1 Majority Judgment

For with what judgement ye judge, ye shall be judged: and with what measure ye mete, it shall be measured to you again.
—Matthew 7:2

1.1 Inputs and Outputs

Throughout the world, voters elect candidates, and judges rank competitors, goods, alternatives, cities, restaurants, universities, employees, and students. How? Schemes, devices, or *mechanisms* are invented to reach decisions. Each defines

· the specific form of the voters' and judges' *inputs*, the *messages* used to exert their wills, and

· the procedure by which the inputs or messages are amalgamated or transformed into a final decision, social choice, or *output*.

In piano competitions, a judge's input message is a grade assigned to each competitor—often in the range from 0 (low) to 25 (high)—and the output is the rank-ordering determined by the competitors' average grades over all judges (sometimes by the means or averages of the grades after the one or two lowest and highest grades of each competitor have been eliminated). In the international standard for wine competitions, obtaining a judge's input message is more complex: each of fourteen attributes of a wine is assigned one of seven mentions (*Excellent, Very Good, Good, Passable, Inadequate, Mediocre, Bad*). The *Excellent*s are given a number score of 6 or 8, the *Bad*s a 0, and the others integer scores in between. The sum of the scores over all the attributes determines the judge's input message. The output is one of four medals (grand gold, gold, silver, bronze) or none, assigned to each wine on the basis of its average total score over all judges. In international figure skating contests, still more involved rules dictate the way the scores given by the twelve judges to each of the many parts of a competitor's performance become the number grades that are their input messages. The output is a rank-ordering determined by first, eliminating the grades of three judges chosen at random; second, eliminating the highest and lowest of the grades that are left; third, ranking the competitors according to the averages of the seven remaining grades. This complicated procedure is intended to combat judges who manipulate their inputs to favor or disfavor one or another competitor (the piano, wine, and figure skating mechanisms are described in detail in chapter 7).

In Australian elections, a voter's input is a complete rank-ordering of the candidates, and the output is a winner. But in most countries a voter's input message is at most one vote for one candidate, and the output is a winner, the candidate with the most votes; or the output is a ranking determined by the candidates' respective total votes. "Approval voting" is a relatively new mechanism used by several professional scientific societies: a voter's input message is one vote or none for every candidate, the output is a winner or a ranking determined by the candidates' respective total votes. These electoral schemes—each a pure invention to elicit the opinions of voters—offer an extremely limited possibility for voters to express themselves (various voting mechanisms used in practice are described in detail in chapter 2; other traditional methods in chapters 3 and 4; approval voting is discussed and analyzed in chapter 18; point-summing methods in chapter 17).

In fact, every mechanism generates information, notably the candidates' total scores, that in many situations may be viewed as constituting a part of the genuine outputs (see chapter 20).

1.2 Messages of a Common Language

Practice—with the notable exception of elections—suggests that letters (e.g., from a high of A to a low of E), descriptive words or phrases (e.g., from *Excellent* to *Bad*), or numbers (e.g., from 100 to 0) that define an ordered scale provide judges with a *common language* to grade and to rank competitors in a host of different settings. Typically, such languages are invented to suit the purpose, and carefully defined and explained. Their words are clearly understood, much as the words of an ordinary language or the measurements of physics (e.g., temperatures in degrees centigrade or Fahrenheit). A judge's input *message* is a grade or word of the common language. These grades or words are "absolute" in the sense that every judge uses them to measure the merit of each competitor independently. They are "common" in the sense that judges assign them with respect to a set of benchmarks that constitute a shared scale of evaluation. By way of contrast, ranking competitors is only “relative”; it bars any scale of evaluation and ignores any sense of shared benchmarks. The common languages used by judges in wine, figure skating, diving, and other competitions are described in chapter 7; their connections with measurement in general are discussed in chapter 8.

Judges and voters have complex aims, ends, purposes, and wishes: their *preferences* or *utilities*. A judge’s or a voter’s preferences may depend on many factors, including his beliefs about what is right and wrong, about the common language, about the method that transforms input messages into decisions, about the other judges’ or voters’ acts and behaviors, in addition to his evaluations of the competitors or candidates. But the judges’ or voters’ input messages—the grades they give—are assuredly not their preferences: a judge may dislike a wine, a dive, or a part of a skater’s performance yet give it a high grade because of its merits; or a judge may like it yet give it a low grade because of its demerits. Rules and regulations define how certain performances are to be evaluated, yet votes can be strategic. The fact that voters or judges share a common language of grades makes no assumptions about their preferences or utilities. Utilities are measures of the judges’ or voters’ satisfaction with the output, the decision of the jury or the society; grades are measures of the merits of competitors used as inputs. A judge’s or a voter’s input message is chosen strategically: depending on the mechanism for transforming inputs into an output, a judge may exaggerate the grades he gives, upward or downward, in the hopes of influencing the final result.

