge is

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

Only a voter who can change Royal's majority-grade or majority-gauge 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<sup>18</sup>). Then the voter's optimal voting strategy is to give the candidate the grade *g*: the majority judgment is *strategy-proof-in-grading*.<sup>19</sup> More is true. The majority judgment is *group strategy-proof-in-grading*. If a group of voters (e.g., belonging to a same political party) believed Royal merited better than *Good* and all raised the grade they gave her, her majority-gauge would remain the same; if all lowered the grades they gave her, her majority-gauge 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-gauge would remain the same; if all raised the grades they gave her, her majority-gauge 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-gauge 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 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.

<sup>18</sup> The voter's preferences in grading are said to be "single-peaked."

<sup>19</sup> In an entirely different context, a related technical result is proved in Moulin (1980).

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.”<sup>20</sup> But the majority judgment best resists such manipulation. Take the example of Bayrou with a *Good*<sup>+</sup> and Royal with a *Good*<sup>−</sup>, their respective majority-gauges being,

Bayrou: (44.3%, *Good*, 30.6%)    Royal: (39.4%, *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-gauge and to raise Royal’s majority-gauge. But the majority judgment is *partially strategy-proof-in-ranking*: those voters who can lower Bayrou’s majority-gauge cannot raise Royal’s, those who can raise Royal’s majority-gauge 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-gauge. Symmetrically, a voter who can raise Royal’s majority-gauge must have given to her a *Good* or worse, so to Bayrou a worse than *Good*, so the voter cannot lower Bayrou’s majority-gauge. Compared with mechanisms that sum or average, the majority judgment 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 2.24. 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-gauge 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-gauge 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).

**Table 2.24** Strategic voting: could Royal have won in Orsay’s three precincts?

Type	Percentage ballots	<i>Excellent</i>	<i>Very Good</i>	<i>Good</i>	<i>Good</i>	<i>Acceptable</i>	<i>Poor</i>	<i>to Reject</i>	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

(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)

<sup>20</sup> In the context of the traditional model, this is the Gibbard-Satterthwaite theorem.

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 2.23 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 2.24 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 poll on election day showed 31% of Royal supporters voted strategically). The result is to change the candidates' majority-gauges to

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

so both have the majority-grade\* *Good*<sup>+</sup>, but Bayrou remains ahead in the majority-ranking. This shows how the majority judgment 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.

**Voters' utilities.** In theory the motivations of voters and their satisfaction are modelled by their "utilities." Given the decision mechanism and whatever information that is available, a rational voter chooses a message that maximizes his or her utility. But 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, etc. 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. It is "strategy-proof" for large classes of absolute utility functions. When the utilities of voters depend solely on the winner – a hypothesis often made – no mechanism is "strategy-proof." The majority judgment is not only partially-strategy-proof when utilities are relative but the analysis of the "game" of voting shows its behavior dominates that of other methods at Nash equilibria Balinski and Laraki (2010).

**Grades for candidates.** Voters who participated in the experiment were delighted with the idea that the majority judgment 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 judgment. 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 judgment may be used to disclose the electorate's evaluation of that candidate.



The majority judgment 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 Young (1975) 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.<sup>21</sup> But scoring (and point-summing) methods are all highly manipulable. The majority judgment 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 judgment: 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.<sup>22</sup>

**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? Should not someone who sees no discernable difference between two or more candidates be allowed to record this? Should not a voter who believes his or her second ranked candidate is merely acceptable or worse be allowed to express this? The majority judgment gives voters complete freedom of expression (within the bounds of the language).

<sup>21</sup> See Balinski and Laraki (2010). In *point-summing methods*, voters assign points from an interval to candidates and they are ranked according to the sum of their points.

<sup>22</sup> See Balinski and Laraki (2007), or Balinski and Laraki (2010) for proofs.