Arrow’s theorem plays an important role in this approach as well. It proves in theory what practitioners intuitively have learned by trial and error: without a common language there can be no consistent collective decision (see chapter 11, theorem 11.6a). Its true moral is that judges and voters *must* express themselves in a common language.

1.3 Majority-Grade

The fundamental problem is to find a *social decision function*: a method whose inputs are the grades of a common language and whose outputs are jury or electoral decisions, namely, final-grades and/or rank-orderings of competitors or candidates. Our theory shows there is one best method of assigning a final grade to each competitor or candidate—the *majority-grade*—and one best method of assigning a “generalized final grade” that ranks the competitors or candidates—the *majority-ranking*, where the winner is the first-place competitor or candidate. The definitions of these two terms are simple and have already been accepted in practice (e.g., for wine tasting; see Peynaud and Blouin 2006, 104–107). A host of different arguments prove they are the only social decision functions that satisfy each of various desirable properties.

Two supplementary concepts explained below are linked to the majority-ranking. The *majority-value* is a sequence of grades that determines the majority-ranking. The *majority-gauge* is a simplified majority-value that is sufficient to determine the majority-ranking when the number of judges or the electorate is large.

To begin, the aim is to decide on a final grade, given the individual messages of all the judges. Suppose the common language is a set of ten integers {0, 3, 5, 6, . . . , 11, 13} (from worst to best), the system of school grades previously used in Denmark (it is amusing that 1, 2, 4, and 12 are missing, the reasons for which are explained in chapter 8; but this has no influence on the present discussion). Imagine that the grades given to a competitor by all the judges are listed in ascending order from worst to best. When the number of judges is odd, the *majority-grade* α is the grade that is in the middle of the list (the median, in statistics). For example, if there are nine judges who give a competitor the grades {7, 7, 8, 8, **8**, 9, 10, 11, 11}, the competitor's majority-grade is 8. When the number of judges is even, there is a *middle-interval* (which can, of course, be reduced to a single grade if the two middle grades are the same), and the *majority-grade* α is the lowest grade of the middle-interval (the "lower median" when there are two in the middle). For example, if there are eight judges who give a competitor the grades {7, 7, 8, **8**, 11, 11, 11, 13}, the middle-interval goes from 8 to 11 and thus is the set of grades {8, 9, 10, 11}, and the competitor's majority-grade is 8.

The majority-grade α of a competitor is the highest grade approved by an absolute majority of the electors: more than 50% of the electors give the competitor at least a grade of α, but every grade lower than α is rejected by an absolute majority. Thus the majority-grade of a competitor is the final grade wished by the majority. In the first example, {7, 7, 8, 8, **8**, 9, 10, 11, 11}, only two (of nine) judges would vote for a lower grade, and only four for a higher grade. In the second example, {7, 7, 8, **8**, 11, 11, 11, 13}, only two (of eight) judges would vote for a lower grade, and only four for a higher grade.

The choice of the smallest grade of the middle-interval when the number of judges is even is the logical consequence of a principle of consensus. Compare two competitors A and B when there is an even number of judges: if all of A's grades strictly belong to the middle-interval of B's grades, then since there is a greater consensus among the judges for A's grade than for B's grade, A should be ranked at least as high as B. For example, if B's grades are {7, 7, 8, **8**, 11, 11, 11, 13} (the second example) and all of A's grades are either 9 or 10 (e.g., {9, 9, 9, 9, 9, 10, 10, 10}) and thus strictly belong to B's middle-interval, then A should rank higher than B.

The majority-grade is necessarily a word that belongs to the common language, and it has an absolute meaning. When an absolute majority of the judges give a competitor a particular grade α, then the competitor's majority-grade must necessarily be α, for if the number of judges is odd, the middle grade is necessarily α, and if the number of judges is even, the two middle grades are necessarily α.

It is reasonable to suppose that if a judge wishes that a competitor be accorded a certain grade—say, a 9 in the Denmark school scale—then the more the competitor's final grade deviates from 9, the greater will be the judge's discontent. When this is true, the best strategy for a judge is always to assign the grade that she believes the competitor merits, neither more nor less. For suppose that a judge believes that a candidate merits a grade of 9. If the majority-grade was higher, say, 11, she might be tempted, in anticipation of such an outcome, to assign a lower grade than 9. But doing so would change nothing because the majority-grade 11 would resolutely remain in the middle whatever lower grade she chose to give. If, on the other hand, the majority-grade was lower, say, 7, she might anticipate the outcome and be tempted to assign a higher grade than 9. Again this would change nothing because the majority-grade 7 would stay in the middle. The only other possibility is that the judge assigns a grade of 9 and the majority-grade is 9; in this case the judge is completely content. Thus in any case honesty is the best policy.

1.4 Majority-Ranking

In some applications, a complete ranking among all competitors or alternatives is not sought. For example, most wine competitions only wish to discern one of four medals (grand gold, gold, silver, bronze) or none. In many applications, however, notably in sports competitions, a complete rank-ordering is essential.

When two competitors have different majority-grades, the competitor with the higher grade is naturally ranked higher. Suppose, however, that two competitors A and B have the same majority-grade. For example, A's grades are {7, 9, **9**, 11, 11} and B's are {8, 9, **9**, 10, 11}, so they both have a majority-grade of 9. How are they to be compared? Their common (first) majority-grade is dropped (a single one) because it has already yielded all the information it can give relevant to comparing A with B, and the majority-grades of the grades that remain to each competitor—their second majority-grades—are found. In this example, A's remaining grades are {7, **9**, 11, 11} and B's are {8, **9**, 10, 11}, so their second majority-grades are both again 9. If one were higher than the other, it would designate the competitor who is ranked higher. If, as here, the second majority-grades are the same, they are discarded, and the third majority-grades of the competitors—the majority-grades of the grades that remain—are found, and so on, until one competitor is ranked ahead of the other. In this case, A's third majority-grade is 11 and B's is 10, so A ranks above B. One of the two must be ranked ahead of the other unless the competitors have identical sets of grades. This defines the *majority-ranking*.

1.5 Majority-Value

A competitor's *majority-value* is the sequence of his (first) majority-grade, his second majority-grade, his third majority-grade, down to his nth majority-grade (if there are n judges). Continuing with the example of section 1.4, A's majority-value is the ordered sequence of grades $\{09, 09, 11, 07, 11\}$ and B's is $\{09, 09, 10, 08, 11\}$. The lexicographic order of the majority-values gives the majority-ranking of the competitors. Thus A ranks higher than B because A's first grade in the sequence where their grades differ is higher than B's. When the common language is a set of integers, as in this case, the majority-value may simply be written as a number—in this example, A's is 09.09110711 and B's is 09.09100811—and the magnitudes of the majority-values determine the majority-ranking of the competitors. Moreover, dividing by 1.01010101 rescales the majority-values so that the minimum is 0, the maximum 13, and a competitor assigned the same grade α by all judges has a rescaled majority-value of exactly α. In this case A's rescaled majority-value is 9.000196 and B's is 9.000098. And if the set of grades were $\{09, 09, 09, 09, 09\}$, the rescaled majority-value would be 9.

The *majority-values* summarize all the results of an election or a competition. The first term of the sequence of grades of a competitor is his majority-grade (and when the grades are integers and the majority-value is written as a value or number, its integer part is his majority-grade). The lexicographic order among the sequences of grades that are the majority-values is the majority-ranking (and when the grades are integers and the majority-values are written as values or numbers, the majority-values determine the majority-ranking).

To more clearly understand these assertions, suppose four competitors, A, B, C, and D receive the Danish-style grades from nine judges 1, 2, …, 9 (table 1a). Reorder the grades of each competitor from highest to lowest (table 1.1b). The competitors' majority-grades are 10 for A, 9 for B and C, and 8 for D. Their majority-values (written as numbers with only the needed precision to

Table 1.1a
Hypothetical Example: Four Competitors, Nine Judges

Judge	*A*	*B*	*C*	*D*
1	13	09	07	08
2	10	09	13	05
3	09	08	11	13
4	10	11	09	08
5	04	10	09	09
6	13	08	00	02
7	11	07	10	13
8	10	11	09	08
9	10	11	11	07

Table 1.1b
Hypothetical Example: Grades Ordered from Best to Worst

A	*B*	*C*	*D*
13	11	13	13
13	11	11	13
11	11	11	09
10	10	10	08
10	**09**	**09**	**08**
10	09	09	08
10	08	09	07
09	08	07	05
04	07	00	02

Note: Majority-grades are shown in boldface.

distinguish their order) determine the majority-ranking among them (where $X \succ_S Y$ means the society or jury S prefers X to Y):

$$\begin{array}{llllllll} \text{Majority-ranking:} & A & \succ_S & C & \succ_S & B & \succ_S & D \\ \text{Majority-value:} & 10.\ldots & > & 9.091009\ldots & > & 9.091008\ldots & > & 8.\ldots \end{array}$$

Suppose, instead, that wines are to be judged, and the common language goes from *Excellent, Very Good, Good, Passable, Inadequate*, and *Mediocre* to *Bad*. The usual standard is five judges, so three wines—a St. Amour, a Bourgueil, and a Cahors—could receive the (already reordered) mentions shown in table 1.2. They all have the same majority-grade: *Good*. Their majority-values (written as a sequence of mentions to the needed precision to distinguish their order) determine the majority-ranking among them:

St. Amour	$\succ_S$	Bourgueil	$\succ_S$	Cahors
Good–Good–Very Good–Passable	>	*Good–Good–Very Good–Mediocre*	>	*Good–Passable*

Table 1.2
Hypothetical Example: Three Wines, Five Judges

St. Amour	Bourgueil	Cahors
Very Good	*Excellent*	*Excellent*
Very Good	*Very Good*	*Excellent*
Good	***Good***	***Good***
Good	*Good*	*Passable*
Passable	*Mediocre*	*Mediocre*

Note: Majority-grades are shown in boldface.