## *An Application to American Primaries*

American presidential primaries leap to mind as an immediately realistic application: not only would it be relatively easy to implement, but it would permit a much more complete expression of the voters' opinions. With as many as five to ten candidates, the first-past-the-post system drastically curtails expressions of the voters' opinions. Moreover, a "big winner" often garners as little as 25% of the total vote, hardly a mandate to be singled out as the principal candidate. A very small scale experiment was conducted on the web in late September, early October 2008. Members of INFORMS<sup>23</sup> were asked:

"Suppose that instead of primary elections in states to designate candidates, then national elections to choose one among them, the system was one national election in which all eligible candidates are presented at once. Or, suppose you are in a state holding a primary where you are asked to evaluate the candidates of all parties (at least one state primary votes on all candidates at once). A possible slate of candidates for President of the United States could be: [followed the names of the eight candidates given in Table 2.25 below together with their affiliations.]"

They were then instructed: "You will be asked to evaluate each candidate in a language of grades. A candidate's majority-grade is the middlemost of her/his grades (or the median grade). The candidates are ranked according to their majority-grades. The theory provides a natural tie-breaking rule." The ballot was the same as in Table 2.7.

Then they were invited to vote. The experiment was certainly not representative of the US electorate (nor was it meant to be). The results are nevertheless of interest.

In this case, the winner stands out as the only candidate with a *Very Good*, and the collective opinion of those who voted is quite clear.

**Table 2.25** INFORMS web experiment, mid-September to mid-October, 2008

	$p$ better than the majority-grade	$\alpha$ the majority-grade	$q$ worse than the majority-grade	
1st	Barack H. Obama	35.9%	<i>Very Good</i> +	32.0%
2nd	Hillary R. Clinton	45.0%	<i>Good</i> +	33.6%
3rd	Collin L. Powell	32.8%	<i>Good</i> —	41.2%
4th	Michael R. Bloomberg	42.0%	<i>Acceptable</i> +	31.3%
5th	John R. Edward	36.6%	<i>Acceptable</i> +	32.8%
6th	John S. McCain	33.4%	<i>Acceptable</i> —	44.2%
7th	W. Mitt Romney	46.6%	<i>Poor</i> +	22.9%
8th	Michael D. Huckabee	33.5%	<i>Poor</i> —	47.3%

<sup>23</sup> The Institute for Operations Research and the Management Sciences, a scientific society. A large majority of the members are US citizens, but many members are citizens of other nations.

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,<sup>24</sup> conducted under the same general conditions as the experiment of 2007, in five of Orsay’s twelve precincts<sup>25</sup> and the one precinct of Gy-les-Nonains, a small country town in Loiret. Of the 3,346 voters, 2,597 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.<sup>26</sup>

On average the voters cast 3.15 crosses per ballot (the distribution is given in Table 2.26). The actual system offered voters 17 possible messages, approval voting offered more than 65 thousand.<sup>27</sup> 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.

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

**Table 2.26** Number of ballots with  $k$  crosses,  $k = 0, 1, \dots, 16$ , approval voting experiment, five voting precincts of Orsay and Gy-les-Nonains, first round, April 21, 2002

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
Percentage Ballots	1.4	11.1	22.0	30.3	19.0	10.0	3.6	1.5	0.6	0.2	0.2

<sup>24</sup> 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 Balinski et al. (2003).

<sup>25</sup> 1st, 5th, 6th, 7th, and 12th.

<sup>26</sup> 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.”

<sup>27</sup> With 16 candidates there are  $2^{16} = 65,536$  possible messages. With the majority judgment, there are  $6^{16}$  or some 2.8 trillion possible messages.

**Table 2.27** Approval voting results, five precincts of Orsay and Gy-les-Nonains, first round, April 21, 2002

	Percentage ballots with crosses	Percentage 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%

spectrum, i.e., there exists no alignment of the candidates by which 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.<sup>28</sup>

The outcomes in the six voting precincts with approval voting and with the official voting are given in Table 2.27. 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 [Brams and Fishburn 1983](#)). Whereas we had entered into this experiment persuaded by the

<sup>28</sup> 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.

**Table 2.28** Percentages of both crosses or both no crosses, five precincts of Orsay and Gy-les-Nonains, first round, April 21, 2002

	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%	–

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.<sup>29</sup>

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 2.28).<sup>30</sup>

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 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 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 – do not mean the voter is indifferent among the corresponding candidates. This shows that the approval voting

<sup>29</sup> For a different analysis of this experiment, see Laslier and Van Der Straeten (2004).