The first mention where two majority-values differ determines the order: thus the St. Amour and the Bourgeuil have the same mentions in the first three positions, and the fourth determines which ranks ahead of the other.

Two important points may be made immediately. First, if one or several competitors withdraw, the majority-ranking among the remaining competitors necessarily agrees with the majority-ranking among all competitors. So the majority-ranking satisfies Arrow's independence of irrelevant alternatives (IIA) condition: the relative positions of two competitors in the majority-ranking do not depend on the merits of another competitor. This is decidedly not the case with most voting mechanisms used in practice, as the United States learned in the presidential election of 2000 (see chapter 2), nor is it the case with the methods traditionally used to rank figure skaters (see chapter 7).

Second, what is the aim of a competition (or of an election)? It is to reach a *consensual decision*. The jury (or the society) seeks to find agreement. The majority-ranking makes the effect of middle grades more decisive. A few extreme or "cranky" evaluations should have a less decisive effect, though of course, every grade counts. If after the k best and the k worst grades of two competitors are dropped, the grades of a woman rank her ahead of a man by the majority-ranking, then she is ranked ahead of him by the majority-ranking of the jury (or the society). To see this in the first example of this section, where there are nine judges and four competitors, drop the two best and the two worst grades of competitors B and C. C is ranked ahead of B, $C \succ_S B$, because on the basis of the remaining five grades the majority-ranking puts C ahead of B (in this case their grades are all the same except for one, an 8 for B and a 9 for C). It has already been pointed out that the majority-ranking satisfies another such property, namely, when the number of judges is even and all the grades of one competitor A strictly belong to the middle-interval of the grades of another competitor B, then A should be ranked ahead of B. It is proven that the majority-ranking is the *only* method that satisfies these two properties.

1.6 Majority-Gauge

Juries or committees usually have a small number of judges or members: five, nine, perhaps twenty. The method and theory are the same when juries are any numbers of judges or voters. However, the majority judgment for "juries" composed of hundreds or millions of judges—nations electing presidents, cities electing mayors, congressional districts electing representatives, institutions and societies electing officers—has an easier and more compelling description in that context.

Majority judgment was tested in a field experiment on April 22, 2007, in parallel with the first round of the French presidential election, in Orsay, a town close to Paris. In French presidential elections, an elector casts her vote for one candidate (or none). If no candidate receives an absolute majority of the votes, a second round is held two weeks later between the two candidates who had the most votes in the first round. The results of the first round are used to explain the approach.

The experiment took place in three of the twelve voting precincts of Orsay that together had 2,695 registered voters. Of these, 2,383 voters cast official ballots (88% of those registered), of which 2,360 were valid. After voting officially, the voters were asked to participate in the experiment using the majority judgment. They had been informed about it by mail, printed flyers, and posters. It was conducted in accordance with usual French voting practice: ballots were filled out and inserted into envelopes in voting booths with curtains, then deposited in transparent ballot boxes.

The ballot is reproduced in figure 1.1 (the names of the candidates are given in the official order, the result of a random draw). A serious and solemn question was posed to the voters,

To be president of France,
after having taken every consideration into account,
I judge in conscience that this candidate would be

and asked to give an answer for every candidate in a common language of grades—absolute evaluations—common to all French voters:

Très Bien, Bien, Assez Bien, Passable, Insuffisant, or *A Rejeter*.

The first five designations are known to all those who have been school children in France; the last is clear enough. Reasonable translations are:

Excellent, Very Good, Good, Acceptable, Poor, or *To Reject.*

Bulletin de vote:
Élection du Président de la République 2007

Pour présider la France,
ayant pris tous les éléments en compte,
je juge en conscience que ce candidat serait:

	Très Bien	*Bien*	*Assez Bien*	*Passable*	*Insuffisant*	*A Rejeter*
Olivier Besancenot						
Marie-George Buffet						
Gérard Schivardi						
François Bayrou						
José Bové						
Dominique Voynet						
Philippe de Villiers						
Ségolène Royal						
Frédéric Nihous						
Jean-Marie Le Pen						
Arlette Laguiller						
Nicolas Sarkozy						

Cochez une seule mention dans la ligne de chaque candidat.
Ne pas cocher une mention dans la ligne d'un candidat revient à le Rejeter.

Figure 1.1
Ballot, Orsay experiment, 2007 French presidential election.

The meanings of the grades are directly related to the question posed.

The sentences at the bottom of the ballot say, "Check one grade in the line of each candidate. No check in the line of a candidate means To Reject him." We believe that every voter must be required to evaluate every candidate. A voter having no opinion concerning a candidate has not even taken the time to evaluate him and thus has implicitly rejected him (other possibilities are discussed in chapters 13 and 14).

Of the 2,383 persons who cast official ballots, 1,752 participated in the experiment (74%). In fact, the rate of participation was slightly higher because in France a voter is permitted (under certain conditions) to ask someone else to vote in his place, but no one was allowed to vote twice in the experiment. Nineteen ballots were invalid, usually because more than one grade was assigned to a candidate, leaving 1,733 valid majority judgment ballots. The results are given in table 1.3.

Table 1.3
Majority Judgment Results, Three precincts of Orsay, April 22, 2007

	Excellent	*Very Good*	*Good*	*Acceptable*	*Poor*	*To Reject*	No Grade
Bayrou	13.6%	30.7%	25.1%	14.8%	8.4%	4.5%	2.9%
Royal	16.7%	22.7%	19.1%	16.8%	12.2%	10.8%	1.8%
Sarkozy	19.1%	19.8%	14.3%	11.5%	7.1%	26.5%	1.7%
Voynet	2.9%	9.3%	17.5%	23.7%	26.1%	16.2%	4.3%
Besancenot	4.1%	9.9%	16.3%	16.0%	22.6%	27.9%	3.2%
Buffet	2.5%	7.6%	12.5%	20.6%	26.4%	26.1%	4.3%
Bové	1.5%	6.0%	11.4%	16.0%	25.7%	35.3%	4.2%
Laguiller	2.1%	5.3%	10.2%	16.6%	25.9%	34.8%	5.3%
Nihous	0.3%	1.8%	5.3%	11.0%	26.7%	47.8%	7.2%
de Villiers	2.4%	6.4%	8.7%	11.3%	15.8%	51.2%	4.3%
Schivardi	0.5%	1.0%	3.9%	9.5%	24.9%	54.6%	5.8%
Le Pen	3.0%	4.6%	6.2%	6.5%	5.4%	71.7%	2.7%

Note: When there was no grade, it was counted as a *To Reject*, as per the instructions on the ballot (so Bayrou's *To Reject* was counted as 7.4%, Royal's as 12.6%, . . . , Le Pen's as 74.4%). There were few.

The Orsay experiment is discussed in greater detail in chapters 6, 15, and 19. It is used here as a vehicle to explain the procedure when there is a large electorate or jury, and to remark on several of its features.

When there are a very large number of voters or judges, as in this case, generically (almost surely) the middle-interval—a single grade if the number of voters is odd, two grades if the number of voters is even—will be one and the same grade. Thus it is safe to simply say that a candidate's *majority-grade* is the median of his grades: it is at once the highest grade approved by a majority and the lowest grade approved by a majority. Alternatively, only a minority would be for a higher grade or for a lower grade. For example, the majority-grade of D. Voynet (see table 1.3) is *Acceptable* because a majority of $53.4\% = 2.9\% + 9.3\% + 17.5\% + 23.7\%$ of the voters judge that she merits at least an *Acceptable*, and a majority of $70.3\% = 23.7\% + 26.1\% + 16.2\% + 4.3\%$ of the voters judge that she merits at most an *Acceptable*. Or, only a minority of 29.7% would be for a higher grade and only a minority of 46.6% for a lower grade.

The *majority-ranking* is calculated more directly in the case of a large number of voters. When the numbers or percentages of grades higher or lower than the candidates' majority-grades are different, which is almost surely true, the majority-ranking is obtained from three pieces of information concerning each candidate:

- p, the number or percentage of the grades better than a candidate's majority-grade;

- α, the candidate's majority-grade;
- q, the number or percentage of the grades worse than a candidate's majority-grade.

The triple (p, α, q) is the candidate's *majority-gauge*. S. Royal's majority-gauge (see table 1.3) is (39.4%, *Good*, 41.5%) since 39.4% = 16.7% + 22.7% of her grades are better than *Good*, and 41.5% = 16.8% + 12.2% + 10.8% + 1.8% are worse than *Good*. If the number or percentage p of the grades better than a candidate's majority-grade α is higher than the number or percentage q of those worse than the candidate's majority-grade, then the majority-grade is completed by a plus (+); otherwise the majority-grade is completed by a minus (−). Thus S. Royal's majority-grade is *Good*−. The plus or minus attached to the majority-grade is implied by the majority-gauge, so it is not necessary to show it, but for added clarity it is most often included, so that, for example, Royal's majority-gauge may be written (39.4%, *Good*−, 41.5%).