<sup>30</sup> The analyses are confined to the more important candidates.

mechanism does not permit the voters to correctly express their preferences or their indifference. Crosses have different senses: it is not meaningful to aggregate them.

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 judgment 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 2.20): "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 judgment 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 [Baujard and Igersheim \(2007\)](#) tested two mechanisms at once<sup>31</sup> – approval voting (and a point-summing mechanism with points 0, 1 or 2, discussed below) – in six different voting precincts<sup>32</sup> 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, etc. 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 2.29](#). 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.”

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

**Table 2.29** Approval voting results, Illkirch/Louvigny/Cigné, April 22, 2007  
[Baujard and Igersheim \(2007\)](#)

	Percentage ballots with circles	Percentage 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%

<sup>31</sup> One ballot contained both. This permits analyses of potential interest. On the other hand, the participants expressed themselves twice simultaneously, which may have induced interdependencies.

<sup>32</sup> Three precincts in Illkirch (Alsace), two in Louvigny (Basse-Normandie), and one in Cigné (Mayenne).

**Table 2.30** Average number of highest, second highest, and third highest grades, three precincts of Orsay, April 22, 2007

Grades:	Three precincts	1st precincts	6th precincts	12th precincts
Average number highest	1.64	1.51	1.62	1.80
Average number second highest	2.19	2.08	2.16	2.34
Average number third highest	2.76	2.73	2.78	2.76

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, etc. But (see Table 2.8) 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 judgment 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 2.30): 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 more. It is not, suggesting that the behavior is not purely relative.

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*.
- 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 1st, of 2.27 in the 6th, and of 2.43 in the 12th. This is in substantial agreement with the 2.33 observed in the 2007 approval voting experiment.<sup>33</sup>

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 2.31).

<sup>33</sup> Applying this behavior to the majority judgment 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%, and Schivardi 2.6%.



**Table 2.31** Average number of grades per ballot: all and four candidates (Bayrou, Le Pen, Royal, and Sarkozy, normalized to sum to 12)

	<i>Excellent</i>	<i>Very Good</i>	<i>Good</i>	<i>Acceptable</i>	<i>Poor</i>	<i>to Reject</i>	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

Behavior that sees voters approving of some 20% of the candidates suggests 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 judgment 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 judgment 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., [Krantz et al. 1971](#)). “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 measures” 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, etc. – define what is meant by the numbers. Denmark’s new seven-grade number language adopted for the academic year

2006–2007 (to conform with the new European Credit Transfer Accumulation System’s ECTS grading scale<sup>34</sup>) 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,
- 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).<sup>35</sup> 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 [Baujard and Igersheim \(2007\)](#).

<sup>34</sup> 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>, December 5, 2007.

<sup>35</sup> This analysis results from a theoretical argument developed in [Balinski and Laraki \(2010\)](#).

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 etc. 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 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 about the 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*, etc. 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 2.30), 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 0–100 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 0–20 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.

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 2.31). (This change is unimportant to the majority judgment 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.

A Statistical Comparison of Methods

The traditional mechanisms are Condorcet’s, Borda’s, and their derivatives and combinations. They have never been used in elections.<sup>36</sup> The mechanisms used in the USA, the UK, and France are first- and two-past-the-post. Approval voting is a relative new comer. None offers the voters the freedom of expression allowed by the majority judgment, none asks or yields the electorate’s evaluations of the candidates.

The database of the ballots of the 2007 Orsay experiment permits a statistical comparison of the behavior of methods by deducing the votes between pairs of candidates as follows: when their grades differ, a vote is given the candidate with the higher grade; when their grades are the same, each is given 1/2 vote. The experiments 1,733 ballots, the *representative base* refers to 501 ballots that are “representative” of the votes cast in the first round in all of France [considerably more extensive analyses have been made Balinski and Laraki (2010)]. The 501 ballots were drawn randomly from the database of the 1,733 valid ballots. Assuming that when *k* candidates receive the highest grade on a ballot each is accorded 1/*k* votes, Table 2.32 shows how they compare with the national vote.