Naturally a *majority-grade*+ is ahead of a *majority-grade*− in the majority-ranking. Of two *majority-grade*+'s, the one having the higher number or percentage of grades better than the *majority-grade* is ahead of the other; of two *majority-grade*−'s, the one having the higher number or percentage of grades worse than the *majority-grade* is behind the other. For example, in table 1.4, S. Royal and N. Sarkozy both have the majority-grade *Good*−. Royal has 41.5% worse than *Good*, and Sarkozy has 46.9% worse than *Good*, so Royal finishes ahead of Sarkozy. O. Besancenot and M.-G. Buffet both have the majority-grade *Poor*+. Besancenot has 46.3% better than *Poor*, and Buffet has 43.2% better than *Poor*, so Besancenot finishes ahead of Buffet. To see a candidate's majority-gauge, imagine a see-saw or teeterboard with all the voters lined up according to the grades they give, from best to worst. Assuming each voter's weight is the same, the grade given by the voter who stands at the fulcrum where the board is in perfect balance is the majority-grade. Remove, now, all voters who gave the majority-grade and place the fulcrum at the juncture between the better than and worse than majority-grade. If the board tilts to the better grades, a plus (+) is accorded, and the more it tilts, the better is the majority-gauge. If it tilts to the worse grades, a minus (−) is accorded, and the more it tilts, the worse is the majority-gauge.

The majority-gauges and the majority-ranking of the experiment are shown in table 1.4. It may be seen that the majority-ranking is quite different than the order of finish given by the official vote in these three voting precincts (which are *not* representative of the vote in all of France; see the last two columns of table 1.4). This is due to the fact that the majority judgment allows voters to

Table 1.4
The Majority-Gauges (p, α, q) and the Majority-Ranking, Three Precincts of Orsay, April 22, 2007

		p Better Than Majority-Grade	α Majority-Grade	q Worse Than Majority-Grade	Official Vote, 3 Precincts	Official National Vote
1st	Bayrou	44.3%	*Good+*	30.6%	25.5%	18.6%
2d	Royal	39.4%	*Good−*	41.5%	29.9%	25.9%
3d	Sarkozy	38.9%	*Good−*	46.9%	29.0%	31.2%
4th	Voynet	29.7%	*Acceptable−*	46.6%	1.7%	1.6%
5th	Besancenot	46.3%	*Poor+*	31.2%	2.5%	4.1%
6th	Buffet	43.2%	*Poor+*	30.5%	1.4%	1.9%
7th	Bové	34.9%	*Poor−*	39.4%	0.9%	1.3%
8th	Laguiller	34.2%	*Poor−*	40.0%	0.8%	1.3%
9th	Nihous	45.0%	*To Reject*	–	0.3%	1.1%
10th	de Villiers	44.5%	*To Reject*	–	1.9%	2.2%
11th	Schivardi	39.7%	*To Reject*	–	0.2%	0.3%
12th	Le Pen	25.7%	*To Reject*	–	5.9%	10.4%

express their opinions on all the candidates rather than simply singling out one among them.

The reasons for believing in the validity of the experiment are given later, but several salient observations are made here.

- More than one of every three participants gave their highest grade to two or more candidates.
- Only half of the voters used a grade of *Excellent*.
- On average, a voter gave a grade of *To Reject* to over one-third of the candidates.

This proves that voters do not have in their minds rank-orderings of the candidates (and still more evidence supports this claim). A rank-ordering does not allow a voter to express an equal evaluation of candidates, or an intensity of appreciation, or an outright rejection. It also shows that the actual system—cast one vote for one among several or many candidates—forced one-third of the voters to opt for a candidate when in fact they saw no difference among two or more of them. These observations are convincing because voters had no incentive to vote strategically since they were only participating in an experiment. It is precisely such experiments that can elicit the true opinions of voters.

Strategic voting played an important role in the French presidential election of 2007. Voters had in mind what had happened in 2002, when the vote to the left was so widely distributed among some eight candidates that instead of a second round between Jacques Chirac (the incumbent president and major candidate of the right) and Lionel Jospin (the standing prime minister and major candidate of the left), Chirac was pitted against Jean-Marie Le Pen, the perennial candidate of the extreme right (see chapter 2). It seems safe to assert that in the first round of 2007 a significant number of voters did not vote for the candidate they preferred. Instead of voting for their favorite—an ecologist, a communist, or a Trotskyist—many voters of the left opted for Ségolène Royal, the socialist, the major candidate of the left. And the same phenomenon occurred on the right: it seems that many voters abandoned the extreme right to vote for the major candidate of the right, the U.M.P. candidate, Nicolas Sarkozy. In contrast, the majority judgment encourages a voter to express his convictions (as is proven via several precise criteria that are defined in later chapters).

To see how the majority judgment resists strategic manipulation in the context of elections, take a candidate, say Ségolène Royal, whose majority-grade is *Good*− and whose majority-gauge is

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

Then, 39.4% of her grades are better than *Good*, 41.5% are worse than *Good*, so 19.1% are *Good*. Who are the voters who can change Royal's majority-gauge by changing the grades they give her, and what are their motivations to change?