- The following methods are compared:
- First-past-the-post,
  - Two-past-the-post,

**Table 2.32** National first-round vote and estimates based on the representative base

	Sarkozy	Royal	Bayrou	Le Pen	Besancenot	Voynet	Others
National	31.2%	25.9%	18.6%	10.4%	4.1%	1.6%	Difference
501 sample	30.7%	25.9%	18.7%	9.3%	2.5%	3.2%	<0.6%

<sup>36</sup> Condorcet’s was, for a very short time, used to rank figure skaters, doubled – in case of an intransitivity – by Borda’s rule (see Balinski and Laraki 2010; in fact, the exact rule has been proposed and defended Dasgupta and Maskin 2004). Borda’s method was adopted in about 1,784 to elect members of France’s Academy of Sciences until a newly elected member, Napoléon Bonaparte, insisted it to be discarded in 1,800, presumably because it is highly manipulable, as Laplace had argued. It violates IIA, it ignores intensities, in Laplace’s words it gives “a big advantage to candidates of mediocre merit.” Arguments for it, alone or in convolutions, continue to be made to the present day Saari (2001).

- Condorcet’s,
- Borda’s,
- Approval voting where a ballot gives a cross, a tick, or a 1 whenever the grade is at least *Good*,
- Approval voting, where a ballot gives a cross, a tick, or a 1 whenever the grade is at least *Very Good*,
- Point-summing, where 5 points is given for *Excellent*, 4 for *Very Good*, 3 for *Good*, 2 for *Acceptable*, 1 for *Poor*, and 0 for *to Reject*,
- Majority judgment.

Two experiments investigate the manipulability of methods. Take a method. Ten thousand random samples are drawn from one of the bases, given that there is a unique winner *A* and a unique runner-up *B*. Two different strategies are applied. *Strategy 1*: All those ballots that give a grade to *B* two levels above the grade given to *A* are changed to raise *B* as much as possible and lower *A* as much as possible. Thus, for example, a ballot where *B* is *Good* and *A* is *Acceptable* nothing is changed, but if *A* is at most *Poor* then the change is made. *Strategy 2*: 30% of those ballots that give *B* a higher grade than *A* are changed to raise *B* as much as possible and lower *A* as much as possible. Tables 2.33 and 2.34 show how often the manipulation is successful in the sense that *A* is no longer the winner.

Note that if the Condorcet-winner *A* is no longer the winner, then there must be a Condorcet-cycle in the changed ballots. For, *A* has a higher grade than *B* on a

**Table 2.33** Numbers of successful manipulations in 10,000 random samples of 101 ballots drawn from both bases, with each of seven methods<sup>37</sup>

Total base	Point-summing	Borda	First-p-post	Apprvl $\geq$ Good	Apprvl $\geq$ Very good	Condorcet	Majority judgment
Strategy 1	9,418	8,145	8,435	4,536	3,559	5,071	3,138
Strategy 2	8,657	6,829	6,372	5,643	3,966	1,702	3,852
Rep base							
Strategy 1	9,965	9,313	8,699	8,569	8,407	7,042	6,142
Strategy 2	9,769	7,864	4,411	8,849	8,557	4,641	5,369

**Table 2.34** Numbers of successful manipulations in 10,000 random samples of 201 ballots drawn from both bases, with each of seven methods

Total base	Point-summing	Borda	First-p-post	Apprvl $\geq$ Good	Apprvl $\geq$ Very good	Condorcet	Majority judgment
Strategy 1	9,797	8,121	8,737	3,557	2,012	6,173	2,612
Strategy 2	9,233	9,711	8,801	5,213	2,465	8,215	3,807
Rep base							
Strategy 1	9,998	9,199	8,731	9,633	9,345	8,953	7,548
Strategy 2	9,974	9,917	7,860	9,830	9,296	9,378	6,380

<sup>37</sup> With these strategies, voters cannot manipulate the two-past-the-post method.

majority of the ballots, and that cannot change; thus, some candidate  $C$  must have a higher grade than  $A$  on a majority of the changed ballots. But  $B$  had a higher grade than  $C$  on a majority of the ballots to begin with, so also in the changed ballots, implying a Condorcet-cycle must exist among the three in the changed ballots ( $B \succ C \succ A \succ B$ ).

The statistics clearly show that the majority judgment is more stable against strategic manipulation than the other methods.