Suppose a voter believes a candidate merits a grade of α, and the further the majority-grade is from α, the less he likes it (a reasonable motivation: the voter's preferences in grading are then said to be single-peaked). Then, as was seen, the voter's optimal voting strategy is simply to give the candidate the grade α: the majority judgment is *strategy-proof-in-grading*.

Similar reasoning shows that the majority-grade mechanism is *group strategy-proof-in-grading*. A group of voters who share the same beliefs (e.g., they belong to the same political party) has the same optimal strategy, namely, to give to the candidates the grades it believes they merit. For if the group believed that Royal merited better than *Good*, and all raised the grade they gave her, her majority-gauge would remain the same (p does not change). If all lowered the grade they gave her, her majority-gauge would decrease (q increases), and perhaps her majority-grade would be lowered (not their intent). If the group believed that Royal merited worse than *Good*, and all lowered the grade they gave her, her majority-gauge would remain the same (q does not change). If all raised the grade they gave her, her majority-gauge would increase (p increases), and perhaps her majority-grade would be raised (not their intent).

These strategy-proof-in-grading properties are *not* true of any of the mechanisms currently used today. The strategy of a voter may, however, focus on the final ranking of the candidates rather than on 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 he thinks the candidate merits: there is no mechanism that is strategy-proof-in-ranking. The majority judgment does not escape the Gibbard-Satterthwaite impossibility theorem (see chapters 5 and 13), but it best resists such manipulation.

One means by which it resists is easy to explain. Take the example of Bayrou with a *Good*+ and Royal with a *Good*− (see table 1.4); their respective majority-gauges are

Bayrou: (44.3%, *Good*, 30.6%)   Royal: (39.4%, *Good*, 41.5%).

How could a voter who graded Royal 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, and those who can raise Royal's majority-gauge cannot lower Bayrou's. For suppose a voter can lower Bayrou's. Then 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 she cannot raise Royal's majority-gauge (cannot raise her p). Symmetrically, a voter who can raise Royal's majority-gauge must have given her a *Good* or worse and thus to Bayrou a worse than *Good*; so the voter cannot lower Bayrou's majority-gauge (cannot increase his q).

Compared with other mechanisms, the majority judgment cuts in half the possibility of manipulation, however bizarre a voter's motivations or whatever her utility function. The majority judgment resists manipulation in still other ways that other methods do not, but to see how requires information found in voters' individual ballots that is not shown in the elections results of table 1.4. For example, significant numbers of voters cannot contribute at all either to raising Royal's majority-gauge or to lowering Bayrou's (28% of those who graded Royal above Bayrou). Moreover, those who can manipulate have no incentive to exaggerate very much in any case, for it does not pay to do so (a more detailed analysis is given in chapter 19).

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

Voters who participated in the Orsay experiment were delighted with the idea that a candidate could be assigned a final grade. The majority-grade is an important signal that expresses the electorate's appreciation of a candidate. Chirac's "triumph" against Le Pen in 2002—a majority of over 80% in the second round—would have been very different had the majority judgment been used: he surely would have won, but with a middling grade—perhaps an *Acceptable* or a *Good* to Le Pen's *To Reject*—that would have given a more sober sense to his reelection. In the election of 2007, Voynet's majority-grade of *Acceptable* and her fourth place in the majority-ranking more clearly express the electorate's concern with environmental issues than her eighth-place finish in the official national vote. Le Pen's last place in 2007 and solid *To Reject* evaluation shows the electorate's strong rejection of his ideas, whereas the official vote makes him one of the four major contenders. Even when there is only one contender, a not infrequent occurrence—in the 2002, 2004, and 2006 U.S. congressional elections, respectively 81, 66, and 59 candidates were elected with no Democratic or Republican opponent—the majority judgment establishes the esteem in which the candidate is held.

U.S. presidential primaries leap to mind as an immediately realistic application. The majority judgment would be relatively easy to implement since the decision to do so may be taken at the state level. It would permit a real expression of the voters' opinions versus sending a message consisting of one name. With as many as five to ten candidates, the first-past-the-post system drastically curtails expression of the voters' opinions: a "big winner" often garners as little as 25% of the total vote, hardly a mandate to be singled out as the principal candidate. Indeed, the luck of the draw may determine the "winner" due to the mere presence of strategic candidates who have no real chance of emerging as real contenders. Finally, and of real importance, the current system is divisive for a political party: it opts for one candidate and throws out the others. With the majority judgment, candidates are not rejected: many, perhaps all, receive good majority-grades, yet one is singled out because he is first in the majority-ranking.

A very small-scale experiment was conducted on the Web in late September, early October 2008. Members of the Institute for Operations Research and the Management Sciences (INFORMS), a scientific society, were asked, Election of the President of the United States of America 2008

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

	Excel-lent	*Very Good*	*Good*	*Accept-able*	*Poor*	*To Reject*	*No Opinion*
Michael R. Bloomberg, Ind.							
Hillary R. Clinton, Dem.							
John R. Edwards, Dem.							
Michael D. Huckabee, Rep.							
John S. McCain, Rep.							
Barack H. Obama, Dem.							
Colin L. Powell, Ind.							
W. Mitt Romney, Rep.							

You must check one single grade or *No Opinion* in the line of each candidate.
No Opinion is counted as *To Reject*.

Figure 1.2
Ballot, U.S. presidential election, INFORMS experiment, September–early October 2008.

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 [here followed the names of the eight candidates given in the ballot together with their affiliations.]

They were 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.

Then they were invited to vote with the ballot shown in figure 1.2.

This experiment was certainly not representative of the U.S. electorate (nor was it meant to be). A large majority of the members of INFORMS are U.S. citizens, but many members are citizens of other nations. The results, shown in table 1.5, 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 1.5
The Majority-Gauges (p, α, q) and the Majority-Ranking, INFORMS Experiment, September–Early October 2008

		p Better Than Majority-Grade	α Majority-Grade	q Worse Than Majority-Grade
1st	Barack H. Obama	35.9%	*Very Good+*	32.0%
2nd	Hillary R. Clinton	45.0%	*Good+*	33.6%
3rd	Colin L. Powell	32.8%	*Good−*	41.2%
4th	Michael R. Bloomberg	42.0%	*Acceptable+*	31.3%
5th	John R. Edwards	36.6%	*Acceptable+*	32.8%
6th	John S. McCain	33.4%	*Acceptable−*	44.2%
7th	W. Mitt Romney	46.6%	*Poor+*	22.9%
8th	Michael D. Huckabee	33.5%	*Poor−*	47.3%

The descriptions of the majority judgment given in this chapter should permit its use in any application—with few judges or many voters—given that a common language of grades has been defined and explained.

1.7 Nomenclature

Majority judgment is the name we have chosen to give to the method we advocate. It encompasses several key ideas, each endowed with a name that is used again and again throughout the book.

- *Majority-grade* The middlemost or median of the grades given to a competitor by the judges or voters; when there is an even number of judges or voters, the lower middlemost of the grades (see section 1.3 and chapter 12).
- *Majority-ranking* The majority judgment ranking of all competitors on the basis of their grades (see section 1.4 and chapter 13). They are ranked according to their majority-values or majority-gauges.
- *Majority-value* The sequence of a competitor's grades consisting of his first majority-grade, his second majority-grade, …, and so on, up to his nth majority-grade (when there are n judges or voters). Competitor A ranks higher than competitor B in the majority-ranking if and only if A's majority-value is lexicographically above B's (see section 1.5 and chapter 13).
- *Majority-gauge* A simplification of the majority-value from which may be deduced the majority-ranking among the competitors in many cases; in particular, when there are many judges or voters such as in most elections. However, when two competitors are tied with the same majority-gauge, they are not necessarilty tied with their majority-values (see section 1.6 and chapter 14).

· *Abbreviated majority-value* An abbreviated but complete expression of the majority-value (see chapter 14).

· *kth-order function*. It singles out a competitor's kth-highest grade. When there are n judges or voters, there are n order functions. The first-order function is the highest grade, the nth-order function is the lowest grade. Among them the majority-grade is the $\left(\frac{n+1}{2}\right)th$-order function when n is odd and the $\left(\frac{n+2}{2}\right)$ th-order function when n is even.

1.8 The Thesis

The intent of this book is to show why the majority judgment is superior to any known method of voting and to any known method of judging competitions.

To do so, it presents the fundamentals of the traditional theory of social choice, describing the principal known methods, together with simple proofs of the most important results. It also proposes new characterizations, new incompatibility theorems, and new methods in the traditional model of social choice. Actual voting systems and methods used in practice to judge competitors (e.g., wines, divers, figure skaters) are also described to show how they are in fact subject to all the paradoxes and failures that are identified in theory. Throughout, theory, experiments, and practical evidence in voting and in judging competitions are provided to support the first central point: *the traditional model is a bad model, in theory and in practice.*

The new model is then developed. It is shown that a host of properties uniquely characterize the "order functions" as the only methods that can be used. This is established from a variety of points of view. Practice again plays a central role: experimental evidence is given that shows the majority judgment is a practical method and that common languages in voting and in judging competitions do in fact exist and can be meaningfully defined. Statistical comparisons that depend on real data are made with other methods to show why the majority judgment is better in voting; in particular, approval voting is shown to fail. Often, judging competitions (such as wine, ski jumping, or figure skating) invoke several different criteria: the majority judgment is generalized to such situations. Throughout, theory, experiments, and practical evidence are provided to support the second central point: *the majority judgment is a better alternative to all other known methods, in theory and in practice.*