The database of 1,733 ballots confirms that there is no alignment of candidates according to which the “preferences” of all voters are “single-peaked.” However, the grades reveal a great deal of evidence about the preferences of each voter for the various candidates. One can calculate estimates of how voters favorable to one candidate might transfer their votes to others. It may be deduced from the numbers alone that statistically, the voters’ transfers are almost single-peaked among the important candidates Balinski and Laraki (2010). This may well be the case for other countries as well as France. In particular, Bayrou emerges as the single centrist candidate. This may be seen for other reasons as well (e.g., see Table 2.18). Thus, it becomes possible to compare the methods with regard to how they favor or penalize a centrist candidate.

Two experiments investigate how a centrist candidate fares under the various methods. This is an important question. The majority judgment has been attacked as a method that would be very favorable to centrists, and many political scientists, journalists, politicians, and voters believe that systematically electing centrists is not good for society. That allegation is shown to be wrong by the experiments. In one, the methods are used to obtain the results among only the three principal candidates, Bayrou, Royal, and Sarkozy. In the other, the methods are used to obtain the results among all twelve candidates: it turns out that in every case one of the three principal candidates is the winner. The results for the representative database are given in Table 2.35 (the results for the total database give the same ranking of the methods, but Bayrou is of course elected more frequently by each).

Several conclusions may be drawn from these results. First, the first- and two-past-the-post methods systematically eliminate centrist candidates, even when

**Table 2.35** How the centrist candidate (Bayrou) fares under different methods: numbers of wins in 10,000 random samples of 201 ballots drawn from the representative database<sup>38</sup>

	Royal		Bayrou		Sarkozy		Ties	
	(3)	(12)	(3)	(12)	(3)	(12)	(3)	(12)
First-past-the-post	656	977	0	0	9,261	9,022	83	5
Two-past-the-post	1,078	1,146	172	98	8,154	8,197	596	559
Approval $\geq$ Very Good	472	467	651	658	7,919	7,947	958	928
Majority judgment	587	606	4,402	4,326	5,008	5,065	3	3
Condorcet	138	142	8,390	8,329	954	974	389	441
Approval $\geq$ Good	36	23	9,436	9,465	30	40	498	472
Point-summing	132	139	9,444	9,463	260	239	164	159
Borda	51	12	8,659	9,976	1,122	0	168	12

(3) indicates the experiment with three candidates, (12) that with 12 candidates

they are highly regarded by the electorate (as was Bayrou in 2007). Second, the Condorcet method, and still more the point-summing and Borda methods, are extremely favorable to centrist candidates. In particular, notice that the more there are minor (unelectable) candidates, the more Borda guarantees the election of a centrist candidate. Third, approval voting is extremely sensitive to the question posed. When voters are asked to interpret “approval” as at least *Good* (in French, *Assez Bien*), the centrist is elected; when asked to interpret “approval” as at least *Very Good* (in French *Bien*), the centrist is eliminated. Imagine what would have happened if the threshold had been either higher or lower. Once again, this shows that approval voting’s two-word language is insufficient and arbitrary.

The majority judgment does not eliminate the centrist, yet neither does it necessarily elect the centrist. Statistically, Sarkozy wins more often than Bayrou. A method that is very favorable to the center will in the long run push all candidates to a centrist position. This is not desirable. Inversely, a method that systematically eliminates centrists will in the long run polarize society into two blocks. Something in between would seem to serve society better: a wider spectrum of political expression would be opened [Balinski and Laraki \(2010\)](#).

## Conclusion

The majority judgment experiment proves that the model on which the theory of social choice and voting is based is inadequate: 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 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 judgment 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 resists but is not impervious to manipulation. But there exists no method that is. The majority judgment best resists manipulation by several criteria, as the

---

<sup>38</sup> In the experiment with three candidates, for example, Royal had 656 wins, Bayrou 0 wins, Sarkozy 9,261 wins, and there were 83 ties: the sum is 10,000 (and similarly for the other methods in both experiments). However, to Condorcet must be added 129 Condorcet-cycles in the experiment with three, and 114 Condorcet-cycles in the experiment with 12. Ties with the majority judgment means ties in the majority-gauges

experimental evidence has illustrated and mathematical arguments have proven Balinski and Laraki (2010). 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 judgment 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 judgment.

*Ecclesiastes* poses the question:

“Is there *any* thing whereof it may said, See, this *is* new?”

Indeed, one century ago, Sir Galton (1907) had the germ of the idea. 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.*<sup>39</sup>

**Acknowledgments** We are deeply indebted to Cheng Wan whose final project as an undergraduate at the École Polytechnique (April–July, 2008) was devoted to statistical analyses of the various

---

<sup>39</sup> Our emphasis.



methods based on the 2007 Orsay experiment. The experience itself 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 Laraki, David Chavalarias, 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

- K.J. Arrow, *Social Choice and Individual Values*, Yale University Press, New Haven, CT, USA, 1951, 2nd ed., 1963.
- M. Balinski and R. Laraki, "A theory of measuring, electing and ranking," *Proceedings of the National Academy of Sciences, U.S.A.*, vol. 104 (2007) pp. 8720–8725.
- M. Balinski and R. Laraki, *Majority Judgment: Measuring, Ranking and Electing*, M.I.T. Press, Cambridge, MA, 2010, in press.
- 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*, (2003) No. 2003–013.
- A. Baujard and H. Igersheim, "Expérimentation du vote par note et 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.
- Jean Charles le Chevalier de Borda, "Mémoire sur les leçons au scrutin," *Histoire de l'Académie royale des sciences* (1784) pp. 657–665.
- S.J. Brams and P.C. Fishburn *Approval Voting*, Birkhäuser, Boston, 1983.
- Jean Antoine Caritat le Marquis de Condorcet, *Essai sur l'application de l'analyse la probabilité des décisions rendues à la pluralité des voix*, Paris (1785) l'Imprimerie royale.
- P. Dasgupta and E. Maskin, "The fairest vote of all," *Scientific American* (2004) pp. 92–97.
- E. Farvaque, H. Jayet, and L. Ragot, "Quel mode de scrutin pour quel 'vainqueur'? : une expérience sur le vote préférentiel transférable," working paper, Laboratoire Equipe, Universités de Lille, 2007.
- F. Galton, "One vote, one value," Letter to the editor, *Nature*, vol. 75 (1907) p. 414.
- A. Gibbard, "Manipulation of voting schemes: a general result," *Econometrica*, vol. 41 (1973) pp. 587–601.
- G. Hägele and F. Pukelsheim, "Llull's writings on electoral systems," *Studia Lulliana*, vol. 41 (2001) pp. 3–38.
- 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, 2008, pp. 229–249.
- D.H. Krantz, R.D. Luce, P. Suppes and A. Tversky, *Foundations of Measurement, Vol. 1*, Academic Press, New York, 1971.
- P. Kurrild-Klitgaard, "An empirical example of the Condorcet paradox of voting in a large electorate," *Public Choice*, vol. 107 (1999) pp. 1231–1244.
- Pierre-Simon le Marquis de Laplace, *Théorie analytique des probabilités*, 3rd ed., in *Œuvres Complètes de Laplace*, 1820, t. 7, pp. v and clii–cliii.
- 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 (2004) pp. 99–130.
- I. McLean, "The Borda and Condorcet principles: three medieval applications," *Social Choice and Welfare*, vol. 7 (1990) pp. 99–108.

- H. Moulin, "On strategy-proofness and single peakedness," *Public Choice*, vol. 35 (1980) pp. 437–455.
- D.G. Saari, *Chaotic elections! A Mathematician Looks at Voting*, American Mathematical Society, Providence, RI, 2001.
- M.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.
- A. Sen, *Collective Choice and Social Welfare*, Holden-Day, Inc., San Francisco, CA, 1970.
- H.P. Young, "Condorcet's theory of voting," *American Political Science Review*, vol. 82 (1988) pp. 1231–1244.
- 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.
- H.P. Young, "Social choice scoring functions," *SIAM Journal of Applied Mathematics*, vol. 28 (1975) pp. 824–838.

In Situ and Laboratory Experiments on Electoral Law  
Reform

French Presidential Elections

Dolez, B.; Grofman, B.; Laurent, A. (Eds.)

2011, X, 174 p., Hardcover

ISBN: 978-1-4419-7538-6